



Ysgoloriaethau Sgiliau Economi Gwybodaeth  
Knowledge Economy Skills Scholarships



# The Effect of Post Translational Modification and Oligomerisation on the Structure-Function Relationship of Horseradish Peroxidase and Fluorescent Proteins.

Rachel Louise Johnson

A Dissertation Submitted to Cardiff University  
in candidature for the degree of  
Doctor of Philosophy  
September 2020

Molecular Biosciences Division  
School of Biosciences  
Cardiff University

## i. Acknowledgements

First and foremost, I would like to thank my primary supervisor Dr Dafydd Jones. A big thank you is due for the advice which I have been given in this writing up period, and I truly appreciate the time, speed and effort in which feedback was given. I would also like to thank my secondary supervisors, Dr Chris Thomas, and Dr Mark Young, for the advice, resources, and guidance they have given me in the duration of this project.

Aside from my supervisors, this work would not have been possible without the resources provided by the technical staff at Cardiff School of BIOSI and by the assistance given in mass spectrometry by Cardiff School of Chemistry. Additionally, I would also like to thank Dr Chris Pudney and his team at Bath, especially Sam Winter, for the use of their research facilities, and, Dr Pierre Rizkallah, for his time in setting up crystals, and for the opportunity to visit the Diamond Light Source.

Thanks, must be paid to every member of the DDJ research group over the last 4 years in their many varied and valued ways. I have greatly enjoyed working with a group of individuals whom I could learn so much from whilst enjoying a cup of tea and a crossword. Special thanks must be paid to the members of the lab who were present when I first joined, in particular, Husam, Ben, Jacob, and Harley, who taught me the basics which were vital to me developing my research. Next, I would like to thank Khyati, Valentine, and Shweta for Fun-Friday lunches, and the master's students Hayley and Tom for their collaboration on the cytochrome project. Lastly, I would like to thank the most recent members of the DDJ lab, Rochelle and Beckie, for their encouragement during my final year in the lab.

Large thanks must be paid to each of my friends and family who supported me in many immeasurable ways during this process. Emma and Dan, thank you for always giving me a place to stay when I wanted a break, and Sky, it has been great to be able to discuss our shared experience as we've gone through this process together. Mum and Dad, thank you for the many sacrifices you have both made to boost my education over the years, especially during this final year of writing. I do not think I would have gotten to this point if I had not been given the help with the times-tables and spellings during those early years (even if the spellings didn't stick as well), but more importantly, I certainly would not have got through this last year without your generosity, and it is very much appreciated.

Finally, Conor, this thesis would have certainly had far fewer commas (none), if you had not taken the time to proof the segments I was stuck on. I would definitely not be at the point of submission if it were not for the support and motivation you have given me at every stage along the way. I cannot wait to start our new life together in Salisbury, it is going to be an adventure which I am very excited to share.

## ii. Abstract

Protein function is inherently linked to amino acid sequence; however, this vocabulary is innately limited. One natural method to supplement protein chemistry is the application of PTM. This modification has significant implications for both the structure and function of protein and can now be emulated with non-natural chemistry. This thesis examines the influence PTM has, on protein stability (Chapter 3), function (Chapter 4), structure (Chapter 5), and how nnPTM can be applied to induce functional communication (Chapters 6 and 7).

This thesis first explores the expression of HRP in *E. coli* (Chapter 3), wherein it is demonstrated that soluble HRP expression, is achieved by the inclusion of the full-length proto-HRP sequence. These proto-HRP regions are then shown to cleave, and additionally, it is also shown that the removal of unoccupied N-linked glycosylation sites from the surface of HRP decreases the rate of precipitation observed in recombinant HRP.

Next, recombinant *apo*-HRP is investigated as a corrective agent in a commercial immunoassay (Chapter 4), to establish if the corrective function of glycosylated *apo*-HRP can be achieved without glycosylation. Evidence generated indicates that glycosylation of *apo*-HRP is essential for the removal of false-positive rogue signalling. Subsequently, Chapter 5 explores the influence of PTM on HRP's structural rigidity. Using fluorescence emission to assess the red edge excitation shift of W117, it is observed that both haem binding and glycosylation increase protein structural rigidity, and this is confirmed by CD spectra.

The latter two chapters of this thesis explore the potential of chromophore communication which can be induced by nnPTM. Together, both *in silico* interface modelling, and nnAA incorporation, is utilised for the formation of two fluorescent dimers. Firstly, this system was applied to the structurally similar combination of sfGFP and mCherry (Chapter 6). In which the enhanced chromophore proximity resulted in a significant decrease in function of mCherry, contrary to expectation. In Chapter 7 sfGFP and cytochrome *b*<sub>562</sub> were dimerised to generate dimeric protein which could transfer energy by way of FRET (Chapter 7). Lastly, both nnPTM and natural PTM were combined in the formation of trimeric protein which was linked by SPAAC and disulphide linkage.

### iii. Abbreviations

In this thesis, a standard one-letter amino acid code followed by the residue number was used to describe amino acids within a protein. While a single point mutation is denoted as follows: residue number and mutant amino acid code, in superscript after the protein name. For example, the substitution of lysine residue 198 with non-canonical amino acid AZF in mCherry is denoted as mCherry<sup>198AZF</sup>.

Dimeric protein, formed of two non-natural protein mutants were named as follows: name of first protein followed by the name of second protein, then in superscript, the residue of linkage in first protein and the residue of linkage in second protein. For example, dimer formed by linkage of GFP<sup>148SCO</sup> and mCherry<sup>198AZF</sup> is referred to as GFPCH<sup>204-198</sup>.

Abbreviation	Meaning
AA	Amino Acid
aaRS	Aminoacyl-tRNA-synthetase
<i>apo</i>	Protein lacking metal porphyrin saturation
AZF	p-azido-L-phenylalanine
bp	Base Pair
C-Terminus	Carboxyl-terminus
CD	Circular Dichroism
Cyt <i>b</i> <sub>562</sub>	Cytochrome <i>b</i> <sub>562</sub>
Da	Dalton
ELISA	Enzyme Linked Immunosorbent Assay
ε	Molar absorbance
FRET	Förster resonance energy transfer
GFP <sup>WT</sup>	GFP wildtype as obtained from <i>Aequorea victoria</i>
HIS TAG	poly-histidine Tag
<i>holo</i>	Protein which has been saturated with porphyrin
HRP	Horseradish Peroxidase
HRP <sup>EC1</sup>	HRP expressed from the first construct designed for protein expression
HRP <sup>EC2</sup>	HRP expressed from the second construct designed for protein expression
IEC	Ion Exchange Chromatography
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kB	Kilobase
mature-HRP	HRP which has been cleaved of N and C-terminal sequences
mCherry	Red Fluorescent Protein derived from DSRED
nnAA	Non-natural Amino Acid
N-Terminus	Amine Terminus of Protein
OCD	Ortho Clinical Diagnostics
OD <sub>600</sub>	Optical Density at 600 nm

OPD	o-phenylenediamine dihydrochloride
PCR	Polymerase Chain Reaction
pI	Isoelectric Point, the pH at which protein charge is 0.
proto-HRP	HRP which has not been subjected to proteolytic cleavage
PTM	Post Translational Modification
QY	Quantum Yield
R <sub>Z</sub>	Reinheitszahl Value
REES	Red edge excitation shift
RMSD	Root mean square deviation
RMSF	Root Mean Square Fluctuation
RT	Room Temperature (20 °C)
SCF	Serum Correction Factor
SCO-K	S-cyclooctyne-L-lysine
SDS-PAGE	Sodium dodecyl sulphate Polyacrylamide gel electrophoresis
SEC	Size Exclusion Chromatography
sfGFP	Super folder Green Fluorescent Protein derived from Aequorea victoria
Soret Peak	Wavelength of maximum haem porphyrin absorbance
SPAAC	Strain Promoted Azide-Alkyne Cycloaddition
TEV	Tobacco etch virus
T <sub>m</sub>	Mid-point of thermal denaturation
T <sub>opt</sub>	The optimum temperature of catalysis
tRNA	Transfer-RNA
tRNAAA	Aminoacyl-tRNA
v/v	Volume per Volume
w/v	Weight per Volume
WT	Wildtype
λ <sub>EM</sub>	Wavelength of maximum fluorescence emission
λ <sub>EX</sub>	Wavelength of maximum fluorescence excitation
λ <sub>MAX</sub>	Wavelength of maximum absorbance

## iv. Table of Contents

i. Acknowledgements .....	ii
ii. Abstract .....	iii
iii. Abbreviations .....	iv
iv. Table of Contents .....	vi
Chapter 1. Introduction .....	10
1.1. The Principles of Protein Structure and Function .....	11
1.1.1 Sequence-Structure-Function Relationship of Proteins .....	11
1.1.2. Post Translational Modification of Protein .....	12
1.2. The Implications of Native PTM on the Recombinant Expression of HRP .....	16
1.2.1. The Fundamentals of Recombinant Production .....	16
1.2.2. The Generation of Recombinant Protein .....	16
1.2.3. The Limitations of Recombinant Protein Production .....	18
1.2.4. Horseradish Peroxidase: A Target of Recombinant Expression .....	19
1.2.5. The Importance of PTM to HRP Expression .....	21
1.2.6. The Importance of Cofactor Binding of HRP in Recombinant Expression .....	24
1.2.7. The Influence of HRP Recombinant Expression on Glycosylation .....	26
1.2.8. The Implications of HRP Proteolytic Cleavage for Recombinant Expression .....	27
1.2.9. Disulphide Bond Formation in Recombinant Expression .....	28
1.3. Protein Engineering .....	30
1.3.1. A Historical Perspective of Protein Engineering .....	30
1.3.2. Protein Mutagenesis .....	30
1.3.3. Green Fluorescent Protein: A Model for Protein Engineering .....	31
1.3.4. The Relationship Between Structure and Function of GFP <sup>WT</sup> .....	32
1.3.5. Generation of GFP Variants by Protein Engineering .....	35
1.3.6. The Mutagenesis of GFP to Generate Functional Change .....	36
1.3.7. The Use of GFP As A Probe of Protein Interaction .....	37
1.3.8. Förster Resonance Energy Transfer .....	38
1.3.9. Emulation of Natural PTM to Modify GFP Structure and Function .....	40
1.3.10. The Application of PTM to Induce GFP Oligomerisation .....	41
1.4. Expanding the Genetic Code to Enhance Control of Protein Interaction .....	43
1.4.1. Why use non-natural Chemistry .....	43
1.4.2. The Incorporation of nnAA Into Proteins During Cellular Protein Synthesis .....	44
1.4.3. The Modification of Protein Reactivity by The Incorporation of nnAA .....	46
1.4.4. Strain Promoted Azide-Alkyne Cycloaddition .....	47

1.5. The Scope of Project and General Aims .....	50
Chapter 2. Materials and Methods .....	51
2.1. Materials .....	52
2.1.1. Water Source and Buffer Preparation .....	52
2.1.2. Chemicals .....	52
2.1.3. Bacterial Cell Strains .....	53
2.1.4. Plasmid Vectors .....	53
2.1.5. Bacterial Growth Media.....	55
2.2. Molecular Biology.....	55
2.2.1. Construction of HRP Expression Vectors .....	56
2.2.2. Design of Mutagenesis Oligonucleotides.....	56
2.2.3. Site-Directed Mutagenesis .....	57
2.2.4. Bacterial transformation .....	58
2.2.5. Extraction and Quantification of DNA from Cells .....	59
2.3. Bacterial Growth and Protein Expression .....	60
2.4. Protein Purification.....	62
2.5. Protein Characterisation.....	66
2.6. Trial of <i>apo</i> -HRP <sup>EC2</sup> in Immunoassay .....	75
2.7. <i>In silico</i> Molecular Modelling.....	77
2.8. Click Chemistry .....	80
Chapter 3. The Soluble Expression of Recombinant Horseradish Peroxidase in <i>E.coli</i> . .....	81
3.1. Introduction.....	82
3.1.1. The Commercial Production of HRP .....	82
3.1.2. The Influence of Host on the Expression of HRP .....	83
3.1.3. Problems Faced by The Expression of HRP in <i>E. coli</i> .....	83
3.1.4. Expression of the Full-Length Sequence of Proto-HRP .....	84
3.1.5. Chapter Aims .....	85
3.2.1. Structural Influence of Proto-Peptide Regions on HRP.....	86
3.2.2. Design of Expression Construct HRP <sup>EC1</sup> .....	87
3.2.3. Expression of HRP <sup>EC1</sup> .....	88
3.2.4. Optimisation of HRP <sup>EC1</sup> Expression. ....	90
3.2.5. Periplasmic Localisation of HRP <sup>EC1</sup> .....	92
3.2.6. Prediction of Pre-Peptide Cleavage .....	93
3.2.7. Purification of HRP <sup>EC1</sup> .....	94
3.2.8. HRP <sup>EC1</sup> Expression and Purification .....	97
3.2.9. Mutagenesis to Reduce Recombinant HRP Precipitation .....	98

3.2.10. Design of Construct HRP <sup>EC2</sup> .....	100
3.2.11. Expression of HRP <sup>EC2</sup> .....	101
3.2.12. The Solubility of Expressed HRP <sup>EC2</sup> .....	103
3.2.13. Nickel Affinity of HRP <sup>EC2</sup> .....	103
3.2.14. Post-Lysis Cleavage of HRP <sup>EC2</sup> .....	105
3.2.15. Cleavage of Purified HRP <sup>EC2</sup> .....	108
3.2.16. The Activity of HRP <sup>EC2</sup> .....	112
3.3. Discussion .....	114
3.3.1. Expression of Soluble Proto-HRP .....	114
3.3.2. Comparison of HRP <sup>EC1</sup> and HRP <sup>EC2</sup> .....	115
3.3.3. Proteolytic Cleavage of Recombinantly Expressed HRP .....	116
3.3.4. Conclusion .....	117
Chapter 4. The Identification and Analysis of Serum Corrective Agent .....	118
4.1. Introduction.....	119
4.1.1. Immunoassays .....	119
4.1.2. False Positive Reporting .....	120
4.1.3. Potential Mechanisms of SCF Action .....	120
4.1.4. Production of Serum Correction Factor.....	122
4.1.5. Chapter Aims .....	123
4.2. Results and Discussion.....	125
4.2.1. Characterisation of SCF .....	125
4.2.2. Crystallisation of SCF .....	133
4.2.3. Recombinant Protein Selection .....	135
4.2.4. Haem Extraction of Recombinant Protein.....	136
4.2.5. Assay Calibration .....	138
4.2.6. Corrective Action of Recombinant Protein .....	140
4.3. Discussion.....	144
4.3.1. Identity of SCF .....	144
4.3.2. Replication of SCF by Recombinant Expression .....	144
4.3.3. The Importance of Glycosylation.....	145
4.3.4. Conclusion .....	146
Chapter 5. The Impact of Post Translational Modification on the Structure and Function of Horseradish Peroxidase.....	147
5.1 Introduction.....	148
5.1.1. The Structural Influence of PTM.....	148
5.1.2. The Significance HRP PTM .....	149

5.1.3. Red Edge Excitation Shift.....	150
5.1.4. Using REES to infer Protein Structural Characteristics.....	151
5.1.4. Chapter Aims and Objectives.....	152
5.2. Results and Discussion.....	153
5.2.1. Protein Used in Investigation.....	154
5.2.2. Tryptophan 117 Emission .....	155
5.2.3. REES Measurement .....	157
5.2.4. Circular Dichroism. ....	160
5.2.5. Thermal Induced Unfolding of HRP .....	162
5.2.6. Steady-State Kinetics of holo-HRP .....	164
5.2.7. Temperature Dependence of Activity .....	166
5.2.8. Molecular Dynamics .....	168
5.3.1. The influence of PTM on HRP Flexibility .....	171
5.3.2. Influence of Glycosylation on Peroxidase Structure .....	172
5.3.3. Influence of Haem Binding on Peroxidase Structure .....	174
5.3.4. Concluding Statement .....	175
Chapter 6. The Effect of Oligomerisation on The Function of Disparate Fluorescent Proteins. ....	177
6.1. Introduction.....	178
6.1.1. The Importance of Protein Oligomerisation .....	178
6.1.2. Previous Synthetic Oligomerisation of Monomeric Proteins. ....	178
6.1.3. Oligomeric History of Fluorescent Proteins .....	179
6.1.4. Structural and Functional Comparison of mCherry and sfGFP .....	181
6.1.5. Functional Communication between sfGFP and mCherry.....	183
6.1.6. Aims and Objectives. ....	184
6.2. Results and Discussion.....	185
6.2.1. Selection of an sfGFP and mCherry <sup>WT</sup> Oligomerisation Interface .....	185
6.2.2. Residue selection for AZF Incorporation .....	188
6.2.3. Site-Directed Mutagenesis of mCherry <sup>WT</sup> to mCherry <sup>198TAG</sup> .....	190
6.2.4. Expression and Purification of mCherry <sup>198AZF</sup> .....	191
6.2.5. Functional Characterisation of mCherry <sup>198AZF</sup> .....	193
6.2.6. UV Irradiation of mCherry <sup>198AZF</sup> .....	194
6.2.7. Molecular Attachment of mCherry <sup>198AZF</sup> with Cy3 DBCO .....	195
6.2.8. SPAAC between mCherry <sup>198AZF</sup> and sfGFP SCO-K .....	197
6.2.9. Separation of Monomeric Protein from GFPCH <sup>204-198</sup> .....	199
6.2.10. Absorbance of GFPCH <sup>204-198</sup> .....	199
6.2.11. Fluorescence of GFPCH <sup>204-198</sup> .....	201

6.2.12. FRET of GFPCH <sup>204-198</sup> .....	204
6.2.13. Functional Characterisation of GFPCH <sup>148-198</sup> .....	206
6.2.14. Unfolding of mCherry <sup>198AZF</sup> .....	209
6.3.1. Discussion .....	212
6.3.1. Explanation of mCherry Chromophore Functional Loss .....	212
6.3.2. Functional Communication between Chromophores.....	216
6.3.3. Conclusion .....	217
Chapter 7. The Generation of Structurally Diverse Protein Dimers and Oligomers.....	218
7.1 Introduction.....	219
7.1.1. The Importance of Chromophore Communication .....	219
7.1.2. Cytochrome <i>b</i> <sub>562</sub> .....	220
7.1.3. Communication between Cytochrome <i>b</i> <sub>562</sub> and GFP .....	222
7.1.4. Aims.....	223
7.2. Results and Discussion.....	225
7.2.1. <i>In Silico</i> Simulation of Interface between sfGFP and Cytochrome <i>b</i> <sub>562</sub> .....	225
7.2.2. Residue Selection for nnAA Incorporation in Cytochrome <i>b</i> <sub>562</sub> .....	227
7.2.3. TAG Mutagenesis of Cytochrome <i>b</i> <sub>562</sub> .....	229
7.2.4. Purification of Cytochrome <i>b</i> <sub>562</sub> <sup>71AZF</sup> .....	230
7.2.5. Absorbance of Cytochrome <i>b</i> <sub>562</sub> <sup>71AZF</sup> .....	232
7.2.6. Cycloaddition of sfGFP <sup>148SCO</sup> and Cytochrome <i>b</i> <sub>562</sub> <sup>71AZF</sup> .....	233
7.2.7. The Influence of Temperature on Dimer Formation .....	234
7.2.8. Separation of Heterodimer from Monomer .....	235
7.2.9. Absorbance Spectrum of Oxidised GFP <i>b</i> <sup>148-71</sup> and GFP <i>b</i> <sup>204-71</sup> .....	238
7.2.10. Absorbance of GFP <i>b</i> <sup>204-71</sup> in Reducing Conditions.....	240
7.2.11. Excitation and Emission of GFP <i>b</i> <sup>148-71</sup> and GFP <i>b</i> <sup>204-71</sup> .....	241
7.2.12. The Sensitivity of sfGFP <sup>WT</sup> Emission to Haem .....	244
7.2.13. Redox Sensitivity of GFP <i>b</i> <sup>204-71</sup> Emission .....	245
7.2.14. The Rationale for the Double Mutagenesis of Cytochrome <i>b</i> <sub>562</sub> <sup>71AZF</sup> .....	246
7.2.15. Mutagenesis, Expression and Purification of Cytochrome <i>b</i> <sub>562</sub> <sup>DM</sup> .....	249
7.2.16. Oligomeric Assembly of sfGFP and Cytochrome <i>b</i> <sub>562</sub> <sup>DM</sup> .....	250
7.3. Conclusion .....	253
7.3.1. Discussion of Oligomerisation of Cytochrome <i>b</i> <sub>562</sub> and sfGFP .....	253
7.3.2 Functional Communication between Chromophores.....	254
7.3.3. Sensitivity of GFP <i>b</i> <sup>204-71</sup> to Reducing Conditions.....	256
7.3.4. Concluding Statement .....	257
Chapter 8. Discussion .....	258

8.1. General Overview .....	259
8.2. The importance of native PTM .....	259
8.3. The Use of Non-Natural PTM for Protein Attachment .....	261
8.4. Future Work .....	263
8.4.1. Formation of a Novel Biomolecule Detection Assay Using nnPTM of HRP .....	263
8.4.2. Oligomerisation of HRP to Form a Novel Biosensing Dimer .....	265
8.5 Concluding Statement. ....	267
8.6. Publication Associated with This Thesis.....	268
Chapter 9. Bibliography.....	269

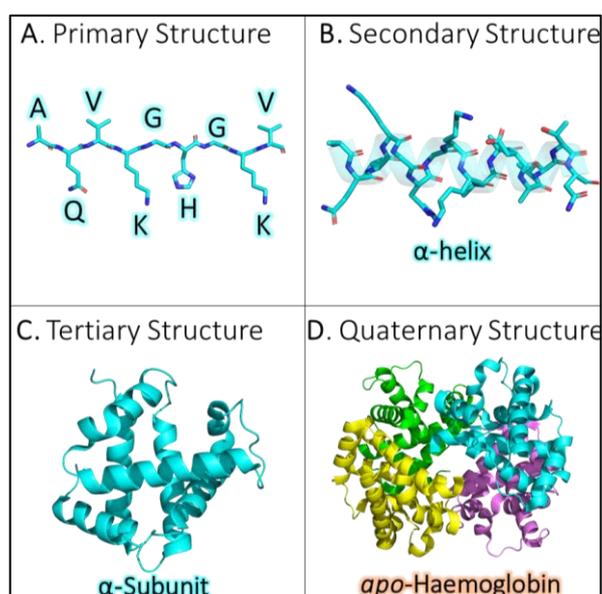
# Chapter 1. Introduction

# 1.1. The Principles of Protein Structure and Function

## 1.1.1 Sequence-Structure-Function Relationship of Proteins

Proteins are molecular machines, which are essential for life <sup>1</sup>. The function of proteins in nature is diverse, with roles including energy conversion, catalysis, signalling, structural regulation, and cell motility. Modern understanding of the translation of the genetic code into the amino acid sequence is fundamental to our perception of biology <sup>2</sup>. The “central dogma” is that the DNA sequence of a gene encodes for a protein’s amino acid sequence and therefore, its primary structure <sup>3</sup>.

The amino acid sequence is unique to each protein and forms the basis of protein structure and function <sup>4,5</sup> (Figure 1.1, A). Protein structure is derived by a hierarchical organisation which transforms a linear protein sequence to form a fully folded polypeptide. This arrangement commences when amino acids in the peptide chain, interact to form a three-dimensional arrangement, such as  $\alpha$ -helices or  $\beta$ -strands <sup>6</sup> (Figure 1.1, B), before the progression into a more complex, non-symmetrical, tertiary structure <sup>7</sup>(Figure 1.1, C). Typically, this forms by the formation of a protein core which is driven by the isolation of hydrophobic residues from solvent. Finally, a protein’s quaternary structure is driven by the interaction of



**Figure 1.1. The Different Levels of Protein Structure, As Seen in Haemoglobin Biosynthesis.** **A**, the sequential arrangement of amino acid (sticks), which forms primary structure. **B**, folding of the polypeptide chain to form  $\alpha$ -helical arrangement, shown as both sticks and cartoon. **C**, the contracting of the formed peptide chain to form the globular structure of haemoglobin  $\alpha$ -subunit (cartoon). **D**, the association between two  $\alpha$ -subunits and two  $\beta$  subunits to form *apo*-haemoglobin polypeptide (PDB 3B75).

---

side chains on multiple peptide strands, which align together to form one polypeptide entity. For example, in the interaction of haemoglobin  $\alpha$  and  $\beta$  subunits, this interaction is stabilised by the formation of several salt bridges, hydrophobic interactions and hydrogen bonds, for the formation of multiple polypeptides in one complex, haemoglobin (Figure 1.1, D). Protein folding itself is not always a sequential process with secondary, tertiary, and even quaternary structure forming concurrently during the folding process.

There are 20 primary proteogenic amino acids in nature, which form the basis of all chemical diversity sampled in naturally evolved proteins<sup>6</sup>. Selenocysteine and pyrrolysine are also found in nature but are uncommon exceptions to this rule<sup>8,9</sup>. Although the variety of chemistry sampled in canonical amino acids is limited, it is sufficient to generate all-natural protein diversity. For example, for every 200-residue protein, there are  $20^{200}$  possible combinations of amino acids<sup>10</sup>. However, the formation of both tertiary and quaternary structure and subsequent protein function is hitherto restricted by the side-chain association of these amino acids<sup>11</sup>.

### 1.1.2. Post Translational Modification of Protein

Post-translational modification (PTM) is the process by which covalent amendments are made to a protein after ribosomal synthesis. Most commonly, side chains are chemically modified with new adducts, cofactors are bound, and proteolytic cleavage occurs<sup>12</sup>. PTM is ubiquitous in all forms of life and is the mechanism by which protein chemical diversity can be supplemented in natural systems<sup>12</sup>. PTM is distinct from natural proteinogenesis as the modification is not exclusively controlled by the genome, as in the case of amino acid sequence, but instead, is regulated by the recognition and modification of specific protein regions which are themselves encoded from the genetic code. Many diverse forms of PTM exist in nature, which allows for the expansion of protein structural and functional diversity, by specific peptide interaction (Table 1.1).

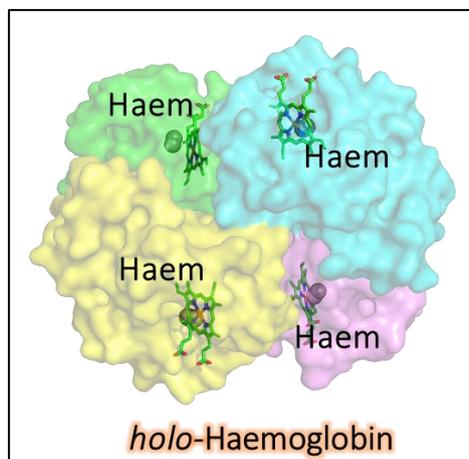
Table 1.1. Common Forms of Post Translational Modification

Common forms of PTM*	Example of Function	Characterisation
Acetylation	Regulation of transcription.	Attachment of acetyl group to a protein's N terminus and lysine residues.
ADP ribosylation	Addition of ADP-ribose moieties for cell signalling.	Attachment of ADP ribose to a variety of amino acid side chains.
Amidation	Receptor Recognition.	Attachment of amide group to a protein's C-terminus.
Carboxylation	Localization of Protein.	Carboxylic group is added to <b>glutamate</b> residues.
Cofactor Addition	Enable activity and increase stability.	Binding of either a metal ion or porphyrin to protein.
Disulphide Bonding	Link two polypeptide regions.	Formation of covalent attachment between two <b>cysteine</b> residues.
Glycosylation	Aid protein stability.	Covalent attachment of carbohydrate to, <b>asparagine</b> , <b>serine</b> , or <b>threonine</b> residues.
Methylation	Epigenetic regulation of gene expression.	Addition of methyl group to typically an arginine or lysine amino acid.
Nitration	Regulation of gene expression.	Attachment of nitro group. Typically, this is added to <b>tyrosine</b> , <b>tryptophane</b> , <b>cysteine</b> , and <b>methionine</b> residues.
Non-Natural Oligomerisation	Assembly of oligomeric protein from monomeric units	Incorporation and interaction of non-natural chemistry either through attachment or translation
Peptide Cleavage	Enzymatic removal of specified peptide regions from formed protein	Incorporation of specific amino acid sequence for recognition by protease enzyme
Phosphorylation	Signal transduction and apoptosis.	Addition of phosphoryl group.
Prenylation	Anchoring protein to membranes.	Addition of isoprenyl lipid, to the C-terminal <b>cysteine</b> of a target protein.
SUMOylation	Covalent attachment of a Small Ubiquitin-like Modifier	SUMO proteins are like ubiquitin and attach to protein by linkage of C terminal glycine to an existing lysine.
Ubiquitination	Signalling of Protein degradation	Attachment of ubiquitin, commonly to a <b>lysine</b> residue on the protein.

\* List does not feature an exhaustive list of PTM but rather references the diversity of PTM in nature.

Many proteins undergo PTM in nature <sup>13</sup>; for example, an estimated 25-50% of all proteins are post-translationally modified to bind cofactors <sup>14</sup>. Haemoglobin, for instance, must

bind the prosthetic cofactor haem to function as an oxygen carrier (Figure 1.2). Therefore, in the absence of PTM, haemoglobin is inactive, as the folded protein lacks the functional capacity to both bind and transport molecular oxygen.



**Figure 1.2. Post Translational Activation of Haemoglobin by Haem Saturation.**

The surface of haemoglobin coloured in accordance with subunit identity. Coordinated haem is shown as sticks (Green) whilst associated molecular oxygen is shown as spheres (Grey).

PTM can be applied for the modification of a protein by three distinct mechanisms: natural, engineered, or non-natural in origin. Natural PTM describes PTM which arose natively alongside the evolution of the protein in a specific organism. This process is precise and is regulated by many additional cellular proteins which regulate site-specific chemical alteration<sup>15</sup>. Virtually all eukaryotic proteins undergo a form of PTM post-synthesis, in a process which can be transformative to both function and structure<sup>13</sup>. This process is complex and often depends upon the presence and availability of many additional cellular enzymes to regulate modification for defined, site-specific, chemical alteration<sup>15</sup>. Consequently, the recombinant expression of protein in cells which do not feature native cellular conditions can have significant implications for both protein structure and function.

Engineering of the genetic code can be used to introduce new PTM into a specific site or target protein using natural protein chemistry. Examples include: the introduction of phosphorylation sites, to modulate receptor interaction<sup>16</sup>; the linkage of two proteins, by insertion of two surface-exposed thiol groups in cysteine mutagenesis<sup>17</sup>; and the inclusion of protein cleavage sequences, to induce enzymatic cleavage<sup>18</sup>. The area has been expanded in recent years for the application of highly specific natural modifying enzymes, which facilitate target specific modification both *in vitro* and *in situ*. Examples of such an approach are the

---

HaloTags and SpyTag/Spycatcher systems of attachment. These two systems are formed by the inclusion of natural protein sequences into a protein of interest so that the protein be modulated post-translationally for specific conjugation <sup>19,20</sup>.

An extension of the engineering of PTM is the application of non-natural PTM (nnPTM), which enables the modification of a target protein with chemistry which is not observed in nature. Synthetic approaches to PTM have expanded significantly with the advent and application of non-natural protein labelling and amino acids. This expansion can include the attachment of synthetic chemical adducts, such as azide and alkyne chemistry to native side-chains for protein attachment or oligomerisation. One such example is the conjugation of an enzyme to an antibody for diagnostic analysis <sup>21</sup>. In this example HRP and antibody are typically linked by covalent attachment of two surface-exposed amino groups *via* an intermediary molecular bridge, such as glutaraldehyde <sup>22</sup>. A recent advancement to this mechanism is the development of non-natural amino acids (nnAA) which allow for specific covalent attachment without the need for an intermediary molecular bridge. For example, the incorporation of two click-chemistry amenable amino acids, can induce the molecular anchoring of protein (see Section 1.4).

The relationship between protein amino acid sequence and function is underpinned by our understanding of the genetic code as a template for protein production <sup>23</sup>. However, given the wide-spread use of PTM in nature, and their significance and implications on the protein structure-function relationship, the modification holds many considerations for modern biotechnology. Thus, the modification of PTM, either by recombinant expression<sup>24</sup>, protein engineering <sup>25</sup> or through the expansion of the genetic code <sup>26</sup>, can exert significant implications on protein function and structure, both of which will be explored in this thesis.

---

## 1.2. The Implications of Native PTM on the Recombinant Expression of HRP

### 1.2.1. The Fundamentals of Recombinant Production

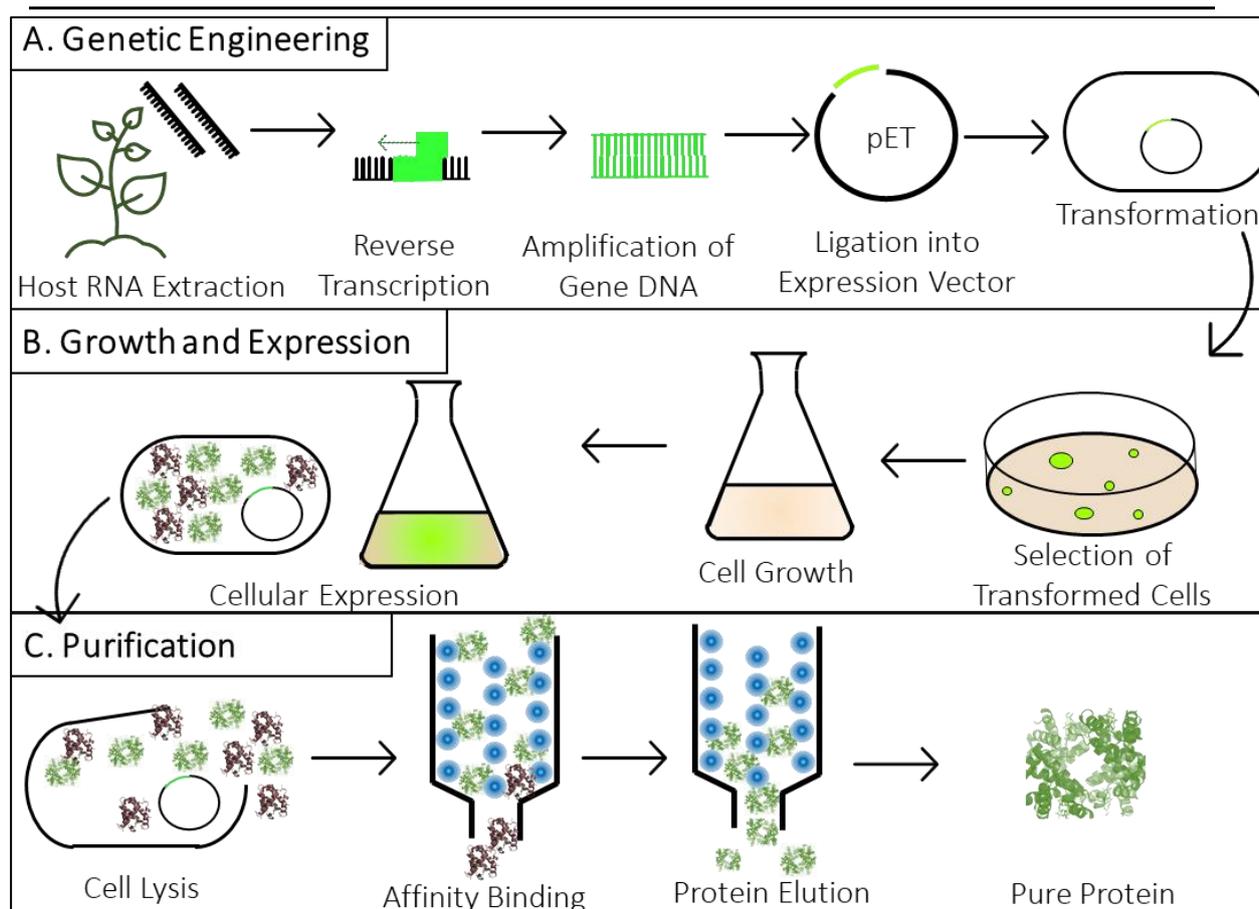
Since its inception 40 years ago, recombinant protein expression has revolutionised molecular biology<sup>27</sup>. The first functional protein produced by this technique was by Itakura *et al.* (1977), who demonstrated that the expression of human hormones could be achieved by heterologous expression in *Escherichia coli* (*E.coli*)<sup>28</sup>. In the decades since, this principle has developed, now many protein therapies<sup>29,30</sup> and drug discovery targets<sup>27,31</sup>, are produced by recombinant expression.

Recombinant, or heterologous, expression, in principle, is the transfer of a gene from one organism to another, to produce, engineer, modify, or study the encoded protein of interest<sup>6</sup>. The process can be used to generate both an increase in protein expression and a reduction in the time necessary for protein accumulation. This increase in abundance, in turn, improves the availability of protein for study<sup>23</sup>, whilst additionally providing a means of reducing the challenges presented by downstream processing<sup>32</sup>.

Due to their rapid growth, low cost, and their capability for continuous fermentation; microbial cells are widely used for recombinant expression<sup>23,33</sup>. For example, *E. coli* cells can reproduce rapidly, with a maximal doubling time of approximately 20 minutes<sup>34</sup>, reaching an approximate cell density of 200 g/1L of growth<sup>35</sup>. Organism engineering has, in recent decades, enabled an increased supply of protein for study<sup>23</sup>. One such example on an engineered organism is that of commercially available *E. coli*, which has been engineered to the point wherein up to 50% of the cell's metabolism to be directed towards recombinant expression<sup>36</sup>.

### 1.2.2. The Generation of Recombinant Protein

A fundamental principle in recombinant protein expression is the combination of a gene sequence from one organism with the machinery necessary for expression by another<sup>37,38</sup>. This process requires three key stages of research: genetic engineering, transformation, and purification (Schematic 1.1). In genetic engineering, the sequence of the target protein must first be identified and isolated before insertion into a vector of choice<sup>37</sup> (Schematic 1.1,



**Schematic 1.1. The Basic Methodology of Recombinant Protein Production in *E. coli*.**

**A**, genetic engineering, the identification, and isolation of the target gene sequence before vector assembly and transformation of plasmid for host. **B**, growth and expression, of cells transformed with gene to acquire protein. **C**, purification, the extraction, and isolation of expressed protein for generation of pure product.

A). Cells are then grown to high density and stimulated to express the protein (Schematic 1.1 B). In turn, this expressed protein is extracted from cells and purified ready for use (Schematic 1.1, C).

For example, expression of recombinant protein in *E. coli* cells requires the cloning of the DNA sequence of choice into an extrachromosomal plasmid compatible for use with the host. The constructed genetic vector is then supplied to cells by transformation, which is the process by which cells are subjected to mild stress to promote uptake of the plasmid vector<sup>39,40</sup>. Commonly, several additional genetic features are incorporated with the gene of interest to control expression in the host organism. These sequences control protein induction<sup>23</sup>, isolation<sup>41</sup>, localisation<sup>42</sup>, and cleavage<sup>43</sup>(Table 1.2).

The pET plasmid system is one such example of a common vector used in the production of protein from *E. coli* cells<sup>44,45</sup> This system utilises the kanamycin resistance gene

**Table 1.2. Genetic Features of a Generic Recombinant Construct.**

Expression Vector Feature	Function	Purpose
Ori	The origin of plasmid replication is initiated.	Amplification of plasmid within cells.
Selection marker	Provides resistance to antibiotic.	Selection of transformed cells.
Promotor	Initiates induction of gene of interest.	Controlled gene expression.
Terminator	Termination of protein sequence.	Ensure expression of only desired protein.
Affinity Marker (Optional)	Binding of expressed protein to specific resins <sup>41</sup> .	Isolation of expressed Protein <sup>37</sup> .
Cleavage Site (Optional)	Remove N or C-terminus from protein <sup>43</sup> .	Removal of affinity marker.
Signal Sequence (Optional)	Locate protein to specific cellular compartment <sup>42</sup> .	Allows for the formation of disulphide bonds in non-reducing environment.

for cell selection and incorporates the T7 promotor for induction of target protein<sup>46</sup>. Consequently, the use of kanamycin in growth media selects for only transformed cells whilst protein induction can be stimulated by the presence of either lactose<sup>35,47</sup> or Isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG)<sup>48</sup>. This plasmid will be used in Chapter 3 for *E. coli* expression.

### 1.2.3. The Limitations of Recombinant Protein Production

Although the benefits of recombinant expression are well-established, limitations to protein expression endure. It is estimated that approximately half of all target proteins are not soluble when expressed heterologously<sup>38</sup>. Whilst the expression of the largest and most complex proteins is impeded by the limited number of chaperone proteins in microbial host cells<sup>49</sup>. Consequently, just 10% of eukaryotic proteins are expressed as soluble entities in *E. coli*<sup>37</sup>.

A limitation of recombinant expression is the deviation in PTMs from that applied in native expression<sup>50</sup>. As PTM is not encoded genetically, instruction for its application is not transferred with the genetic sequence of the protein. Therefore, extraction and expression of a protein's genetic sequence in an alternative host organism does not replicate PTM due to the divergence in host cell environment<sup>51</sup>. As protein organisation is dynamic and sensitive to

chemical change, the absence of PTM can induce a significant impact on activity, stability and interaction of a recombinantly expressed protein<sup>52,53</sup>. One such example of the importance of PTM is the recombinant expression of the enzyme horseradish peroxidase, which will form the central focus of Chapter 3-5.

#### 1.2.4. Horseradish Peroxidase: A Target of Recombinant Expression

Horseradish peroxidase (HRP) is a haem binding glycoprotein, which functions in the breakdown of hydrogen peroxide to form water. HRP is a widely applied and studied enzyme<sup>54</sup>. The enzyme's primary use is as a reporter within commercial immunoassays<sup>55</sup>, which is a rapidly growing commercial market. Consequently, the enzyme's global market value is predicted to double from \$49 million to \$100 million within the next decade<sup>56</sup>. However, the utility of the enzyme is varied and its catalytic function has been utilised in many additional systems, including the detection of environmental contaminants, biofuel generation, and as a cancer therapy<sup>57-60</sup>.

The name "horseradish peroxidase" does not apply to a single enzyme but rather a broad group of peroxidase enzymes which are extracted from the roots of *Armoracia rusticana*. HRP is the most widely available commercial peroxidase due to the ease of growth of *Armoracia rusticana* and the abundance of the enzyme within plant roots. These enzymes collectively are implicated in various essential processes within the plants such as lignification, metabolism of reactive oxygen species, and cell wall metabolism<sup>61, 62</sup>. A total of 34 of these isoenzymes have been characterised and feature on the UniProt database (Table 1.3)<sup>63,64</sup>. Of

**Table 1.3. Amino Acid Percentage identity similarity of HRP isoform sequences, obtained from the NCBI database.**

	A2	A2A	C1A	C1B	C1C	C1D	C2	C3	N	E5
A2		98.36%	55.74%	55.74%	55.08%	55.41%	53.29%	55.92%	54.10%	57.57%
A2A	98.36%		55.18%	56.82%	56.17%	56.49%	54.22%	53.94%	55.23%	57.57%
C1A	55.74%	55.18%		88.95%	90.66%	89.52%	76.36%	68.67%	50.93%	69.61%
C1B	55.74%	56.82%	88.95%		94.26%	94.03%	75.91%	69.78%	49.10%	68.95%
C1C	55.08%	56.17%	90.66%	94.26%		99.10%	76.06%	71.43%	52.44%	69.93%
C1D	55.41%	56.49%	89.52%	94.03%	99.10%		76.36%	71.96%	51.55%	70.92%
C2	53.29%	54.22%	76.36%	75.91%	76.06%	76.36%		64.97%	50.15%	66.78%
C3	55.92%	53.94%	68.67%	69.78%	71.43%	71.96%	64.97%		50.47%	90.52%
N	54.10%	55.23%	50.93%	49.10%	52.44%	51.55%	50.15%	50.47%		51.31%
E5	57.57%	57.57%	69.61%	68.95%	69.93%	70.92%	66.78%	90.52%	51.31%	

these 34 isoenzymes many share significant similarities on both the N and C termini. Frequently, the presence of glutamine within the N-terminal region often accompanied by a preceding cleavage site (Table 1.4). This cleavage site will be explored in detail in Chapter 3. Meanwhile, the C terminal region of most characterised isoenzyme also features the presence of both a double valine and a double serine (Table 1.5).

**Table 1.4. The first 30 amino acids of a selection of characterised HRP. Sequences are obtained from the NCBI Database.**

Isoform Identity	Amino Acid Sequence on N-terminus	Cleavage site
A2	QLNATFYSGT CPNASAIVRS TIQQAFQSDT	No Cleavage
A2A	MAVTNLSTTC DGLFIISLLV IVSSLFGTSS A/Q	C31/32
C1A	MHFSSSTLF TCITLIPLVC LILHASLSDA/ Q	C28/29
C1B	MHSPSSTSFT WILITLGCLA FYASLSDAQ/L	C29/30
C1C	MLHASFSNAQ LTPTFYDNSC PNVSNIVRDI	No Cleavage
C1D	MHSPSSTSFT WATLITLGCL MLHASFSNA/Q	29/30
C2	MHSSSLIKL GFLLLLNVLS LSHA/QLSPSF	24/25
C3	MGFSPLISCS AMGALILSCL LLQASNSNA/Q	29/30
N	MKTQTKVMGG HVLLTVFTLC MLCSAVRA/QL	28/29
E5	QLRPDFYSRT CPSVFNIKN VIVDELQTD	No Cleavage

**Table 1.5. The final 30 amino acids of a selection of characterised HRP. Sequences are obtained from the NCBI Database.**

Isoform Identity	Amino Acid Sequence on N-terminus
A2	N M G N I S P L T G S N G E I R L D C K K V D G S
A2A	I N M G N I S P L T G S N G E I R L D C K K V N G S
C1A	C R V V N S N S L L H D M V E V V D F V S S M
C1B	V V N S N S L L H D I V E V V D F V S S M
C1C	R V V N S N S L L H D I V E V V D F V S S M
C1D	R V V N S N S L L H D I V E V V D F V S S M
C2	I R L N C R V V N S K P K I M D V V D T N D F A S S I
C3	T Q G E I R Q N C R V V N S R I R G M E N D D G V V S S I
N	F T C S M I R M G S L V N G A S G E V R T N C R V I N
E5	I R M G N L R P L T G T Q G E I R Q N C R V V N S R

Commercially, HRP is produced when the peroxidase enzyme is extracted from the roots of its native host species. This means of production commonly yields a heterologous assembly of isoforms, and their separation can involve complex downstream processing.

Growth of *A. rusticana* occurs over the globe at a range of locations with cultivars in temperate regions far from the enzyme processing centres in industrialised nations <sup>65</sup>. Furthermore, the expression of HRP isoforms can vary based on abiotic factors, such as the season and location of harvest <sup>66</sup>. These isoforms are encoded by different genes and vary in similarity with isoforms C1C and C1D sharing a 99 % amino acid percentage identity homology, whilst isoform C1C shares only 52 % homology with the isoform named HRP N <sup>67,68</sup> (Table 1.3). Due to the heterogeneity observed in isoform characteristics (Table 1.6) and the identity of residues within each active site <sup>69</sup>, a substantial degree of downstream processing is necessary to generate enzyme product <sup>62</sup>. To date, most studies into the recombinant expression of HRP<sup>60,70-72</sup> have focused on the expression of isoenzyme C1A HRP; however, commercially available enzyme products extracted from plant roots can feature as little as 75% enzyme from group C <sup>73,74</sup>. For the remainder of this thesis HRP isoenzyme, C1A will be referred to exclusively as HRP.

**Table 1.6. Physiochemical Diversity of Several HRP Isoforms Available on the UniProt Database**

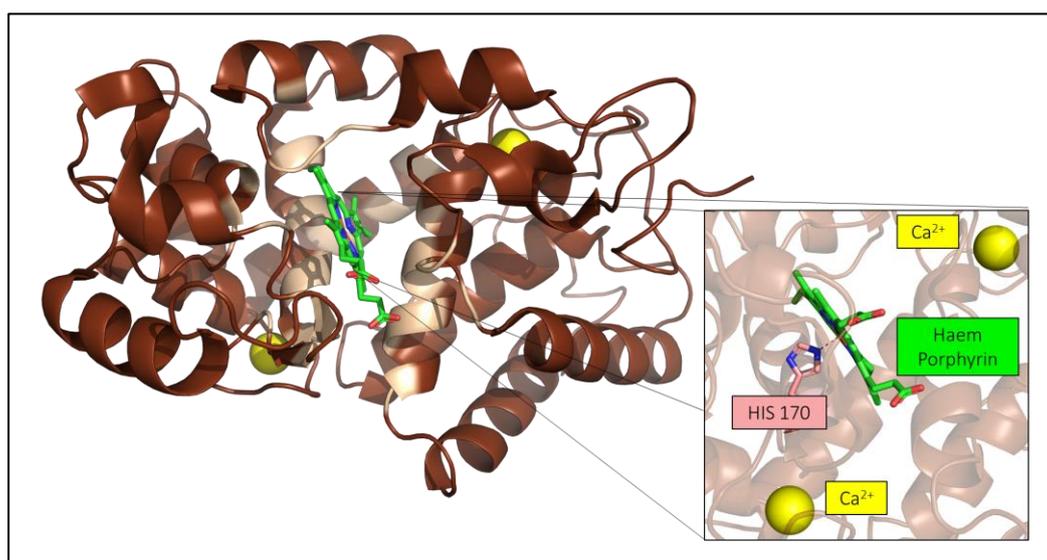
Isoenzyme Group	UNIProt Entry Code	Surface Charge	Length (Amino Acids)	Mass (Da):	pI
A2	P80679	Acidic	305	31,899	4.72
A2A	K7ZW28	Acidic	336	35,029	4.82
C1A	P00433	Neutral	353	38,825	5.67
C1B	P15232	Neutral	351	38,646	5.74
C1C	P15233	Neutral	332	36,548	6.21
C1D	K7ZW56	Neutral	352	38,782	6.98
C2	P17179	Neutral	347	38,035	8.70
C3	P17180	Neutral	349	38,180	7.50
N	Q42517	Neutral	327	35,126	7.48
E5	P59121	Basic	306	33,722	9.13

### 1.2.5. The Importance of PTM to HRP Expression

The complex downstream processing of native expression of HRP and its high commercial value make the enzyme an attractive target for recombinant production. Heterologous expression presents an opportunity to increase the speed and quantity of HRP production while also reducing both the carbon footprint and isoform heterogeneity of

enzyme product <sup>66</sup>. Therefore the PTM of HRP has significant implications for heterologous expression, and the choice of expression host selected dictates if several PTMs are applied to the enzyme (Table 1.7) <sup>66</sup>.

The first reported expression of recombinant HRP occurred 30 years ago by Smith *et al.* (1990). This study produced insoluble HRP expression sequestered in *E. coli* inclusion bodies <sup>75</sup>. Since then, many studies <sup>65,67,72,76–78</sup> have investigated alternative host organism strategies for HRP expression (Table 1.7), and there are now many techniques and species hosts which can emulate active HRP expression. HRP can be expressed in *E. coli* in under 24 hours; however, expressed protein is insoluble and must be refolded for use <sup>75</sup>. In comparison, HRP expressed in *P. pastoris* is soluble but typically hyperglycosylated with heterogeneous glycans to that of native expression <sup>79,80</sup>. Therefore, despite the range of microbial hosts available for HRP expression, no expression host comes without limitation <sup>24</sup>.



**Figure 1.3. Structure of Non-Glycosylated C1A HRP.**

HRP shown as cartoon (**Brown**) with the regions comprising the active site highlighted (**Light Brown**). Insert display components internal to the active site. Haem protoporphyrin IX is shown as sticks (**Green**) coordinated to HIS 170 (**Peach**) with both proximal and distal calcium ions shown as spheres (**Yellow**)(PDB 1H58).

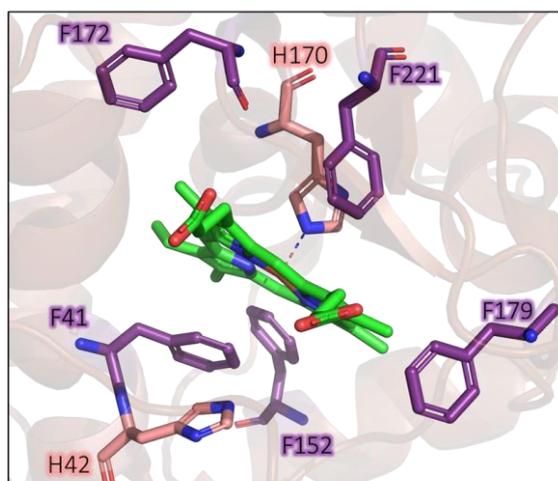
**Table 1.7. Previously Reported Recombinant HRP Expression.**

Expression Species	Strategy	Isoform Produced	Problem Presented	Study
<i>Armoracia rusticana</i>	Extraction of protein from hair root culture of native plants	Isoform Mix	Heterogeneous Product	Flocco <i>et al.</i> (1998) <sup>484</sup>
	Commercial protein production currently applied by Sigma Aldrich	Isoform mix	Only 75% of HRP produced is HRP – C type	Arnold and Zhanglin (2002) <sup>73</sup>
<i>Escherichia coli</i>	Refolding from inclusion bodies	HRP C	2-3% recovery of active protein	Smith <i>et al.</i> (1990) <sup>75</sup>
	Exporting to periplasm for soluble active protein production	HRP C1A	Only a fraction of protein is exported to periplasm	Gundinger and Spadiut (2017) <sup>79</sup>
<i>Saccharomyces cerevisiae</i>	Expression in eukaryotic host to enhance protein activity	HRP C	Decreased activity	Morawski <i>et al.</i> (2000) <sup>77</sup>
<i>Pichia pastoris</i>	Production of glycosylated HRP	Isoenzyme C1A	Hyperglycosylated protein	Spadiut <i>et al.</i> (2012) <sup>80</sup>
<i>Nicotiana tabacum</i>	Expression of protein with both N and C cleavage sequences to enhance production	10-fold increase in peroxidative activity of extract when N-terminal sequence is incorporated	Slower plant growth	Kis <i>et al.</i> (2004) <sup>76</sup>
<i>Nicotiana benthamiana</i>	Expression with both N and C cleavage constructs	Enhanced expression with N terminus present	Slower growth of plant to achieve high yield	Huddy <i>et al.</i> (2018) <sup>65</sup>
<i>Spodoptera frugiperda</i>	Production of soluble glycosylated protein	Unspecified	Low recovery and differences in accessory glycan units	Hartmann <i>et al.</i> (1992) <sup>485</sup>
Human T24 bladder carcinoma cells	Direct expression of HRP in cancerous cells as a targeted therapy	Unspecified peroxidase with cancer therapy potential	Direct cell therapy and not a source of commercial protein	Greco <i>et al.</i> (2000) <sup>486</sup>

HRP illustrates that while there are multiple effective strategies of expressing protein recombinantly, recombinant protein expression is not straightforward, and each expression host features a trade-off in PTM<sup>64,66</sup>. However, one clear achievement produced from recombinant HRP expression was the determination of the first crystal structure of HRP. This structure was produced from the non-glycosylated heterologous enzyme produced from *E. coli* cells and is an example of the insight which can be gleaned by recombinant protein expression, even in the absence of native PTM (Figure 1.3.)<sup>81,82</sup>.

#### 1.2.6. The Importance of Cofactor Binding of HRP in Recombinant Expression

When produced natively, HRP binds the tetrapyrrole cofactor protoporphyrin IX together with two Ca<sup>2+</sup> ions, to form *holo*-HRP<sup>81</sup>. The presence of haem enables *holo*-HRP to function in catalysis, while the affinity of nearby residues to the calcium ions increases the stability of enzyme<sup>83</sup>. Conjugation of HRP occurs by a coordinate bond at the five-coordinate haem position and the imidazole side chain of histidine 170<sup>82</sup>(Figure 1.4)<sup>84</sup>. Once bound, several further non-covalent interactions between the porphyrin and the surrounding amino acids then stabilise the localisation of haem within the active site<sup>84</sup> (Figure 1.4).



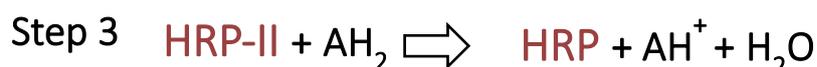
**Figure 1.4. Residues Within the Active Site of HRP.**

Labelled residues shown as sticks and coloured according to residue identity (F: Purple, H, Peach), with haem shown centrally (Green) (PDB 1H58).

Functionally, *holo*-HRP like all plant peroxidases oxidises an electron-rich substrate in a combined reaction with the decomposition of H<sub>2</sub>O<sub>2</sub> to form H<sub>2</sub>O (Equation 1.1)<sup>62</sup>. The enzyme catalyses the oxidation of a range of both chromogenic and luminescent substrates, such as the oxidative coupling of *o*-phenylenediamine (OPD) to form the red coloured 2,3-

diaminophenazine (DAP), or the breakdown of 5-amino-2,3-dihydro-1,4-phthalazinedione (luminol) to stimulate the emission of blue light <sup>85,86</sup>.

Equation 1.1. Turnover of Substrate by HRP in the Presence of Peroxide



Peroxidase action of HRP occurs in a 3-step process when the enzyme is first oxidised by  $\text{H}_2\text{O}_2$  and then later reduced back to native form by the simultaneous oxidation of two substrate molecules (Equation 1.1) <sup>84</sup>. Within the active site two key residues R38 and H42 initiate catalysis by peroxide activation <sup>87</sup>. This polarises the oxygen-oxygen bond within the peroxide molecule and initiates histidine mediate deprotonation of the peroxide <sup>82</sup>. The hydrogen atom from the peroxide is delivered via H42 to the terminal oxygen of the peroxide

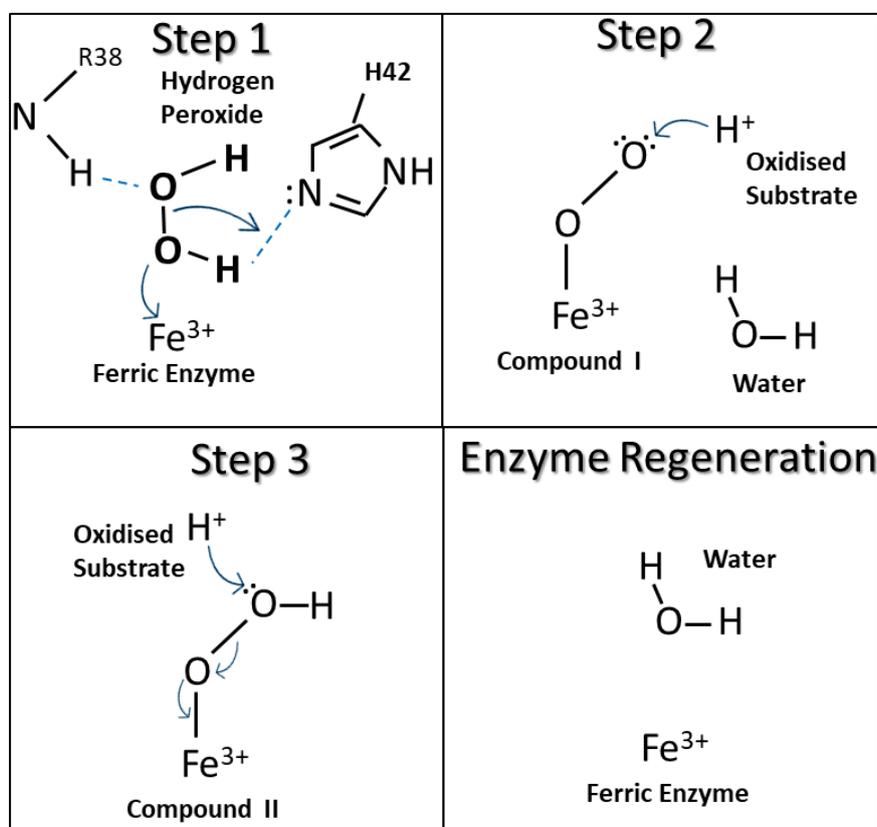


Figure 1.5. Mechanism of HRP activity adapted from Berglund *et al.* (2002) <sup>82</sup> and Azevedo *et al.* (2003) <sup>84</sup>.

---

(Step 1, Figure 1.15)<sup>88</sup>. This cleaves the oxygen backbone of peroxide and activates the enzyme to give Compound I and release water in the process (Figure 1.5)<sup>89</sup>.

Compound I is then able to oxidise a wide range of substrates including phenols, aromatic amines, thioanisoles and iodide, by single electron transfer and the formation of Compound II (Step 2, Figure 1.5). Compound II is then reduced by way of a second substrate molecule to form molecular water and return enzyme to its native ferric state (Step 3 Figure 1.5).

Critically, the expression of HRP in microbial cells results in a reduced activity of enzyme product, due to the low availability of haem porphyrin. Saturation of HRP with haem is often hindered in microbial cells as the haem biosynthesis pathway does not match that of protein synthesis<sup>90,91</sup>. Thus, the recombinant expression of HRP typically results in the expression of mostly *apo*-protein, which must be later haem saturated by the addition of haemin to purified protein<sup>66</sup>.

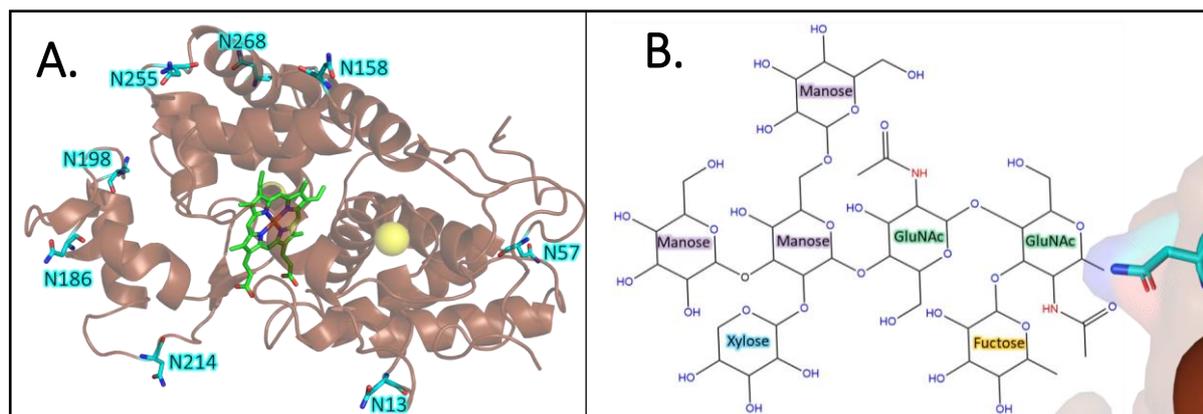
#### 1.2.7. The Influence of HRP Recombinant Expression on Glycosylation

Glycosylation is the addition of one or several carbohydrate chains to a protein<sup>92</sup>. The process is the most common PTM event in eukaryotes<sup>93,94</sup> and the process is controlled by the number and availability of glycosylase enzymes present within the cell<sup>95</sup>. Therefore, glycosylation is not regulated genetically and is instead variable between different yields of the same protein. Additionally, glycosylation can be heterogeneous between can be heterogeneous upon specific proteins or between successive protein batches, as is the case in glycans characterised upon HRP<sup>70,96</sup>

The presence of surface glycans exert a significant role in protein-protein recognition and, when modified, can both signal and stimulate disease<sup>97</sup>. For example, the hyper-glycosylation of haemoglobin HbA1c is a diagnostic marker for the occurrence of diabetes mellitus<sup>97</sup>. Whilst the modification to glycosylation of protein can trigger the expression of many glycans targeting antibodies in autoimmune diseases such as motor neuron disease<sup>98</sup>. Thus, the risk of the false recognition of glycans ultimately limits the use of glycosylated proteins, such as glycosylated HRP, as biotherapeutics<sup>51</sup>.

Glycosylation of HRP is significant and accounts for 20% of the protein's dry weight. Glycosylation occurs at eight surface asparagine residues: N13, N57, N158, N186, N198, N214,

N255, and N268 (Figure 1.6.)<sup>55,99</sup>. Several studies indicate that each glycan chain of HRP is identical in composition and weighs 1296 Da<sup>55,66</sup>. However, mass spectrometry performed by Gray *et al*, (1998) indicated that the glycans present on each N residue are not homologous and several possible carbohydrates can be applied<sup>100</sup>. No structure of glycosylated HRP, in either *apo* or *holo* form, currently exists in the PDB database. Therefore, the exact identity and structural influence of HRP glycans remain uncertain.



**Figure 1.6. N-Linked Glycosylation Pattern of HRP**

**A**, Protein shown as cartoon (**Brown**) whilst each N-linked glycosylation site shown as sticks and labelled by residue (**Teal**). **B**, representation of a single glycan unit typically attached to HRP.

The influence of glycosylation on the activity of HRP is complex. Evidence indicates that the removal of glycans from plant-produced protein exerts a negative influence upon HRP catalysis but has little influence on the thermostability of the enzyme<sup>101</sup>. However, mutagenesis of non-glycosylated asparagine residues for recombinant expression appears to reverse these adverse effects and increase both substrate turnover and thermostability<sup>54,63,70</sup>, which indicates that glycosylation is neither essential to protein expression or activity<sup>75</sup>. Furthermore, mutagenesis of recombinant non-glycosylated HRP to remove non-glycosylated asparagine residues increases both stability and activity of the enzyme<sup>63</sup>. This effect is not common in recombinant expression but removal of surface asparagines is also reported to improve stability of lipase proteins by reducing the occurrence of de-amination<sup>102</sup>.

### 1.2.8. The Implications of HRP Proteolytic Cleavage for Recombinant Expression

Proteolytic cleavage is an irreversible PTM, whereby a section of amino acids are cleaved after expression<sup>103–105</sup>. In protein expression, the inclusion of native termini elements which are then cleaved is essential for the maturation of many proteins. For example, the

---

formation of insulin is reliant on the cleavage acidic residues on the N-terminal pre-peptide of pro-insulin during protein maturation for correct protein folding<sup>106,107</sup>. This is important as recombinant protein expression often relies on the inclusion of heterologous cleavage sites for the controlled separation of specified regions<sup>18</sup>.

The genetic sequence of C1A HRP, first described by Fujiyama *et al.* (1990)<sup>68</sup>, contains both N and C-terminal peptide sequences, which are cleaved during protein maturation<sup>66</sup>. The N-terminal peptide localises expressed HRP to the endoplasmic reticulum in plants<sup>76</sup>, whereby three crucial PTM reactions can occur (N-linked glycosylation, disulphide bond formation, and the removal of the N-terminus)<sup>108</sup>. In comparison, the exact function of the C-terminus sequence is uncertain, but it is not essential for the formation of active protein<sup>76</sup>. To date, no study has attempted to express HRP with either the N or C-terminus intact within prokaryotic cells<sup>75</sup>. Instead, prokaryotic gene constructs of HRP C1A have omitted the first 30 amino acids on the N-terminus of HRP and the final 15 amino acids on the C-terminus, to express only residues which are present in mature HRP<sup>75,99,109</sup>.

### 1.2.9. Disulphide Bond Formation in Recombinant Expression

Disulphide bonds form when sulfhydryl groups of two neighbouring cysteine residues react together, forming a covalent linkage<sup>110</sup>. Commonly the bonds form post-translationally for the stable attachment of two regions within the same polypeptide or between polypeptide subunits<sup>111</sup>. However, the bond can be broken in reducing conditions, such as those observed within a cell's cytosol. Consequently, disulphide bonds are scarce in cytosolic proteins<sup>112</sup>, but can have significant implications for the cytosolic expression of recombinant proteins.

HRP contains four disulphide bonds (C11-C91, C44-C49, C97-C301, and C177-C209) (Figure 1.7.). Therefore, recombinant expression of HRP within *E.coli* must account for the presence of these four disulphide bonds to generate natively folded protein. Several previous methods of disulphide bond formation in *E. coli* expressed HRP have been explored, such as, the export of expressed protein to the periplasm<sup>79</sup>, and the overexpression of the enzyme disulphide isomerase in combination with HRP<sup>113</sup>. However, the most common form of HRP expression in *E.coli* features the formation of disulphide bonds during recovery of protein from the insoluble cell fraction. This typically occurs *via* the breaking of bonds in a reducing unfolding

---

buffer and then titration of protein back into an oxidising refolding solution, although this typically results in a low yield of active protein <sup>64,75,99,114</sup>.

---

## 1.3. Protein Engineering

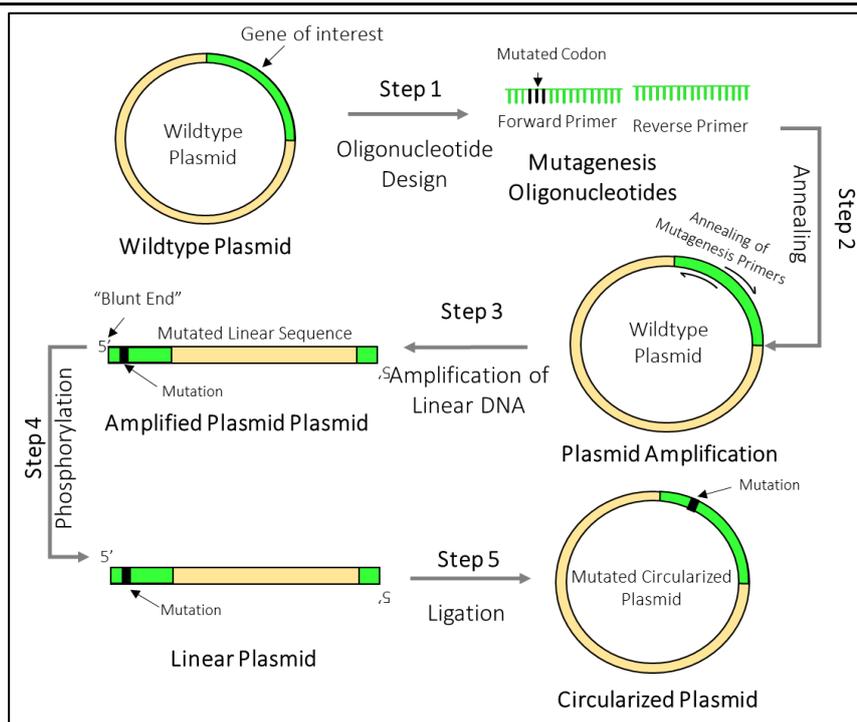
### 1.3.1. A Historical Perspective of Protein Engineering

Genetic adaptation in response to a selective pressure is slow, even in organisms with the fastest generation times <sup>115</sup>. Modern demand for improved protein function for specific technological applications or even for fundamental studies of the structure-function relationship is greater than that which is feasible through evolution alone <sup>116,117</sup>. Hutchison *et al.* (1978) first described the process by which an intentional change was introduced to the sequence of a bacteriophage at a specified point, in the first example of genetic protein engineering <sup>118</sup>. Therefore, protein engineering is a process by which desired changes can be introduced to a protein's amino acid sequence which subsequently modifies its structural or functional characteristics <sup>119</sup>.

Initially, protein engineering began small, with the selection of specific amino acids for substitution <sup>120</sup>, insertion <sup>121</sup>, or deletion <sup>122</sup>. However, as these methods proved successful, the process expanded to target larger regions of protein and later, the generation of mutant protein libraries <sup>123–125</sup>. The latter has a basis in the natural evolutionary process (random mutations followed by selection) however the process of engineering change forms a key process in directed evolution <sup>126–128</sup>. Modern protein engineering now allows for the substitution of entire protein domains <sup>129</sup>, *de novo* protein assembly <sup>130</sup>, and selective induction of non-native PTM to a protein <sup>131</sup>.

### 1.3.2. Protein Mutagenesis

Current approaches and technologies allow for the genetic modification of a protein to be performed quickly <sup>132</sup>, at low cost <sup>133</sup>, and with high fidelity <sup>134</sup> to the existing genetic template <sup>129</sup>. The technique is based on the amplification of a template sequence by a polymerase chain reaction, using mutagenic oligonucleotides to introduce the desired mutations into the amplified DNA <sup>135</sup>. The process can be used randomly over many repeated cycles of non-faithful replication to induce generational diversity by error-prone PCR<sup>136</sup>. Conversely, the process can be site-specific and exert specific, targeted modification to the genetic sequence, termed site-directed mutagenesis (Schematic 1.2) <sup>137</sup>. This introduction will focus on site-specific mutation approaches, as this is the primary approach used in this thesis.



**Schematic 1.2. General Protocol for The Site Directed Mutagenesis**

Site-directed mutagenesis (SDM) is the method by which intentional, targeted change to genetic code is achieved. Typically, missense mutations are incorporated so as they change the amino acid sequence of the target protein, but more recently nonsense mutations have been used to code for the incorporation of non-natural amino acids through codon reprogramming approaches (see Section 1.4.2). In SDM, two short-chain DNA oligonucleotides of  $\sim 30$  bases in length are synthesised which are modified at a specific location from the original gene sequence. These alterations are situated at a desired location and reflect the specified mutations which are desired within the target sequence (Schematic 1.2, Step 2). Consequently, when PCR is used to amplify the template using these mutagenic oligonucleotides to initiate replication the sequence modification is ensured (Schematic 1.2, Step 3). This technique forms the basis for a variety of approaches for SDM, including splice-by-overlap PCR (recombination without reliance on restriction sites)<sup>138</sup>, Quickchange<sup>TM</sup> (mutagenesis of short gene regions)<sup>139</sup>, and whole plasmid (inverse) mutagenic PCR<sup>140</sup>. In this thesis whole plasmid PCR will be used for the generation of protein mutants.

### 1.3.3. Green Fluorescent Protein: A Model for Protein Engineering

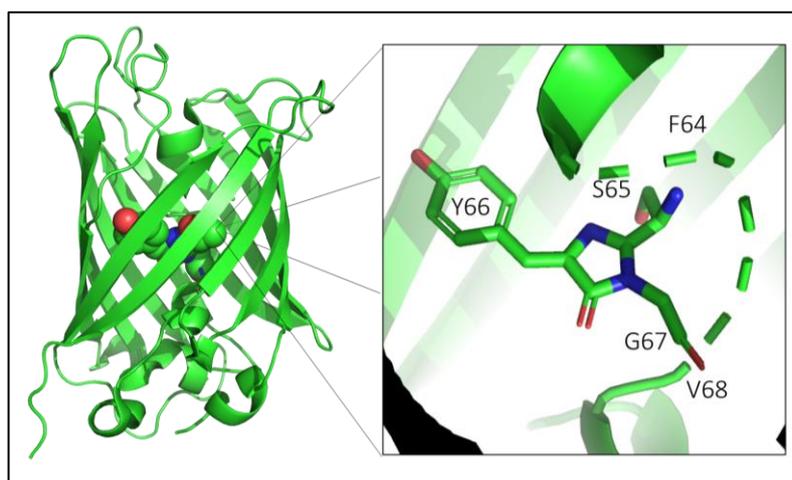
The fluorescent protein (FP), Green Fluorescent Protein (GFP<sup>WT</sup>), is an important example of the significance and utility of protein engineering. GFP<sup>WT</sup> was first isolated in 1962 by Shimomura *et al.* from the jellyfish *Aequorea victoria*<sup>141</sup>. The native function of GFP<sup>WT</sup> is to

enhance the emission efficiency of light from the molecule aequorin, absorbing blue light emitted from aequorin and increasing the wavelength of emission into the green region <sup>142</sup>. GFP<sup>WT</sup> is inherently fluorescent with no requirement of additional small-molecule cofactors <sup>143</sup>. Under aerobic conditions, all the information needed for the protein to become fluorescent is encoded within the amino acid sequence. Fluorescent function is therefore possible in a wide range of recombinant organisms and this function has therefore enabled GFP<sup>WT</sup> to become the first widely used, genetically encoded, fluorescent probe, for cell imaging *in situ*.

The application of GFP<sup>WT</sup> has revolutionised our ability to monitor specific processes within the cell from gene expression <sup>144</sup> to protein localisation <sup>145</sup>, and has been utilised both on earth <sup>146</sup> and in space <sup>147</sup>. As a result of its importance to the life sciences, the Nobel Prize for Chemistry was awarded for the discovery and development of the molecule in 2008 <sup>148</sup>.

#### 1.3.4. The Relationship Between Structure and Function of GFP<sup>WT</sup>

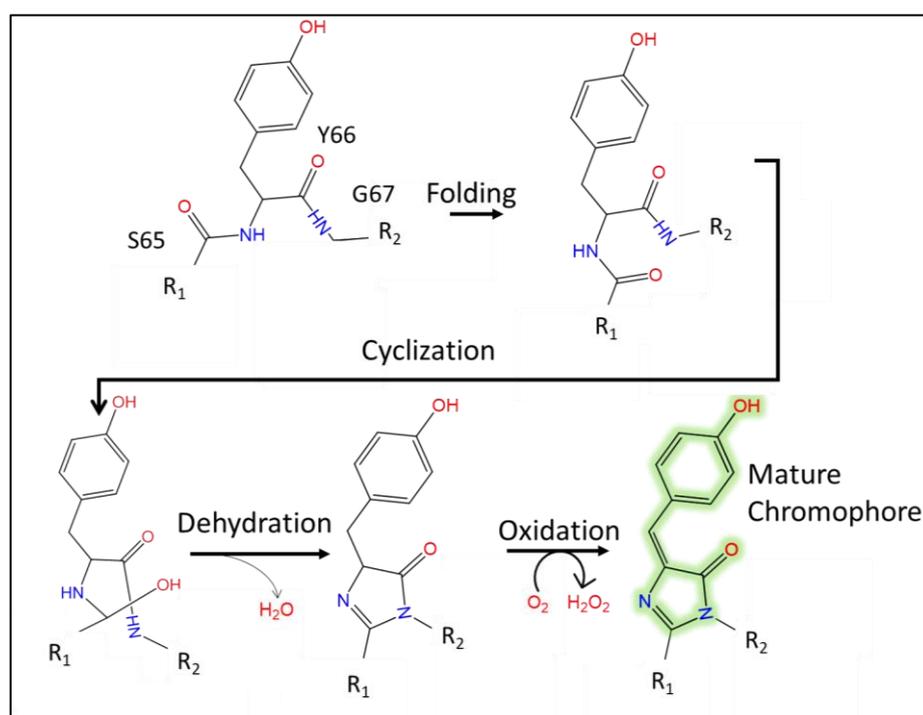
Structurally, GFP<sup>WT</sup> consists of 11  $\beta$ -strands, arranged to form a curved sheet termed a  $\beta$ -barrel/ $\beta$ -can that surrounds a central  $\alpha$ -helical core <sup>149</sup> (Figure 1.8.). Isolation of GFP's chromophore from the surrounding solvent by the  $\beta$ -barrel is critical to its function <sup>150</sup>. Unfolded protein is not fluorescent, due to the immediate transfer of excitation energy from the chromophore to the solvent <sup>151</sup>. Thus, the chromophore function is dependent on solvent shielding by the  $\beta$ -barrel <sup>152</sup>.



**Figure 1.8. Structure of GFP<sup>WT</sup>.**

The cartoon structure of sfGFP shown with central chromophore represented as spheres. Insert shows the chromophore (S65, Y66, G67) as sticks coloured by element (PDB 1GFL).

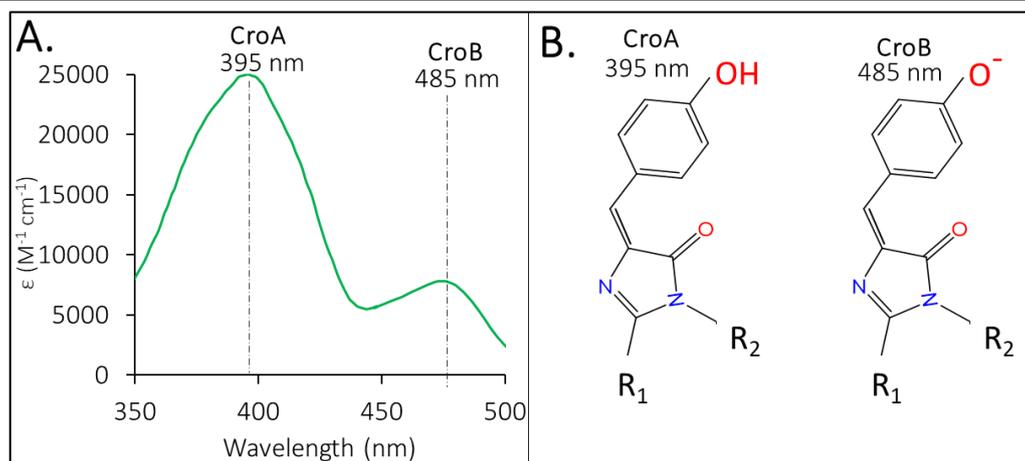
The chromophore is formed by the covalent rearrangement of three contiguous amino acids, 65-Ser-Tyr-Gly-67 in the presence of  $O_2$ , to generate *p*-hydroxybenzylideneimidazolinone<sup>153</sup> (Figure 1.8, Insert). The first stage in chromophore maturation is folding of the polypeptide chain to a near-native conformation, bringing S65 and G67 into proximity (Schematic 1.3). Then the carbonyl group of S65 forms the imidazolinone by reaction with the amine group of G67 and the release of molecular water. The final oxidation step requires molecular oxygen and brings the phenol ring from Y66 into conjugation with the rest of the system to form a comprehensive delocalised system of electrons<sup>154</sup>. However, one alternative description of chromophore maturation proposed is that the order of oxidation and dehydration steps are switched, with oxidation believed by some<sup>155,156</sup> to occur before dehydration.



**Schematic 1.3. Maturation of sfGFP chromophore.**

Covalent rearrangement between S65 and G67 to form mature chromophore as described by Barondeau *et al* (2005)<sup>487</sup>.

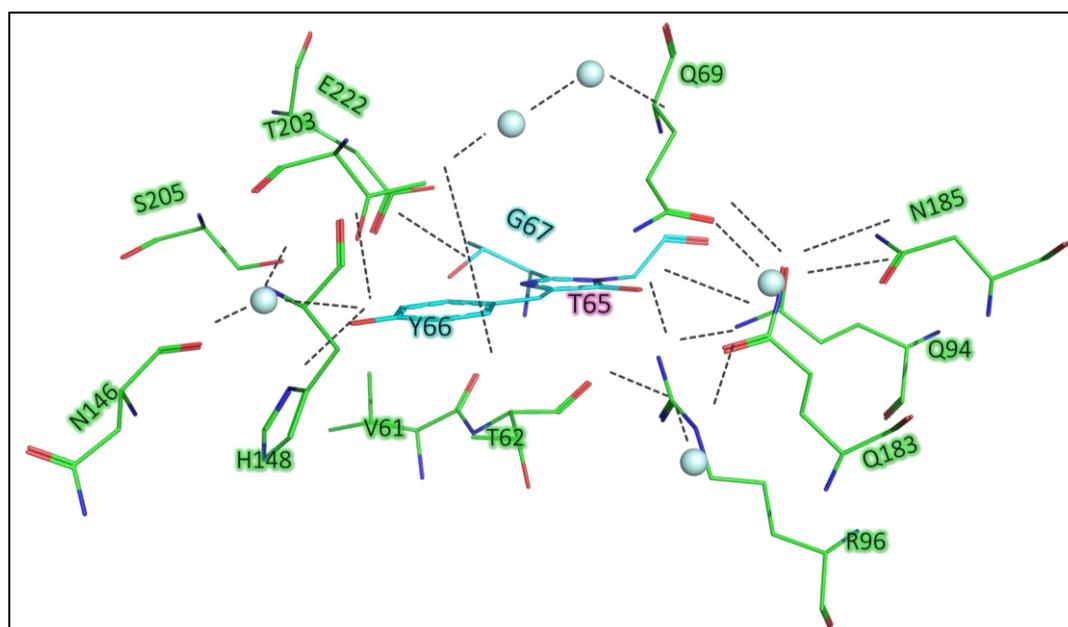
Due to the occurrence of two potential protonation states on the ionisable phenol OH group of Y66, the chromophore of GFP<sup>WT</sup> can exist in two possible forms, named CroA and CroB<sup>154</sup> (Figure 1.9). CroA is the neutrally protonated chromophore which is excited at 395 nm. Whilst CroB is the charged phenolate which excites at 475 nm. In GFP<sup>WT</sup>, the chromophore exists predominantly as the protonated CroA form at a ratio of 6:1 with the unprotonated CroB<sup>157</sup>.



**Figure 1.9. Absorbance Spectrum of GFP<sup>WT</sup>**

**A**, the molar extinction coefficient of GFP<sup>WT</sup>, with absorbance peaks at both 395 nm and 485 nm labelled in accordance with chromophore identity. Spectrum obtained from [www.fpbse.org](http://www.fpbse.org). **B**, the protonation state of Y66 within both CroA and CroB<sup>488</sup>.

Surrounding the chromophore, are several residues (V61, T62, Q69, Q94, R96, N146, H148, R168, Q183, T203, S205 and E222) common to both sfGFP and WT GFP, which determine the vibrational relaxation of excited electrons. This then determines the point from which electrons return to their ground state by the emission of light and therefore controls the fluorescent properties of each chromophore (Figure 1.10).<sup>154</sup> Residues H148, T203, and S205 form hydrogen bonds directly with the hydroxyl of the chromophore, whilst R96 and Q94 both



**Figure 1.10. Polar Contacts of sfGFP Chromophore.**

Mature chromophore (Teal) with adjacent polar contacts shown; residues shown as lines (Green) and water shown as spheres (Pale Cyan). T65 (Pink) is shown as this is common to sfGFP, in the case of WT GFP this amino acid is S65. Hydrogen bonds represented by dashed lines. Adapted from Tsien *et al* (2008)<sup>154</sup> (PDB 2B3P).

bond with the imidazolidinone ring<sup>153</sup>. Final polar contacts are formed when E222, R96, Q183, and T62, interact directly with the chromophore<sup>158</sup>. Mutagenesis of these key residues is a known stimulus of functional change in GFP, and is used frequently in the engineering of new GFP variants<sup>159</sup>.

### 1.3.5. Generation of GFP Variants by Protein Engineering

While the cloning of GFP<sup>WT</sup> was a significant scientific discovery, it soon became clear that the protein had inherent limitations including its low stability<sup>160</sup>, slow folding kinetics at 37°C<sup>161</sup>, tendency to dimerise at high concentration<sup>162</sup> and slow chromophore maturation<sup>163,157</sup>. Thus, GFP<sup>WT</sup> became a target for protein engineering.

Introduction of a single point mutation, F64L, increased GFP<sup>WT</sup>'s folding efficiency at 37 °C by increasing the protein's solubility at high concentration<sup>145</sup>. Whilst a subsequent single point mutation, S65T, resulted in the suppression of the neutral chromophore, CroA, and favoured excitation at 485 nm over 395 nm<sup>164</sup>. The deprotonation of the chromophore maintained fluorescent function but enhanced both the emission intensity and photostability of GFP fluorescence<sup>164</sup>. The combination of both mutations formed a double mutant protein dubbed "enhanced GFP" (EGFP) which exhibited a further 35-fold increase in fluorescence brightness in comparison to GFP<sup>WT</sup><sup>165</sup> (Table 1.6). Therefore, EGFP has one excitation maximum (485 nm)<sup>166</sup> whilst GFP<sup>WT</sup> has two distinct excitation maxima which reflect the two protonation states of Y66 (Figure 1.9)(Table 1.8). Although, *in situ* in *Aequorea victoria*, EGFP like GFP<sup>WT</sup> is weakly homodimeric<sup>167</sup>.

Building on the mutations introduced in the formation of EGFP, directed evolution was used to generate a further structurally homologous fluorescent protein termed "super folding" GFP (sfGFP)<sup>168</sup> (Table 1.8). This variant was modified from the starting protein at several key

**Table 1.8. Spectral Properties of Popular GFP Variants Obtained from FPbase.org**

Variant	Excitation	Emission	Molar Extinction Coefficient ( $\epsilon$ )	Quantum Yield
GFP <sup>WT</sup>	395 nm/475 nm	508 nm	25,000 M <sup>-1</sup> cm <sup>-1</sup> /7,000 M <sup>-1</sup> cm <sup>-1</sup> .	0.79 $\Phi$
EGFP	488 nm	507 nm	55,9000 M <sup>-1</sup> cm <sup>-1</sup> .	0.6 $\Phi$
sfGFP	485 nm	510 nm	83,300 M <sup>-1</sup> cm <sup>-1</sup> *	0.65 $\Phi$

\*the reported molar extinction coefficient of sfGFP has been repeatedly measured at 49,000 M<sup>-1</sup> cm<sup>-1</sup> by the Jones group<sup>431</sup>

---

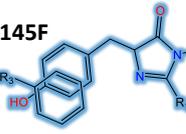
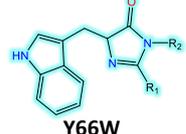
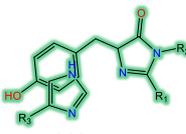
residues outside of the chromophore environment (S30R, F64L, Y99S, N105T, T145F, M153T, V163A, I171V, and A206V), and displayed improved folding kinetics and stability together with an increased chromophore maturation rate<sup>169</sup>. This thesis will utilise sfGFP as the GFP variant of choice in Chapters 6 and 7.

### 1.3.6. The Mutagenesis of GFP to Generate Functional Change

The interactions between the chromophore and surrounding residues (Figure 1.10) directly influence the fluorescent properties of GFP variants, including sfGFP (Table 1.9). The mutation of these residues alters the chromophore environment, and thus generates a spectral change in the protein. For example, breaking the polar contact between Y66 and H148 ensures the chromophore favours the protonated CroA after maturation (Figure 1.9)<sup>170</sup>. This, in turn, promotes the excitation of the chromophore, at 395 nm, and reverts both EGFP and sfGFP to the spectral characteristics of GFP<sup>WT</sup> due to the presence of both CroA and CroB present (excitation at both 395 nm and 485 nm)<sup>164,171</sup>.

Meanwhile the T203Y mutant initiates  $\pi$ - $\pi$  electron interaction (stacking) with the chromophore which red-shifts excitation and emission so that the chromophore is excited at 514 nm and emits at 524 nm (Table 1.9) and is a key mutation in the formation of the yellow fluorescent protein Venus. The direct substitution of Y66 with other aromatic amino acids such as W, F, or H modifies the conjugated double bond network of the chromophore can either lengthen or shorten the path over which delocalised electrons can travel across the chromophore. Modification to the network of delocalised electrons then modifies the vibrational relaxation of the excited chromophore and varies the point from which electrons relax back to ground state. The difference in electronic relaxation therein varies the wavelength of light which is emitted from the chromophore and enables the formation of several unique fluorescent protein variants, such as blue fluorescent protein (BFP), and cyan fluorescent protein (CFP)(Table 1.9). Therefore, GFP<sup>WT</sup> and derived fluorescent proteins form a broad family of spectroscopically distinct molecular probes which consist of over 200 unique FPs<sup>172</sup>.

**Table 1.9. A Selection of Colour Variants derived from GFP (data obtained from FPbase.org)**

Variant Name	Mutations to form Variant	Excitation (nm)	Emission (nm)	Chromophore Arrangement	Study
Sirius	T65Q, Y66F, Y145G, H148S and T203V	<b>355</b>	<b>424</b>	 <b>Y66F</b>	489
BFP	Y66H	<b>382</b>	<b>448</b>	 <b>Y66H</b>	490
Sapphire	S72A, Y145F, T203I, H23L	<b>399</b>	<b>511</b>	 <b>Y145F</b>	491
CFP	Y66W	<b>456</b>	<b>480</b>	 <b>Y66W</b>	490
Clover	S30R, Y39N, S65G, Q69A, F99S, N105T, Y145F, M1553T, V163A, I171V, T203H	<b>505</b>	<b>515</b>	 <b>T203H</b>	492
Venus	S65G, V68L, S72A, T203Y, H231L, F64L, M153T, V163A, S175G, and F46L	<b>515</b>	<b>528</b>	 <b>T203Y</b>	493

### 1.3.7. The Use of GFP As A Probe of Protein Interaction

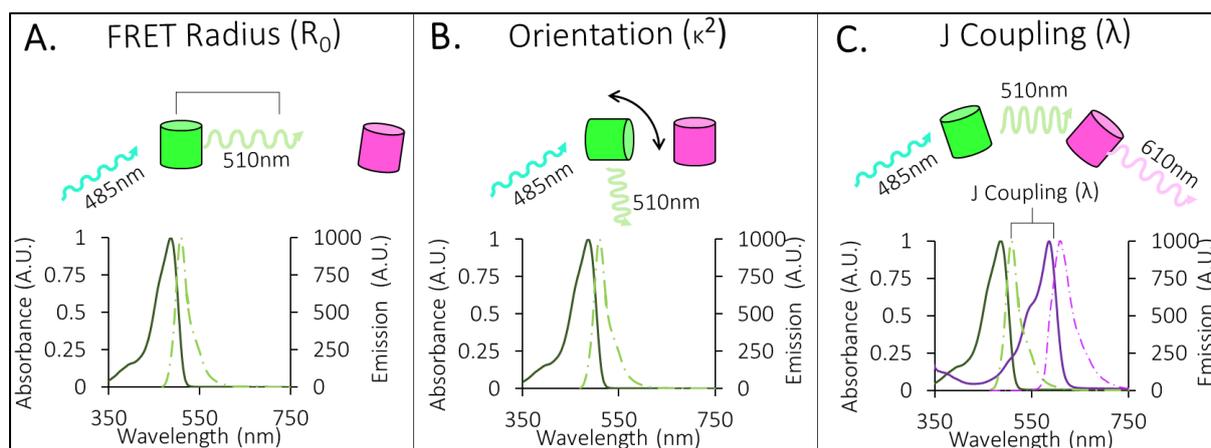
The relatively small size of GFP (238 amino acids) makes it an ideal fluorescent probe for protein fusion<sup>173</sup>. At the same time, the generation of spectroscopically distinct variants of GFP (see Table 1.9) allows for the tagging of multiple proteins simultaneously. GFP<sup>WT</sup> and its derivative variants can be attached to many molecular targets. By far the most common form of molecular attachment is the direct fusion of the FP to the N- or C-terminal of the target protein<sup>174</sup> although other approaches are available (e.g. *via* disulphide bonding<sup>175</sup>, or domain insertion<sup>176</sup>).

The formation of fusion proteins is not always without issue, as both folding kinetics and weak interactions, that are intensified by high local concentrations of FP on complexation

result in fluorescent artefacts within analysis. With regards to the folding kinetics, the development of proteins such as sfGFP has benefited the formation of many fusion proteins. Reducing GFP's tendency to oligomerise has additionally, been addressed through mutating residues thought to promote weak intermolecular interaction (e.g. A206K<sup>162</sup>) and reducing the protein's tendency to align at this interface. Reducing sfGFP's tendency to oligomerise further enables fusion to potential target proteins of interest which enables approaches, such as live-cell fluorescent imaging<sup>177</sup>. Furthermore, the combination two spectroscopically distinct fluorescent proteins within the same system and facilitates insight into protein association, *via* fluorescent probe functional interaction<sup>159</sup>.

### 1.3.8. Förster Resonance Energy Transfer

Fluorescent proteins can report on the dynamic nature of biomolecular interactions, especially protein-protein interactions, through a process known as Förster resonance energy transfer (FRET). FRET allows for non-radiative energy transfer from the chromophore of the donor FP to that of an acceptor FP chromophore by dipole-dipole coupling. Essential to the occurrence of FRET is the proximity, compatible orientation, and spectral overlap of fluorescent chromophores (Figure 1.11). Energy transfer occurs when chromophores are proximal ( $\sim 8$ - 10 nm), chromophore dipoles are suitably aligned, and the emission of the donor chromophore overlaps with the excitation of the acceptor<sup>178</sup>.



**Figure 1.11. Representation of Variables Which Influence FRET Transmission Between Chromophores.**

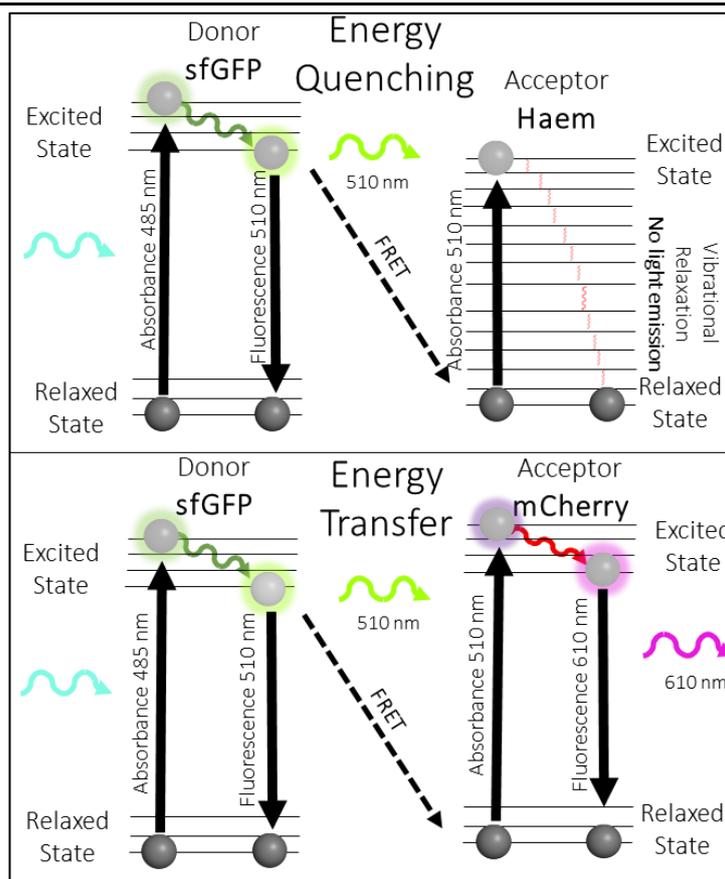
**A**, FRET Radius ( $R_0$ ), the distance between chromophores at which 50% energy transfer occurs. **B**, Orientation ( $\kappa^2$ ), the orientation at which emission wavelengths interact with the acceptor chromophore. **C**, J Coupling ( $\lambda$ ), if the emission of the donor chromophore overlaps with emission of the acceptor FRET is observed.

---

The distance at which 50% of the donor chromophore emission initiates the excitation of the acceptor, is defined as the FRET radius ( $R_0$ )(Figure 1.11, A). However, if the two chromophores are not aligned in compatible orientations, the second chromophore cannot be excited to maximum efficiency (Figure 1.11, B) <sup>179</sup>. Thus, the orientation factor ( $\kappa^2$ ) is an important influence of FRET efficiency observed. Furthermore, the extent of spectral overlap between the emission of the donor chromophore and the absorbance of the acceptor chromophore is proportional to the energy transfer observed and is known as the J coupling constant ( $\lambda$ ). Therefore, the greater the spectral overlap observed the greater potential for energy transfer (Figure 1.11, C). For example, the combination of sfGFP ( $\lambda_{EX}$  485nm  $\lambda_{EM}$  510 nm) and the DsRed derived mCherry ( $\lambda_{EX}$  587 nm  $\lambda_{EM}$  610 nm)(see Chapter 6.1.3)<sup>180</sup>, results in the emission of mCherry at 610 nm when the pair are excited at 485 nm (Figure 1.11, C).

FRET does not occur exclusively between two fluorescent chromophores, however, and the principles of FRET can also be used to describe the energy transfer observed in fluorescence quenching. FRET quenching is a non-productive form of energy transfer in which excited electrons released in GFP emission promote the excitation of electrons within a secondary non-fluorescent chromophore, such as haem porphyrin which is then lost a vibrational relaxation across the network of conjugated double bonds (Figure 1.12, A). As the secondary chromophore cannot emit photons, excited electrons are transferred across the tetrapyrrole ring. Subsequently, the excitation energy is released to the environment in the form of heat, which returns electrons to their ground state. As the fluorescence energy transferred to haem porphyrin is not emitted in the form of light, the fluorescence of GFP is quenched in the presence of haem.

This transfer of energy follows the same FRET principles, as is observed between fluorescent proteins (Figure 1.12, B), and is dependent upon distance, orientation, and the spectral overlap of chromophores. Consequently, the spectral overlap between emission of GFP<sup>WT</sup>, and GFP variants, with that of haem porphyrin within haemoglobin (Figure 1.2), often results in the reduction, or quenching, of GFP emission <sup>181</sup>. Therefore, energy transfer via FRET between GFP and haem limits the *in vivo* application of GFP. This limitation of GFP has stimulated mutagenesis efforts for the generation of redshifted variants which do not spectroscopically overlap with haemoglobin absorbance, thereby reducing emission quenching



**Figure 1.12.** Jablonski diagram for comparison of resonance energy transfer of GFP.

**A**, light energy emitted from sfGFP at 510 nm is absorbed by haem which cannot emit light, so relaxes back to ground state by vibrational relaxation across the porphyrin ring, resulting in a quenching of emission at 510 nm. **B**, light energy emitted by sfGFP at 510 nm is used to feed the excitation of mCherry which results in a larger Stokes shift in emission from 510 nm to 610 nm.

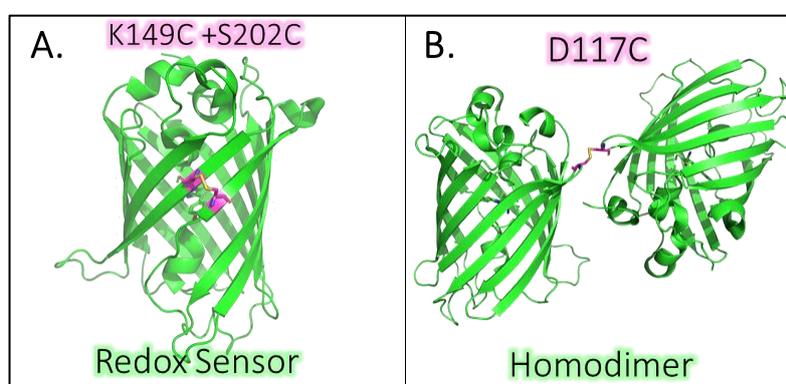
### 1.3.9. Emulation of Natural PTM to Modify GFP Structure and Function

While many FPs are used as passive probes in protein fusions, the mutagenesis of GFP has formed several proteins that can be modified post-translationally to alter their intrinsic structural and functional properties. Subsequently, this modification has produced a series of active fluorescent probes which can be used as functional biosensors. These biosensing fluorescent proteins respond to a post-translational external stimulus which modifies their inherent characteristics. This modification can, in turn, infer the localised environment of the FP, and to date, has been employed to induce GFP sensitivity to: calcium<sup>182</sup>, voltage<sup>183</sup>, chloride<sup>184</sup>, reducing conditions<sup>185</sup>, and pH<sup>186</sup>.

One such example of post-translational modulation of function is the mutagenesis of two GFP surface residues, K149C and S202C (Figure 1.13, A). In this example, the disulphide

bridge between these two residues under oxidising conditions results in a decrease in GFP emission at 510 nm<sup>187</sup>. However, when the bond is broken in reducing conditions, fluorescence is enhanced. Therefore, environmental oxidising conditions can be inferred *via* measured fluorescence output.

Alternatively, the mutagenesis of the surface residue D117C induces the formation of a homodimeric interface between two identical monomeric units, which can be broken in reducing conditions<sup>188</sup>(Figure 1.13, B). This mutagenesis triggers the formation of a fluorescent protein oligomer by the formation of a sulphhydryl bridge<sup>189</sup>. This in turn modifies the protein's quaternary structure in a way which is sensitive to environmental reducing conditions.



**Figure 1.13. Modification of GFP by Surface Cysteine Mutagenesis**

**A**, formation of redox sensitive GFP probe by linking of residues K149 and S202 (**Pink**) *via* cysteine mutagenesis and disulphide formation (PDB 1H6R). **B**, formation of a redox sensitive fluorescent dimer by mutagenesis of D117C and disulphide bond formation (**Pink**) (PDB 4W6J).

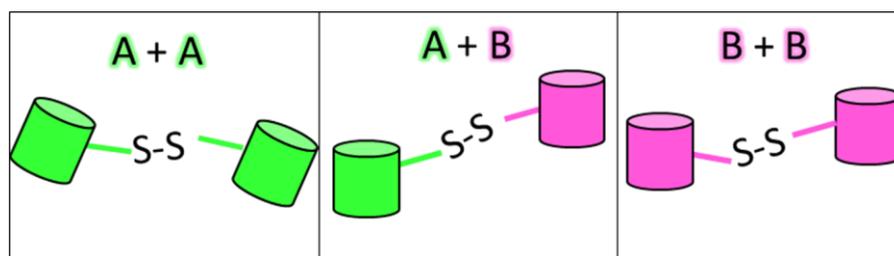
### 1.3.10. The Application of PTM to Induce GFP Oligomerisation

Oligomerisation typically allows for the linkage of two disparate proteins and can be achieved numerous ways by the application of PTM. In nature, oligomers can arise from a variety of natural PTM strategies, including hydrophobic interaction<sup>190</sup>, electrostatic attraction<sup>191</sup>, metal cofactor sharing<sup>192</sup>, disulphide bridging<sup>188</sup>, and hydrogen bonding<sup>193</sup>. These strategies can be utilised in the engineering of protein<sup>188,194–200</sup> and allow for the combination of protein functions which are not currently paired in nature<sup>176,201–203</sup>.

Post-translational oligomerisation is of research significance as proteins rarely act in isolation in nature<sup>204</sup>. It is estimated that approximately 35% of all proteins in a cell are oligomeric<sup>205</sup>, and, approximately two-thirds of human enzymes are thought to be multimers *in situ*<sup>206</sup>. As a result, it has been reasoned that oligomerisation is the simplest mechanism available to regulate localised protein concentration<sup>201</sup>, active site orientation<sup>205,207</sup>, and

reduce a protein's surface area conferring a greater resistance to both denaturation and degradation<sup>205,208</sup>. However, the engineering of protein often disregards the significance of interaction on protein function, in favour of generating mechanistically simpler protein monomers<sup>205</sup>. For example, in nature, many fluorescent proteins such as GFP<sup>WT</sup> and DsRed, exist as oligomeric assemblies<sup>154,209</sup>, however, each has been engineered to form a monomeric fluorescent probe<sup>168,180</sup>.

Nevertheless, two significant limitations exist in the application of natural PTM in the formation of engineered protein oligomers<sup>189</sup>. The first issue is that natural chemistry, such as the thiols in disulphide bonds, is pervasive in natural systems<sup>210</sup>. Therefore, incorporation of natural chemistry can stimulate cross-reactivity of the target protein to others, or an intramolecular mismatch of disulphide bonds. Consequently, it is feasible that for every heterodimeric complex formed two unintentional dimeric waste products will be additionally generated (Figure 1.14)<sup>189</sup>. This convolutes molecular assembly, and therefore, limits the formation of complex oligomers.



**Figure 1.14. The Lack of Control of Heterodimeric Linkage Offered by Cysteine Mutagenesis.** The three potential dimeric associations which can arise when disulphide linking two protein monomers.

The second limitation of disulphide linkage is that association is often temporary and in the case of thiol linkage, environmentally sensitive. This sensitivity can be of benefit, as dimerisation can be used to infer environmental conditions, such as oxidation<sup>188</sup>. However, the environmental influence on association often restricts the environments in which dimers can be formed. Thus, alternative approaches of oligomerisation, which are not observed in nature, must be explored to ensure both interface control and permanence of oligomerisation. This thesis will explore the use of true biorthogonal non-natural PTM in the linkage of evolutionarily distinct proteins in Chapters 6 and 7.

---

## 1.4. Expanding the Genetic Code to Enhance Control of Protein Interaction

### 1.4.1. Why use non-natural Chemistry

Despite the existence of over 300 naturally occurring amino acids only 20 are widely utilised in natural protein synthesis <sup>211</sup> (with the exception of the less frequently used selenocysteine and pyrrolysine). This set of amino acids is sufficient for cellular life but does not sample the full range of side-chain chemistry which is available in nature <sup>212</sup>. Therefore, all proteins have the same limited chemistry and their structure, function, and interaction are hitherto restricted in natural systems <sup>211,213</sup>.

Synthetic approaches to the PTM of protein has drastically increased the scale of PTM related experimentation in recent years <sup>13</sup>. The application of non-natural PTM (nnPTM) has the potential to expand protein function by mimicking the impact of natural modification but without the need for various subsidiary enzymes needed to implement the desired change <sup>214</sup>. The modification of PTM with non-natural methods can now be used to modulate activity <sup>215</sup>, induce conformational change <sup>216</sup>, and introduce metal cofactors <sup>217</sup>, fluorophores <sup>218</sup>, and reactive tags <sup>219</sup>. Due to the wide range of nnPTM which is now available, it is now possible to target specific residue types and modify localised protein regions as desired <sup>220</sup>.

Several clear advantages have arisen from the application of nnPTM. For example, utilising nnPTM to either switch on or switch off protein function eliminates the need to either drastically increase protein expression or protein degradation within cells <sup>131</sup>. Another advantage is that in nature, PTM requires a complex assembly of precise structural/sequence motifs in combination with specific enzymes which recognise and apply PTM. Whereas nnPTM allows for the targeting of specified residue chemistry, such as the supplementation of surface cysteine residues with glycosyl iodoacetamide in a non-natural means of glycosylation <sup>221</sup>.

Many existing natural amino acid residues can be targeted for the specified application of desired functional groups without genetic modification. For example, the tethering of both natural proteins such as antibodies and enzymes can be achieved by application of a range of synthetic adducts which target the natural protein sequence. These adducts can additionally be used to label the natural amino acids with fluorescent probes (e.g. Cy3 dye) <sup>218</sup>, binding

---

agents (e.g. biotin) and to immobilise the protein itself to a specified surface (e.g. streptavidin-coated surfaces) <sup>222,223</sup>. The modification of PTM aids both the discovery of novel drug targets and enhances potential therapeutics <sup>220</sup>, e.g. PEGylation of anti-cancer monoclonal antibodies increases their circulating half-life within the body <sup>224</sup>. However, the major disadvantage of this approach is PTM specificity, both in terms of residue selection and specified protein targeting, due to the prevalence of many target residues on the surface of proteins.

To address this problem PTM has, in recent years, been used in partnership with the incorporation of nnAA. The combination of the two approaches has an advantage as it allows the application of residue-specific chemistry. This specificity, in turn, allows for a highly precise biorthogonal means of protein attachment, which can be controlled to feature specific structural and sequence motifs within the protein of choice. The incorporation of nnAAs during cellular protein synthesis has, therefore, been described as a “powerful tool” for developing our modern understanding of the structure-function relationship of protein <sup>225</sup>.

Currently, over 170 non-natural amino acids (nnAAs) have been developed and applied in protein engineering, expanding the chemistry available in protein synthesis <sup>26</sup>. The utilisation of non-natural protein chemistry is beneficial as it enables the introduction of both new and specific means of PTM, which is biorthogonal (reaction of two different but compatible chemical handles) and does not compete with existing chemistry within living systems <sup>111</sup>. These chemistries allow site-specific binding for precise protein labelling and attachment <sup>226,227</sup>. The incorporation of nnAA into at a specified residue of choice in sfGFP, for example, can be post-translationally modified to enable photo-switching <sup>228</sup>, pH sensitivity <sup>229</sup>, and defined molecular attachment of both small molecules <sup>227</sup> and whole proteins <sup>202</sup>.

Furthermore, nnAA incorporation has application in structural biology wherein heavy atom amino acid derivatives, such as selenomethionine, can be directly incorporated into recombinant proteins to aid with phasing issues in X-ray crystallography <sup>230,231</sup>.

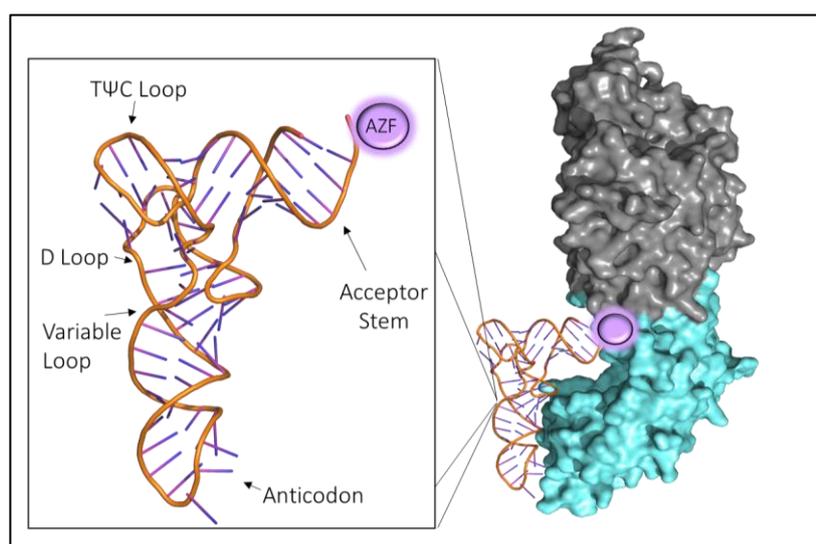
#### 1.4.2. The Incorporation of nnAA Into Proteins During Cellular Protein Synthesis

The incorporation of nnAA into proteins originally started with the use of auxotrophic strains of bacteria. In this method, cultures are grown in minimal medium and supplemented with structurally analogous non-natural amino acids <sup>232,233</sup>. Therefore, when the natural amino acid had been depleted, cells then utilise the nnAA for protein synthesis and incorporated it

throughout the proteome. This approach removed the specificity for both the protein target and residue of choice and resulted in the global replacement of the natural amino acid.

Before the advent of the expanded genetic code approach, specific nnAA incorporation could only be achieved using cell-free lysate transcription and translation approaches<sup>234–236</sup>. In this method, tRNAs had to specifically removed and replaced by chemically modified equivalents. This approach was complex, generated low yield and did not allow for nnAA incorporation *in situ*<sup>237</sup>.

The expanded genetic code addressed this issue by reprogramming a non-sense codon, usually the amber stop codon (TAG/UAG), to code for the incorporation of the nnAA at the specific residue in the target protein<sup>238</sup>. More recently, both the opal and ochre stop codons<sup>239</sup>, in addition to both four<sup>240</sup> and five<sup>241</sup> base codons have been successfully used for incorporation. The simplest means of achieving incorporation is to co-opt the existing tRNA charging system, the aminoacyl tRNA synthase and its cognate tRNA, produced in an alternative organism to that of expression (Figure 1.15). This machinery is engineered to recognise the required nnAA and thus charge it to the associated tRNA and no other<sup>233</sup>. Additionally, the cognate tRNA is also engineered, primarily at the anticodon loop to recognise the anticodon of choice, typically the UAG stop codon, and direct the specific placement of the



**Figure 1.15. The Association Between tRNA<sup>CUA</sup>/tyrosyl-tRNA Synthetase (TyrRS) of *Methanococcus jannaschii***

The association of tRNA<sup>CUA</sup> with aaRS<sup>TYR</sup> necessary for the transferal of nnAA AZF (PDB 1U7D). Insert displays the folded structure of tRNA<sup>CUA</sup> of *M. jannaschii* (PDB 1J1U) charged with AZF.

---

nnAA during natural ribosomal protein synthesis. This process is complementary to native proteogenesis and mimics natural proteogenesis for the controlled placement of nnAA.

The technique was first performed by the Schultz group in 2001 using the tRNA<sup>CUA</sup>/tyrosyl-tRNA synthetase of *Methanococcus jannaschii* (Figure 1.15.) and remains the most widely used approach for nnAA incorporation in *E.coli* <sup>242</sup>. Critically, the *Mj* tRNA/aaRS pair does not feature an editing mechanism which would recognise the nnAA as incorrect, so de-acylation does not occur <sup>243</sup>. Therefore, when supplied to *E.coli*, the pair does not aminoacylate any endogenous tRNAs and is biorthogonal to native protein synthesis <sup>244</sup>. Instead, the enzyme aminoacylates tyrosyl-tRNA<sup>CUA</sup> with nnAA at both high fidelity and efficiency <sup>245</sup>. Due to the technique's proven success rate it has been utilised widely in multiple cell lines from single cellular systems <sup>244,246,247</sup>, to whole organisms <sup>248</sup>, and lastly in cell-free protein synthesis <sup>237</sup>.

When a protein is mutated to contain the amber stop codon, protein expression is inherently linked to nnAA presence. In the absence of either nnAA or tRNA/aaRS, the codon UAG reverts to a stop signal and expression is prematurely terminated, generating truncated and largely non-functional protein product <sup>249</sup>. This dependency ensures that all target protein expressed is incorporated with the non-natural amino acid, and only truncated wild-type protein is co-expressed in the process.

### 1.4.3. The Modification of Protein Reactivity by The Incorporation of nnAA

One common approach for the incorporation of novel chemistry into amino acids is to use derivatives of the natural amino acid phenylalanine <sup>225</sup>. Phenylalanine has been used as a template generate a wide range of orthogonal substitutions to the benzyl ring, including *p*-iodo <sup>250</sup>, *p*-cyano <sup>251</sup>, and *p*- azido-phenylalanine<sup>11</sup>. Iodo and cyano-phenylalanine, offer unique X-Ray diffraction and infra-red characteristics, which aid structural determination, while *p*-azido phenylalanine has the unique benefits of azide reactivity and UV sensitivity <sup>252</sup>.

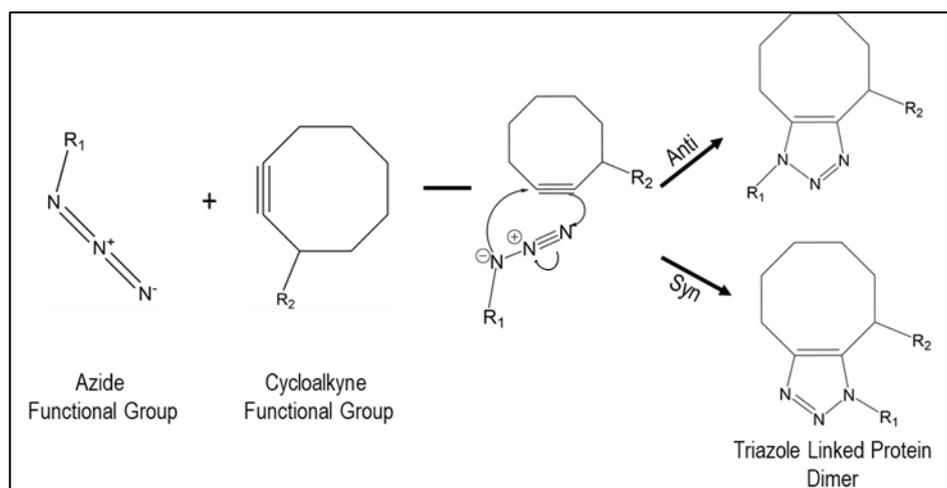
The incorporation of *p*-azido-phenylalanine provides two routes to modify protein structure and function through an nnPTM event, either by photochemical modification or click chemistry <sup>253</sup>. Phenyl-azide photochemistry is well known and has been used classically in biochemistry to induce cross-linking on exposure to UV light <sup>254</sup>. On irradiation, the phenyl azide (F-N<sub>3</sub>) converts to a reactive nitrene radical (F-N:) with the loss of N<sub>2</sub>. The nitrene radical

can then react with electron-rich protein regions, to cross-link with proximal carbon-carbon double bonds in an addition reaction <sup>255</sup>.

The advent of the expanded genetic code has enabled precise incorporation of phenyl azide chemistry to proteins, which allows the photochemistry to be used in a variety of new biological situations. For example, phenyl azide photochemistry has been used in both sfGFP and mCherry to induce photo-control of function <sup>252</sup> and enables specific attachment of sfGFP to carbon nanotubes <sup>256</sup>. Critically, as there is no similar side-chain chemistry in nature, there is no competition or interference to AZF reactivity in natural systems <sup>244</sup>.

#### 1.4.4. Strain Promoted Azide-Alkyne Cycloaddition

The term click chemistry describes a highly efficient synthetic reaction between a range of functional groups and was first coined by Klob in 2001 <sup>257</sup>. Strain Promoted Azide-Alkyne Cycloaddition (SPAAC) is considered part of the general “click chemistry” class of reactions as the reaction is; biocompatible, orthogonal and can occur without the need of a catalyst <sup>258</sup>. SPAAC is the permanent linkage of two proteins or molecules by the reaction between an azide and a strained alkyne to form a covalent triazole linkage (Schematic 1.4).

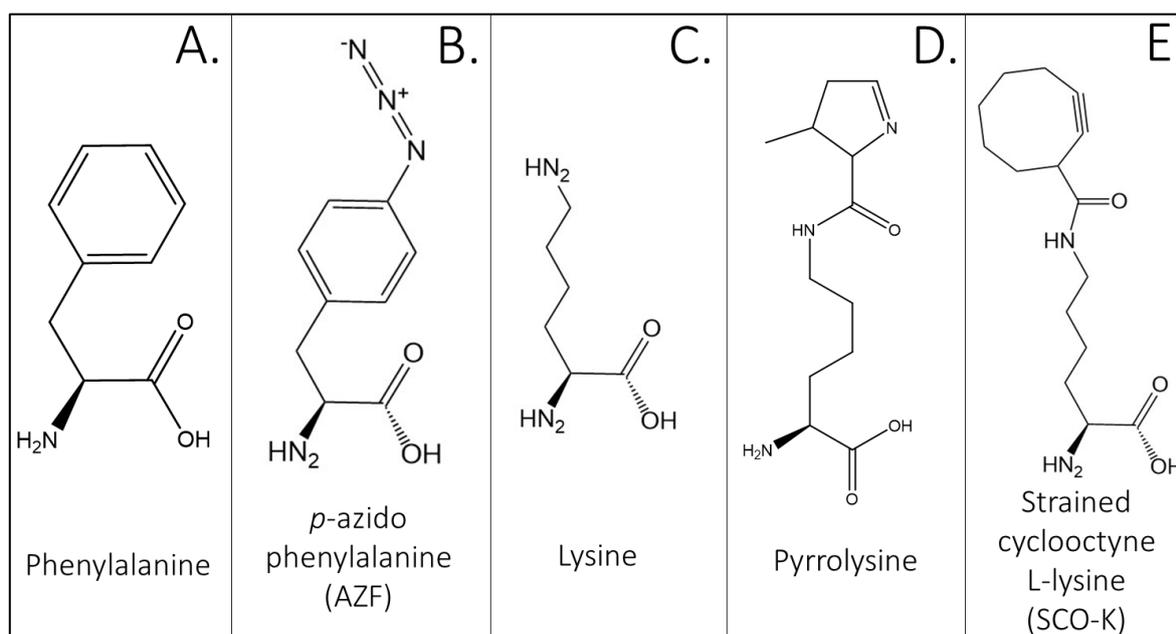


Scheme 1.3. Strain Promoted Azide Alkyne Cycloaddition for the oligomerisation of proteins, where  $R_1$  and  $R_2$  represent two separate proteins.

SPAAC linkage can post-translationally modify a protein with a range of adducts for both bioconjugation <sup>258</sup>, and oligomeric assembly <sup>259,260</sup>. To induce controlled oligomerisation by way of SPAAC three key developments must occur. The first development is the specified placement of a nnAA, either azide and alkyne, into two separate proteins by SDM (see Section 1.3.2). The second development is then nnAA incorporation (see Section 1.4.2). Then finally,

the third development is the formation of a triazole linkage which is achieved via spontaneous reaction between the incorporated azide and alkyne when the protein pair is mixed <sup>261</sup>. One such pair of SPAAC compatible nnAAs is that of strained cyclooctyne L-lysine (SCO-K), and *p*-azido phenylalanine (AZF)<sup>261</sup> (Figure 1.16).

The incorporation of AZF and SCO-K is achieved by the pairing of each nnAAs exclusively



**Figure 1.16. Comparison of Natural And non-natural Amino Acids.**

The structure of amino acids phenylalanine (A), Lysine (C), and Pyrrolysine (D) and their non-natural derivatives *p*-azido phenylalanine (B), and strained cyclooctyne L-lysine (E).

with an appropriately engineered tRNA/aaRS pair <sup>233,262</sup>. The current methodology of AZF incorporation in *E. coli* uses an engineered tyrosyl-tRNA/aaRS<sup>CUA</sup><sub>tyr</sub> pair from *M. jannaschii* (*Mj*), while, SCO-K incorporation utilises the pyrrolysine derivative aaRS/tRNA pair of *Methanosarcina barkeri* (tRNA<sup>CUA</sup><sub>PYR</sub>/pyrrolysyl-aaRS) <sup>111</sup>. When each protein is stored in isolation oligomerisation cannot occur as neither azide nor alkyne can form a covalent linkage. However, once the pair are mixed, the process of SPAAC commences, and the strained carbon-carbon triple bond ring within SCO-K promotes reaction without a catalyst when in proximity to the azide of AZF forming a permanent stable triazole linkage between the pair (Schematic 1.4) <sup>263,264</sup>

Oligomerisation by click chemistry has many advantages over non-specific means protein assembly by natural PTM <sup>111,189</sup>. As, the unique chemistry observed ensures protein A, always forms a linkage with protein B, and no waste products are formed in the process, as

---

observed in disulphide bonding (Figure 1.14). However, the linkage is dependent upon the incorporation of nnAA into two mutually compatible protein interfaces, which have an existing tendency to interact<sup>202,265</sup>. Thus, interface prediction must be employed to select the best regions of nnAA incorporation and maximise the existing affinity between species proteins. Recent work by Worthy *et al.* (2019), used this technique of interface prediction and demonstrated the significance of SPAAC attachment through the assembly of functionally enhanced sfGFP dimers<sup>202</sup>. Investigation in Chapters 6 and 7, will build on these findings and assess if two disparate protein monomers can interface and form SPAAC induced heterodimeric assembly.

---

## 1.5. The Scope of Project and General Aims

The presence of post-translational modification is inherently linked to both protein structure and function. The central objective of this work is to explore how both native and non-natural means of protein PTM can influence protein production, function, and enable new covalent interactions. This work will focus on two central protein models for investigation, HRP and GFP. HRP is used as a model to assess the influence and importance of native PTM, while GFP will be used to investigate if protein linkage by nnPTM, enables the oligomerisation of two functionally diverse proteins. Finally, Chapter 7 will link both these themes in the formation of protein oligomers by both natural and non-natural means of PTM.

**Chapter 3** investigates if the recombinant expression of soluble HRP is possible in *E.coli*. It will assess if the inclusion of the N-terminal and C-terminal peptide regions aid soluble protein expression, and if the removal of glycosylation sites enhances recombinant protein solubility.

**Chapter 4** explores if the action of a known corrective agent, glycosylated *apo*-HRP, can be replicated by recombinant expression. It will be investigated if the corrective properties can be duplicated in the absence of PTM, or if the corrective function is inherently linked to glycosylation presence.

**Chapter 5** is a study of the influence PTM exerts on the structural rigidity of HRP. HRP will be compared in four different PTM states, to address the effects both glycosylation and haem binding exerts on the emission of W117. Results will then be contrasted to protein thermostability and turnover, to assess if a correlation can be identified between PTM states.

**Chapter 6** utilises nnPTM oligomerisation to assess the impact on the function of the direct linkage of sfGFP with mCherry. *In-silico* simulations of protein interfacing will be employed to identify a compatible region of protein interface which will then be utilised for nnAA incorporation and subsequent protein attachment.

**Chapter 7** - uses the methodology applied in Chapter 6 to explore if functional cooperation of two functionally disparate but FRET compatible proteins, sfGFP and the haem binding cytochrome *b*<sub>562</sub> by nnPTM. Finally, both non-natural and natural means of PTM will be combined in one system, to investigate the formation of higher-order protein oligomers.

## Chapter 2. Materials and Methods

---

## 2.1. Materials

### 2.1.1. Water Source and Buffer Preparation

All water used in this study was either ddH<sub>2</sub>O (for the preparation of growth media), or Milli Q™ (MQ) grade (for buffer preparation in analytical methods). All buffers prepared, were first measured using a pH probe (Hanna Instruments, Luton, UK), before adjusting to the correct pH by titration of either 5 M HCL (Fisher Scientific, Leicestershire, UK), or 5 M NaOH (Melford, Suffolk, UK). All buffers used were sterilised by drawing through a 0.22 µm filter (VWR, Leicestershire, UK).

### 2.1.2. Chemicals

Antibiotics were prepared to a 1000x working concentration stock, and filter sterilised using 22 µm filter (VWR) before storage at -20 °C. Antibiotics used were: carbenicillin (Melford, Suffolk, UK), kanamycin (Duchefa Biochemie, Haarlem, NL), chloramphenicol (Melford), and tetracycline (Melford), at a working concentration of 50 µg/mL, 30 µg/mL, 35 µg/mL and 25 µg/mL, respectively. All antibiotics were prepared by dissolving of powder in MQ water except for Tetracycline and Chloramphenicol, which were prepared by dissolving in 100% ethanol (VWR).

Routine laboratory chemicals were purchased from Fisher Scientific (Loughborough, UK) and Sigma-Aldrich (Dorset, UK) while the suppliers of specialised chemicals are noted in the appropriate sections. Haem, in the form of haemin (Sigma-Aldrich, Dorset, UK), used in protein saturation was prepared to 100 mM concentration in 0.5 M NaOH (Sigma-Aldrich) before dilution in ddH<sub>2</sub>O to desired working concentration. Dithiothreitol (DTT )(Melford, Suffolk, UK) was prepared as a 10 mM stock before dilution in ddH<sub>2</sub>O to a 1 mM working concentration.

Non-natural amino acids (nnAA) were added at a 1 mM working concentration to the culture medium. *P-azido-phenylalanine* (AZF)(Bachem, Weil am Rhein, Germany) was prepared in dark conditions and dissolved in 1 M NaOH before dilution 1:5 in ddH<sub>2</sub>O immediately before addition to culture. Strained-cyclooctyne-L-lysine (SCO- K)(Sichem, Bremen, Germany) was prepared as a 100 mM stock solution by dilution in 0.2 M NaOH/15% DMSO and stored at -80 °C as described by the manufacturer. For use, a stock solution was diluted 1:4 in 1M 4-(2-

hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7 buffer before addition to the culture medium.

Both glycosylated *apo* and *holo*-HRP used in this investigation were generously supplied by Ortho Clinical Diagnostics (OCD), (Pencoed, UK) as lyophilised protein. Protein was prepared fresh to 100  $\mu$ M stock concentration in Tris 50 mM pH 8.0 before dilution to the appropriate working concentration. Rehydrating protein was left to equilibrate for 30 minutes under gentle agitation at room temperature before centrifugation at 150,000  $\times$ g for 1 min to remove any residual powder.

### 2.1.3. Bacterial Cell Strains

Three cell strains of *Escherichia coli* were used in this investigation, and their genotypes are recorded in Table 2.1. Plasmid amplification and DNA propagation were achieved in NEB<sup>®</sup> 5- $\alpha$  (New England Biolabs, Hertfordshire, UK) due to their reduced DNase activity. Expression of pBAD vector genes utilised the *araB* promoter and was achieved in TOP10<sup>™</sup> (New England Biolabs) and induced with arabinose. Whilst pET vector plasmids were transformed into SHuffle<sup>®</sup> cells (New England Biolabs) to utilise the T7 promoter and utilise the extra chromosomal copy of disulphide bond isomerase to increase disulphide bond formation.

**Table 2.1. Genotypes of *E. coli* Competent Cells**

E. coli Cell Strain	Genotype
NEB <sup>®</sup> 5- $\alpha$	<i>fhuA2</i> $\Delta$ ( <i>argF-lacZ</i> )U169 <i>phoA</i> <i>glnV44</i> $\Phi$ 80 $\Delta$ ( <i>lacZ</i> )M15 <i>gyrA96</i> <i>recA1</i> <i>relA1</i> <i>endA1</i> <i>thi-1</i> <i>hsdR17</i>
SHuffle <sup>®</sup>	<i>F'</i> <i>lac</i> , <i>pro</i> , <i>lacIq</i> / $\Delta$ ( <i>ara-leu</i> )7697 <i>araD139</i> <i>fhuA2</i> <i>lacZ::T7 gene1</i> $\Delta$ ( <i>phoA</i> )PvuII <i>phoR</i> <i>ahpC*</i> <i>galE</i> (or <i>U</i> ) <i>galK</i> $\lambda$ att::pNEB3-r1- <i>cDsbC</i> ( <i>SpecR</i> , <i>lacIq</i> ) $\Delta$ <i>trxB</i> <i>rpsL150</i> ( <i>StrR</i> ) $\Delta$ <i>gor</i> $\Delta$ ( <i>malF</i> )3
TOP10 <sup>™</sup>	<i>F-</i> <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ <i>lacX74</i> <i>recA1</i> <i>araD139</i> $\Delta$ ( <i>ara-leu</i> )7697 <i>galU</i> <i>galK</i> <i>rpsL</i> ( <i>StrR</i> ) <i>endA1</i> <i>nupG</i>

### 2.1.4. Plasmid Vectors

Protein production was induced by stimulation of either the T7 promoter of pET-24b plasmid vector (HRP<sup>EC1</sup> and HRP<sup>EC2</sup>) or *araB* promoter of the pBAD plasmid vector (sfGFP, mCherry and cytochrome *b*<sub>562</sub>). Both pDULE and pEVOL were co-transformed with pBAD for the incorporation of nnAA into protein. The full list of plasmid vectors and corresponding functions are listed in Table 2.2 with maps of significant vector regions displayed in Figure 2.1.

Table 2.2. Vectors Used in This Thesis and Their Function

	Encoding Gene	Gene Size (bp)	Vector size (bp)	Function	Antibiotic Resistance	Source
pET-24b	HRP <sup>EC1</sup> HRP <sup>EC2</sup> HRP <sup>68C</sup>	1109 1134 1134	5309	Protein Expression	kanamycin	Purchased with gene from GenScript
pBAD	Cytochrome <i>b</i> <sub>562</sub> (and Variants) mCherry (and Variants) sfGFP (and Variants)	324 735 714	4100	Protein Expression	ampicillin/ carbenicillin	Donated from DDJ group, Cardiff School of Biosciences, Cardiff University, Cardiff, UK
pEVOL	tRNA <sub>CUA</sub> /aaRS <sup>SCO</sup> pair	900	5400	Incorporation of SCO-K nnAA in cells	chloramphenicol	
pDULE	tRNA <sub>CUA</sub> /aaRS <sup>AZF</sup> pair	1000	5000	Incorporation of AZF nnAA derivatives in cells	tetracycline	

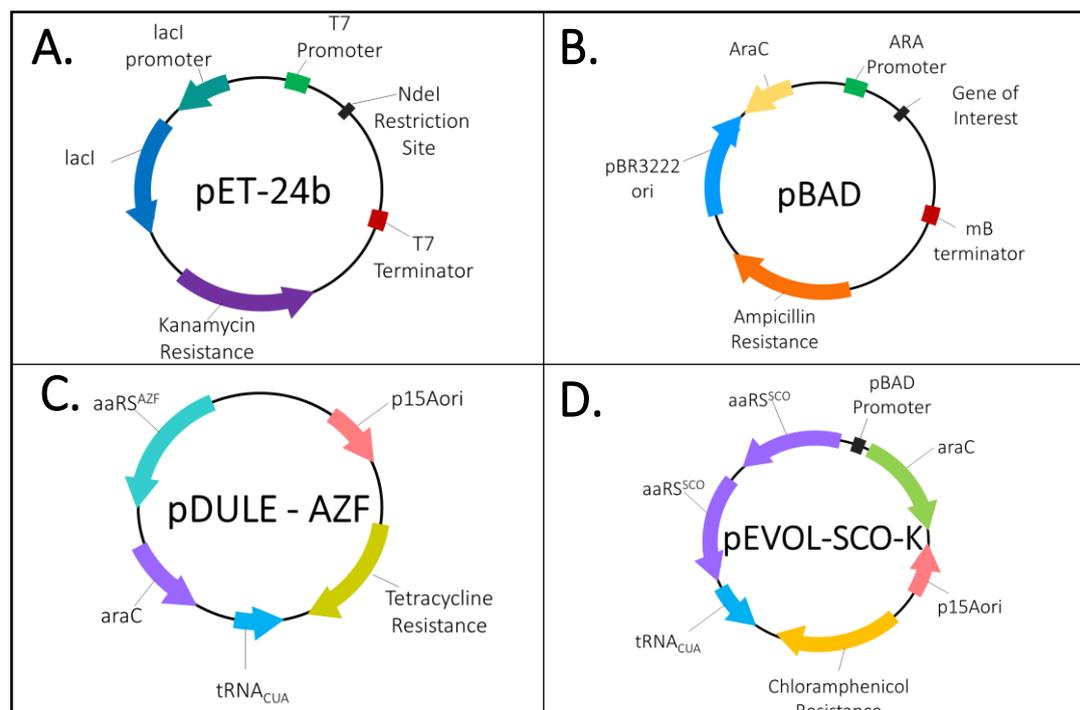


Figure 2.1. Plasmid Map of Each Vector used.

Plasmid maps of **A**, pET-24b, **B**, pBAD, **C**, pDULE-AZF, and **D**, pEVOL-SCO-K. Constructed genetic sequences of for recombinant HRP expression were inserted into the NdeI restriction site of pET- 24b (**A**). Whilst all other proteins expressed in this were inserted into the pBAD plasmid (**B**). Paired tRNA<sub>CUA</sub>/aaRS for nnAA incorporation observed in both pDULE (**C**) and pEVOL (**D**), with pEVOL containing two copies of aaRS<sup>SCO</sup>.

### 2.1.5. Bacterial Growth Media

All media was sterilised by autoclave at 121°C for 15 minutes on the day of use. Lysogeny broth (LB) Agar (Melford) was prepared according to the manufacturer's instructions (35 g/L) in ddH<sub>2</sub>O and stored at 60 °C after autoclaving until use. Lysogeny broth (Melford) was used as a standard growth medium for all starter cultures and prepared as per the manufacturer's instruction (25 g/L) in ddH<sub>2</sub>O. Protein expression utilised 2x Yeast Tryptone Broth (2xYT) (Melford, Ipswich, UK), which was again prepared per manufacturer's instructions (31 g/L) in ddH<sub>2</sub>O.

Auto-induction of protein was achieved by supplementing of media with appropriate chemicals after the media had cooled to room temperature (Table 2.3) and the inclusion of either arabinose (for induction of pBAD vector genes) or IPTG ( for the induction of pET vector genes). All chemicals used were prepared to stock concentration and sterilised by 0.22 µM filter.

**Table 2.3. Ingredients Necessary for the Generation of Culture Medium for the automatic induction of protein. Adapted from Faust *et al.* 2015 <sup>48</sup>.**

Component	Stock Concentration	Final Concentration in Growth Media	
		Arabinose Expression	IPTG Expression
Glycerol	50%	0.5% (w/v)	0.5% (w/v)
Glucose	20%	0.05% (w/v)	1% (w/v)
Lactose	10%	0.2% (w/v)	0.2% (w/v)
Na <sub>2</sub> HPO <sub>4</sub>	1M	25 mM	25 mM
KH <sub>2</sub> PO <sub>4</sub>	1M	25 mM	25 mM
NH <sub>4</sub> Cl	1M	50 mM	50 mM
Na <sub>2</sub> SO <sub>4</sub>	1M	5 mM	5 mM
MgSO <sub>4</sub>	1M	2 mM	2 mM
Trace Metals*	5000x	1 x	1 x
Arabinose	20%	0.5% (w/v)	-
IPTG	1M	-	4 mM
Water	-	Up to 1L	Up to 1L
Antibiotic	1000x	1x	1x

\* working concentration of Trace Metals were as follows; 4µM CaCl<sub>2</sub>, 2µM MnCl<sub>2</sub>, 2µM ZnSO<sub>4</sub>, 0.4 µM CoCl<sub>2</sub>, 0.4 CuCl<sub>2</sub>, 0.4µM NiCl<sub>2</sub>, 0.4µM Na<sub>2</sub>MoO<sub>4</sub>, 0.4µMH<sub>3</sub>BO<sub>3</sub>, and 10µM FeCl<sub>3</sub>.

## 2.2. Molecular Biology

### 2.2.1. Construction of HRP Expression Vectors

Expression constructs were generated to optimise production of soluble recombinant HRP. The sequence was assembled in Serial Cloner 2-6-1 utilising the genetic sequence of HRP C1A (Uniprot Entry P00433). The signal sequence of HRP C1A was identified using the server SignalP 5.0 (<http://www.cbs.dtu.dk/services/SignalP/>). The periplasmic localisation sequence (GRKSLAILAVSSLVSSASFA) of cytochrome  $b_{562}$  was incorporated on the N-terminus of HRP<sup>EC1</sup>. While the TEV cleavage (ENLYFQ/G) sequence was incorporated within the new N-terminus component of HRP<sup>EC2</sup>. Synthetic gene constructs optimised for *E. coli* expression were generated by GenScript (NJ, USA) and cloned into the pET-24b plasmid for expression.

### 2.2.2. Design of Mutagenesis Oligonucleotides

*In silico* design of oligonucleotides was performed using Serial Cloner 2.6.1 ([http://serialbasics.free.fr/Serial\\_Cloner.html](http://serialbasics.free.fr/Serial_Cloner.html)) and AmplifX 2.0.7. (Marseille, France).

**Table 2.4. Oligonucleotides Used in Investigation for Mutagenesis and Sequencing of Plasmid. (Mutated codons highlighted in bold).**

Name	5' Sequence 3'	Tm
Sequencing of T7 induced Gene		
T7 - F	TAA TAC GAC TCA CTA TAG GG	58 °C
T7 - R	CTA GTT ATT GCT CAG CGG T	58 °C
Sequencing of araB induced Genes		
pBAD - F	ATG CCA TAG CAT TTT TAT CC	51 °C
pBAD - R	GAT TTA ATC TGT ATC AGG	51 °C
Creation of mCherry <sup>198AZF</sup>		
mCherry <sup>198TAG</sup> - F	C GGC GCC TAC AAC GTC AAC ATC <b>TAG</b> T	72 °C
mCherry <sup>198TAG</sup> - R	GG CAG CTG CAC GGG CTT CTT	72 °C
Creation of Cytochrome $b_{562}$ <sup>I67TAG</sup>		
Cyt $b_{562}$ <sup>I67TAG</sup> - F	GTT TCG <b>ACT</b> <b>AGC</b> TGG TCG GT	64 °C
Cyt $b_{562}$ <sup>I67TAG</sup> - R	CGT GGC GGA AAT CTT TCA T	64 °C
Creation of Cytochrome $b_{562}$ <sup>Q71TAG</sup>		
Cyt $b_{562}$ <sup>Q71TAG</sup> - F	C GGT <b>TAG</b> ATT GAC GAC G	61 °C
Cyt $b_{562}$ <sup>Q71TAG</sup> - R	AC CAG AAT GTC GAA ACC G	61 °C
Creation of Cytochrome $b_{562}$ <sup>Q93TAG</sup>		
Cyt $b_{562}$ <sup>Q93TAG</sup> - F	TGC AGA <b>GTA</b> <b>GCT</b> GAA AAC GA	65 °C
Cyt $b_{562}$ <sup>Q93TAG</sup> - R	GCA GCC TGC GCT TCT TTT AC	65 °C
Creation of Cytochrome $b_{562}$ <sup>D21CYS</sup>		
Cyt $b_{562}$ <sup>D21CYS</sup> - F	A AAA GCG <b>TGT</b> AAC GCG GCG CAA	62 °C
Cyt $b_{562}$ <sup>D21CYS</sup> - R	TC GAT CAC TTT TAA ATT GTC GTT G	62 °C

Annealing temperatures were determined using New England Biolabs T<sub>m</sub> Calculator v 1.12.0 (<https://tmcalsculator.neb.com/>). Oligonucleotides were synthesised by Integrated DNA Technologies (Iowa, USA) (<https://eu.idtdna.com/>), a full list of which are displayed in Table 2.4. Once obtained, oligonucleotides were diluted to 10 mM working concentration before storage at - 20 °C.

### 2.2.3. Site-Directed Mutagenesis

#### 2.2.3.1. Mutagenesis PCR with Q5 Polymerase

Whole plasmid amplification, for site-directed mutagenesis, was executed with one pair of mutagenesis oligonucleotides. Amplification was achieved with Q5<sup>®</sup> High Fidelity DNA Polymerase (New England Biolabs) per manufacturers instruction (Table 2.5.) in a 50 µL reaction volume (Table 2.6), with a negative (with no template) DNA control. Amplification was initiated using Techne Thermocycler (Staffordshire, UK).

**Table 2.5. Q5 PCR Reaction Conditions**

Reaction Stage		Time Length	Temperature
Initial Denaturation		30 sec	98 °C
30 Repeat Cycles	Denaturation	10 sec	98 °C
	Annealing	30 sec	X °C <sup>a</sup>
	Extension	30 s per kB of product	72 °C
Final Extension		3 min	72 °C
Hold		∞	4 °C

<sup>a</sup> X = The T<sub>m</sub> of primer pairs as listed in Table 2.4.

**Table 2.6. Standard Reactants for Site Directed Mutagenesis.**

Reagent	Final Concentration
dNTPs	0.2 mM
Q5 Polymerase x5 Buffer	1x
Forward Primer	0.5 µM
Reverse Primer	0.5 µM
Template Plasmid	1 ng/ µL
Q5 Polymerase	1 unit
Q5 x5 GC Enhancer	1 x
Total Volume	50 µL <sup>a</sup>

<sup>a</sup>Total volume was made up with ultra-pure water

---

### 2.2.3.2. Agarose Gel Electrophoresis

To assure the quality of DNA stocks in addition to confirming correct plasmid amplification, agarose gel electrophoresis was implemented. Gels were prepared to 1% w/v by dissolving of 1 g of agarose (Melford) in 100 mL of TAE (40 mM Tris-acetate, 1 mM EDTA pH 8.8) and supplementing with 1  $\mu$ L ethidium bromide (0.5  $\mu$ g/mL). Each sample was prepared by mixing 10  $\mu$ L with 2  $\mu$ L of 5x loading dye (New England Biolabs) and run against 3  $\mu$ L of either 1 kB DNA ladder (New England Biolabs) or 100 bp ladder (New England Biolabs). Samples were run for 40 minutes at 100 V using a PowerPac (BioRad CA, USA). Bands were finally visualised and recorded by GelDoc IT UV-Transilluminator (Ultra-Violet Products Ltd, Cambridge, UK).

### 2.2.3.3. PCR Clean-Up

Once the size of the amplified DNA product had been confirmed by agarose gel electrophoresis, the methylated plasmid template was removed from samples by digestion with DpnI enzyme. For this, 10  $\mu$ L of PCR product was mixed with 1 unit of DpnI and 5  $\mu$ L of 10 x NEBuffer and incubated for 1 hour at 37 °C. After digestion sample was cleaned up using the QIAquick PCR purification kit (Qiagen, Venlo, NL) following the manufacturer's protocol.

### 2.2.3.4. Quantification, Phosphorylation, Ligation, and Clean-Up of PCR Product

Purified PCR product was quantified using NanoDrop® ND1000 Spectrophotometer (Thermo Fisher Scientific, MA, USA) after blanking. Phosphorylation of the 5' end of amplified plasmid DNA was achieved using T4 Polynucleotide Kinase (10 Units) (New England Biolabs), in 1x Quick Ligase Buffer (66 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM ATP, 15% (w/v) PEG 6000; pH 7.6), as this provided samples with the ATP necessary for ligation. The reaction mixture was incubated for 30 minutes at 30 °C before heat inactivation of the enzyme at 65 °C for 20 minutes. The reaction mix was subsequently supplied with 10 units of Quick Ligase enzyme (New England Biolabs), for ligation. The ligation mix was incubated for 15 minutes at room temperature before confirmation of sample ligation by gel electrophoresis. The sample was then purified by QIAquick PCR purification kit (Qiagen, Venlo, NL) before storage at - 20 °C.

## 2.2.4. Bacterial transformation

### 2.2.4.1. Preparation of Competent Cells

For the preparation of competent cells, 10  $\mu$ L of the original stock of desired cell strain was streaked on an Agar plate (without antibiotic) and incubated overnight at 37 °C. A single

---

colony was selected for inoculation of 1 L LB medium and allowed to grow at 37 °C until reaching an OD<sub>600</sub> of 0.4 A.U. As soon as cells reached the desired OD<sub>600</sub> the flask was removed from the incubator and placed on ice. Cells were recovered by centrifugation (Beckman Coulter rotor, JLA-16.25) at 5000 xg for 20 minutes at 4 °C. Cells were resuspended in 200 mL of autoclaved MQ water. Electrocompetent cells were washed twice more to remove any salt present before resuspension in 2 mL of 15% glycerol. Chemically competent cells were washed twice before resuspension in 2 mL of 0.1 M CaCl<sub>2</sub>, 15 % glycerol. All cells were finally separated into 50 µL aliquots and flash-frozen in liquid nitrogen before storage at - 80 °C until use.

#### 2.2.4.2. Transformation by electroporation.

Before cell transformation, competent cells and plasmid DNA were thawed on ice for 1 hour. Once thawed, 1 ng of DNA was mixed with 50 µL of competent cells and incubated for 5 minutes before the transfer to electroporation cuvette (VWR). Cells were shocked with a 4.5 – 5 ms electrical pulse at 12.5 kV/cm field strength by Gene Pulser (BioRad). Shocked cells were immediately transferred to 1 mL of recovery medium (LB medium, see section 2.1.5, supplemented with 1% (w/v) glucose) and incubated at 37 °C for 1 hour to allow recovery. After which cells were plated on an LB agar plate with the appropriate antibiotic and left to grow overnight at 37 °C. Agar plates which produced colonies were stored at 4 °C.

#### 2.2.4.3. Transformation by heat shock

To transform chemically competent cells by heat shock, 50 µL of cells were once more thawed on ice. Thawed cells were mixed with 1 ng of DNA and chilled for a further 15 minutes on ice before transferal to 42 °C water bath for 45 seconds. Cells were placed once more on ice for a further 5 minutes before transferal to recovery media (1% glucose) and incubation at 37 °C for 1 hour to allow recovery in medium (LB medium, see section 2.1.5, supplemented with 1% (w/v) glucose). After which cells were plated on LB agar plate with the appropriate antibiotic and left to grow overnight at 37 °C. Cells which produced colonies were stored at 4 °C.

#### 2.2.5. Extraction and Quantification of DNA from Cells

The plasmid was extracted from cells which had been grown in 10 mL of LB media overnight. Cells were harvested by centrifugation at 4100 rpm for 20 minutes before plasmid was extracted using the Qiagen Miniprep Kit (Qiagen) and quantified *via* NanoDrop® ND1000

---

Spectrophotometer (Thermo Fisher Scientific). Finally, the plasmid stock was stored at -20 °C in nuclease-free water. DNA was quantified by NanoDrop® ND1000 Spectrophotometer (Thermo Fisher Scientific) and prepared to a concentration of 50 ng/μL and sent to Eurofins MWG Biotech (Eurofins, Wolverhampton, UK), for sequencing *via* their Tube Seq service.

## 2.3. Bacterial Growth and Protein Expression

### 2.3.1. Expression of IPTG induced protein

Expression of pET-24b vector protein, HRP<sup>EC1</sup> and HRP<sup>EC2</sup>, was achieved by IPTG induction in IPTG auto-induction media (Table 2.3). A single colony was selected from cells transformed with plasmid and used to inoculate 5 mL of LB media supplemented with kanamycin. Following overnight growth at 37 °C, cells were used to inoculate 1 L flask and incubated at 16 °C under shaking conditions for 24 hours for expression. Expressing cells were harvested by centrifugation for 20-minute at 5000 xg at 4 °C (Beckman Coulter rotor, JLA-16.25) and protein was extracted either by periplasmic extraction (Section 2.3.4.) or French pressure cell lysis (Section 2.3.5.). Protein expression was confirmed by SDS-PAGE of the whole-cell lysate (Section 2.5.1).

### 2.3.2. Expression of Arabinose induced protein

Expression of all pBAD vector proteins (cytochrome *b*<sub>562</sub>, sfGFP, and mCherry) was executed using arabinose-based autoinduction. Again, a scraping of a single cell colony was used to inoculate 5 mL of LB media supplemented with the appropriate antibiotic. Following overnight incubation at 37 °C cells were added to 1 L flask of arabinose auto-induction media and the cultures left overnight at 37 °C (Table 2.3). Cells were harvested by centrifugation at 5000 xg at 4 °C (Beckman Coulter rotor, JLA-16.25) and SDS-PAGE of the whole-cell lysate confirmed the expression of the protein of interest.

### 2.3.3. Production of nnAA containing Protein

Incorporation of nnAA into protein was implemented by the transformation of cells with both pBAD plasmid and either pEVOL-SCO-K or pDULE-AZF. The placement of the nnAA within the sequence was directed by SDM of the pBAD plasmid to incorporate amber stop codon at the residue of interest (see Section 2.2.3.1). Cells were induced as described (Section 2.3.2.) with the addition of 1 mM of nnAA after 1 hour of growth. For the initial expression of

---

each TAG mutant, cells were grown in the presence and the absence of nAA (to ensure dependence of expression on nAA presence). For expression with AZF, nAA was prepared in dark conditions and cultures were subsequently grown in the dark to prevent UV exposure. All cells and protein produced using AZF were kept in foiled tubes and purified in dark conditions for the same purpose.

#### 2.3.4. Periplasmic Extraction of Protein

Extraction of the periplasmic cell fraction was achieved by osmotic shock. Induced cells were centrifuged for 20 minutes at 5,000 xg before being resuspended in 200 mL of TES buffer (20 mM Tris-HCl, 1 mM EDTA, 20% Sucrose) and gently stirred on ice for 15 minutes. Cells were centrifuged once more at 5,000 xg for 20 minutes, and the supernatant was removed for analysis by SDS PAGE. Cells were resuspended in 100 mL of 5 mM MgSO<sub>4</sub> and stirred on ice for a further 15 minutes to release periplasmic protein. The sample was centrifuged for a final time (20 minutes 5,000 xg) to remove whole cells from the released periplasmic protein fraction before analysis by SDS-PAGE.

#### 2.3.5. Cell Pressure Lysis by French Press

Following the collection of expressing cells from the culture medium by centrifugation at 5000 xg for 20 minutes, cells were resuspended in 20 mL of 50 mM Tris-HCL pH 8.0. Cells were lysed by 1250 psi pressure using French pressure cell press, and the lysate was collected. Soluble protein was removed from the insoluble cell fraction by centrifugation at 20,000 xg (Beckman Coulter-rotor JA-25.50) for 40 minutes. Both soluble and insoluble cell fractions were analysed by SDS-PAGE to determine if the expressed protein was soluble. The soluble cell lysate was subsequently filtered by Corning® 70 µm cell strainer (Sigma) to ensure that all large debris had been removed from the sample before applying to an ÄKTA for purification (GE, Healthcare).

#### 2.3.6. Ammonium Sulphate Precipitation

To separate protein based on solubility, the lysate was subjected to ammonium sulphate precipitation before purification. The soluble lysate was mixed with ammonium sulphate to the desired percentage saturation (w/v) and allowed to dissolve at room temperature under gentle agitation. Once all ammonium sulphate had dissolved protein

precipitate was collected by centrifugation at 20,000 xg for 20 minutes and the supernatant (soluble protein) was separated from the precipitant (insoluble protein).

## 2.4. Protein Purification

In all instances where the purification of protein was desired, the protein was separated by ÄKTA Pure FPLC (GE Healthcare, IL, USA) and use of the appropriate purification column (Table 2.7) (GE Healthcare). Elution of protein was monitored by absorbance at 280 nm, with SDS PAGE used to assess the composition of each fraction collected. Before purification, both column and system were equilibrated in Tris buffer (50 mM Tris-HCL pH 8.0).

Following confirmation of purity by SDS PAGE, the protein of interest was concentrated by spin column (VWR) at 4100 rpm until the desired protein volume was reached. Protein was buffer exchanged using a PD10 column (GE Healthcare) and flash-frozen in liquid nitrogen for storage at -80 °C either as a lyophilised powder or in Tris Buffer (50 mM Tris-HCL pH 8.0).

**Table 2.7. Columns Used in Protein Chromatography.**

Column Name	When Used	Bed Volume
HisTrap™ HP Ni <sup>2+</sup>	Purification affinity tagged protein from lysate.	5 mL
Hiload™ 16/600 Superdex™ S75	Separation of cytochrome b <sub>562</sub> from lysate (separation range 3 kDa to 75 kDa)	120 mL
Hiload™ 16/600 Superdex™ S200	Separation of HRP <sup>EC2</sup> and fluorescent protein from contaminants. (separation range 10 kDa to 600 kDa)	120 mL
HiLoad™ 26/600 Superdex™ S200	Separation of monomeric protein from dimeric protein (separation range 10 kDa to 600 kDa)	320 mL
HiTrap™ Q	Purification of protein based on surface charge using anion exchanger	5 mL

### 2.4.1. Immobilised Metal Affinity Chromatography (IMAC)

Protein featuring a poly-histidine affinity Tag (HIS-TAG) were purified by immobilised metal affinity chromatography (IMAC). First, the Ni<sup>2+</sup> column was equilibrated with 4 column volumes of Tris buffer (50 mM Tris pH 8) before the application of cell lysate at a rate of 1 mL/min. All lysate was run through the column until the absorbance at 280 nm reached an intensity of 0 A.U. The column was washed with 4 column volumes of wash buffer (50 mM Tris,

---

200 mM NaCl, 20 mM Imidazole, pH 8.0) to remove non-specific binding to the column. Elution commenced with the application of a gradient elution buffer (50 mM Tris, 200 mM imidazole, pH 8.0) at a rate of 1% / 1 mL flow. Upon the application of elution gradient, fractions were collected and analysed by SDS PAGE to assess for target protein presence.

#### 2.4.2. Size Exclusion Chromatography (SEC)

Size exclusion chromatography (SEC), was used for the separation of protein by hydrodynamic radius. Column selection was undertaken based upon the size of the protein of interest (Table 2.7) and equilibrated in Tris buffer (50 mM Tris pH 8). Before application, the sample was concentrated to a total volume of 2 mL and then applied to the column at a flow rate of 1 mL/min. Elution of protein was monitored by absorbance, and 2 mL fractions were collected and analysed for purity by SDS-PAGE.

#### 2.4.3. Anion-Exchange Chromatography (IEX)

When preparing protein for anion-exchange chromatography, all protein samples needed to have minimal buffer salts present. Where necessary protein was buffer exchanged by application to PD-10 desalting column (GE Healthcare) before application to the column. Wherein all lysate was run through the column until the absorbance at 280 nm reached an intensity of 0 A.U. Protein was eluted by the application of high salt buffer (2M NaCl, 50 mM Tris HCL, pH 8.0) at a rate of 1%/ 1 mL of flow. Fractions were, once more, collected after application of elution buffer and analysed by SDS-PAGE.

#### 2.4.4. Purification of HRP<sup>EC1</sup>

Following growth and induction (Section 2.3.1.) of cells transformed with pET-24b HRP<sup>EC1</sup> (Section 2.3.1), HRP<sup>EC1</sup> was extracted from cells through periplasmic extraction (Section 2.3.4). Purification of the periplasmic cell fraction was achieved by IEC (Section 2.4.3) and SEC (Section 2.4.2) before SDS PAGE of collected fractions (Section 2.5.1) and storage at -80 °C.

#### 2.4.5. Purification of HRP<sup>EC2</sup>

Cells transformed with pET-24b-HRP<sup>EC2</sup> were grown in IPTG autoinduction medium at 16 °C for induction of HRP<sup>EC2</sup> (Section 2.3.1.). Induced cells were lysed by French press (Section 2.3.5) before soluble protein was separated from the insoluble lysate by centrifugation at 20,000 xg for 40 minutes. Protein was purified by nickel affinity chromatography (see Section

2.4.1.) and subsequently SEC (Hiload™ 16/600 Superdex™ S200) at a flow rate of 1 mL/min (see Section 2.4.2.). The purity of the eluted protein was finally confirmed by SDS PAGE (purity > 99%) before haem saturation (if desired). Haem saturation was achieved by mixing protein with a 100 x molar excess of haemin (Sigma-Aldrich) overnight at room temperature, and removal of free porphyrin by SEC. Finally, purified HRP<sup>EC2</sup> was stored by freeze-drying in VirTis BenchTop Pro (SP Scientific, PA, USA), and storage as a lyophilised powder at - 80 °C.

#### 2.4.6. Purification of sfGFP and mCherry variants

Purification of fluorescent proteins sfGFP and mCherry together with their variants was undertaken as follows. Fluorescent protein was expressed from cells transformed with either pBAD-sfGFP or pBAD-mCherry (Section 2.3.2) and lysed (Section 2.3.5.). The soluble lysate was purified by IMAC after application and elution from HisTrap™ HP Ni<sup>2+</sup> as described previously (Section 2.4.1.). Eluted fractions were further polished to remove contaminants by SEC (Section 2.4.2) before confirmation of protein purity by SDS PAGE (Section 2.5.1.). Protein was concentrated to 100 μM by spin filtration (15,000 MWCO) and separated to 100 μL samples before flash freezing and storage at -80 °C.

#### 2.4.7. Purification of Cytochrome *b*<sub>562</sub> and Variants

Following expression (Section 2.3.2.) and lysis (Section 2.3.5.) of cytochrome *b*<sub>562</sub> and variants, the soluble lysate was collected by centrifugation. The soluble lysate was subjected to 30% (w/v) ammonium sulphate precipitation (Section 2.3.6) to remove some contaminant protein as precipitant. The soluble supernatant was subjected to a further round of ammonium sulphate precipitation to a concentration of 90% (w/v) to precipitate all protein.

The precipitated lysate was resuspended in 2 mL 50 mM Tris pH 8.0 and if desired was mixed with 100 x molar excess of haem porphyrin (section 2.1.2.) for the generation of *holo*-cytochrome *b*<sub>562</sub>. Protein was applied to Hiload™ 16/600 Superdex™ S75 at a flow rate of 1 mL/min for purification and SDS-PAGE analysis of fractions. Finally, cytochrome *b*<sub>562</sub> was applied to a Q anion exchange column for IEC (Section 2.4.3) before final buffer exchange into 50 mM Tris pH 8.0 using a PD-10 desalting column (see Section 2.4.3). Pure cytochrome *b*<sub>562</sub> was concentrated to 100 μM concentration and separated to 100 μL samples before flash freezing and storage at - 80 °C until use.

---

Haem was extracted from *holo*-cytochrome *b*<sub>562</sub> as follows by Dr Ben Bowen of Cardiff School of BIOSI. Firstly, the pH of the sample reduced to pH 2 to dissociate haem from protein. Free haem was then removed from the sample by organic extraction. For this, an equal volume of butanone (methyl ethyl ketone) which had been chilled on ice for 1 hour was added and agitated. The sample was subsequently centrifuged at 16,000 ×g for 2 minutes to separate butanone, with dissolved haem, from protein. Haem extracted *apo*-cytochrome *b*<sub>562</sub> was finally exchanged into 50 mM Tris buffer pH using PD-10 desalting column (see Section 2.4.3).

## 2.5. Protein Characterisation

### 2.5.1. SDS-PAGE

Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse both protein size and purity <sup>266</sup>. Samples were prepared by boiling for 10 minutes by mixing of sample 1 in 4 with 4 x reducing loading buffer (40% glycerol (v/v), 240 mM Tris-HCL pH 6.8, 8.8% SDS (w/v), and 0.04% bromophenol blue (w/v), 5%  $\beta$ -mercaptoethanol (v/v)). For non-reducing gels, 4 x non-reducing loading buffer (40% glycerol (v/v), 240 mM Tris-HCL pH 6.8, 8.8% SDS (w/v), and 0.04% bromophenol blue (w/v)) was mixed with the sample instead. Whole-cell samples were prepared after resuspension of cell pellets to an OD<sub>600</sub> absorbance value of 1 A.U. After boiling samples were centrifuged for 1 minute at 15,400 xg to remove insoluble components before gel application. Both the separating and stacking gels were created fresh on the day of analysis according to specification as listed in Table 2.8.

Gels were prepared and run using the mini-PROTEAN electrophoresis system (BioRad, Hertfordshire, UK). Running buffer was composed of 25 mM Tris-HCL pH 8.3, 192 mM Glycine, and 0.1% (w/v) SDS. Gels were run at 200 V for 40 minutes, and BLUeye Pre-stained protein ladder (resolution 10-245 kDa)(GeneFlow, Staffordshire, UK) was used a reference of protein size. Gels were stained with 50 mL of protein stain solution (50% MQ water (v/v), 40% methanol (v/v), 10% acetic acid (v/v), and 0.1% Coomassie blue, R250 (w/v) for 1 hour. Gels were destained overnight with 50 mL destaining solution (50% MQ water (v/v), 40% methanol (v/v), and 10% acetic acid (v/v)). Image capture was performed using Gel Imagine Doc (UVP, CA,USA). Image exposure set to ensure the best dynamic range of the protein banding observed ensuring no overexposure or underexposure of bands. Where it was necessary quantification of each protein band observed was achieved using ImageJ software was used to analyse captured gel images <sup>267</sup>.

**Table 2.8. Composition of Separating and Stacking Gel for SDS-PAGE.**

Component	Separating Gel	Stacking Gel
Acrylamide <sup>a</sup>	12.5% (w/v)	5.00% (w/v)
Tris HCl	0.375 mM <sup>b</sup>	65 mM <sup>c</sup>
SDS	0.10% (w/v)	0.20% (w/v)
Ammonium Persulfate (APS)	0.05% (w/v)	0.10% (w/v)
TEMED	0.02% (w/v)	0.02% (w/v)

<sup>a</sup> Acrylamide/bis Acrylamide 37.5:1 40% solution, <sup>b</sup> pH 8.8, <sup>c</sup> pH 6.8.

### 2.5.2. Quantification of Protein Concentration

Concentration determination of new protein variants was undertaken using DC-Protein Assay (BioRad), per manufacturers instruction. Absorbance at 750 nm was recorded after mixing with 25  $\mu\text{L}$  reagent A (alkaline copper tartrate solution) and 200  $\mu\text{L}$  reagent B (Folin reagent) for 15 minutes at room temperature. Variant absorbance was subsequently compared with that of several standard concentrations of WT protein as appropriate. Absorbance was measured using a CLARIOstar microplate reader (BMG LABTECH, UK) and concentration was determined in Microsoft Excel 2019.

### 2.5.3. UV-Visible Absorption Spectroscopy and Calculation of Molar Absorbance Extinction Coefficient

UV-visible absorption spectra were recorded with Cary Eclipse spectrophotometer using 1 cm pathlength quartz cuvettes (Hellma, Müllheim, Germany). Absorbance was recorded between 200-800 nm at a scan rate of 300 nm/min To generate WT samples of known concentration, the absorbance spectrum of each WT protein was recorded and converted to concentration by use of Beer-Lambert Law (Equation 2.1) and extinction coefficient (Table 2.9).

**Table 2.9. Extinction Coefficients of WT Protein Standards.**

Protein	Extinction Coefficient ( $\text{M}^{-1} \text{cm}^{-1}$ )	$\lambda_{\text{MAX}}$
Holo-HRP (Plant)	100,000	403 nm
sfGFP <sup>WT</sup>	83,300	485 nm
mCherry <sup>WT</sup>	72,000	587 nm
Cytochrome <i>b</i> <sub>562</sub> <sup>WT</sup>	117,000	418 nm

To establish an extinction coefficient of each new variant, the Beer-Lambert Law (Equation 2.1), was applied to samples of known concentration to measure absorbance at  $\lambda_{\text{max}}$ . This value and the value of concentration were substituted into the rearranged Beer-Lambert law equation (Equation 2.1) to determine the molar extinction coefficient ( $\epsilon$ ). Proteins were subsequently re-quantified (see Section 2.5.1) to confirm the concentration of the diluted sample.

#### Equation 2.1. Beer-Lambert Law

$$A = \epsilon \times c \times l$$

Where **A** is absorbance at  $\lambda_{\text{max}}$ ,  $\epsilon$  is molar extinction coefficient ( $\text{M}^{-1} \text{cm}^{-1}$ ), **c** is concentration (**M**) and **l** is pathlength of light(**cm**).

### 2.5.4. Determination of Protein Haem Saturation

Haem saturation of protein was calculated by way of absorbance spectra as described (Section 2.5.3). The absorbance of protein at each haem protein's Soret peak (403 nm for HRP, and 418 nm for cytochrome  $b_{562}$ ) was compared with absorbance at 280 nm to calculate a Reinheitszahl ( $R_z$ ) value (Equation 2.2). An  $R_z$  value of 3 was taken as the value for complete saturation of HRP whilst a value of 4 was taken for complete saturation of cytochrome  $b_{562}$ . If a protein was not fully saturated with haem, a 100-fold molar excess of haem was incubated

#### Equation 2.2. Calculation of Reinheitszahl Value ( $R_z$ )

$$\text{Reinheitszahl Value} = \frac{\text{Soret Peak Absorbance}}{\text{Absorbance at 280 nm}}$$

with the protein for overnight at room temperature under gentle agitation. Free haem was removed by desalting by SEC (Section 2.4.2) and absorbance was recorded once again. This process was repeated until the generation of  $R/Z$  value equal to that of saturated protein.

#### 2.5.5. Fluorescence Spectroscopy

Excitation and emission spectra of fluorescent proteins were recorded on Cary Eclipse Fluorimeter (Agilent, CA, USA), using 5 mm x 5mm QS quartz cuvette (Hellma, Kruibeke, BE). A sample volume of 400  $\mu$ l was prepared to a working concentration of 0.5  $\mu$ M for each auto-fluorescent protein (sfGFP and mCherry), and 10  $\mu$ M for non-fluorescent proteins (HRP). Samples were excited using the  $\lambda_{MAX}$  of absorbance spectra (Section 2.5.3.) and emission was recorded at every 1 nm from the point of excitation to 700 nm. A scan rate of 120 nm/min was used for all spectra recorded with a 5 nm slit width and voltage set to medium. To measure the emission protein in reducing conditions protein was first incubated with a 10 fold molar excess of DTT overnight, before measurement.

#### 2.5.6. Quantum Yield Determination

Quantum yield (QY) of fluorescent proteins were calculated by comparison of emission with that of a control. Fluorescein (QY 0.95,  $\lambda_{max}$  496 nm) was used as a reference for sfGFP

#### Equation 2.3. Calculation of Quantum Yield

$$\frac{\Phi S_m}{\Phi S_t} = \left( \frac{\int S_m}{\int S_t} \right) \times \left( \frac{\eta S_m}{\eta S_t} \right)$$

Where  $S_t$  relates to fluorescein and  $S_m$  relates to sample.  $\int$  is the area of measurement and  $\eta$  is the refractive index of the solvent.

derivatives, and mCherry<sup>WT</sup> (QY 0.22,  $\lambda_{\max}$  587 nm) as a reference for mCherry variants. Fluorescein (Fisher Scientific) was prepared fresh in 0.1 M NaOH for each series of measurements. The concentration of both the reference sample and the protein variant was prepared to a  $\lambda_{\max}$  absorbance of 0.05 A.U. before the recording of emission spectra after excitation at each  $\lambda_{\max}$  as described previously (see Section 2.5.5.). Fluorescence intensity of each sample and standard was used in Equation 2.3. to calculate the QY of each variant.

### 2.5.7. Calculation of FRET Efficiency

The efficiency of Förster resonance energy transfer (FRET) between combined protein dimers or protein conjugated with a fluorescent probe was calculated experimentally by excitation of the fluorophore pair using the  $\lambda_{\text{EX}}$  of the donor fluorophore and recording of emission spectra up to 800 nm. The intensity of emission at the  $\lambda_{\text{EM}}$  of the donor and the acceptor fluorophore was recorded and utilised in Equation 2.4 to determine fluorescence energy transfer.

#### Equation 2.4. Calculation of FRET Efficiency

$$\text{FRET} = \frac{I_A}{(I_A + I_D)}$$

Where  $I_A$  is emission intensity of acceptor FP and  $I_D$  is emission intensity of donor FP.

### 2.5.8. Measurement of Red Edge Excitation Shift

Determination of protein red edge excitation shift (REES) was achieved by measurement of intrinsic protein fluorescence. Fluorescence was measured as described (Section 2.5.5.) using 10  $\mu\text{M}$  protein samples which were prepared in 50 mM Tris pH 8.0 Protein was then excited at 19 different excitation wavelengths (292-310 nm) and emission was recorded between 325 nm to 500nm at a scan rate of 30 nm/min. For all readings, excitation and emission slit width was set to 5 nm with a detector voltage set at high. A total of 12 repeat readings were recorded for each wavelength of excitation before averaging in Microsoft Excel 2019.

The average emission values for each wavelength of excitation was fitted to a skewed gaussian plot in OriginPro v 2020. Subsequently, the fitted plots were used to determine the fluorescence intensity of each protein at each wavelength of excitation (Equation 2.5). Then,

the centre of spectral mass (CSM) of each variant at each excitation wavelength was calculated (Equation 2.6) and plotted against the change in excitation wavelength. Whilst finally, REES curvature ( $R$ ) was calculated (Equation 2.7) and plotted against CSM and amplitude REES effect each emission spectra ( $A$ )<sup>268</sup>.

#### Equation 2.5. Calculation of Fluorescence Intensity

$$f_i = f_{max} \exp(-\ln 2) \left( \frac{\ln \left( 1 + \frac{2b(\lambda_{Em} - \lambda_{Em}^{max})}{w} \right)}{b} \right)^2$$

Where  $f_i$  is the measured fluorescence intensity,  $f_{max}$  is the maximum emission intensity at wavelength  $\lambda_{Em}^{max}$ , with a full width at half maximal of  $w$  and the 'skewness' is controlled by  $b$ .

#### Equation 2.6. Calculation of the centre of spectral mass (CSM)

$$CSM = \frac{\sum(f_i \times \lambda_{Em})}{\sum(f_i)}$$

Where  $f_i$  is the measured fluorescence intensity and  $\lambda_{Em}$  is wavelength of emission.

#### Equation 2.7. REES Magnitude.

$$CSM = CSM_0 + A e^{R\Delta\lambda_{Em}}$$

Where the amplitude and curvature of the exponential is described by  $A$  and  $R$  respectively and  $CSM_0$  is the CSM value independent of  $\lambda_{Em}$ .

#### 2.5.9. Steady-State Protein Kinetics

The enzymatic activity of HRP was determined using the substrate o-phenylenediamine dihydrochloride, (OPD) (Sigma). Each spot test of HRP activity was performed by adding 10  $\mu$ L of test protein sample to 300  $\mu$ L of substrate and peroxide solution (1 mM OPD, 1% H<sub>2</sub>O<sub>2</sub>, Tris 50 mM pH 8) and monitoring the formation of a visibly red precipitate by eye. The stock solution of OPD was prepared fresh for each set of data recorded at a concentration of 1 mg/mL (9.24 mM) and disposed of after 1 hour to prevent ambient oxidation. Stock concentrations of glycosylated *holo*-HRP and non-glycosylated *holo*-HRP<sup>EC2</sup> were prepared fresh for experimentation.

HRP and hydrogen peroxide (Fisher) were first added to cuvette as appropriate for the desired working concentration of protein (10 pM glycosylated *holo*-HRP, 4  $\mu$ M *holo*-HRP<sup>EC2</sup>, 3% (v/v) H<sub>2</sub>O<sub>2</sub>). The substrate, prepared fresh in Tris buffer (50 mM pH 8) was added for a working concentration ranging from 0.1mM to 5 mM, for a total reaction volume of 400  $\mu$ L. Immediately after the addition substrate, the sample was mixed by pipette. After a delay of 5 seconds to allow the escape of any air bubbles present absorbance measurement was initiated. Single wavelength absorbance at 450 nm was recorded every 0.25 seconds over a total duration of 30 seconds using Cary Eclipse spectrophotometer and disposable cuvettes (VWR). Absorbance change over the initial 30s of measurement was converted to substrate turnover per second using the extinction coefficient 59,000 M<sup>-1</sup> cm<sup>-1</sup> 269,270. Results were monitored to ensure a linear rate of turnover and recorded in quintuplets for each substrate concentration. Data for each variant were then fitted to Michaelis Menten plot in Origin V 2020 (MA, USA) to determine V<sub>MAX</sub>, K<sub>M</sub>, and k<sub>cat</sub>.

### 2.5.10. Temperature Dependence of HRP Enzyme Kinetics

Temperature dependent enzyme kinetics was achieved by calculating the rate of turnover of OPD by HRP as described previously (Section 2.5.9.). The reaction temperature was controlled by a circulating water bath for temperature regulation ( $\pm 1$  °C) (Aligent, CA, USA) and substrate turnover was measured between 20 °C to 40 °C. For each temperature of measurement, all reaction components were incubated at the desired temperature for 1 hour before measurement. OPD concentration equal to 10 x that of calculated k<sub>m</sub> as calculated previously (Section 2.5.8) was used as the substrate working concentration. Substrate turnover was recorded at 450 nm for a total of 2 minutes, and results were recorded in triplicate. Substrate turnover per minute was calculated by the change in absorbance per minute. Data were fitted to a curve of thermal turnover by way of Equation 2.8 in OriginPro v 8 using

**Equation 2.8. Optimum temperature of kinetic turnover.**

$$\ln k = \ln \frac{k_B T}{h} - \left[ \frac{\Delta H_{T_0}^\ddagger + \Delta C_P^\ddagger (T - T_0)}{RT} \right] + \left[ \frac{\Delta S_{T_0}^\ddagger + \Delta C_P^\ddagger (\ln T - \ln T_0)}{R} \right]$$

Where  $T_0$  is an arbitrary reference temperature.  $\Delta C_P^\ddagger$  is the difference in heat capacity between the ground and transition states.  $\Delta C_P^\ddagger$  determines the change in  $\Delta H^\ddagger$  and  $\Delta S^\ddagger$  with temperature and thereby defines the non-linearity of the temperature dependence of the Gibbs free energy difference between the ground state and the transition state ( $\Delta G^\ddagger$ ).

---

Equation 2.8. The temperature at which the peak rate was observed by thermal efficiency curve was taken as the optimum temperature ( $T_{opt}$ ) of turnover.

### 2.5.11. Mass Spectroscopy

Protein samples (50  $\mu$ M) were prepared in Tris 50 mM pH 8.0 for mass spectrometry analysis by Cardiff University School of Chemistry's Mass Spectrometry service by way of time-of-flight mass spectrometry (TOF-MS). Data was collected on a Waters Synapt G2-Si QT in positive Electrospray ionisation mode. Proteins were passed through a Waters Acquity UPLC CSH 130 C18 (80°C) and eluted using a gradient of acetonitrile (5-95%). Mass predictions were made using the ExpASY Peptide Mass tool<sup>271,272</sup>. Glycosylation mass of protein was determined by subtraction of predicted protein mass of P00433<sup>81</sup> from observed protein weights. Identity of each glycan unit was then calculated using ExpASY GlycanMass and ExpASY GlycoMod tools<sup>273,274</sup>.

### 2.5.12. Growth of Protein Crystals

The crystallisation of HRP was achieved using 100 mg/mL samples of glycosylated *apo*-HRP in Tris 50 mM pH 8.0. The handling of all liquid was executed by Art Robbins Phoenix robot (Alpha Biotech, UK) at Cardiff University School of Medicine. Two forms of commercially available crystal screen were used for the growth of crystals: PACT *premier*<sup>TM</sup> HT-96 screen and JCSG-plus<sup>TM</sup> HT-96 screen (Molecular Dimensions, Suffolk, UK). Sitting drop vapour diffusion was used in the incubation of 0.2  $\mu$ L of protein with 0.2  $\mu$ L of reagent at 20°C. The growth of crystals was monitored by use of light microscope over the course 0, 1, 3, and 7 days of growth, and then once again eight weeks later.

To aid formation of crystals of 100 mg/mL of HRP was mixed with 12.5%, 25%, 37.5% and 50% (w/v) sucrose. Conditions which were shown to yield protein crystals were that of 1.4 M  $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7$ , 100 mM HEPES, pH 7.5 when protein was mixed with 37.5% (w/v) sucrose concentration. Crystals were harvested by mounted litho-loop (Molecular Dimensions, Sheffield, UK) and plunged immediately into liquid nitrogen for transportation. Diffraction of formed crystals was accomplished at Diamond Light Source, (Harwell, UK), by Dr Pierre Rizkallah, of Cardiff University School of Medicine.

### 2.5.13. Circular Dichroism

Circular Dichroism measurements were achieved using 10  $\mu\text{M}$  protein in 5 mM Tris pH 8.0. Measurements used were obtained at Bath University School of Chemistry on Chirascan™ CD Spectrophotometer (Applied Photo Physics, Surrey, UK), using 0.1 cm pathlength quartz cuvette after blanking the spectrophotometer with 5 mM Tris pH 8.0. Light polarisation by HRP was measured between 190 and 260 nm, at 1 nm intervals at a scan rate of 7.5 nm/min. Data was recorded in millidegrees and converted to molar ellipticity in Excel using Equation 2.9. Far UV spectra of each protein variant were recorded initially at 293 K and repeated after the thermal unfolding of protein at 363 K.

The mid-point of thermal denaturation ( $T_m$ ) of HRP was calculated by heating of protein over a temperature range. The temperature of protein was increased at a ramp rate of 1 K per minute by Quantum Northwest Peltier and polarisation at 222 nm was recorded at every K change up to 363 K. Data obtained was converted to molar ellipticity by way of Equation 2.9 and fitted to a Boltzmann distribution function by way of Equation 2.10 in Origin V 2020 (MA,

**Equation 2.9. Conversion of millidegrees to molar ellipticity.**

$$\theta_{222nm} = \frac{100 \times \theta}{(C \times L)}$$

Where  $C$  is concentration in molar,  $L$  is the pathlength in cm and  $\theta$  is ellipticity.

**Equation 2.10. Fitting of Molar Ellipticity at 222 nm to Boltzmann Distribution Function**

$$Y = \frac{A1 + (A1 - A2)}{(1 + \exp\left(\frac{(x - Tm)}{dx}\right))}$$

Where  $A1$  max molar ellipticity is and  $A2$  is minimum molar ellipticity.  $Dx$  is change in temperature divided by 20.

USA).

#### 2.5.14. UV Irradiation of mCherry

Photolysis of 1  $\mu\text{M}$  mCherry<sup>WT</sup> and 1  $\mu\text{M}$  mCherry<sup>198AZF</sup> was executed in 5 mm x 5 mm quartz cuvettes. Protein was irradiated for defined time periods at 302 nm (spectrum from 275 – 380 nm) by UVM-57 mid-range 6W UV lamp (UVP, Cambridge, UK) at 1 cm distance. Post

---

irradiation absorbance and fluorescence measurements were made as described previously (Section 2.5.3, and 2.5.5).

#### 2.5.15. De-Glycosylation of Plant Protein

Trial cleavage of glycans from glycosylated *apo* and *holo*-HRP was achieved using PNGase A (New England Biolabs) per manufacturer's instruction. Glycoprotein (50 µg) was combined with 10 µL of glycoprotein denaturing buffer (10 x) and heated at 100 °C for 10 minutes. Subsequently, the protein was chilled on ice and centrifuged to collect all liquid to the base of the tube. The reaction mix (10 µL of denatured HRP with 20 µL GlycoBuffer 3 (10 x), 2 µL of 1 x NP-40, 60 µL H<sub>2</sub>O, and 50 units PNGase A) was formed and incubated at 37 °C for one hour. A negative control sample, not supplied with PNGase A, was additionally prepared and incubated at 37 °C for one hour. Following incubation SDS-PAGE of both treated and the negative control was performed to determine if a change in size had been initiated.

#### 2.5.16. TEV Cleavage

Trial TEV cleavage of the TEV recognition site (ENLYFQ/G) engineering into HRP<sup>EC2</sup> was implemented by mixing 100 µL 50 µM TEV protease (Generated by Dr Husam Al-Maslookhi, Cardiff University School of BIOSI)<sup>275</sup> with 100 µL 10 µM HRP<sup>EC2</sup>. Protein and protease were incubated for overnight at room temperature before the sample was analysed by SDS PAGE (Section 2.5.1.). The sample was run against a control non-treated sample of HRP<sup>EC2</sup> to determine if a size decrease had occurred.

---

## 2.6. Trial of *apo*-HRP<sup>EC2</sup> in Immunoassay

### 2.6.1. Sample Preparation

The action of HRP<sup>EC2</sup> in commercial immunoassay was investigated to determine if SCF's function was replicated by recombinant protein. HRP<sup>EC1</sup> was produced and purified in the absence of haem (Section 2.3.1. and 2.4.5), and the calculation of  $R_z$  (Section 2.5.4) of pure protein was used to confirm the presence of apo-protein. Protein was lyophilised to match that of commercial SCF so that powdered protein could be spiked into assay reagents as desired. Protein was transported to OCD (Pencoed, UK) as lyophilised powder on dry ice and immediately stored at -80 °C until use.

Lyophilised batches of *apo*- cytochrome  $b_{562}^{WT}$  and mCherry<sup>WT</sup> were also used as experimental controls. *Apo*-cytochrome  $b_{562}^{WT}$  was generated to serve as a control for the application of a haem binding agent, whilst mCherry<sup>WT</sup> was used to indicate the effects of the addition of recombinant protein in general to immunoassay.

### 2.6.2. Calibration of Assay

The cardiac troponin diagnostic assay TROP I (OCD) was implemented as instructed in the manufacturer's protocol on Vitros 250 Chemistry Analyser (OCD). Assay was calibrated in both the presence and absence of SCF to ensure consistent readings of each reference calibration sample (0.03 ng/mL, 0.07 ng/mL, 1 ng/mL, 15 ng/mL, 55 ng/mL serum cardiac troponin). Once the standard curve had been created, additional quality control reference samples, of a known troponin concentration, were analysed to ensure the correct quantification was achieved experimentally. Then serum samples were measured for cardiac troponin presence both in the presence and absence of SCF. Those which showed false reporting of troponin concentration were selected for investigation.

### 2.6.3. Method of Immunoassay

Each recombinant protein (Section 2.6.1) was spiked directly into aliquots of two serum separate false serum samples and allowed to equilibrate for 1 hour before investigation. As human serum is acellular no ethical approval was necessary for this investigation. Protein was spiked to a concentration of 2.1875 mg/mL equal to a working concentration of 10 x that of SCF (Table 2.10).

**Table 2.10. Concentration of Corrective Agent.**

Component	Volume	SCF Concentration within Commercial Assay	Experimental Concentration
Biotin Conjugate Reagent	35 $\mu$ L	0.5 mg/ mL	-
HRP Conjugate Reagent	35 $\mu$ L	-	-
Sample	80 $\mu$ L	-	2.1875 mg/mL
Working Concentration	150 $\mu$ L	0.117 mg/mL	1.17 mg/mL

Quantification of serum experimentally was achieved per standard operating procedure using Vitros 250 Chemistry Analyser (OCD) Trop I protocol <sup>276</sup>. The assay was implemented by mixing 35  $\mu$ L biotin conjugate reagent (without SCF), 35  $\mu$ L HRP conjugate reagent, and 80  $\mu$ L of test serum into a streptavidin-coated well (OCD). The sample and reagent mix was incubated for 10 minutes to allow binding. Then wells were washed with VITROS Immunodiagnostic Products Universal Wash Reagent (OCD) to remove all non-bound entities.

Signal was generated from wells by application of VITROS Immunodiagnostic Products Signal Reagent (OCD) and measurement of light at 428 nm emitted by the turnover of luminol. The intensity of light emitted for each sample was subsequently converted to signalled troponin concentration by use of the previously established calibration curve (Section 2.6.2). Reported concentration of troponin of each sample was compared to the negative control, a serum which had not been spiked with any corrective agent, to determine the change in reported troponin concentration. The percentage of signal change was then calculated between each spiked protein and the negative spiked control reference in Microsoft Excel 2019.

## 2.7. *In silico* Molecular Modelling

### 2.7.1. Robetta Modelling

Structure prediction of full-length proto-HRP with the pre and pro sequences intact was obtained using the Robetta online server ([rosetta.bakerlab.org](http://rosetta.bakerlab.org))<sup>277,278</sup>. For this, the PDB File for non-glycosylated *holo*-HRP, 1H58, was submitted as a PDB template of the complete amino acid sequence of HRP C1A (ProtParam P00433). The obtained output PDB was subsequently aligned with 1H58 in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC) so that any structural variation could be identified.

### 2.7.2. Molecular Dynamics

Molecular dynamic simulations were run by Dr Harley Worthy (Cardiff University, School of BIOSI) using the PDB entry for HRP, 1H58, both with and without haem porphyrin incorporated. The protein simulation was achieved over a 500 ns total duration at 100 ns intervals with each successive simulation taking up where the previous finished, under standard temperature and pressure (25 °C, 1 atmosphere). Cardiff University supercomputer was used to run the simulation using GROMACS MD software<sup>279</sup>. The difference in RMSF between *apo* and *holo* simulation was then converted into 10 different percentile categories in Microsoft Excel 2019, and the colour was assigned to each category (Table 2.11). This colouration was subsequently applied to residues within 1H58 in PyMOL to visualise regions of protein flexibility which are influenced in the presence of haem binding.

**Table 2.11, Colouration of Residues Selected to Present Calculated Difference In RMSF**

Difference in RMSF Percentile		Colouration Applied to Residue
More Flexible in <i>apo</i> -HRP	0% - < 10%	Density
	10% - < 20%	Deep Blue
	20% - < 30%	Sky Blue
	30% - < 40%	Slate
	40% - < 50%	Grey
More Flexible in <i>holo</i> -HRP	50% - < 60%	Grey
	60% - < 70%	Salmon
	70% - < 80%	Dark Salmon
	80% - < 90%	Red
	90% – 100%	Firebrick

### 2.7.3. Prediction of Potential Regions of Protein Interface

The ClusPro Protein-Protein server “Dock” function<sup>280</sup> was utilised to predict sites of the potential interface between proteins. For this, a reference PDB of each protein investigation (Table 2.12) was supplied to the server. In the combination of sfGFP (2B3P) with either mCherry (2H5Q) or cytochrome *b*<sub>562</sub> (1QPU), sfGFP was used as the receptor for simulation.

ClusPro simulation was initiated by the rotation of the “ligand” around the “receptor.” Rotation occurred on three planes of rotational axis to sample all possible orientations with potential to interface together. Ranking of potential interfaces was performed by ClusPro by way of Fourier correlation to eliminate all regions of interface with were not compatible based on intermolecular repulsion. Resultant alignments of protein were subjected to “clustering,” in which interfaces were amassed as one potential region interface if occurring within a 9 Å radius. The models were ranked by the number of clustered simulations as calculated in balanced simulation (recognition of all intermolecular electrostatics), and the top ten ranked interfaces were downloaded as PDBs from the server for further analysis. Finally, both the cofactor of cytochrome *b*<sub>562</sub> and the chromophores of each fluorescent protein were reincorporated back into each structure in PyMOL.

**Table 2.12, PDB Reference Entries for ClusPro simulation**

Protein	PDB Entry
sfGFP	2B3P
mCherry	2H5Q
Cytochrome <i>b</i> <sub>562</sub>	1QPU

### 2.7.4. Scoring of Generated Molecular Interfaces

Dynamic modelling of the molecular interfaces identified in the top 10 models obtained from ClusPro was further refined in ROSETTA’s high-resolution docking protocol<sup>281–283</sup>. RosettaDock protocol ranked the predicted interfaces obtained within ClusPro simulation (Section 2.7.3) by both “Total Energy” and “Interface Energy.” In both rankings, the lower predicted scores the greater the likelihood of protein alignment at the specified interface *in vivo*. The predicted protein alignments obtained by this method were therefore taken as the most energetically favourable site of nnAA incorporation.

### 2.7.5. PyMOL Modelling of Protein

Visualisation of protein was routinely performed in PyMOL Molecular Graphic System v 2.0<sup>284,285</sup>. Representation of protein glycosylation was achieved by way of Azahar plugin, which was used to build carbohydrate appendages for 1H58<sup>286</sup>. Whilst incorporation of nnAA into PyMOL was achieved *via* Swiss Sidechains plugin<sup>287</sup>(Figure 2.2) for the representation of nnAA incorporating proteins generated in this investigation (Figure 2.2).

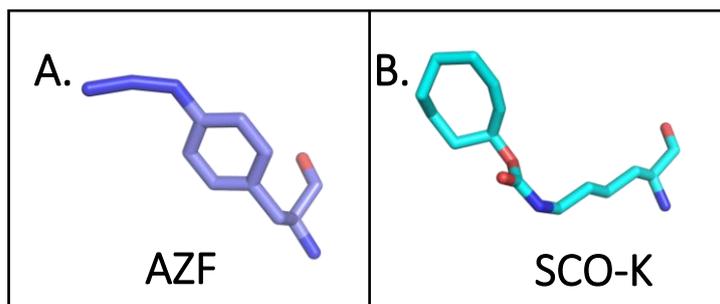


Figure 2.2. 3D Stick Representation of Each nnAA in PyMOL.

A, representation of p-azido-L-phenylalanine. B, representation of s-cyclooctyneL-lysine.

---

## 2.8. Click Chemistry

### 2.8.1. Detection of AZF incorporation *via* Click Chemistry

Incorporation of AZF to protein was assessed with a Dibenzocyclooctyne-amine (DBCO) conjugated fluorescent probe. Cy3 DBCO probe (Click Chemistry Tools, AZ, USA) was mixed with protein at 10 x molar concentration and left to react overnight in dark conditions at room temperature. Reaction with protein was confirmed after SDS PAGE when a band of the specified protein size was observed by UV light (Bio-Imaging Systems). This band was later observed by Coomassie stain to confirm the presence of the appropriately sized protein (Section 2.5.1)

### 2.8.2. Formation of Oligomeric Protein by Strain-Promoted Azide-Alkyne Cycloaddition (SPAAC)

Strain promoted azide-alkyne cycloaddition (SPAAC) of protein was achieved by mixing of appropriate nAA protein variants at 100  $\mu$ M overnight in dark conditions at the desired temperature (room temperature for the combination of sfGFP and mCherry and 37 °C for the combination of sfGFP and cytochrome *b*<sub>562</sub>). After incubation, the formation of oligomeric protein was determined by SDS-PAGE (Section 2.5.1.). Purification of formed oligomeric protein was achieved by size exclusion chromatography using Hiload™ 26/600 Superdex™ S200 gel filtration column (Section 2.4.2) and the purity of the resultant protein oligomer was assessed by SDS PAGE (Section 2.5.1).

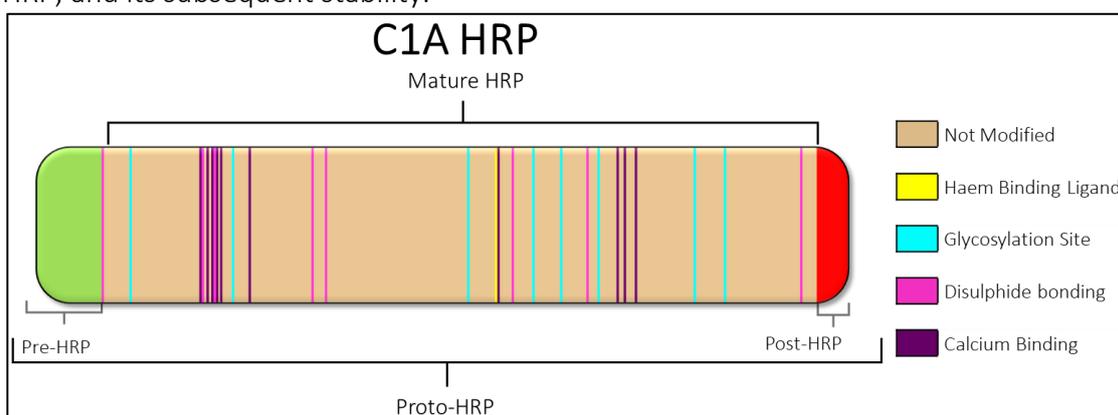
## **Chapter 3.** The Soluble Expression of Recombinant Horseradish Peroxidase in *E.coli*.

## 3.1. Introduction

### 3.1.1. The Commercial Production of HRP

Commercially available HRP is classically obtained from the cultivation of the native host *Armoracia rusticana*. Consequently, produced HRP is heterogeneous, due to the variation of both isoenzyme presence and glycosylation pattern <sup>55</sup>(see Chapter 1.2.4.). This has established a commercial enzyme product in which cultivation is both seasonally and spatially inconsistent <sup>65</sup>, which varies between both batch and supplier <sup>73</sup>. Furthermore, the expression of HRP from alternative plant hosts is unfavoured due to their long cultivation time, cyclical availability and low recovery yield <sup>54</sup>. It is proposed that microbial recombinant expression of HRP will avert heterogeneity of the enzyme product. Nevertheless, expression of HRP is complex and multiple trade-offs to PTMs are necessary when producing an active enzyme in microbial cells (see Chapter 1, Table 1.4).

The existence of several forms of complex PTMs on HRP is a hindrance to prokaryotic expression. Of the 353 amino acids in C1A HRP, 22% (70 amino acids) are either cleaved or modified post expression (Figure 3.1)<sup>288</sup>. When expressed with native PTM, HRP is highly thermostable, and is peroxide tolerant<sup>289</sup>. However, when expressed with either modification or elimination of PTM expression yield, recovery yield and peroxidase efficiency is known to decrease <sup>66</sup>. Furthermore, as enzymes within a cell control the application of PTM, and the process is not genetically regulated, the cell choice for expression greatly influences the PTM of HRP, and its subsequent stability.



**Figure 3.1. Representation of PTM of C1A HRP.**

Locations of the 70 amino acids which are cleaved or modified by PTM, each amino acid coloured in accordance with PTM form. Diagram made to scale with regions labelled according to location of native cleavage site. Bands vary in thickness in accordance with the number of amino acids modified at each location

---

### 3.1.2. The Influence of Host on the Expression of HRP

Numerous research teams have selected to remove many forms of PTM when producing recombinant HRP. For example, expression of HRP in *E. coli* cells prevented the application of glycans and the formation of disulphide bonds within the cytosol<sup>75</sup>. Alternatively, expression of HRP in *P. pastoris* retains glycosylation of the protein, however, glycosylation is not consistent with that applied in *A. rusticana*<sup>80</sup>. Both strategies have previously produced active protein; however, each has also produced low yield with HRP product of comparably lower catalytic efficiency than that of native expression.

To date, the maximum yield of HRP C1A from a plant host is 3 mg/L of protein obtained from *N. tabacum* after 7 days of growth<sup>289</sup>. This is markedly smaller than the yield of C1A obtained by microbial expression, which produces a larger quantity of enzyme, and does so over a shorter time. Comparatively, the yield of C1A produced from *P. pastoris* exceeds that of plant systems with 7.3 mg/L from 87 hours of growth<sup>77</sup>. However, the greatest expression reported is that of 10 mg/mL from the overnight expression of *E. coli*<sup>79</sup>. Although, C1A which is formed in *E. coli* is typically isolated within cellular inclusion bodies and must be re-solubilised to obtain active enzyme.

The many benefits of increased growth and expression offered from *E. coli*, comparatively offset the advantage of a decreased protein solubility of the non-glycosylated enzyme. *E. coli* cells, for example, offer a maximal generation time of 20-minutes<sup>34</sup> and at maximum efficiency can direct 50% of their metabolic resources to target recombinant protein expression<sup>36</sup>. Therefore, *E. coli* cells are the most commonly selected recombinant host for HRP expression<sup>66</sup>.

### 3.1.3. Problems Faced by The Expression of HRP in *E. coli*

The expression of HRP in *E. coli* produces an enzyme which is devoid of glycans, unable to form disulphide bonds when expressed in the cytosol and is commonly produced as inactive apo-protein<sup>66</sup>. Furthermore, expression of HRP in *E. coli* typically does not feature the expression of the full-length proto-HRP. Expression constructs, to date, have only featured the amino acids which are present in mature-HRP with the omission of both pre- and post-HRP regions (See Figure 3.1 and Chapter 1.2.8.). Thus, the expression of HRP from the truncated sequence in *E. coli* cells is now commonplace<sup>54,64,79,290</sup>.

---

When expressed in *E. coli*, HRP is sequestered in insoluble inclusion bodies<sup>74,75,109</sup> and must be both refolded<sup>64</sup> and haem saturated<sup>74</sup> to generate active protein. One potential solution to this is the supplementation of media with 5-aminolevulinic acid (ALA) which increases the efficiency of the haem biosynthesis pathway. However, this technique does not eliminate the necessity of haem saturation at a later stage as approximately 32 % of protein is still expressed in *apo* form<sup>291</sup>.

The widespread omission of PTMs suggests a general ethos shared by research groups is to express a high yield of misfolded HRP as quickly as possible and to then maximise the recovery of the active enzyme form from the insoluble cell fraction. This approach, however, overlooks the significance of PTM, namely glycosylation, on protein activity, solubility, peroxide tolerance and thermostability, and recovers just 2-10% of all HRP expressed<sup>75,79</sup>. Nevertheless, the method does allow for the recovery of the greatest quantity of active enzyme, rather than the greatest stability and activity of expressed HRP.

#### 3.1.4. Expression of the Full-Length Sequence of Proto-HRP

The production of HRP natively in *A. rusticana* features the expression of a 353 amino acid proto-peptide. Then post-translationally, excision of the first 30 and the final 15 amino acids occurs for the formation of mature-HRP<sup>66</sup>. The pre-peptide (N-terminal) sequence has been shown to localise HRP and recombinant GFP alike to the ER when N-terminally linked<sup>292</sup>. In the native expression of HRP, expressed protein is exported to the ER by the pre-peptide sequence. Localisation to the ER then allows for the application of glycosylation and the formation of disulphide bonds<sup>65</sup>. Whilst the post-HRP (C-terminal) region of HRP is of disputed functionality<sup>292</sup>, but it is generally considered that the sequence is responsible for vacuolar targeting<sup>289</sup>.

To date, the expression of HRP in *E. coli* has featured an abridged genetic sequence for the replication of initial experimentation by Smith *et al* 1990<sup>75</sup>. This excludes the presence of both pre- and post-HRP regions from the genetic construct and omits the post expression transition from proto-HRP to mature-HRP. Instead, this technique expresses the shorter mature-HRP sequence alone. It was reasoned, that as *E. coli* features no membrane-bound organelles the proto-HRP/mature-HRP transition is redundant. Especially, as Smith *et al* (1990) proved neither termini were essential for the expression of active HRP<sup>75,293</sup>. This suggests that

---

unlike other common prepropeptides, such as serine proteases, HRP is not a true zymogen and is not activated by proto-peptide folding and subsequent cleavage<sup>294</sup>.

Modern understanding and analysis of the pre-peptide regions, however, indicate additional benefit other than that of localisation, when HRP is expressed as a proto-peptide. Experimentation by both Matsui *et al.* (2003), and Kis *et al.* (2004) found that expression of HRP in the plant species *N. tabacum* with the N-terminal pre-peptide sequence was linked to an increased rate of expression<sup>289,292</sup>. The observed increase in expression indicated that the benefits provided by the sequence were not limited to post-translational events within the ER. The assessment given to this by Matsui *et al.* was that the pre-peptide region was important for protein function, as the study observed no HRP activity in the absence of the pre-peptide sequence. This suggested the immediate degradation of expressed protein without the pre-peptide region and not just the expected delocalisation from the ER<sup>289</sup>.

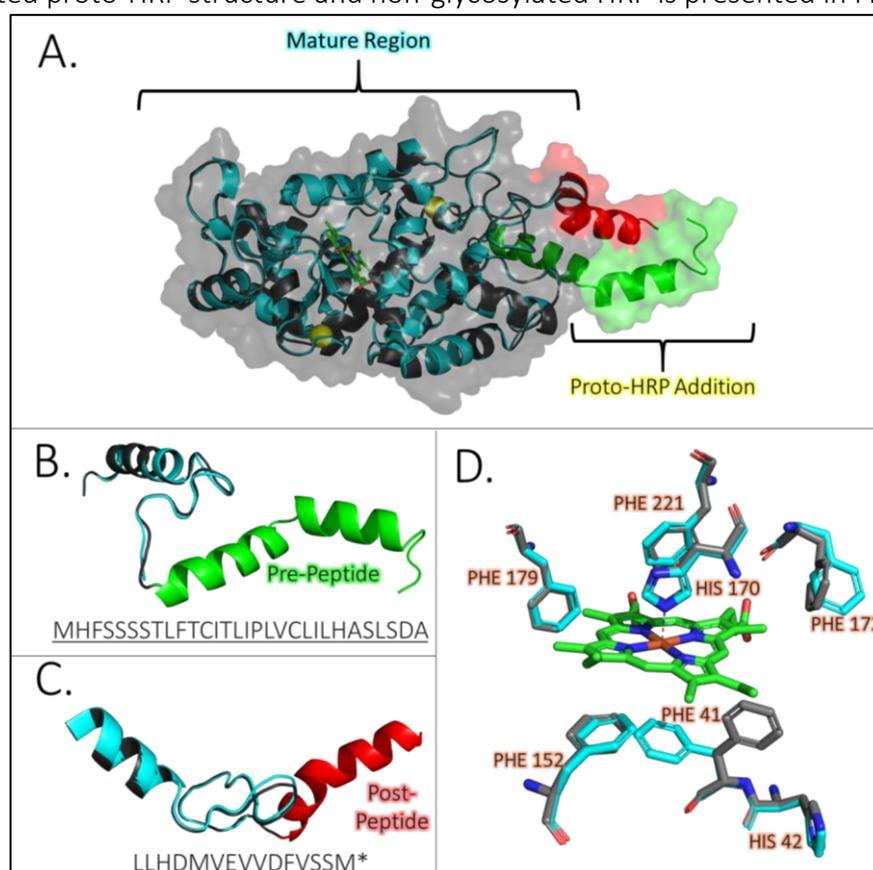
### 3.1.5. Chapter Aims

It is currently unknown if the benefits of proto-HRP expression observed by both Matsui *et al.* and Kis *et al.* will be replicated by expression in prokaryotic systems. Within this chapter, it will be investigated if the expression of full-length proto-HRP aids the formation of soluble HRP when utilised in *E. coli* cells. The presence of the proto-HRP regions is herein reasoned to enhance HRP solubility, and, consistent with the observation by Matsui *et al.* (2003), their absence impedes soluble protein formation. It is proposed that expression of HRP, without proto-HRP regions, in *E. coli* promotes the isolation of protein to inclusion bodies, as observed previously<sup>49,64,109,114</sup>. Thus, the inclusion of the proto-HRP regions in an *E.coli* expression construct will, therefore, benefit the acquisition of a soluble final active enzyme product for later investigation.

## 3.2. Results and Discussion

### 3.2.1. Structural Influence of Proto-Peptide Regions on HRP.

To predict what structural influence the pre- and post-peptide inclusion would exert upon HRP, structure prediction of full-length proto-HRP was performed on the webserver, Robetta. The Robetta Server is a web-based server which predicts protein structure based upon a supplied amino acid sequence and features the option of using an existing PDB file as an initial structure template. The complete sequence for HRP C1A as acquired from the UniProt Database (P00433) was supplied to the Robetta server (<https://robetta.bakerlab.org/>) with non-glycosylated mature HRP (1H58) additionally supplied as a structural reference. Alignment of the predicted proto-HRP structure and non-glycosylated HRP is presented in Figure 3.2.



**Figure 3.2.** The structural influence of the N and C regions on the structure of HRP.

**A**, overlay of cleaved HRP (PDB 1H58)(Teal) with model of the full proto-peptide sequence as output from Robetta Structure Prediction (**Black**). The two additional regions of the proto-peptide, the N-terminus (**Green**), and the C-terminus (**Red**) shown and labelled. **B**, overlap of the N-terminus (sequence given) of the predicted proto-peptide in addition with the first 30 amino acids of the cleaved protein. **C**, overlap of the C-terminus (sequence given) of the predicted proto-peptide in addition to the final 30 amino acids of the cleaved protein. **D**, catalytically active residues of HRP shown as sticks with those relating to mature HRP shown **Teal** and proto HRP shown **Grey**, and haem porphyrin displayed as **Green Sticks**.

---

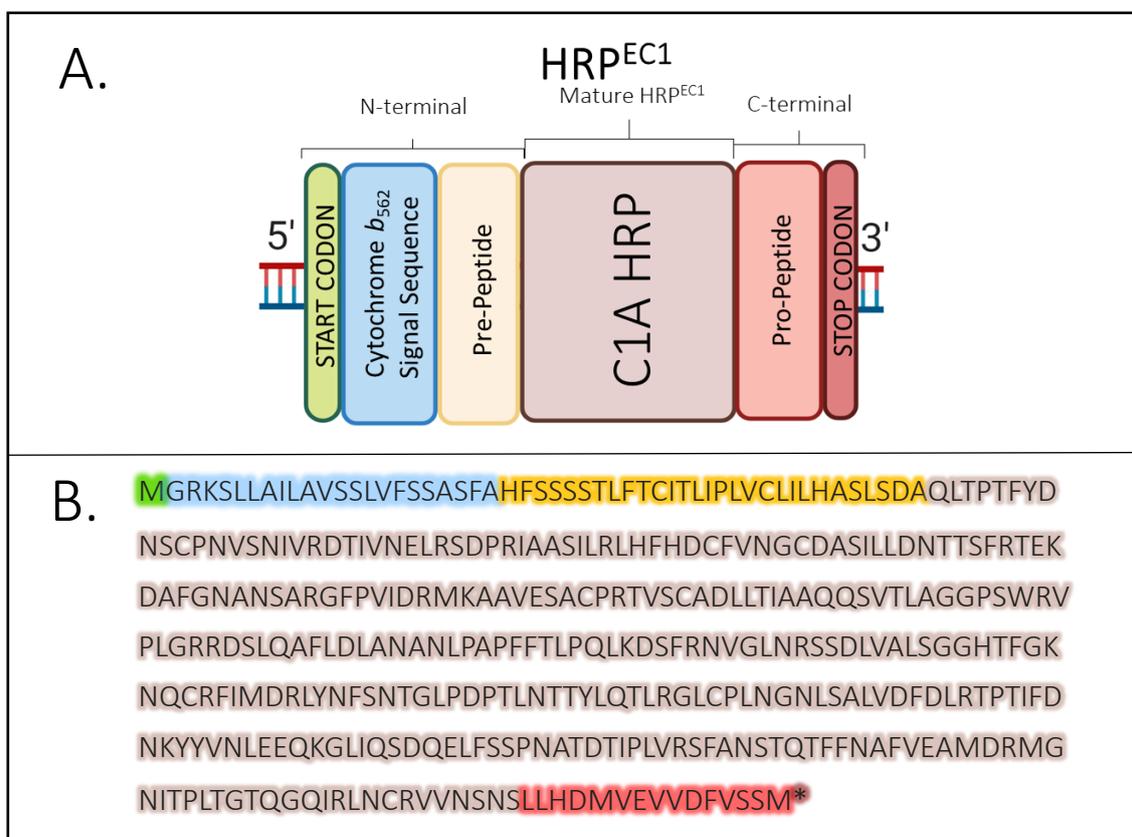
In the predicted structure, the generated model of proto-HRP (with pre- and post-peptide regions) was matched to the known the structure of mature HRP (PDB 1H58). This was expected as both shared the same amino acid sequence and 1H58 was used as a reference template (Figure 3.2). However, the pre-peptide of HRP was observed to be  $\alpha$ -helical in structure with two separate  $\alpha$ -helices (Figure 3.2, B), while the post-peptide region was additionally shown to form a single  $\alpha$ -helical strand (Figure 3.2, C). Crucially, both the cleaved protein termini were predicted to associate together, minimising their anticipated influence upon the structure of HRP (Figure 3.2, A).

Analysis of catalytically active residues indicated some internal modification had been initiated by the inclusion of the proto-peptide sequences (Figure 3.2, D). The residues H42, F152, H170, F179, and F221, were all shown to be unaffected by the addition of the proto-HRP regions and overlay directly with those observed in mature-HRP. However, residues F41 and F172, both show alignment shifts between their aromatic rings with haem. F41, which is aligned 31 Å from the pre-peptide and 35 Å from the post peptide, displays a shift in its aromatic ring orientation, from a parallel alignment with haem to a perpendicular alignment (Figure 3.2, D). It is not known what the functional implications of this shift will be, as F41 is a residue which has been previously hypothesized to have substrate-binding capability<sup>295,296</sup>. However, it is likely, that the change in F41 alignment is not an influence of the proto-peptide presence but rather, an artefact of Robetta simulation which excludes cofactors and predicts only the structure of *apo*-protein.

### 3.2.2. Design of Expression Construct HRP<sup>EC1</sup>

It was rationalised that to commence investigation both pre- and pro-HRP sequence regions would be included in the expression construct of HRP. Although neither region was predicted to cleave or localise protein intracellularly, as *E.coli* does not feature membrane-bound organelles. It was reasoned that expression of the full proto-HRP sequence would test the hypothesis that the pre and post peptide regions aid protein solubility. The proposed gene construct (Figure 3.3) of HRP expression (HRP<sup>EC1</sup>), was designed in Serial Cloner as described in Chapter 2.2.1 and cloned into the pET-24b plasmid (Chapter 2, Figure 2.1) by GenScript. The gene was placed under the control of the T7 promoter for IPTG induction, with the kanamycin

resistance gene used as a selection marker. The constructed gene was 1380 bp in size and formed a plasmid sized at 6689 bp.



**Figure 3.3. The Generated Gene Construct for the Initial Expression of HRP<sup>EC1</sup>.**

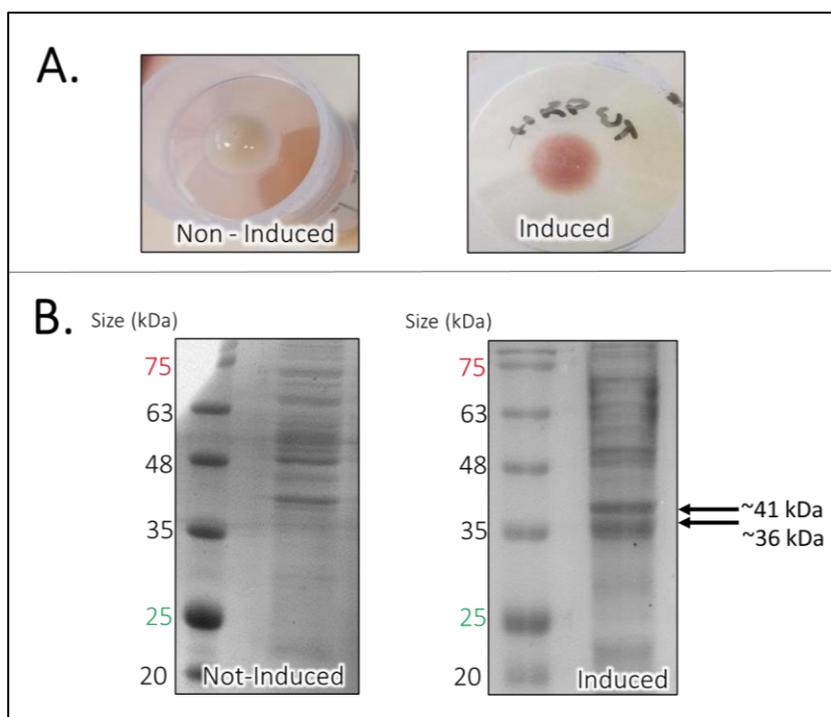
**A**, graphic of the gene construct for expression of HRP<sup>EC1</sup>. **B**, amino acid sequence of construct with regions coloured according to sequence feature as displayed in A.

In addition to the genetic sequence of C1A HRP, the periplasmic export sequence, of cytochrome *b*<sub>562</sub>, as described by Sockolosky *et al.*<sup>297</sup>, was incorporated upstream of HRP (Blue, Figure 3.3, B). The inclusion of the periplasmic export sequence was rationalised, as it would firstly, increase the ease and efficiency of protein purification as cell lysis would not be necessary; and secondly, to aid in the formation of disulphide bonds, by the localisation of HRP to the non-reducing periplasm. The analysis of the constructed sequence on ProtParam<sup>271</sup> indicated HRP<sup>EC1</sup> would weigh 41,032 Da, or if the cytochrome *b*<sub>562</sub> signal sequence was removed, 38,694 Da .

### 3.2.3. Expression of HRP<sup>EC1</sup>

HRP<sup>EC1</sup> cloned into the pET-24b plasmid obtained from GenScript was transformed into *E. coli* SHuffle<sup>®</sup> cells. SHuffle<sup>®</sup> cells were selected for expression as the cells featured a chromosomal copy of the disulphide bond isomerase enzyme, DsbC, to promote the formation

of disulphide bonds in the non-reducing cell cytoplasm<sup>298</sup>. Cells were initially grown to a density at OD<sub>600</sub> of 0.6 before induction with 1 mM IPTG and incubation for 24 hours at 25 °C for protein induction (see Chapter 2.3.1.). Cells induced with IPTG were observed to change colour from off white to pale pink (Figure 3.4, A). This colour was reasoned to be indicative of expression of a haem binding protein and was consistent with that of the increased haem biosynthesis pathway, mirroring colouration observed from Cytochrome *b*<sub>562</sub> expression<sup>299</sup>.



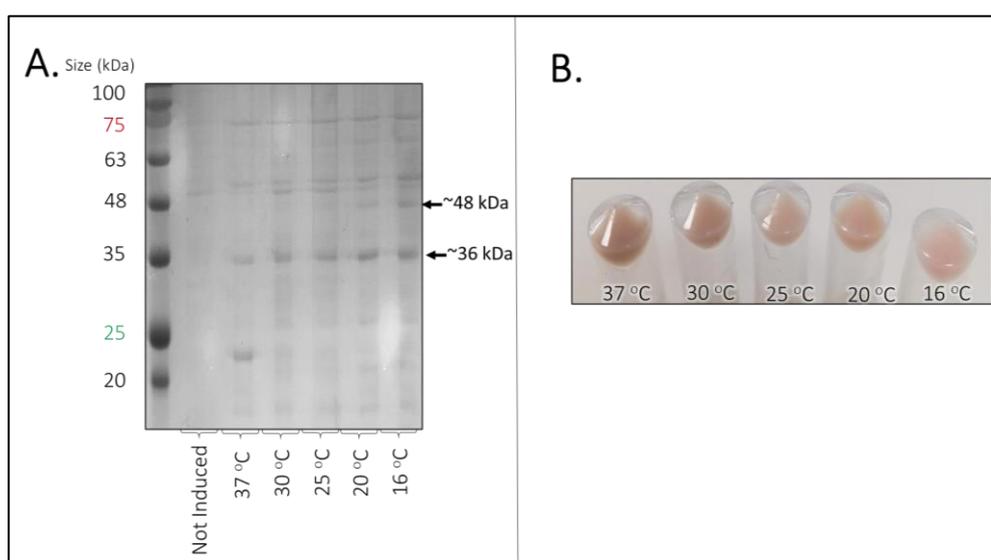
**Figure 3.4. The Overnight Induction of HRP<sup>EC1</sup> Expression With 1 mM IPTG**

**A**, the colour change observed in cells which were induced with 1 mM IPTG. **B**, SDS PAGE of soluble and insoluble protein expressed in the absence and the presence of IPTG. The presence of two bands of over expression are observed in cells which are induced.

Cells induced with IPTG showed the presence of two bands of overexpression at approximately 36 kDa and 41 kDa when viewed by SDS PAGE (Figure 3.4, B). These bands were equal in intensity and the band at ~41 kDa was consistent with that of full length proto- HRP<sup>EC1</sup>. Although, the band at ~36 kDa was smaller than that expected for HRP<sup>EC1</sup> which has been cleaved of the cytochrome *b*<sub>562</sub> signal sequence after export to the periplasmic space, which was predicted to be 39 kDa. In hindsight the application of mass spectrometry to definitively size the protein bands observed would have been of benefit for investigation. If this experiment were to be repeated each SDS PAGE sample presented would additionally be analysed by mass spectrometry to confirm protein identity. The identity of the observed band at ~36 kDa will be discussed in more detail in Section 3.2.5.

### 3.2.4. Optimisation of HRP<sup>EC1</sup> Expression.

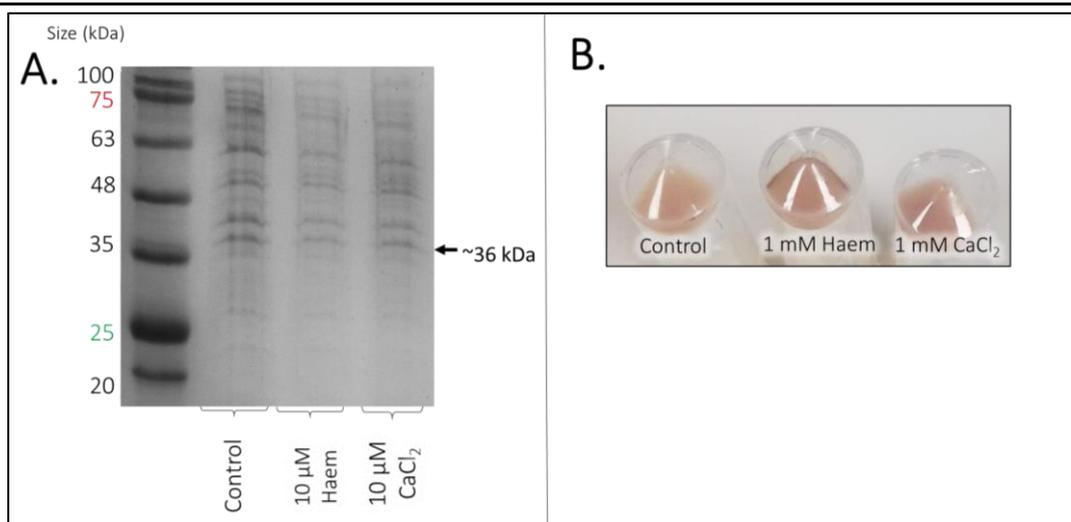
To determine the best temperature for expression of HRP<sup>EC1</sup>, the intensity of expressed protein was investigated after induction with 1 mM IPTG and incubated at several temperatures (16 °C, 20 °C, 25 °C, 30 °C, and 37 °C). Protein sized at approximately 36 kDa on SDS PAGE was observed in all cells induced with IPTG. Cells grown at the lowest incubation temperature (16 °C) were observed to yield the greatest intensity of banding at ~36 kDa (Figure 3.5, A) and the strongest cell colouration (Figure 3.5, B). An additional band of over-expression was observed at ~48 kDa at temperatures 16 °C and 20 °C. Although this band was greater than the predicted size of the full-length protein (41 kDa) and not consistent with any expected size of HRP<sup>EC1</sup>.



**Figure 3.5. The Effect of Temperature on Expression of HRP<sup>EC1</sup>**

**A**, SDS PAGE of soluble and insoluble protein expressed in the presence and absence of IPTG, after 24 hours at various incubation temperatures. Cell pellets corrected to an OD<sub>600</sub> of 1 A.U. prior to SDS PAGE. **B**, the colouration of cells induced with IPTG after incubation at a range of temperatures.

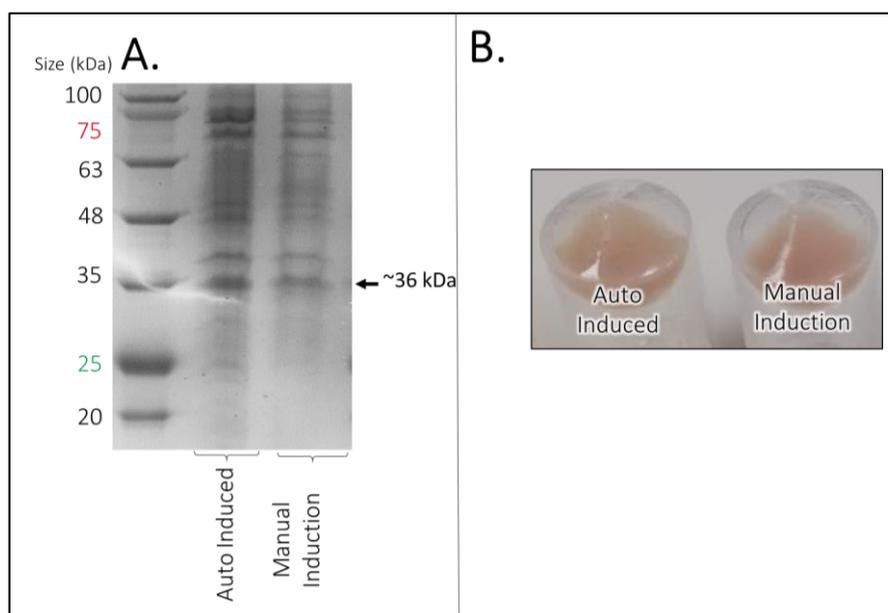
To investigate if protein expression was influenced by cofactor availability, cells were induced in the presence of additional haemin and CaCl<sub>2</sub>. Expression of protein was trialled in media supplemented either with no cofactor (control) and either 1 mM haemin, or 1 mM CaCl<sub>2</sub>. SDS PAGE indicated that the banding at ~36 kDa was not altered by cofactor presence (Figure 3.6, A). Additionally, the cell colouration was not visibly different between conditions (Figure 3.6, B). It was therefore proposed that cofactor presence was not a limitation of protein expression.



**Figure 3.6. The Influence on Cofactor Presence on the Expression of HRP<sup>EC1</sup>**

**A**, SDS PAGE of protein whole cell pellets, and **B**, colouration of cells induced by 1 mM IPTG in the absence (control) or presence of either 1 mM haemin or 1 mM CaCl<sub>2</sub>.

Further to this no change in band intensity (Figure 3.7, A) or colouration of cells (Figure 3.7, B) was observed between methods of expression (manual induction or autoinduction). Cells which were induced automatically by IPTG autoinduction media (see Chapter 2.1.5., Table 2.3), showed no difference the intensity of the ~36 kDa band or cell colouration to that of cells which are induced manually with 1 mM IPTG at OD<sub>600</sub> 0.6. As no advantage was observed by the monitoring of cell growth before manual induction, autoinduction was used as a means of induction for subsequent cultures.



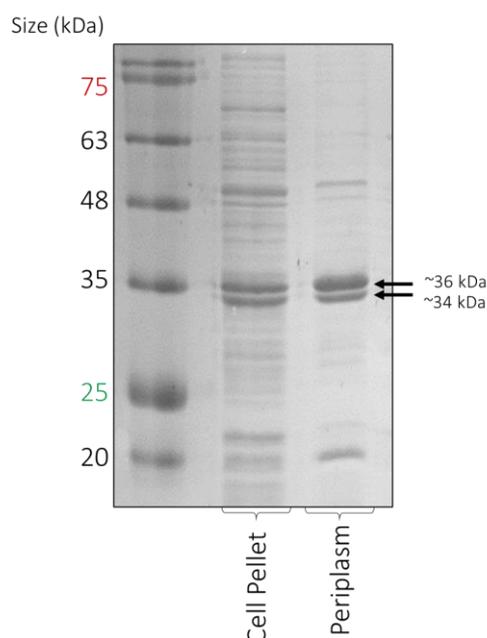
**Figure 3.7. The Influence of Induction Technique on the Expression of HRP<sup>EC1</sup>**

**A**, SDS PAGE of protein whole cell pellets, and **B**, colouration of cells induced by 1 mM IPTG in the absence (control) or presence of either 1 mM haemin or 1 mM CaCl<sub>2</sub>.

A general expression strategy was established based on the results of expression optimisation (Figures 3.5, 3.6, and 3.7). For all subsequent expression of HRP<sup>EC1</sup>, 5 mL of LB media was supplemented with kanamycin and inoculated with transformed cells for overnight growth at 37 °C. The following day cells were used to inoculate a 1 L flask of IPTG auto-induction media (Table 2.3) and incubated overnight for expression at 16 °C before recovery of cells by centrifugation.

### 3.2.5. Periplasmic Localisation of HRP<sup>EC1</sup>

Investigation of protein localisation to the periplasmic cell fraction was performed as described in Chapter 2.3.4. Comparison of protein within the initial cell pellet and that of the periplasmic fraction was performed by SDS PAGE (Figure 3.8). The presence of protein sized at approximately ~34 kDa and ~36 kDa was observed in the periplasmic cell fraction. This was smaller than the predicted full-length protein (~41 kDa) or protein without cytochrome *b*<sub>562</sub> signal sequence (~39 kDa). This observation is of note; however, proteins size is not the only variable which influences how a protein travels in SDS PAGE, and hydrophobicity, glycosylation, and disulphide bond presence can also a protein's interpreted size on SDS PAGE.



**Figure 3.8. Periplasmic Extraction of HRP<sup>EC1</sup> from Cell Pellet.**

HRP<sup>EC1</sup> indicated by arrow at ~36 kDa, with an additional band observed at ~34 kDa.

Instead, the presence of protein sized as 36 kDa is consistent with that of HRP<sup>EC1</sup> which has been subjected to removal of the entire N-terminal region (see Figure 3.3). Banding observed was consistent with the formation of mature-HRP<sup>EC1</sup> in the periplasm and is indicative

---

that cleavage of both the pre-peptide (3221 Da) and post-peptide (1494 Da) of HRP<sup>EC1</sup>, in the periplasm. Furthermore, both bands observed are too small to be that of protein removed of just the periplasmic signal sequence which would be sized at 39 kDa. The additional protein band observed sized at 34 kDa was consistent with that expected from HRP<sup>EC1</sup> cleaved of both pre- and post-peptide regions (34,144 Da). Whilst the band observed at 36 kDa was consistent with protein which was cleaved between the pre-peptide and the mature HRP sequence but retained post-peptide presence (35,621 Da) (Figure 3.3).

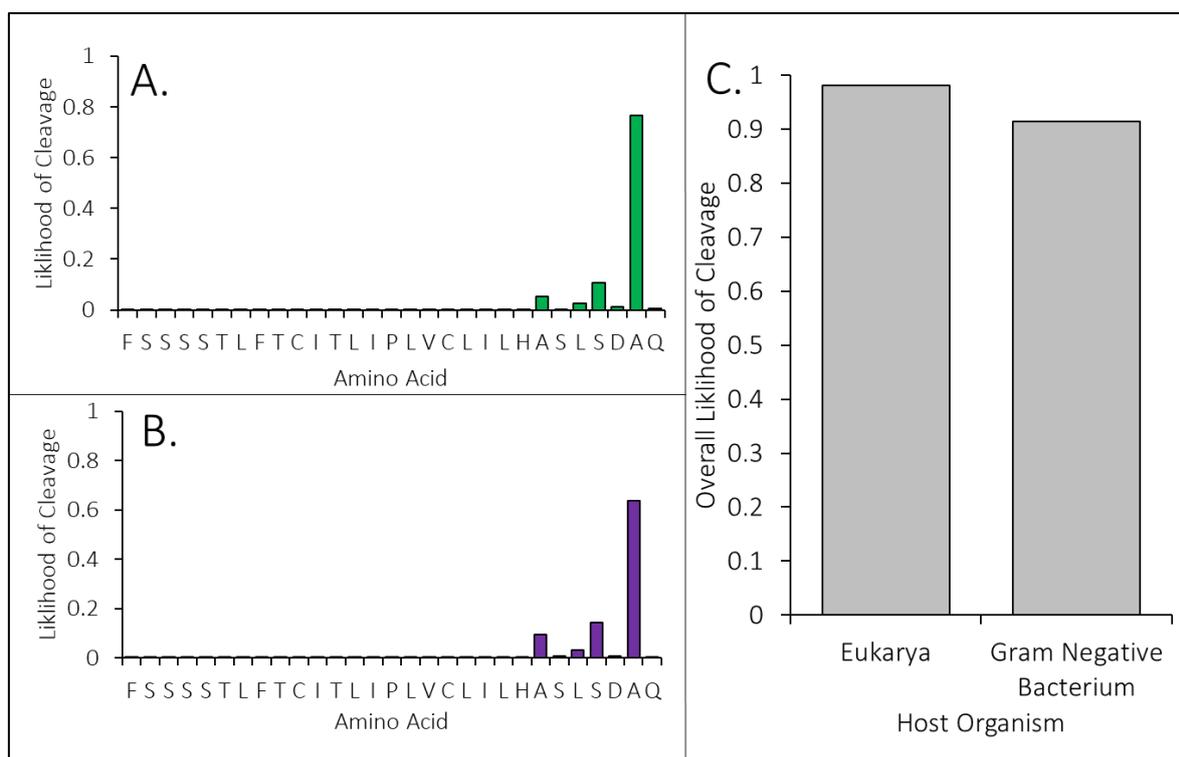
This was unexpected, as it had been previously assumed that cleavage of the N-terminal pre-peptide would not be observed in *E.coli* expression as cells do not feature membrane-bound organelles. Therefore, it was reasoned that *E. coli* would lack the apparatus necessary for the recognition and cleavage of the N-terminal ER localisation sequence of HRP. However, it is theorised, that cleavage of the protein could be feasible if the proto-peptide HRP regions were capable of intein-cleavage. Intein cleavage is found in all domains of life and occurs when a protein self cleaves a specified region, which is typically a conserved motif that features a cysteine, serine, or threonine<sup>300</sup>. This leaves only a defined smaller protein sequence called an intein, and in the case of HRP would remove both proto-peptide regions<sup>301</sup>. It is theorised that the presence of serine residues common to the N and C proto-peptide regions of most HRP isoforms are conserved as they function as intein junctions for self-cleavage of the enzyme.

The size of bands observed suggests that the inclusion of the proto-peptide sequence regions to HRP<sup>EC1</sup> has not prevented the expression and cleavage of protein to form mature-HRP<sup>EC1</sup>. Results indicate that HRP<sup>EC1</sup> is recognised by periplasmic export chaperones and transferred to the periplasm, where it cleaves to form mature-HRP<sup>EC1</sup>, by an unknown mechanism. It is possible that this pathway resembles the SEC translocation pathway, wherein newly synthesised unfolded protein is transported across the cytoplasmic membrane via the SecYEG channel in a process which requires ATP<sup>41</sup>. Critically, the presence of HRP<sup>EC1</sup> within the periplasm indicates that expressed protein is not completely localised to insoluble inclusion bodies<sup>75,109</sup>.

### 3.2.6. Prediction of Pre-Peptide Cleavage

To determine if the cleavage of pre-HRP was, as suspected, viable in *E. coli* cells, the sequence of C1A HRP (UniProt P00433) was analysed on the server Signal P 5.0

(<http://www.cbs.dtu.dk/services/SignalP/>). Signal P 5.0 is a web server which categorises possible signal sequences and identifies the site of possible enzymatic cleavage. Analysis of C1A HRP indicated that the sequence showed potential for cleavage at A30, in both Eukarya and Gram-negative bacterium (Figure 3.9, A and B). In the construct for HRP<sup>EC1</sup> A30, is residue 52, and is located at the beginning of the sequence of HRP-C1A and the pre-peptide sequence (see Figure 3.3), which indicates that the complete N-terminal pre-peptide is likely cleaved in *E.coli*. However, a lower probability of cleavage at both A25, L27, and S28, was additionally predicted, in both cell forms (Figure 3.9, A and B).



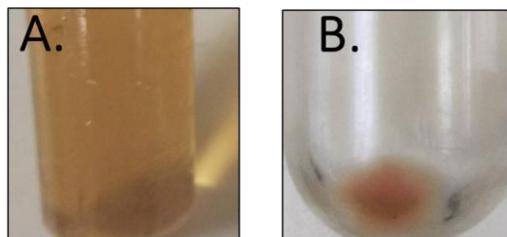
**Figure 3.9. The Likelihood of Pre-peptide cleavage of HRP in Eukarya and Gram-Negative Bacteria**

The probability of cleavage at each amino acid of C1A HRP when expressed in Eukarya (A) and gram-negative bacterium (B). The overall likelihood of cleavage in each cell form shown in C.

These alternative cleavage sites were shown to be marginally more likely in Gram-negative bacterium (Figure 3.9, B). Whilst the overall likelihood of cleavage in eukaryotic host cells was calculated to be 0.98 and 0.91 in Gram-negative bacteria (Figure 3.9, C). When compared it can be reasoned that N-terminal cleavage is highly probable (>90% chance) to occur in each cell host.

### 3.2.7. Purification of HRP<sup>EC1</sup>

Despite the proven expression of soluble HRP<sup>EC1</sup> (Figure 3.8), persistent precipitation of the expressed protein severely hindered the acquisition of purified HRP<sup>EC1</sup>. Frequently, expressed HRP<sup>EC1</sup> was extracted from cells but would gradually precipitate out of solution as



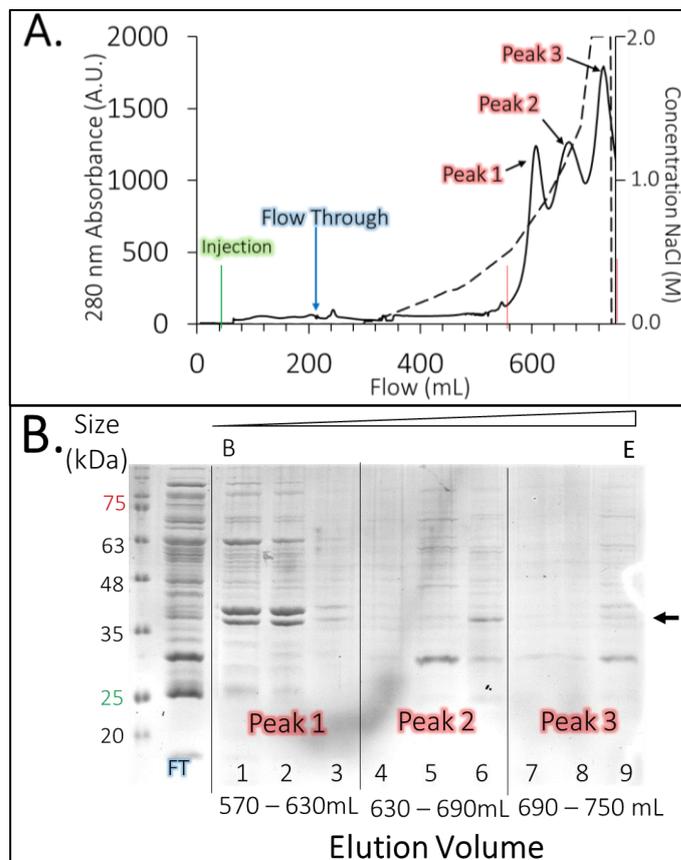
**Figure 3.10. Precipitation of HRP<sup>EC1</sup>**

**A**, Soluble lysate of cells expressing protein immediately following separation of insoluble cell fraction. **B**, protein precipitated from soluble lysate over a duration of 3 hours at 20 °C. time progressed (Figure 3.10). Therefore, the generation of pure HRP<sup>EC1</sup> proved to be complex, despite a reasonable expression of the protein (Figure 3.4). Consequently, the best method of generating pure HRP<sup>EC1</sup> was to minimise the time in which HRP<sup>EC1</sup> was stored at room temperature before flash freezing. Thus, periplasmic extraction of protein was not performed, as attained protein could not be concentrated for SEC and was in an insufficient buffer for direct IEX.

Instead, cell pellets expressing HRP<sup>EC1</sup> were resuspended in IEX binding buffer (Tris 50 mM pH 8) before French pressure cell lysis (see Chapter 2.3.5.). The soluble lysate, separated following ultracentrifugation, was then immediately applied to a Q column for IEX chromatography (Figure 3.11, A) (see Chapter 2.4.3). Once the contaminated protein had been washed from the Q column, the bound protein was eluted over a gradient of Tris 50 mM 2 M NaCl pH 8 and fractions were analysed by SDS PAGE (Figure 3.11, B).

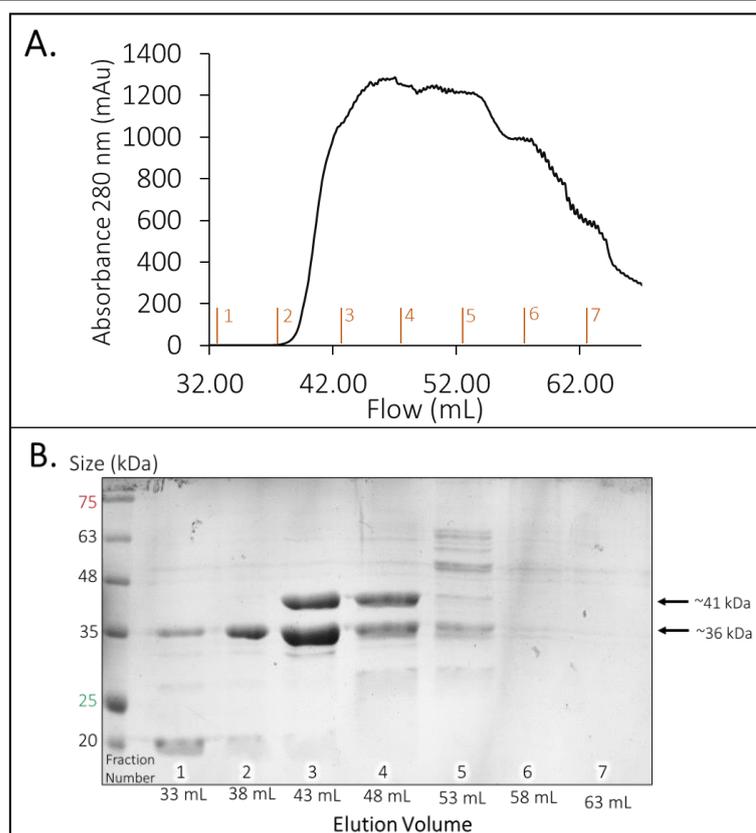
Fractions which were observed to contain protein consistent with HRP<sup>EC1</sup> and sized at ~36 kDa by SDS PAGE (Fractions 1-3, Figure 3.11, B) were pooled and applied directly to a gel filtration column for SEC (see Chapter 2.4.2)(Figure 3.12). Protein was shown to elute after a flow volume of 33 mL (Figure 3.12, A) and samples were assessed for purity by SDS PAGE (Figure 3.12, B). The presence of a band of consistent size with HRP<sup>EC1</sup> was observed in fractions 1-5 which represented an elution volume of 30-55 mL, each gradually reducing in purity. Fractions collected from SEC were observed to be cloudy and were spun at 150,000  $\times g$  for 3 minutes to remove protein precipitant before SDS PAGE. It is reasoned that the lack of peak

definition from SEC chromatography (Figure 3.12, A), is attributable to the precipitation of HRP<sup>EC1</sup>, and represents a further loss of expressed protein.



**Figure 3.11. Purification of HRP<sup>EC1</sup> by IEC.**

**A**, chromatograph of HRP<sup>EC1</sup> demonstrating the elution of protein as 3 distinct peaks, when a NaCl gradient was applied to a Q column. **B**, SDS PAGE of fractions obtained from IEC, labelled according to peak in which they eluted.



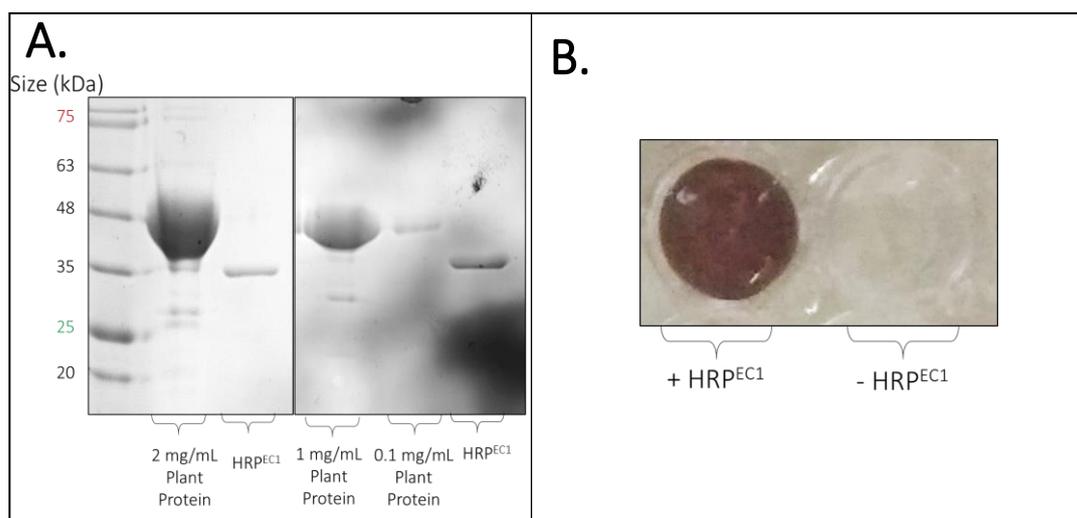
**Figure 3.12. Size Exclusion Chromatography of HRP<sup>EC1</sup>**

**A**, chromatography of SEC on Hiload<sup>TM</sup> 16/600 Superdex<sup>TM</sup> S200 gel filtration column, indicating elution of protein after 40 mL of flow. **B**, SDS PAGE of fractions eluted in A indicating the presence of band consistent with HRP<sup>EC1</sup> at 36 kDa in fractions 1 -5 indicated by arrow.

### 3.2.8. HRP<sup>EC1</sup> Expression and Purification

The application of both IEX and SEC over a single day post cell lysis was sufficient for HRP<sup>EC1</sup> purification and minimised the time in which protein could precipitate before flash freezing. However, this method consistently produced only a small quantity of purified HRP<sup>EC1</sup> regardless of expression efficiency. Pure HRP<sup>EC1</sup> obtained from 1 L of cell growth was observed to be consistently sized at ~36 kDa when viewed on SDS PAGE, consistent with the size of HRP<sup>EC1</sup> expected from pre-peptide cleavage (Figure 3.13, A). Furthermore, pure HRP<sup>EC1</sup> was shown to be catalytically active, as, in the presence of H<sub>2</sub>O<sub>2</sub> purified protein initiated the oxidation of OPD to form a visibly red precipitate (Figure 3.13, B).

Upon the generation of pure HRP<sup>EC1</sup>, a persistent problem faced was that HRP<sup>EC1</sup> could not be concentrated without precipitation. This presented an issue as, despite a reasonable expression of soluble protein, the working sample of HRP<sup>EC1</sup> acquired was prone to degradation once characterisation had commenced. This dramatically impeded the analysis of pure HRP<sup>EC1</sup> and exerted a consistent time pressure for both purification and analysis post cell lysis. It was determined that for the feasible analysis of HRP<sup>EC1</sup> the protein must be acquired at a greater quantity, concentration, or stability.



**Figure 3.13. The Size and Activity of Pure HRP<sup>EC1</sup>**

**A**, 10 mg/mL sample of commercially available HRP in comparison to total sample of pure HRP<sup>EC1</sup> obtained from 1 L culture. Gel is shown twice with the image on the right being a repetition of that on the left with a reduced concentration of plant protein. The concentration of recombinant protein is constant between images. **B**, turnover of OPD catalysed by pure HRP<sup>EC1</sup> in the presence of H<sub>2</sub>O<sub>2</sub>.

### 3.2.9. Mutagenesis to Reduce Recombinant HRP Precipitation

One feature of HRP which allows for recombinant expression without PTM is that glycosylation is neither essential for its activity or folding<sup>54</sup>. However, when expressed without glycans the protein is both less thermotolerant and less active which dramatically impedes non-glycosylated protein yield. Consequently, many studies have targeted the glycosylation sites of HRP for mutagenesis to improve the stability of non-glycosylated HRP<sup>63,70,99,114,290,302–304</sup>.

Several previous studies have linked the presence of unoccupied glycosylation sites with an increased aggregation propensity of purified recombinant HRP <sup>54</sup>. However, several different studies have alternatively described that the substitution of the non-glycosylated asparagine residues of HRP reduce this deleterious effect (Table 3.1). Numerous studies have indicated an increased thermostability, substrate affinity, and peroxide tolerance of recombinant HRP upon the substitution of non-glycated residues N13, N57, N158, N186, N198, N255, and N268 <sup>63,91,99,114,290</sup>. Furthermore, study by Humer and Spadiut (2019), demonstrated that mutation of multiple glycosylation sites at once enabled an 8-fold increase in activity and a two-fold increase in thermostability <sup>54</sup>.

**Table 3.1. Effect of Mutation of HRP N-linked Glycosylation site for recombinant expression.**

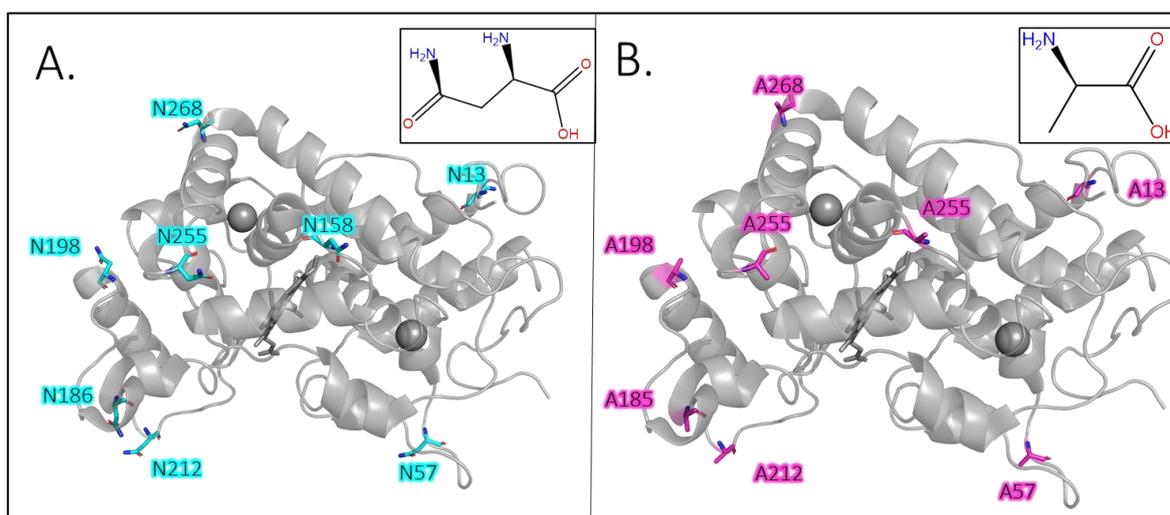
Glycosylation site	Mutated To	Effect of Mutagenesis	Study
N 13	D	Increased peroxide tolerance and thermostability	Asad et al (2011) <sup>63</sup> Capone et al (2014) <sup>70</sup>
N 57	S	Increased peroxide tolerance and thermostability	Capone et al (2014) <sup>70</sup>
N 158	D	Increased peroxide tolerance	Capone et al (2014) <sup>70</sup>
N 186	D	Increased turnover of ABTS	Capone et al (2014) <sup>70</sup>
N 198	D	Increased turnover of ABTS	Capone et al (2014) <sup>70</sup>
N 212	No record of mutation		
N 255	D	Increased folding in E.coli	Lin et al (1999) <sup>99</sup>
	D	Increased peroxide tolerance	Capone et al (2014) <sup>70</sup>
N 268	D	Increased peroxide tolerance, 18- fold increase in substrate affinity and improved thermostability	Asad et al (2011) <sup>63</sup> Asad et al (2016) <sup>290</sup> Capone et al (2014) <sup>70</sup>
	G	Increased peroxide tolerance, substrate affinity and thermostability	Asad et al (2016) <sup>290</sup>

Cumulatively, many studies report improvement to either protein activity or stability upon the mutagenesis of non-occupied glycosylation sites <sup>63,70,99,290</sup>. Furthermore, the study of alternate protein models, monoclonal antibodies, and human prion protein indicate each has an increased propensity to precipitate upon the removal of N-linked glycan units <sup>305,306</sup>. Hydrogen bonding of the non-glycosylated amide group on asparagine residues is a known stimulus of protein aggregation in the formation of amyloid fibres <sup>307</sup>, indicating a potential

destabilising role of these non-glycosylated residues. It was therefore determined that the substitution of all non-glycan linked asparagines on the surface of HRP<sup>EC1</sup> would be beneficial to protein solubility.

### 3.2.10. Design of Construct HRP<sup>EC2</sup>

Alanine was selected as the residue of choice for mutagenesis, as it features a simplified side-chain, in comparison to asparagine, and does not introduce additional, potentially detrimental, chemistry<sup>308</sup>. Alanine was chosen in favour of glycine as the absence of a side chain upon glycine, prevents  $\alpha$ -helical folding, and was expected to impede the primarily  $\alpha$ -helical protein structure of HRP<sup>309</sup>. Instead, the presence of a beta carbon on the alanine's side-chain retains the residue's ability to form an alpha helix and thus would have a minimal influence on the secondary structure of HRP(Figure 3.14)<sup>310</sup>.



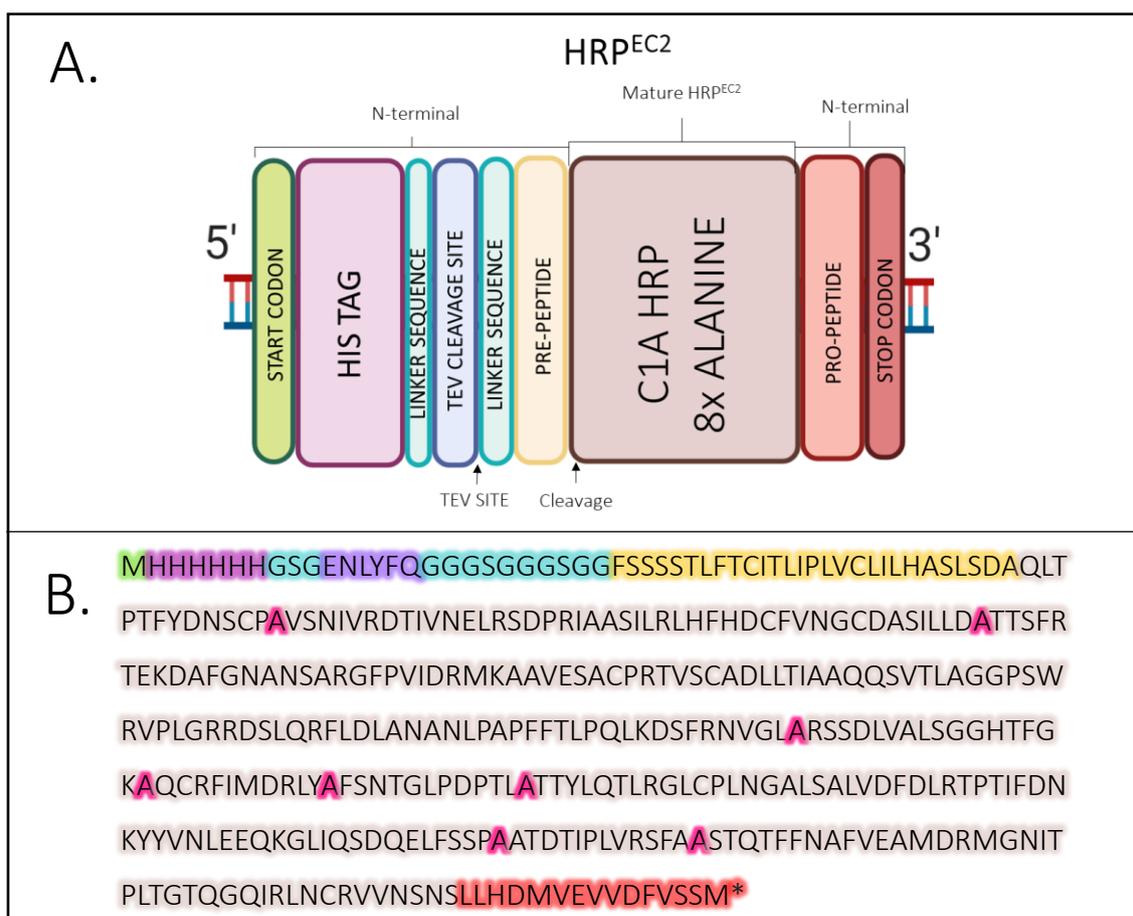
**Figure 3.14. Representation of N-Linked Glycosylation Sites.**

**A**, the native HRP N residues. **B**, the proposed substitution of glycosylation sites with alanine.

To establish if proto-peptide cleavage was a feature only of HRP which had been exported to the periplasm the cytochrome *b*<sub>562</sub> signal sequence was removed from the HRP expression construct. In place of the periplasmic sequence, a poly-histidine purification tag was incorporated on the N-terminus of HRP expression construct 2 (HRP<sup>EC2</sup>)(Figure 3.15). It was reasoned that only HRP<sup>EC2</sup> which was not subjected to N-terminal cleavage would be purified *via* nickel affinity chromatography. Therefore, inclusion would enable comparison of un-cleaved proto-HRP<sup>EC2</sup> with cleaved mature-HRP<sup>EC2</sup>.

Lastly, a TEV cleavage site was incorporated between the poly-histidine tag and the HRP C1A sequence, as an insurance mechanism for HIS Tag removal. A TEV cleavage site is a highly

specific sequence of amino acids (ENLYF/Q) which is acted upon by TEV protease<sup>43</sup>. Inclusion of the site to the expression construct was rationalised so that generated protein could be removed of the haem binding poly-histidine tag<sup>311</sup> before characterisation. To ensure the cleavage site was surface exposed, a flexible linker sequence (GGSGGS) was also incorporated. This linker sequence was formed of repeating glycine serine regions which ensure backbone flexibility without interference to native protein folding<sup>312</sup>. It was reasoned the linker sequence would preserve access to both the poly-histidine site and the TEV cleavage site after protein folding, maintaining their surface accessibility.



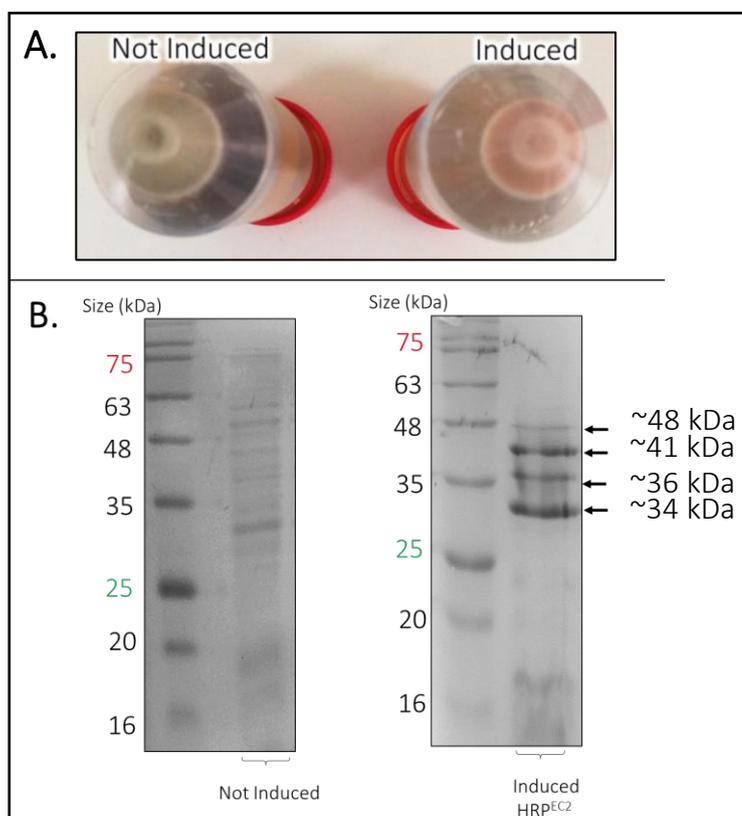
**Figure 3.15. HRP<sup>EC2</sup> Gene Construct for Expression of HRP Without Surface Asparagines.**

**A**, graphic of the gene constructed for expression. **B**, amino acid sequence of construct with regions coloured according to sequence feature as displayed in A.

### 3.2.11. Expression of HRP<sup>EC2</sup>

The expression construct HRP<sup>EC2</sup> (Figure 3.15) was cloned into the pET-24b expression vector by GenScript and transformed into *E. coli* SHuffle<sup>®</sup> cells for expression. Expression of HRP<sup>EC2</sup> was achieved as described previously (see Chapter 2.3.1.) and cells were induced with 1 mM of IPTG overnight at 16 °C. Cells which had been induced with IPTG were once more

shown to turn pink (Figure 3.16, A), and the presence of four bands of overexpression was observed by SDS PAGE (Figure 3.16, B). Three of these bands were above the 35 kDa marker and one was below the 35 kDa marker.



**Figure 3.16. Induction of HRP<sup>EC2</sup> Expression with IPTG at 20 °C.**

**A**, colour change observed in cells which were induced overnight with 1 mM IPTG at 20 °C.

**B**, SDS PAGE of protein expressed in the absence and the presence of IPTG. The presence of four bands of over expression are observed in cells which were induced IPTG and labelled to display approximate size.

The observed protein banding was consistent with the expression of N-terminally cleaved HRP<sup>EC2</sup> (35 kDa), and mature-HRP<sup>EC2</sup> cleaved at both N and C-terminal (34 kDa). The band observed at 48 kDa replicated that noted previously in the expression of HRP<sup>EC1</sup> (Figure 3.6), however, the band size was larger than the maximum size possible for HRP<sup>EC2</sup> (Table 3.2).

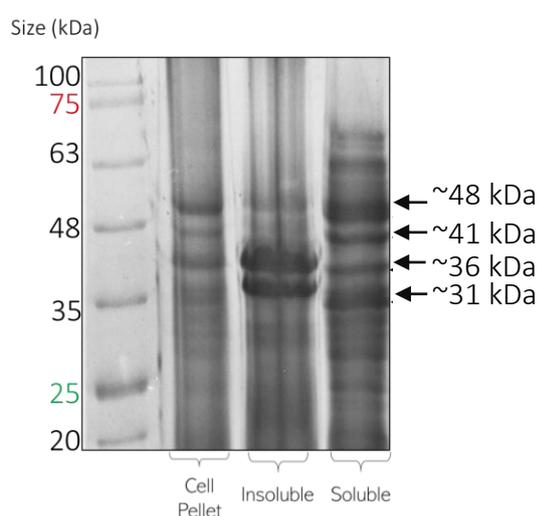
**Table 3.2. Weight of HRP<sup>EC2</sup> as Predicted in ProtParam**

HRP <sup>EC2</sup>	Predicted Size
Full Sequence	40835.29
Cleaved at C-terminus alone	39132.29
Cleaved at TEV site	39013.34
Cleaved at Pre-peptide	35319.15
Both N and C-terminal removed	33616.15

Although, the band might have represented a form of HRP which ran different to expectation on SDS PAGE. This band is described in detail in Section 3.2.13.

### 3.2.12. The Solubility of Expressed HRP<sup>EC2</sup>

To determine if HRP<sup>EC2</sup> was soluble, cells were lysed and the soluble cell fraction was separated by ultracentrifugation (see Chapter 2.3.5.). SDS PAGE was performed on samples to determine in which cellular fraction HRP<sup>EC2</sup> was located (Figure 3.17). In the insoluble cell fraction, two bands of far higher intensity than background were observed. These bands sized at approximately 41 kDa and 35 kDa were consistent with both full-length proto-HRP<sup>EC2</sup> (41



**Figure 3.17. Solubility of Expressed HRP<sup>EC2</sup>.**

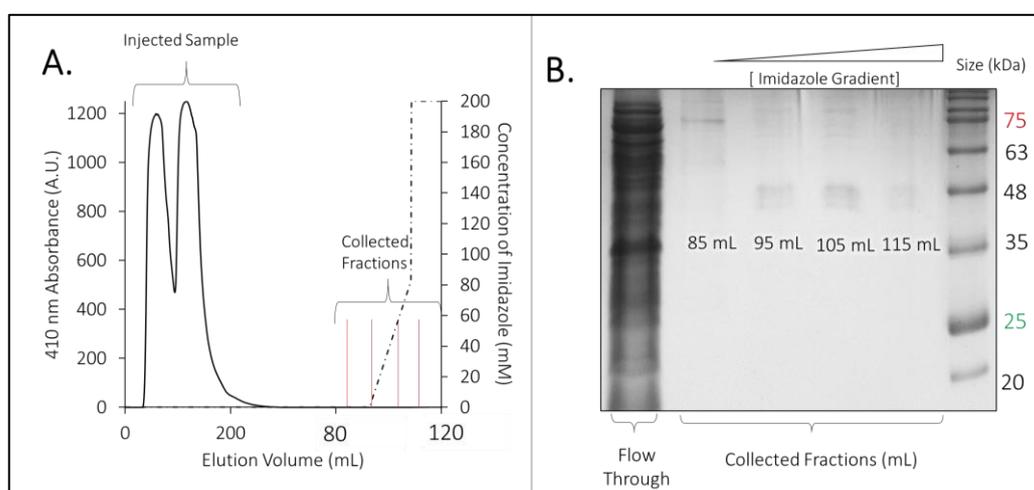
SDS PAGE of cell sample, prior to whole cell lysis (**Cell Pellet**) in comparison to sample post lysis (**Insoluble** and **Soluble**). The presence of four bands consistent with those observed in expressing cells are observed in soluble lysate. Whilst only the smallest two bands are observed in the insoluble cell fraction. Bands appear to be more spread out in this gel due to the incomplete boiling of sample.

kDa) and, HRP<sup>EC2</sup> cleaved of the N-terminus (35 kDa).

The observed presence of bands consistent with full-length proto-HRP<sup>EC2</sup> and N-terminally cleaved HRP<sup>EC2</sup> also within the soluble cell fraction is a critical observation, as these bands demonstrate that the expressed HRP is soluble. For this reason, the insoluble cell fraction was discarded. Soluble HRP expression contrasted the previous study by Smith *et al.* (1990) wherein soluble HRP used in the investigation was recovered from the re-solubilisation of *E. coli* inclusion bodies, and not from the lysate <sup>75</sup>.

### 3.2.13. Nickel Affinity of HRP<sup>EC2</sup>

To establish if expressed HRP<sup>EC2</sup> had undergone N-terminal cleavage soluble cell lysate was applied to a nickel affinity column for IMAC (see Chapter 2.4.1.)(Figure 3.18, A). Within fractions which eluted from the column, two proteins were observed by SDS PAGE sized at ~44 kDa and ~48 kDa (Figure 3.18, B). These two protein forms were consistent with bands observed during expression, and within the soluble cell lysate (Figures 3.16, and 3.17). In comparison, the presence of protein sized at 35 kDa in the column flow-through was not observed in eluted fractions (Figure 3.18, B). This 35 kDa protein displayed no affinity for the nickel resin and was consistent in size with that predicted of HRP<sup>EC2</sup> which had been N-terminally cleaved (Table 3.3).



**Figure 3.18. Affinity Chromatography of Soluble HRP<sup>EC2</sup> lysate .**

**A**, chromatograph of nickel affinity chromatography, indicating that protein eluted over the applied imidazole gradient is substantially lower intensity than protein applied. **B**, SDS PAGE of fractions eluted between 85-120 mL indicating the presence of a protein sized at approximately 46 kDa and 48 kDa.

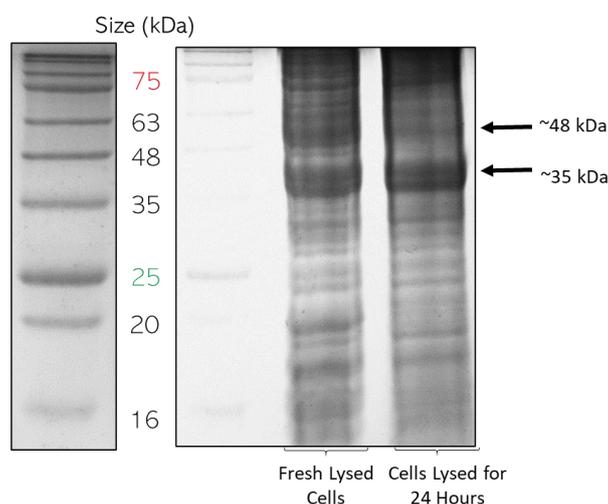
The presence of two bands within eluted fractions at ~44 and ~48 kDa, did not correlate with that expected of HRP<sup>EC2</sup>. However, the size observed was consistent with banding previously noted in Figures 3.16, and 3.17. The presence of these two bands within the eluted protein fractions and not within the column flow-through indicated their affinity for the nickel resin. It is therefore likely, that these bands are that of HRP<sup>EC2</sup> which has not been cleaved on the N-terminus but instead features the N-terminal poly-histidine region (Figure 3.15).

The larger of the two bands (~48 kDa)(observed in Figures 3.16, 3.17, and 3.18) is reasoned to be full-length HRP<sup>EC2</sup> whilst the smaller (~44 kDa) is reasoned to be HRP<sup>EC2</sup> which has been cleaved of only its C-terminus. The banding observed at ~44 kDa, cannot be that of HRP<sup>EC2</sup> which is cleaved at the TEV cleavage site, as this would remove the poly-histidine tag

from HRP<sup>EC2</sup> and prevent binding to the nickel resin. This, therefore, indicates that not all HRP expressed in the cell cytoplasm was subjected to proteolytic cleavage.

### 3.2.14. Post-Lysis Cleavage of HRP<sup>EC2</sup>

To investigate if the cleavage of the HRP<sup>EC2</sup> pre-peptide could occur post cell lysis, cell lysate was stored for 24 hours at 4 °C and compared to freshly lysed cells by SDS PAGE (Figure 3.19). SDS PAGE indicated the intensity of banding at ~48 kDa was greater in cells which had been freshly lysed in comparison to cells which had been lysed for 24 hours. The decrease in intensity of this band indicates full-length proto-HRP decreased in abundance post-lysis. As no protein precipitate was observed, it was subsequently hypothesised that cleavage of the pre-peptide occurred within the soluble cell fraction post cell lysis.

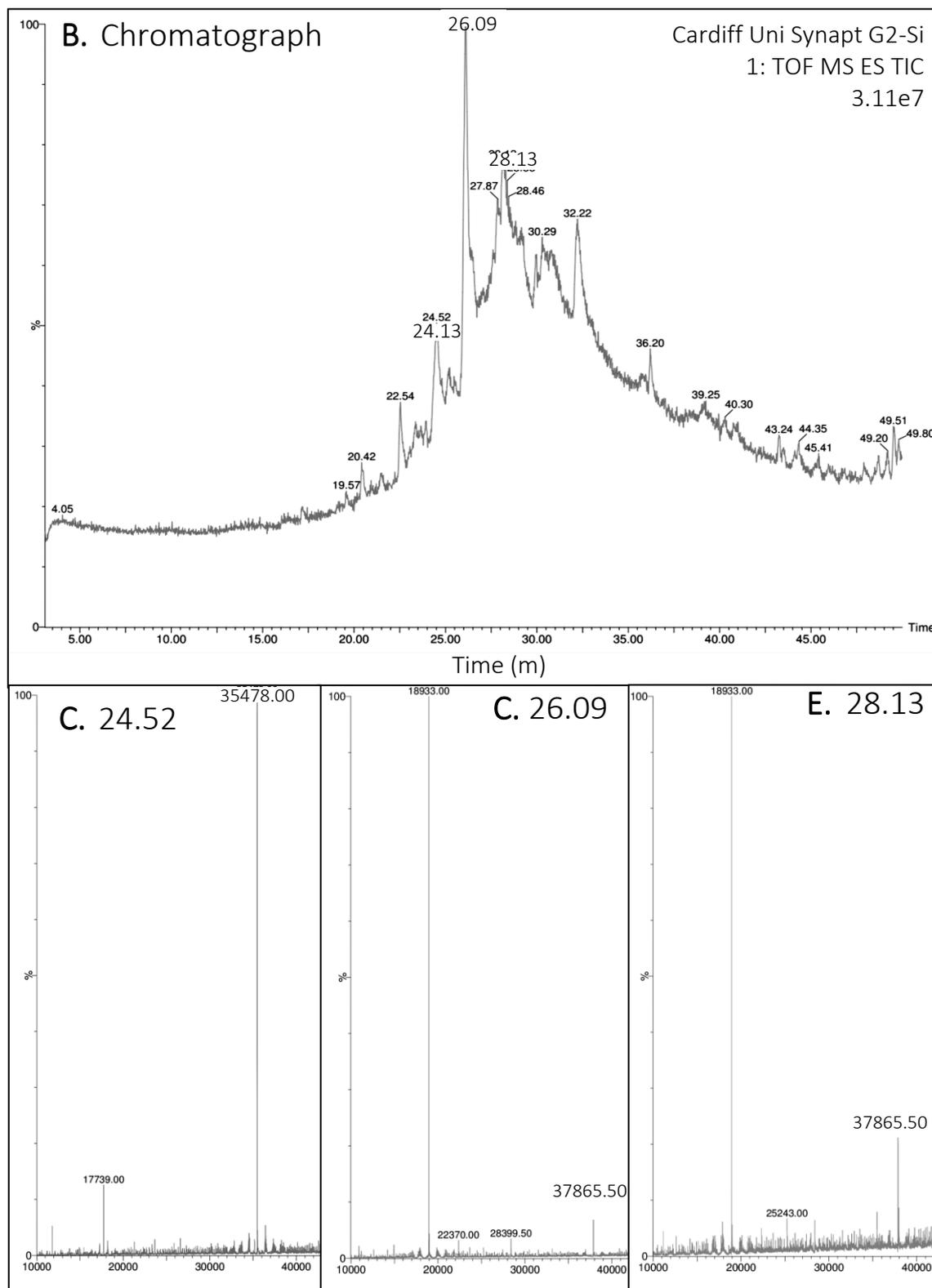


**Figure 3.19. Reduction in size of HRP<sup>EC2</sup> in Soluble Lysate Over 24 Hours.**

SDS PAGE of protein present in soluble lysate immediately after lysis, and 24 hours after lysis for comparison of abundance of protein sized at 48 kDa and 35 kDa over time. Ladder shown twice, by two separate exposures, to give greatest clarity of band size observed.

Mass spectrometry of soluble HRP<sup>EC2</sup> lysate was performed at Cardiff School of Chemistry (see Chapter 2.5.11.) to determine the exact size of the band at ~ 35 kDa. Samples were analysed 72 hours post lysis (to allow for maximum post lysis cleavage of HRP<sup>EC2</sup>) and the results are presented in Figure 3.20. The chromatograph (Figure 3.20, A) produced from mass spectrometry indicated that the sample featured many different mass forms, with major peaks observed after 24.52 min, 26.09 min and 28.13 min, all of which featured mass sizes consistent with HRP<sup>EC2</sup>.

Calculation of the expected mass of HRP<sup>EC2</sup> was performed using the ExPASy Peptide Mass server ([https://web.expasy.org/peptide\\_mass/](https://web.expasy.org/peptide_mass/)) to predict the average mass of the



**Figure 3.20. Mass Spectrometry of HRP<sup>EC2</sup> Lysate.**

**A**, Chromatograph of MS of HRP<sup>EC2</sup> lysate, with peaks of interest labelled accordingly. **B**, **C**, and **D**. Mass identities present in each chromatograph peak, with individual masses recorded in Table 3.4.

constructed sequence<sup>271,313</sup>. Haem porphyrin was not included in the size calculation of these masses as cofactors are typically lost in the ionisation of protein during mass spectrometry, resulting in a trace from only *apo*-protein. Two observed mass peaks (sized at 37,866 Da and 35,486 Da) were shown to be coherent with that expected of HRP<sup>EC2</sup> which had been subjected to N-terminal cleavage (Figure 3.20, B, C, and D).

The smaller of the two masses, sized at 35,478 Da (Figure 3.20, B) was 159 Da larger than that predicted of HRP<sup>EC2</sup> which had been cleaved, as predicted (Table 3.2). This indicated that cleavage was close to that predicted by Signal P analysis and likely occurred at one of the alternative splice regions predicted in Figure 3.9. The most likely region of cleavage (or alternative mRNA splicing) was reasoned to be HRP<sup>EC2</sup> residue S52 (S28 in original HRP C1A

**Table 3.3. The Calculated identity of Mass observed**

Observed Mass	Reasoned Cleavage Site	Size Difference from Observed Mass
35478	MHHHHHHHGSGENLYFQGGGGGGGGSGGFSSS//LFTCITLIPLVCLILHAS//L SDAQLTPTFYDN SCPAVSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCD ASILLDATTSFRTEKDAFGNANSARGFPVIDRMKAAVESACPRTVSCADLLTI AAQQSVTLAGGPSWRVPLGRRDSLQRFLDLANANLPAPFFTL PQLKDSFRN VGLARSSDLVALSGGHTFGKAQCRFIMDRLYAFSNTGLPDPTLATTYLQTLR GLCPLNGALSALVDFDLRPTIFDNKYYVNLEEQKGLIQSDQELFSSPAATDTI PLVRSFAASTQTFFNAFVEAMDRMGNITPLTGTQGQIRLNCRVVNSNLLH DMVEVDFVSS//SM*	10 Da (35488)
37865.00	MHHHHHHHGSGENLYFQGGGGGGGGSGGFSSS//LFTCITLIPLVCLILHASL SDAQLTPTFYDN SCPAVSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGCD ASILLDATTSFRTEKDAFGNANSARGFPVIDRMKAAVESACPRTVSCADLLTI AAQQSVTLAGGPSWRVPLGRRDSLQRFLDLANANLPAPFFTL PQLKDSFRN VGLARSSDLVALSGGHTFGKAQCRFIMDRLYAFSNTGLPDPTLATTYLQTLR GLCPLNGALSALVDFDLRPTIFDNKYYVNLEEQKGLIQSDQELFSSPAATDTI PLVRSFAASTQTFFNAFVEAMDRMGNITPLTGTQGQIRLNCRVVNSNLLH DMVEVDFVSSM*	18 Da (37847)

sequence)(Green, Table 3.3), as the calculated mass of protein cleaved at this site produced a mass 10 Da larger than that observed (Figure 3.20, B). The difference of 10 Da, is close to the value expected from a mixed population of formed disulphide bonds<sup>314</sup>. Therefore, it was reasoned that the mass 35,478 Da, was HRP<sup>EC2</sup> which had been cleaved at S52.

The closest mass calculated from the HRP<sup>EC2</sup> sequence to that of the observed 37,865 Da was 37,847 Da (Table 3.3). This calculated mass features the partial cleavage of the

---

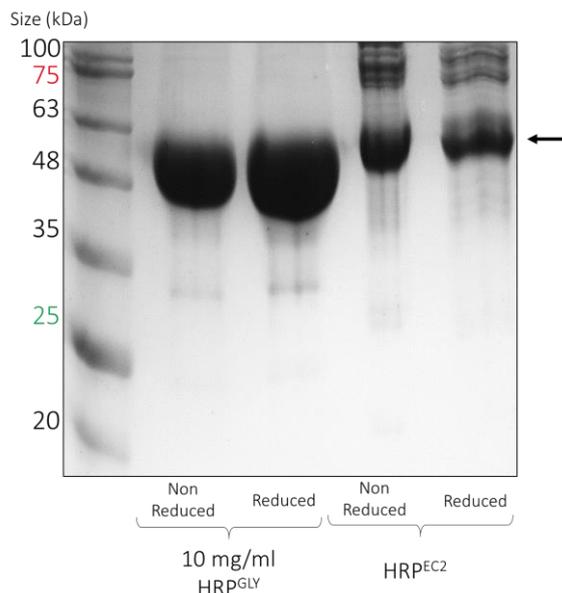
N-terminus within the poly-serine region (Table 3.3). This produced a calculated mass which was 18 Da smaller than that observed in Figure 3.20, C and D, but did not cleave the sequence at the region predicted by Signal P (Figure 3.9). However, regions of poly-serine repeats are common in signal sequences and are believed to function in the recognition of the signal peptide by peptidase enzymes<sup>315</sup>. As the mass of this observed protein form correlates within cleavage at two poly-serine repeat regions this could potentially indicate a secondary enzymatic recognition of the signal sequence. However, as this banding was only observed as a minor peak, the most prevalent cleavage of HRP<sup>EC2</sup> correlates with the predicted cleavage at S52.

Critically, mass spectrometry does not indicate the presence of a protein with a mass of approximately 48 kDa, which was observed previously *via* SDS PAGE (Figures 3.16-3.19). Neither does the analysis indicate the presence of a mass of approximately 41 kDa which is consistent with full-length proto-HRP<sup>EC2</sup>. This supports the hypothesis that the full-length HRP<sup>EC2</sup> can be cleaved post cell lysis.

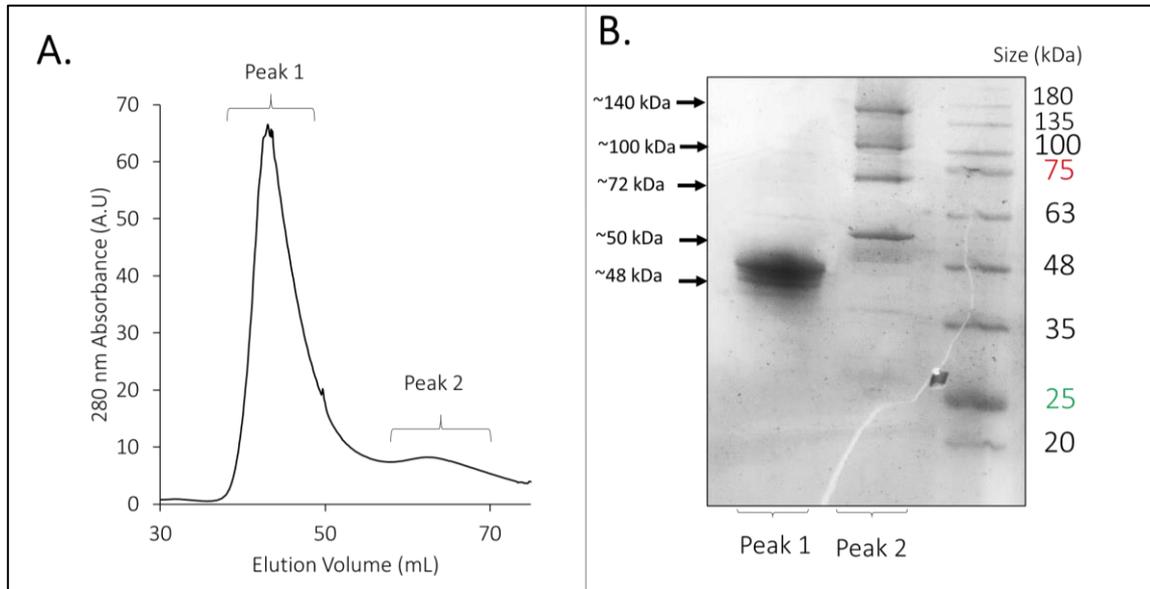
### 3.2.15. Cleavage of Purified HRP<sup>EC2</sup>

One consistent observation was that HRP<sup>EC2</sup> which had been purified by nickel affinity chromatography, spontaneously reduced in size as the time post-purification increased. Initially, HRP<sup>EC2</sup> was observed to be consistent in size with glycosylated, plant produced HRP (~44 kDa), in both reducing and non-reducing conditions (Figure 3.21). This band was buffer exchanged into Tris 50 mM pH 8 by size exclusion chromatography (Figure 3.22) and stored at 4 °C whereby a size reduction was observed. To assess this size change, purified HRP<sup>EC2</sup> was

then monitored daily by SDS PAGE over a duration of 14 days, post SEC, to assess if N-terminal cleavage of purified protein occurred (Figure 3.23).



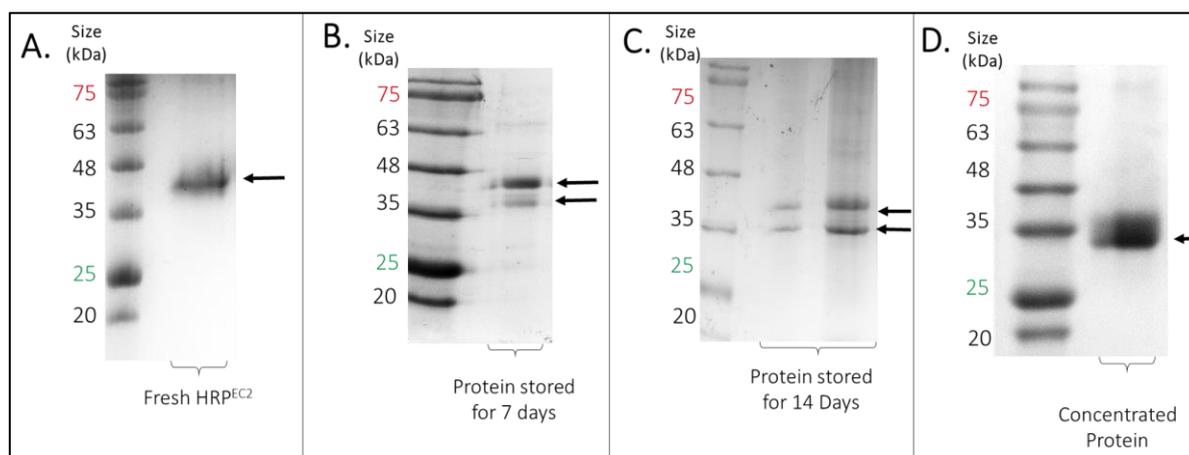
**Figure 3.21. SDS PAGE of both Glycosylated HRP and HRP<sup>EC2</sup> in Non-Reducing and Reducing Conditions**



**Figure 3.22. SEC of HRP<sup>EC2</sup> after Affinity Chromatography.**

**A**, Chromatography of protein eluted by SEC, indicating the elution of two separate peaks of protein with the first found after 40 mL elution volume and the second after 60 mL flow volume. **B**, SDS PAGE of protein eluted from SEC showing the presence of a 48 kDa protein in peak one and several larger proteins in peak 2.

After a duration of 7 days, the presence of a second, smaller protein form sized at approximately ~35 kDa was observed by SDS PAGE (Figure 3.23, B). This protein form was present at a lower intensity than the band observed at ~48 kDa, however as time progressed the intensity of this band (~35 kDa) increased. After a duration of 14 days at 4 °C, the smaller band (~35 kDa) was observed at a roughly equal concentration with that of the larger first formed band (~48 kDa), indicating that the ratio between the two forms had changed over time, and the sample was no longer predominantly sized at ~48 kDa).



**Figure 3.23. Reduction in size of HRP<sup>EC2</sup>**

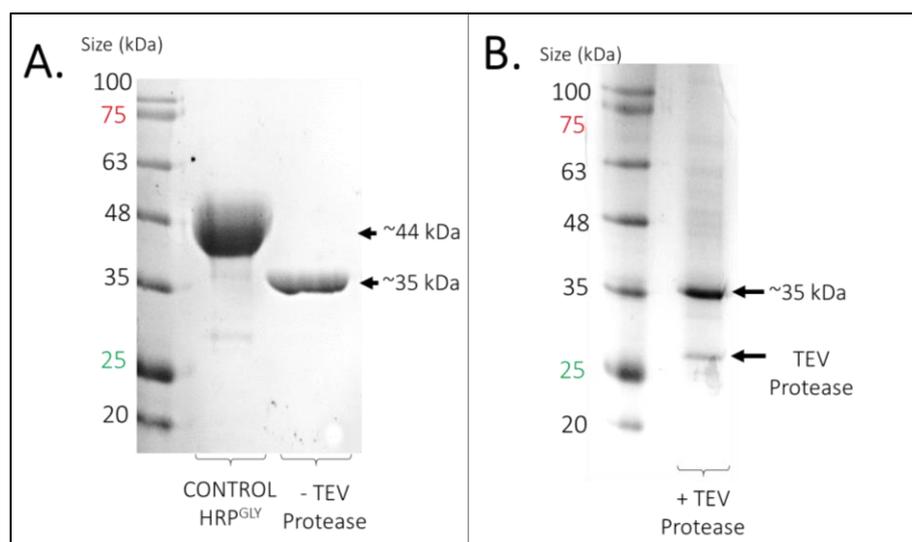
SDS PAGE of HRP<sup>EC2</sup> to indicate the reduction in size of protein observed. **A**, Fresh protein sample obtained by IMAC and SEC. **B**, pure HRP<sup>EC2</sup> which was stored at 4 °C for seven days. **C**, pure HRP<sup>EC2</sup> which was stored at 4 °C for fourteen days. **D**, pure HRP<sup>EC2</sup> which had been concentrated by spin filtration to half initial volume.

Finally, the pure protein sample was concentrated by way of 30,000 MW spin column, until the volume of the sample was reduced by half (1 mL to 0.5 mL). SDS PAGE of this more concentrated protein form indicated that all protein present was sized at approximately 35 kDa, and the larger of the two forms (~48 kDa) was no longer present (Figure 3.23, D).

The reduction in protein size observed was consistent with that expected from N-terminal excision of the pre-peptide from purified protein (a reduction of ~6 kDa). As protein had been previously purified *via* the N-terminal poly-histidine region, this was unexpected, as no additional protease was believed to be present within the sample. The initial protein band observed (Figure 3.23, A) was reasoned to be that of full-length HRP<sup>EC2</sup>, as the form was purified by nickel affinity chromatography. Whilst the secondary form of protein (~35 kDa) (Figure 3.23, D) was consistent with HRP<sup>EC2</sup> which had been cleaved of the N-terminal pre-peptide in cell lysate (Figure 3.19).

The smaller of the two bands, reasoned to be N-terminally cleaved HRP<sup>EC2</sup> was demonstrated to increase in intensity as both storage time and concentration increased (Figure 3.23). To confirm that the final protein form observed had been N-terminally cleaved 10  $\mu$ M sample was incubated with 5  $\mu$ M TEV protease, as described previously (see Chapter 2.5.16). It was reasoned that if a size reduction was observed upon incubation with TEV protease, this would indicate the N-terminal pre-peptide was still present upon HRP<sup>EC2</sup>. Although this is not the only method which could be applied for this purpose and an anti-poly histidine western blot would also serve this purpose. However, if the N-terminus of the protein had been cleaved, the TEV cleavage site would no longer be present on HRP<sup>EC2</sup> and no reduction in size would be observed.

SDS PAGE of incubated protein demonstrated no change in protein size was observed upon incubation with TEV protease (Figure 3.24). HRP<sup>EC2</sup> which was not treated with TEV protease (Figure 3.24, A) was observed to be of consistent size with protein which had been incubated with TEV protease overnight (Figure 3.24, B). Therefore, the removal of the HRP<sup>EC2</sup> N-terminus was confirmed. Whilst the comparison of the control sample HRP<sup>EC2</sup> and glycosylated HRP further demonstrated the reduction in the size of HRP<sup>EC2</sup> which had occurred (Figure 3.24, A). Initially, when HRP<sup>EC2</sup> was first purified, the size of protein which eluted from the nickel column was consistent with that of glycosylated HRP (Figure 3.23). However, purified



**Figure 3.24. Reaction of HRP<sup>EC2</sup> with TEV Protease.**

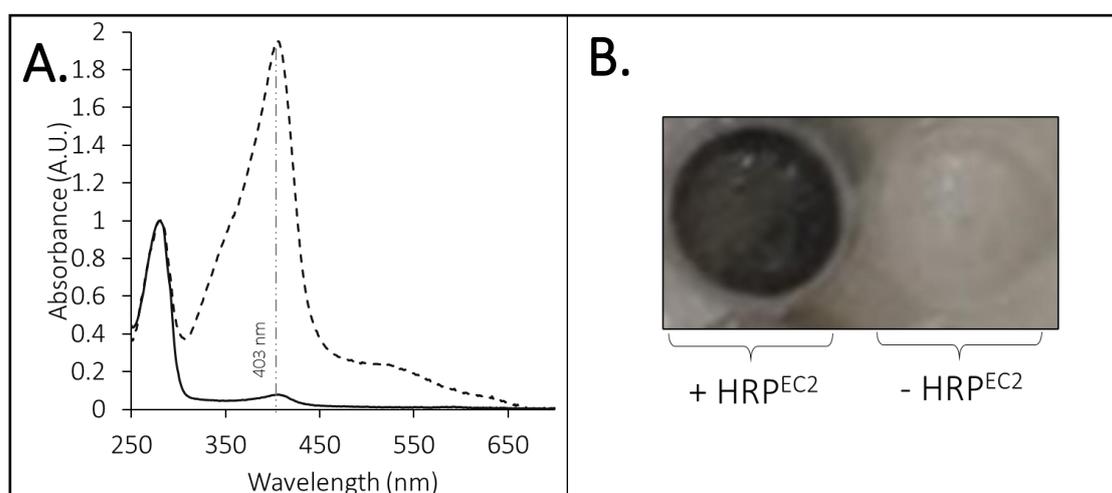
**A**, Size comparison of HRP<sup>EC2</sup> which had undergone suspected cleavage and 2 mg/mL HRP<sup>GLY</sup>. **B**, size of HRP<sup>EC2</sup> observed after incubation with TEV protease for 3 hours.

protein which was first concentrated and then stored at 4 °C was observed to reduce in size and be smaller than glycosylated HRP (Figure 3.24, A).

### 3.2.16. The Activity of HRP<sup>EC2</sup>

Lastly, the absorbance spectrum of HRP<sup>EC2</sup> was recorded to confirm the presence of a haem Soret peak, which is characteristic of a haem binding protein, such as HRP<sup>74</sup> (Figure 3.25, A). The spectrum recorded indicated the presence of a Soret peak at 403 nm consistent with that of plant-produced HRP (see Chapter 5, Figure 5.5). However, pure recombinantly expressed HRP<sup>EC2</sup> was shown to be only partially saturated with haem, with an  $R_z$  value of 0.075 A.U. This was expected as incomplete haem saturation of protein is common in *E. coli* recombinant expression<sup>77</sup>. Subsequently, HRP<sup>EC2</sup> was haem saturated by incubation of 10  $\mu$ M pure protein with a 100 x molar excess of haemin overnight and removal of free haem porphyrin the next day by desalting using an S75 size exclusion column (see Chapter 2.4.). Thus, the absorbance spectrum of saturated HRP<sup>EC2</sup> was characteristic of *holo*-HRP with an  $\lambda_{MAX}$  at 403 nm, although the measured soret peak was only 2/3 the intensity of plant produced HRP at 1.97 A.U. (Figure 3.25, A).

Finally, pure saturated HRP<sup>EC2</sup> was tested for activity by mixing with the substrate OPD in the presence of peroxide (Figure 3.25, B) (Chapter 2.5.9.). Upon the mixing of protein with the substrate, the formation of a visible precipitate was observed, which was not formed in



**Figure 3.25. Absorbance Spectra and Activity of HRP<sup>EC2</sup>.**

**A**, Spectra corrected to a 280 nm absorbance of 1 A.U. *apo*-HRP<sup>EC2</sup> (**Black**) with a calculated  $R_z$  value of 0.46 A.U., and *holo*-HRP<sup>EC2</sup> with an  $R_z$  value of 1.97. **B**, turnover of OPD in the presence of H<sub>2</sub>O<sub>2</sub> by HRP<sup>EC2</sup>. No quantification of activity was performed in this instance.

---

the absence of enzyme. Therefore, the turnover of substrate demonstrated that purified HRP<sup>EC2</sup> contained haem and which enabled catalytic activity.

---

## 3.3. Discussion

### 3.3.1. Expression of Soluble Proto-HRP

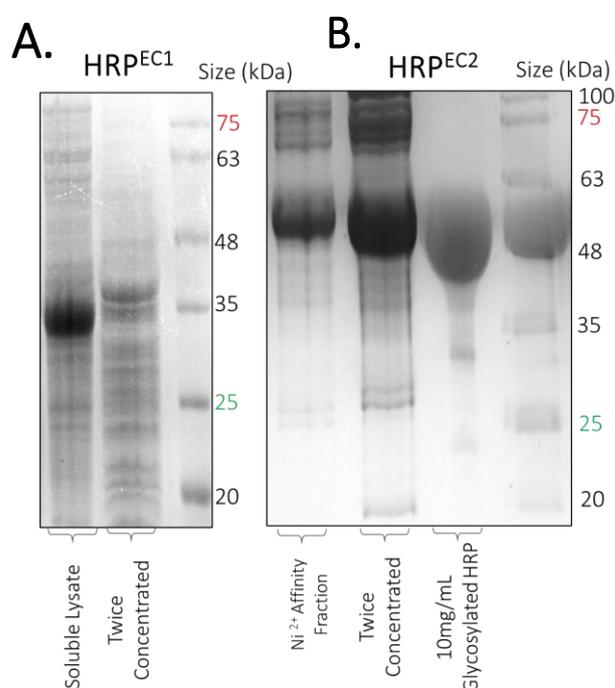
The initial aim of this chapter was to establish if the expression of the full-length proto-HRP sequence produced soluble protein in *E. coli*. Subsequently, expression of both HRP<sup>EC1</sup> (Figures 3.8) and HRP<sup>EC2</sup> (Figure 3.17) was proven to yield soluble protein both in the cell periplasm and within the cell lysate. Thus, the generation of soluble HRP<sup>EC1</sup> and HRP<sup>EC2</sup> eliminated the necessity to refold protein to establish active HRP expression, contrasting previous study<sup>74,75,109</sup>.

As observed in Figures 3.21, 3.22 and 3.23, HRP<sup>EC2</sup> purified by nickel affinity frequently formed a band size on SDS PAGE larger than expected with an observed mass of ~48 kDa. Banding of this size was larger than predicted of the full-length HRP<sup>EC2</sup> (Table 3.2.), however, characteristics observed demonstrate that the protein was consistent with that expected of HRP<sup>EC2</sup>. These attributes include the protein's ability to bind to nickel resin (Figure, 3.18), to uptake haem (Figure 3.25, A), and to oxidise OPD in the presence of H<sub>2</sub>O<sub>2</sub> (Figure 3.25, B). Thus, the previously observed banding within both the cell pellet and the soluble cell fraction (Figures 3.16, and 3.17) at ~48 kDa can be assumed to be HRP<sup>EC2</sup> with the N-terminus intact.

Significantly, this band is observed in greatest intensity within the soluble cell fraction, whilst the smaller band (~35 kDa) is observed at greatest intensity within the insoluble cell fraction (Figure 3.17). Therefore, it is suggested that the expression of the HRP<sup>EC2</sup> with the proto-peptide sequence enhances protein solubility. This ties into findings by Matsui *et al.* (2004), which suggested that protein was degraded in the absence of the N-terminal pre-peptide sequence. It is reasoned that the presence of the pre-peptide in native plant expression prevents degradation and functions to promote solubility before glycosylation. Subsequently, when glycosylation is applied within the ER the solubilising effects of the pre-protein are redundant and the sequence is cleaved. However, in systems of recombinant expression, where glycosylation cannot be used to promote protein solubility, the proto-peptide sequence may not have unnecessary solubilisation effects and may benefit the soluble generation of HRP. It is hypothesised that in the absence of the proto-peptide previous study<sup>49,75</sup> has not observed soluble HRP expression as the protein is isolated in inclusion bodies when the N-terminus is omitted.

3.3.2. Comparison of HRP<sup>EC1</sup> and HRP<sup>EC2</sup>

A common occurrence when working with HRP<sup>EC1</sup> was the inability to establish a sample of high concentration (>5  $\mu\text{M}$ ). Frequently, HRP<sup>EC1</sup> was shown to precipitate (Figure 3.10), and the longevity of the protein was not sufficient for reliable determination of characteristics. This is exemplified in Figure 3.26, which demonstrates the difficulty faced when concentrating HRP<sup>EC1</sup>. The concentration of HRP<sup>EC1</sup> regularly resulted in the complete loss of protein (Figure 3.26, A), which was not observed when the same methodology was applied to HRP<sup>EC2</sup> (Figure 3.26, B). This prevented all practical application of HRP<sup>EC1</sup>, and consistent results were observed from both dialysis and desalting of the sample.



**Figure 3.26. Comparison of the Concentration of HRP<sup>EC1</sup> and HRP<sup>EC2</sup>**

**A**, SDS PAGE of HRP<sup>EC1</sup> lysate, when concentrated by spin filtration. **B**, SDS PAGE of HRP<sup>EC2</sup> purified by affinity chromatography when concentrated by consistent conditions with 10 mg/mL of glycosylated plant protein shown for reference.

In comparison, HRP<sup>EC2</sup> was more resistant to both concentration and storage at 4 °C than HRP<sup>EC1</sup>, making HRP<sup>EC2</sup> a more practical protein variant for later investigation in Chapters 4 and 5. This is reasoned to be evidence that the removal of surface non-glycated asparagine residues is beneficial to protein generation. As, once both HRP<sup>EC1</sup> and HRP<sup>EC2</sup> are cleaved of their N-termini, this is the only difference which exists between the two protein forms. It is hypothesised that removal of the multiple surface-exposed amide side chains, therefore, limits

---

asparagine promoted aggregation by prevention of rogue, non-native, hydrogen bonds, when the protein is non-glycosylated<sup>307</sup>.

### 3.3.3. Proteolytic Cleavage of Recombinantly Expressed HRP

Analysis of the HRP C1A sequence using Signal P 5.0, indicated that the N-terminal protein sequence likely functioned as a signal sequence within Gram-negative bacterium (see Section 3.2.6), which would likely be cleaved by Signal Peptidase I. Signal Peptide I is a class of membrane-bound cleavage enzymes found within *E.coli* and plants alike<sup>315–317</sup>. In *E. coli* the enzymes are typically found upon the periplasmic membrane and cleaves signal sequences as protein is secreted to the periplasm<sup>318</sup>.

Consequently, it is reasoned that the inclusion of the pre-peptide sequence is of additional benefit to the expression of recombinant HRP within *E.coli*. This is rationalised as the periplasm is a non-reducing environment and would thus promote the formation of disulphide bonds in the enzyme. In this investigation, it was not tested if the pre-peptide sequence of HRP alone could initiate export of protein to the periplasm, as HRP<sup>EC1</sup> featured the inclusion of both the pre-peptide sequence and the signal sequence of cytochrome *b*<sub>562</sub>. Therefore, in the future, it is necessary to establish if the HRP pre-peptide sequence can signal localisation of the protein to the periplasm within the non-native host *E.coli*. One proposed means of investigating protein localisation with the pre-peptide sequence is analyse the periplasmic cell fraction of cells which are induced to express HRP<sup>EC2</sup> as this construct does not feature the signal sequence of cytochrome *b*<sub>562</sub>.

Results obtained in this investigation suggest that not all HRP expressed is subjected to cleavage before cell lysis (Figures 3.17, and 3.18), and it is hypothesised that when the rate of enzymatic proto-protein cleavage cannot match the rate of expression, the non-cleaved proto-peptide gradually builds in concentration. Incomplete cleavage of HRP<sup>EC2</sup> is however potentially beneficial to protein purification as it enabled protein purification *via* the N-terminal poly-histidine tag, greatly increasing the rate at which pure HRP could be acquired for investigation. As this pre-peptide sequence was then subsequently removed from the protein, this allows the production of a large yield of HRP<sup>EC2</sup> at fast pace, without the need of TEV cleavage to remove the HIS-TAG prior to haem saturation.

One unexpected observation made in this investigation is the cleavage of the N-terminal peptide sequence from HRP<sup>EC2</sup> post protein purification. Size reduction was not expected when HRP<sup>EC2</sup> was isolated from cell lysate as it was anticipated the protein would no longer be exposed to signal peptidase I. However, cleavage of purified HRP<sup>EC2</sup> contradicts this hypothesis (Figure 3.23). One explanation for this is that the low-level presence of signal peptidase I in purified protein can stimulate proteolysis. This would explain why cleavage was enhanced with the increased concentration of protein, as the reduced volume would likely increase the rate of association between HRP<sup>EC2</sup> and protease. However, another less likely hypothesis is that HRP spontaneously cleaves once purified, which would explain why the process occurs in pure protein samples (Figure 3.23.), and why two separate N-terminal cleavage sites were observed in mass spectrometry analysis (Table 3.3.).

#### 3.3.4. Conclusion

In conclusion, results indicate that the expression of proto-HRP produced both active and soluble HRP. This chapter provides a means of both the expression and purification of recombinant, non-glycosylated, HRP which will be studied in more detail in the two following chapters. The buffer exchange of obtained, mature-HRP<sup>EC2</sup> (sized at ~35 kDa) to ddH<sub>2</sub>O and subsequent freeze-drying indicated an expression yield of 7.1 mg of pure protein from 1 L of overnight cell culture.

This is comparable to the best-reported yield of HRP expression in *P. pastoris* of 7.3 mg/L<sup>77</sup>, but, does not match the 10 mg/mL of protein previously recovered from *E. coli* inclusion bodies<sup>79</sup>. However, it is proposed that with further refinement of both expression and purification conditions this yield may be improved upon in future. The combination of HRP<sup>EC2</sup> with other mutations to improve both solubility and expression may have vast benefits to recombinant production of HRP, and in time increase the yield of purified protein.

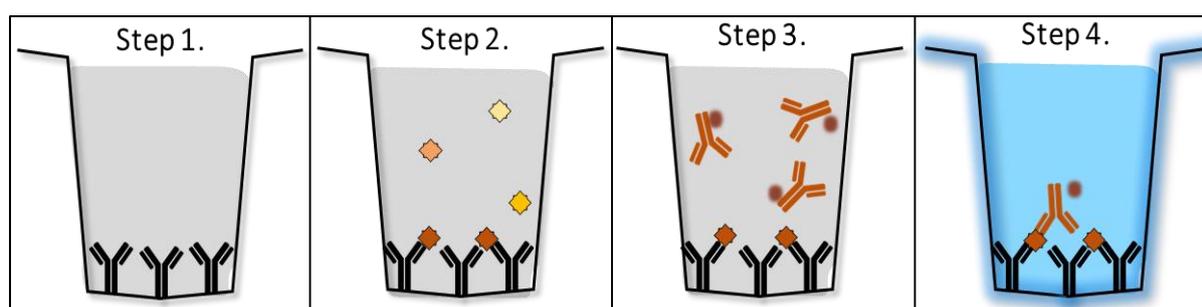
## **Chapter 4.** The Identification and Analysis of Serum Corrective Agent.

## 4.1. Introduction

### 4.1.1. Immunoassays

First conceptualised in 1959, an immunoassay is a process in which a recognition event (between an antibody and an antigen), is enhanced for detection<sup>319</sup>(Figure 4.1). The technique operates through the immobilisation of one antibody, which in turn captures the desired analyte; once analyte is captured, a secondary analyte-specific antibody is applied. This secondary antibody is conjugated with HRP, and where recognition occurs, HRP is localised to wells where the analyte is present. Subsequently, if enzyme turnover is observed in a specific location, analyte presence can be inferred<sup>320</sup>. The process is highly specific and can be employed to identify low concentrations of analyte for medical diagnosis<sup>321</sup>.

HRP is one of the most common enzymes for conjugation to antibodies and is frequently used in immunoassays<sup>9</sup>. The use of HRP-conjugated antibodies has a great increase on the signal observed upon recognition, as enzyme allows for the turnover of multiple substrate molecules from one analyte binding event<sup>322</sup>. This enzyme is typically used in partnership with luminol and can indicate analyte presence through the emission of blue light upon substrate turnover (Figure 4.1.)<sup>319</sup>. Furthermore, immunoassays can be run to completion in as little as 15 minutes, considerably increasing the speed of analyte identification. Consequently, clinicians often rely solely on immunoassay output for analyte detection and diagnosis<sup>323</sup>.



**Figure 4.1. Basic Model of Immunoassay Action.**

**Step 1**, primary antibody is immobilised to well. **Step 2**, patient serum is applied for interaction of analyte (if present) with immobilised antibody. **Step 3**, secondary antibody (conjugated with HRP) is applied to affix to any captured analyte. **Step 4**, well is washed to remove serum before application of luminol and peroxide. Turnover of luminol is observed only in wells where analyte is present and secondary antibody has been captured.

### 4.1.2. False Positive Reporting

A consistent problem within immunoassays is the occurrence of false-positive results. The problem was first reported in the 1970s but persists still today <sup>324</sup>. In a false-positive reaction, the secondary antibody is immobilised without the presence of the analyte <sup>324</sup>. Commonly, false-positive signalling arises from cross-reactive agents native within the patient's serum <sup>325</sup>. The specific cause of false recognition is varied, but the occurrence of a false positive result is common. Current estimates indicate that 40% of human serum samples contain non-specific binding agents which stimulate "rogue" antibody capture. These samples, in turn, then generate output consistent with false-positive diagnosis in 6% of results, despite an absence of test analyte <sup>324</sup>.

Due to the sensitivity of medical diagnosis rogue signalling can lead to considerable consequences to patient wellbeing <sup>323</sup>. Treatment based on false-positive results has previously led to significant outcomes for a patient's wellbeing such as unnecessary chemotherapy and in some cases surgery <sup>326,327</sup>. Due to the severe implications which can arise from each false-positive reaction, elimination of these recognition events is critical to the reliability of results. The solution many commercial companies have turned to is the broadscale blocking of serum. In this approach, many blocking agents are incorporated directly into commercial products to target all potential rogue signalling events <sup>325</sup>.

The company Ortho Clinical Diagnostics (OCD) has developed many agents for use within their immunoassay products (Personal Communication, Dr Chris Thomas, July 2018). These blockers prevent a wide range of reactions and ensure the integrity of resultant diagnosis. One such blocking agent is Serum Correction Factor (SCF)<sup>328</sup>, which is a form of HRP which has been stripped of its functional haem cofactor, by process of butanone extraction <sup>328</sup>. Although, the exact mechanism of SCF's corrective action is unknown, it is hypothesised that the agent corrects either by competitive binding of recognition sites or through haem capture.

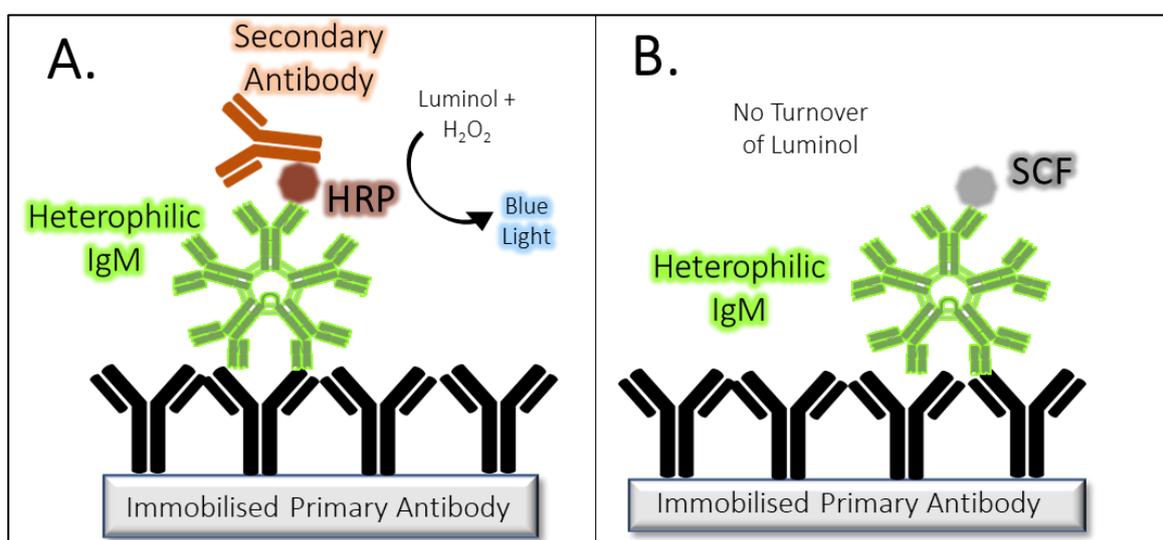
### 4.1.3. Potential Mechanisms of SCF Action

#### 4.1.3.1 Competitive Binding

One possible source of cross-reactivity is the capability of heterophilic antibodies within patient serum to respond to existing immunoassay components (Figure 4.2.) <sup>329</sup>. Heterophilic antibodies are antibodies that have been raised against poorly defined antigens <sup>325,329</sup>.

Ordinarily, these antibodies are of benefit to the immune system as they possess the ability to react with novel foreign antigens before a full immune response is awakened. For example, one such heterophilic antibody is rheumatoid factor which is found within 5% of the adult population. Patients with this antibody are known to exhibit a level of false reactivity within 70% of non-corrected test reactions<sup>276,330</sup>, and, this agent is a known entity of heterophile interference within immunoassays<sup>330</sup>(Figure 4.2, A)<sup>331,332</sup>.

One proposed mechanism of SCF's action is the occupation of the heterophilic "rogue" binding sites which typically react with *holo*-HRP by the inactive *apo*-enzyme. It is hypothesised, that the occupation of sites which would typically bind HRP with the inactive *apo*-protein would effectively silence the false localisation enzyme without extensive serum processing (Figure 4.2, B)<sup>333</sup>. This approach would enable correction as without haem, HRP cannot turnover substrate, and thus, when *apo*-HRP is immobilised over *holo*-HRP no output signal is generated,<sup>334</sup>. However, this proposal also assumes that the structure of *apo*-HRP is similar enough to *holo*-HRP that the same protein recognition events occur. As there is no definitive crystal structure of *apo*-HRP, it is not currently understood if *apo*-HRP is structurally homologous with *holo*-HRP<sup>74</sup>.

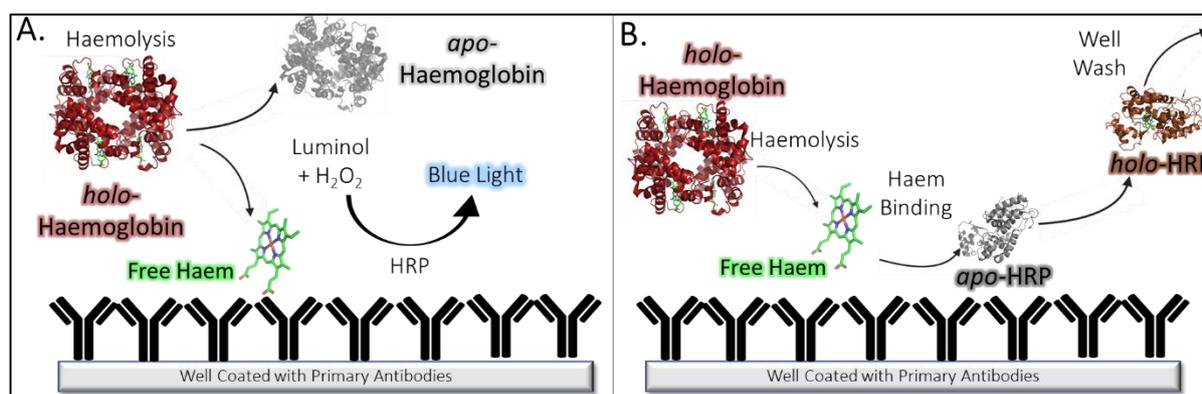


**Figure 4.2. Proposed Mechanism for SCF Action by Competitive Binding**

**A**, a rogue heterophilic IgM antibody, is shown to immobilise secondary antibody. In this instance the IgM antibody had bound both the primary antibody and HRP, which has resulted in a false signalling event. **B**, in the presence of SCF (*apo*-HRP) heterophilic IgM's binding sites are silenced and cannot immobilise HRP.

#### 4.1.3.2. Haem Capture

An alternative explanation for SCF's corrective action is the "mopping up" of rogue haem presence within samples. Haemolysis is the process by which red blood cells burst, releasing both haem and haemoglobin<sup>335</sup>. It is known that haem can induce the catalytic turnover of the substrate luminol in the presence of peroxide, although this process is not as catalytically efficient as that of *holo*-HRP<sup>336–339</sup>. It is also known that haemolysis can be an interfering factor in commercial immunoassays<sup>340–343</sup>. Furthermore, the presence of haemoglobin within the assay is known to reverse the protective effects of SCF<sup>334</sup>. It is possible that if free haem immobilises in test samples, e.g. by cross-reactivity with biotin and subsequent immobilisation to the streptavidin-coated well<sup>344</sup>, a signal would be produced as a result (Figure 4.3, A). Thus, the inclusion of *apo*-HRP would allow for the binding and removal of free haem produced from haemolysis, and removal of the porphyrin, when washed from the well (Figure 4.3, B).



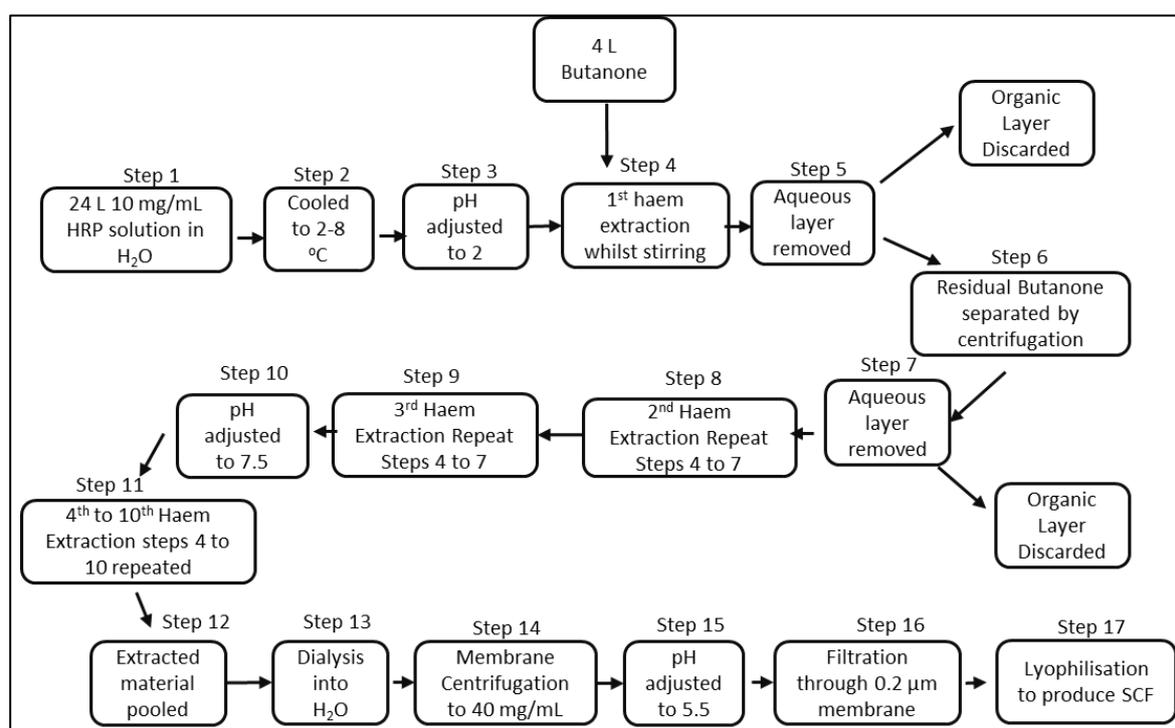
**Figure 4.3. The Proposed Mechanism for SCF Correction by Haem Binding**

**A**, haem released from haemoglobin is introduced to the assay in the event of haemolysis, and binds to the immobilised antibody. Free haem then signals a positive result by turnover of luminol in the presence of peroxide. **B**, application of SCF into test samples removes rogue signalling by the binding and removal of free haem, preventing luminol oxidation.

#### 4.1.4. Production of Serum Correction Factor

Currently, SCF is produced from commercially available *holo*-HRP, by a process of haem extraction. However, as described in Chapter 1.2.4. the production of HRP from the plant expression host features the heterogeneous expression of many peroxidase enzymes<sup>55,66,345</sup>. Consequently, the isoforms present within SCF are uncharacterised, and it is uncertain which isoenzyme(s) exerts corrective function in practice.

SCF is formed from commercial HRP by a process of organic extraction. First, haem is separated from HRP by lowering the pH, and then butanone is applied to phase separate free haem from *apo*-protein<sup>346–349</sup>. This multistep process begins with shipping of HRP following purification of plant cultivars in South Africa to Calzyme, CA, USA. At Calzyme, HRP is processed through 11 rounds of butanone solvation before formed *apo*-HRP is lyophilised and distributed to OCD (Figure 4.4.)<sup>350</sup>. The process is energy-intensive, results in a large-scale loss of protein, and, has an unknown consequence for the glycosylation pattern of HRP. Consequently, an alternative means of production is desired to reduce both the environmental and the financial cost of production.



**Figure 4.4. Protocol for Generation of SCF from Haem Extraction From *holo*-HRP.**

Process performed by Calzyme (CA,USA) for OCD (Pencod, UK).

#### 4.1.5. Chapter Aims

This chapter aims to establish if SCF's corrective effects can be replicated by application of recombinant *apo*-HRP. Firstly, protein SDS PAGE, mass spectrometry, and crystallisation will be employed in SCF analysis to characterise all protein isoforms present and determine if variation exists between batches. If numerous isoenzymes exist, these will be characterised, separated, and compared separately for corrective function.

Secondly, it will be explored if this corrective action can be replicated by recombinant production. The action of recombinant *apo*-HRP will be compared with SCF to determine if the

---

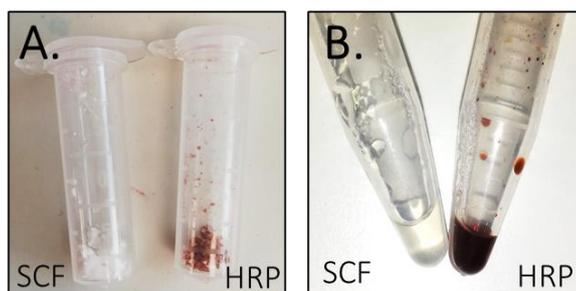
recombinant protein mimics SCF's function and inhibits the false-positive reporting of sera. These results will be contrasted with the alternative recombinant proteins mCherry and the haem binding *apo*-Cytochrome *b*<sub>562</sub>, to determine if corrective action is specific to recombinant *apo*-HRP. It is anticipated that if SCF functions by either competitive binding, its effects will be replicated by recombinantly produced *apo*-HRP. However, if SCF functions by haem capture both recombinant *apo*-HRP and *apo*-Cytochrome *b*<sub>562</sub> will replicate function.

## 4.2. Results and Discussion

### 4.2.1. Characterisation of SCF

#### 4.2.1.1 Appearance of SCF

All batches of SCF and *holo*-HRP used in this investigation were supplied by OCD (Pencoed, UK) in a lyophilised form (Figure 4.5, A). Several batches of both SCF (*apo*-HRP) and *holo*-HRP were supplied so that the two proteins could be accurately compared, and to determine if isoform variation were present between batches. All provided SCF samples were colourless in appearance when rehydrated in Tris 50 mM pH 8 (Figure 4.6, B). The lack of colouration was indicative that most, if not all, haem present in *holo*-HRP had been removed in butanone extraction.



**Figure 4.5. The Physical Appearance of SCF and *holo* - HRP.**

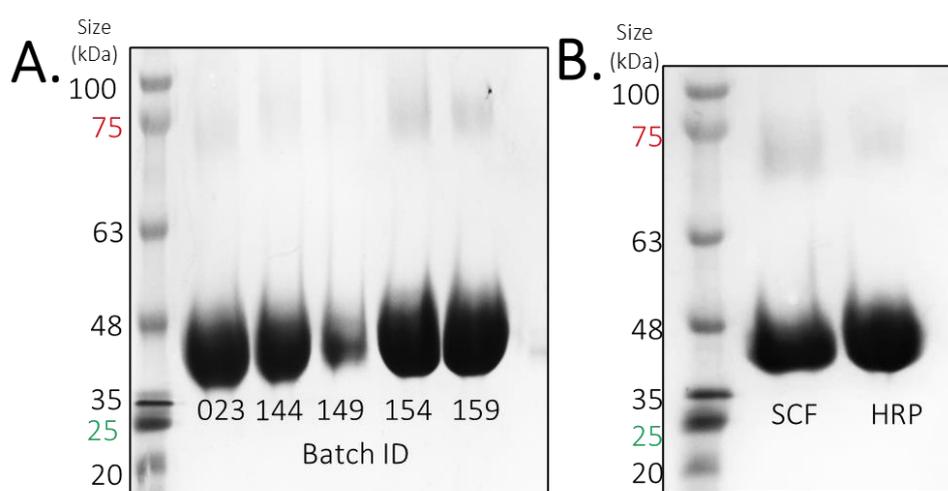
**A**, 50 mg of both SCF and *holo* – HRP are shown in lyophilised form. Noticeably *holo* - HRP is coloured rusty brown whilst SCF is colourless, indicative of the absence of haem in SCF.

**B**, solvated samples after hydration to a concentration of 50 mg/mL in 50  $\mu$ L of Tris 50 mM.

#### 4.2.1.2. SDS PAGE of SCF

To assess if multiple HRP isoforms were present in SCF SDS PAGE was used to analyse the size of protein present in each supplied batch (Figure 4.6, A). Critically, no additional heterogeneously sized isoform contaminants were observed in any batch of SCF provided. The absence of additional isoforms indicated that SCF was homogeneous between batches and was not formed from a mix of widely variant sizes of HRP isozymes. Therefore, it is indicated that corrective action is likely produced from the single protein form sized at  $\sim$ 44 kDa.

The ~44 kDa protein observed, is consistent in size with plant produced *holo*-HRP, which had been supplied by OCD (Figure 4.6, B), and with the expected size of glycosylated-HRP<sup>70</sup>. The consistency in size observed before and after haem extraction indicated that SCF was not degraded in the process of butanone extraction. All samples, additionally, feature the presence of a secondary, larger protein band at approximately 80 kDa. This band was reasoned to be the presence of a disulphide linked protein dimer, as the band was not observed in either low concentration protein preparations or when observed by SDS PAGE in reducing conditions. As Batch 023 was the batch supplied in largest quantity for investigation, this batch was used exclusively for all subsequent analysis.



**Figure 4.6. Non-Reducing SDS PAGE Comparison of SCF Batches, and *holo*-HRP.**

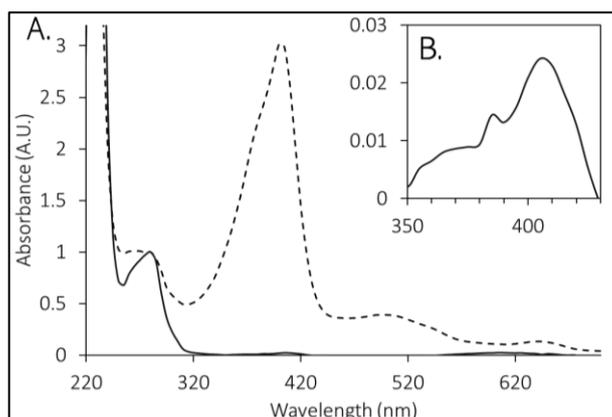
**A**, SDS PAGE of 2 mg/mL of each supplied batch of SCF as provided by OCD after rehydration in 50 mM Tris pH 8. **B**, comparison of SCF (Batch ID 023) and OCD supplied *holo*-HRP.

#### 4.2.1.3. Absorbance Spectrum of SCF

The absorbance spectrum of 2  $\mu$ M SCF was compared with that observed from 2  $\mu$ M *holo*-HRP, to quantify haem saturation (Figure 4.7). Haem absence was evident in SCF, by a calculated  $R_z$  value of 0.02. The low  $R_z$  value is indicative of extremely low haem binding, which is evidence of high efficiency of haem extraction by Calzyme (CA, USA). Haem saturated *holo*-HRP in comparison has a Soret peak absorbance of 2.98 A.U., which is consistent with fully saturated HRP<sup>351</sup>.

Calculation of haem saturation was performed on the spectrum which had been corrected to a 280 nm value of 1 A.U. Absorbance at 403 nm was then compared with absorbance at 280nm to calculate Reinheitszahl value ( $R_z$ ) for each protein (Equation 2.2)(Chapter 2.5.4)<sup>66</sup>. The  $R_z$  was used to indicate haem saturation, where an  $R_z$  value of 3 A.U.,

was used as an indication of 100% haem saturation of protein<sup>352,353</sup>. Therefore, it is established that 99% of the *holo*-HRP sample was haem saturated, and 1% of SCF sample was haem



**Figure 4.7. Haem Saturation of SCF and *holo*-HRP.**

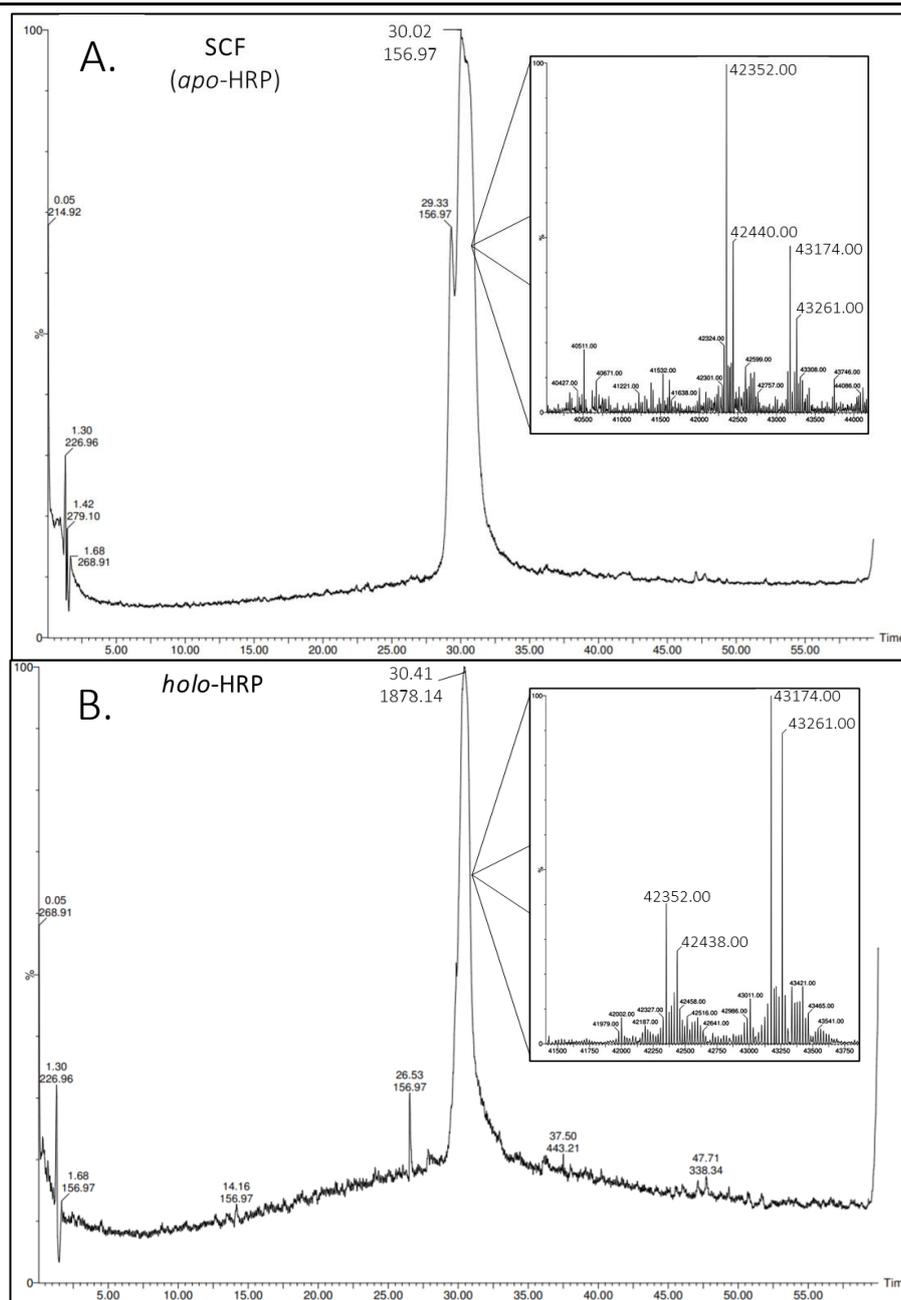
**A**, absorbance spectra of 2 μM SCF (solid line) and 2 μM *holo*-HRP (dashed line), both spectra corrected to 280 nm of 1. **B**, soret peak present at low intensity in SCF, indicating low presence of *holo*-HRP.

saturated.

#### 4.2.1.4. SCF Mass Spectrometry

As discussed in Chapter 1.2.7, eight surface residues of HRP (N13, N57, N158, N186, N198, N214, N255, and N268) are subject to N-linked glycosylation with a glycan chain 1,296 Da in size (Figure 1.4). As surface glycosylation of HRP is extensive, it is necessary to assert if the glycosylation of SCF is influenced by butanone extraction, and to what extent. Also, as SCF is a derivative of commercially supplied *holo*-HRP, SCF may be composed of many isoenzyme forms, any of which may function as a corrective agent. To establish if the glycosylation of HRP is retained in SCF, and to identify which isoforms of HRP are present, mass identification performed by Cardiff School of Chemistry by TOF LC-MS mass spectrometry (see Chapter 2.5.11) (Figure 4.8).

Observed mass identities obtained from SCF (Figure 4.8, A) were contrasted with mass identities of *holo*-HRP to determine if any change in mass was observed upon butanone extraction (Figure 4.8, B). As non-covalently bound cofactors are typically lost when proteins unfold during mass spectrometry<sup>334</sup>, it was reasoned, any change in mass, observed between samples, would be reflective of the influence of butanone extraction.



**Figure 4.8. Mass spectrometry of SCF (A) and *holo*-HRP (B)**

Chromatographs of mass spectrometry with two major peak observed at 30.02 minutes (A) and 30.41 minutes (B). The mass identities within each major peak. Mass identities observed within SCF (A) are, 42,352 Da, 42,440 Da, 43,174 Da, and 43,261 Da. Whilst mass identities observed in *holo*-HRP(B) are, 42,351 Da, 42,438 Da, 43,174 Da, and 43,261 Da.

The mass identities of protein within both SCF and *holo*-HRP were largely homogenous, and a single ionised mass was observed after approximately 30 minutes in each. This ionised mass was formed of four approximately equal mass sizes in both SCF and *holo*-HRP. Three of these peaks (42,353 Da, 43,174 Da and 43,261 Da) were of equal size in each sample. Whilst the fourth mass observed (42,440 Da in SCF and 42,438 Da in *holo*-HRP) was just two Daltons

larger in SCF. This difference was the equivalent to one change in disulphide bond formation, and each was reasoned to be the same entity. Thus, the similarity observed between SCF and *holo*-HRP indicated that butanone haem extract did not influence the glycosylation pattern of SCF. Furthermore, as all protein peaks were larger than the predicted mass of any isoform of HRP listed on the NCBI database, it was evident that all protein presence was glycosylated to some degree.

```
MHFSSSTLFTCITLIPLVCLILHASLSDAQLTPTFYDNPCPNVSNIVRDTIVNELRSDPRIAASILRLHFHDCFVNGC
DASILLDNTTSFRTEKDAFGNANSARGFPVIDRMKAAVESACPRVSCADLLTIAAQSVTLGGPSWRVPLGR
RDSLQAFLDLANANLPAPFFTLPLQKDSFRNVGLNRSSDLVALSGGHTFGKNQCRFIMDRLYNFSNTGLPDPTN
TTYLQTLRGLCPLNGNLSALVDFDLRTPITFDNKYYVNLLEEQKGLIQSDQELFSSPNATDTIPLVRSFANSTQTFN
AFVEAMDRMGNITPLTGTQGQIRLNCRVVNSNSLLHDMVEVVDFVSSM *
```

**Figure 4.9. The Cleavage Site of HRP-C1A, as Reported by Wuhrer *et al.* (2005).**

The complete amino acid sequence of C1A HRP, acquired from UniProt entry P00433, with the residues removed from both N and C terminal cleavage shown in Red. Serine 338, the cleavage site identified by Wuhrer *et al.*, is highlighted in Red <sup>96</sup>.

Two peaks observed in each sample (42,352 Da and 43,174 Da) were observed to be 86-88 Da smaller than the next largest observed peak size. As this size difference was consistent in each sample, it was reasoned that protein present featured two different cleavage sites. The presence of a secondary mass peak, ~87 Da larger and less intense than the HRP's major peak was also observed by Wuhrer *et al.* <sup>96</sup>. Wherein the change was attributed to the presence of HRP-C1A and the non-specific cleavage of the C-terminal at S338 (Figure 4.9). As several sites of N-terminal excision were predicted from the eukaryotic expression of HRP C1A in Chapter 3 (see Chapter 3.2.6), it is therefore likely that the same non-specific cleavage has occurred in this study. Furthermore, utilisation of the Peptide Mass tool on ExPASy ([https://web.expasy.org/peptide\\_mass/](https://web.expasy.org/peptide_mass/)) indicated that loss of a single serine would likely result in a decrease of 87 Da which is consistent with the observed value.

Furthermore, a general consistency across all four mass peaks detected in this investigation, and those observed by Wuhrer *et al.* (2005) is noted. All four peaks observed in this investigation (Figure 4.8) were also reported in the investigation by Wuhrer *et al.* with the exception that peaks observed by Wuhrer were 25 Da smaller than those recorded in this investigation <sup>96</sup>. The size difference of 25 Da cannot be accounted for by a cleavage shift on the N- terminus as the inclusion of alanine would stimulate a change of 73 Da. The difference



Comparison of these mass sizes to existing glycan identities (Table 4.1) produced a mass match with two possible forms of glycan assembly. The smaller of the two glycan masses, 8,520 Da was within 1 Da of the calculated glycan observed from the combination of the chains A + A + A + A + B + B + C + D (Table 4.1). Additionally, the larger of the two observed glycan masses 9,342 Da was calculated to match precisely with the combination of chains A + A + A + A + B + B + B + C (Table 4.1). Both glycan assemblies featured four copies of glycan chain A (Table 4.1), two copies of chain B, and one copy of chain C. However, the smaller of the pair featured the presence of one chain of glycan unit D, whilst the larger mass featured a third copy of glycan chain B. This resulted in a calculated size difference of 821 Da between the pair.

It was determined that both SCF and holo-HRP were composed of mature glycosylated HRP C1A (Table 4.2). This protein form was likely cleaved at the N-terminus in each observed mass at A30, whilst the C-terminal cleavage was heterogeneous. C-terminal cleavage occurred both before and after S338, resulting in the presence of two highly similar protein masses, sized 87 Da apart. All protein observed was glycosylated; however, two forms of glycosylation pattern were present. Combination of the two potential cleavage sites and two forms glycosylation resulted in the presence of four separate peptide masses (Figure 4.8). In SCF (Figure 4.8, A) the major mass peak is 42,352 Da in size, representative of protein cleaved

**Table 4.2. Predicted Peptide and Glycan Identities of the Reported Masses of SCF**

Observed Weight (Da)	Peptide Identity	Glycan Identity <sup>a</sup>	Expected Weight (Da)	Calculated Weight (Da)	Weight Difference Observed (Da)
42352	Mature C1A – Serine	$\frac{(A + A + A + A + B + B + C + D)}{(1201 \times 4) + (1171 \times 2) + 1025 + 350}$	33832 + 8521	42353	1
42440 <sup>b</sup>	Mature C1A	$\frac{(A + A + A + A + B + B + C + D)}{(1201 \times 4) + (1171 \times 2) + 1025 + 350}$	33919 + 8521	42440	0
43174	Mature C1A – Serine	$\frac{(A + A + A + A + B + B + B + C)}{33831 + (1201 \times 4) + (1171 \times 3) + 1025}$	33832 + 9342	43174	0
43261	Mature C1A	$\frac{(A + A + A + A + B + B + B + C)}{33831 + (1201 \times 4) + (1171 \times 3) + 1025}$	33919 + 9342	43261	0

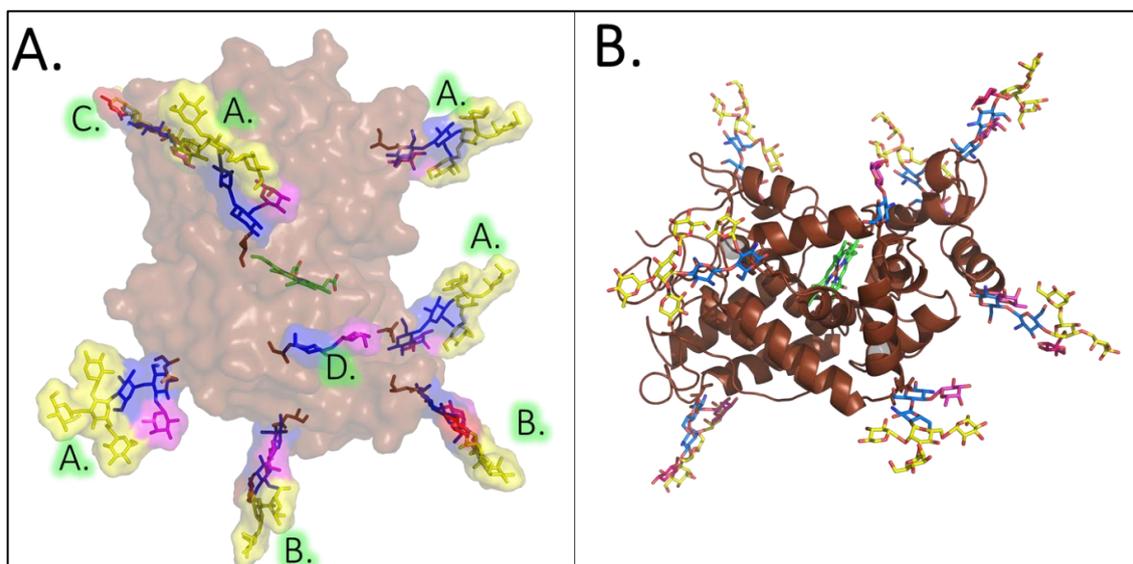
<sup>a</sup> See Table 4.1 for the identity of each glycan chain.

<sup>b</sup> Mass not observed in *holo*-HRP instead mass observed at 42,438 Da, 2 Da smaller

before S338 and featuring the smaller, 8,521 Da, glycosylation pattern. Whilst in holo-HRP, the major peak is 43,174 Da, again indicative of the removal of S338, but with the exception that protein features a larger, 9,342 Da, glycosylation pattern (Table 4.2).

Critically, these masses were consistent in both SCF and holo-HRP (except for a 2 Da difference on one observed mass), indicating that haem extraction did not degrade either the peptide or glycans. The homogenous nature of sample mass identities further verifies that the only difference between the functional holo-HRP as a signal generator and apo-HRP as a corrective agent was the extraction of haem. Furthermore, the presence of glycans on SCF is significant as this cannot be replicated by recombinant means (Chapter 1.2.5)<sup>109</sup>.

The role of glycosylation in the corrective function of SCF is uncertain. However, the extent of glycosylation in SCF is now certain and accounts for a calculated 20-22% of total apo-HRP C1A's weight. Simulated representation of glycans on the surface of HRP using PyMOL (Figure 4.10) demonstrates the significant change in surface morphology which is exerted upon glycosylation. Therefore, it is likely glycosylation influences the protein-protein interactions of SCF, and if SCF functions by competitive binding of rogue agents any heterogeneity may be critical to later recombinant replication efforts.



**Figure 4.10. Visual Representation of Static Glycosylation Pattern of *holo*-HRP.**

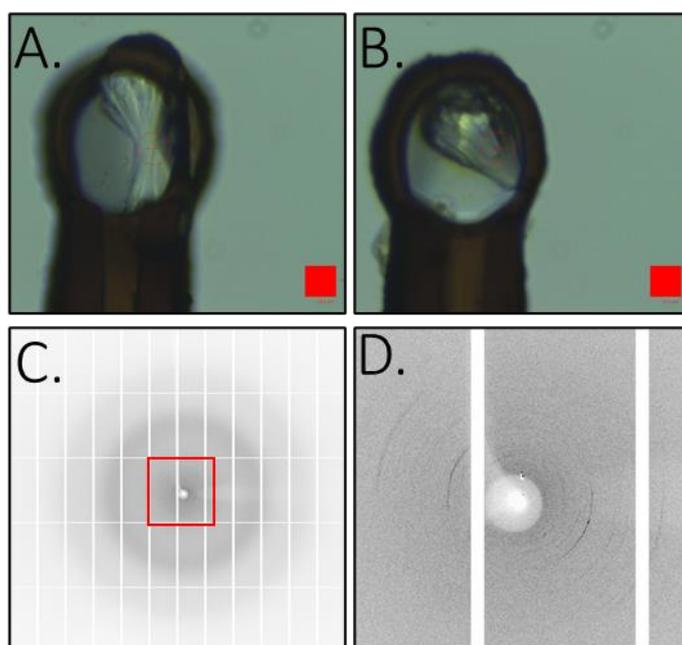
Simulation of identified glycosylation pattern (see Table 4.2) of HRP formed in PyMOL Azahar Plugin, and each individual static glycan unit coloured in accordance with identity (see Table 4.1). Simulated model shown as surface with each static glycan chain labelled by monomeric identity (A). Model also shown as cartoon, with static glycans represented as sticks and coloured by element (B). Carbons of each glycan are coloured in accordance with identity (see Table 4.1).

#### 4.2.2. Crystallisation of SCF

As mentioned previously (Chapter 1.2.7.), no structure currently exists for either the glycosylated or apo HRP. The only structures which exist for the protein are recombinant in origin<sup>81,82</sup>. As mass spectrometry analysis characterised SCF as exclusively glycosylated *apo*-HRP (Table 4.2), the structural influence of both forms of PTM holds significance for our understanding of SCF's corrective action.

To determine the influence the PTMs on SCF, crystallisation of the protein was investigated for later structural determination by X-Ray crystallography (Chapter 2.5.12). Crystallisation, of SCF, however, proved complex. Trials of SCF crystallisation, initially produced no results, as SCF was extremely soluble and exceptionally hard to crystallise in both PACT and JCSG screens. However, the introduction of sucrose to protein samples (ranging from 0-75% w/v sucrose) proved beneficial for the formation of crystals, and several were grown at 37.5 % sucrose concentration.

Titration of 100 mg/mL SCF prepared in Tris pH 50 mM 37.5% sucrose (w/v) pH 8 into 1.4 M  $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7$ , 100 mM HEPES pH 7.5 screening condition generated several protein



**Figure 4.11. Crystallization of SCF and X-Ray Diffraction Grown Crystals**

Magnified image of crystals of SCF grown in 60% sucrose (1.4 M  $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7$ , 100 mM HEPES pH 7.5) observed as colourless double headed broccoli formations (A and B). The overlain red box is used as a guide reference of size and represents  $28 \mu\text{m}^2$ . C, X-ray diffraction of Crystal A, with the central region (Red) enlarged and presented in D. D, extensively disordered diffraction pattern of SCF observed at very low resolution.

---

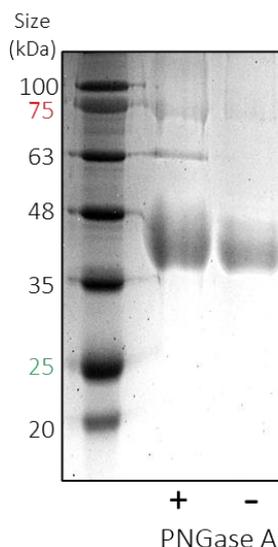
crystals. Crystals were grown for a total of 8 weeks and were observed to be colourless, and morphologically like double-headed broccolis (Figure 4.11, A and B). The colourless nature of crystals was encouraging, as this was consistent with that expected of SCF (Figure 4.6), whilst the complex shape formed was distinctly different from that expected by PEG crystallisation which is typically observed as thin curved flakes (Personal Communication Dr Pierre Rizkallah, April 2018).

Crystals were harvested and transported to Diamond Light Source (Harwell Science and Innovation Campus, Didcot, UK), by Dr Pierre Rizkallah, for X-ray diffraction (Figure 4.11, C and D). The X-ray diffraction pattern obtained was extensively disordered, however, and low resolution. A circular pattern of diffraction was observed, which indicated circular disorder and was representative of layered crystallisation with no correspondence between layers. Due to the absence of register between layers, the cell dimensions could not be established, and no measurement of cell diffraction could be obtained (Personal Communication Dr Pierre Rizkallah, April 2018).

As diffraction was highly disordered, no insight into the structure of SCF could be obtained. Additionally, due to the lack of correspondence between crystallised protein layers, it could not be ruled out that the crystals obtained were not proteinaceous in origin. Likely, the difficulty presented in obtaining both feasible crystal growth and reliable diffraction data was induced by the presence of surface glycosylation on SCF. This is reasoned as the microheterogeneity presented by the presence of glycans, and the increased surface entropy glycosylation induces is a known hindrance to protein crystallisation <sup>355</sup>.

The removal of the protein's glycans was attempted as a solution to this problem. To generate non-glycosylated SCF the enzyme PNGase A was applied to cleave the surface glycans of all surface N-linked glycosylation sites (see Chapter 2.5.15). However, the application of de-glycosylation enzyme PNGase A proved to be ineffective (Figure 4.12). SDS PAGE of SCF, indicated no change in size, between protein treated with PNGase A, and a control sample which had not been treated. No band consistent with de-glycosylated HRP (~35 kDa) was observed after incubation with PNGase A, indicating that the glycans on HRP's surface remained intact and were not cleaved.

Due to the considerable time expense observed in the preparation of crystals for diffraction, attempts of crystallisation were halted. Instead, a more detailed analysis of the biophysical and structural characteristics of glycosylated HRP was performed which will be



**Figure 4.12. PNGase A de-Glycosylation of SCF Observed on non-Reducing SDS PAGE**

The use of PNGase to deglycosylase SCF. No difference in size can be observed from samples which have been treated with PNGase A and those which had not. All protein observed is sized at approximately 44 kDa indicative of monomeric glycosylated HRP. No band is observed at 35 kDa, indicative of non-glycosylated protein.

presented in Chapter 5. Consequently, the computer-generated model for glycosylated HRP (Figure 4.10), was the best current representation of the structure of SCF. Although it is noted that the model is generated *in silico* and the *in vivo* influence of glycans on HRP remains undetermined.

#### 4.2.3. Recombinant Protein Selection

At the beginning of this investigation, it was reasoned that SCF was identical in isoform identity and glycosylation pattern to that of OCD supplied HRP, with the exception that haem porphyrin had been removed. Results acquired thus far support this hypothesis, as SCF featured no absorbance peak at 403 nm and was characteristic of HRP with just 1% haem saturation (Figure 4.7.). Furthermore, mass spectrometry analysis confirmed the presence of surface glycosylation on SCF, was approximately 20% of the total protein weight (Table 4.2) and identified the protein as exclusively isoform C1A, consistent with that of OCD supplied HRP (Table 4.2).

---

As both the *E. coli* expressed recombinant proteins formed in Chapter 3, HRP<sup>EC1</sup> and HRP<sup>EC2</sup>, were both forms of non-glycosylated HRP C1A, both proteins were reasoned to be a suitable agent for recombinant replication of SCF. However, for reasons previously discussed in Chapter 3, HRP<sup>EC1</sup> was not selected for use due to the common precipitation of expressed protein (see Figure 3.10). Therefore, HRP<sup>EC2</sup> was selected as a trial agent of SCF replication, as the protein could be expressed and purified in larger quantity in addition to being more stable when stored at both room temperature and 4 °C. However, one notable difference between HRP<sup>EC2</sup> and SCF was the substitution of the eight surface asparagine residues responsible for glycosylation in HRP<sup>EC2</sup> with alanine residues (Figure 3.15). As HRP<sup>EC2</sup> was not glycosylated due to expression in *E. coli*, this substitution was reasoned to be of minimal functional consequence. Furthermore, as alanine substitution retains the secondary structure of the protein, the structural significance of alanine mutagenesis was also reasoned to have minimal functional influence<sup>312</sup>.

In addition to HRP<sup>EC2</sup>, the recombinantly expressed proteins mCherry and *apo*-Cytochrome *b*<sub>562</sub> were also selected for investigation. The fluorescent protein mCherry was included in the investigation as a control to measure the influence of any general recombinantly expressed and purified protein on the reported intensity of false positive reporting. It was reasoned that inclusion of a protein which was not structurally or functionally homologous to *apo*-HRP would indicate the baseline of correction which can be attained by recombinant protein. In comparison, the inclusion of *apo*-Cytochrome *b*<sub>562</sub> was trialed due to the shared ability to bind haem with *apo*-HRP. It was proposed that if *apo*-HRP were generating corrective function by haem binding within haemolysed samples, this mechanism would be replicated by the application of *apo*-Cytochrome *b*<sub>562</sub>. *Apo*-Cytochrome *b*<sub>562</sub> was selected for use, as it has a high affinity for free haem, with a relatively simple tertiary structure, and is both stable and soluble when overexpressed in *E. coli*<sup>356</sup>.

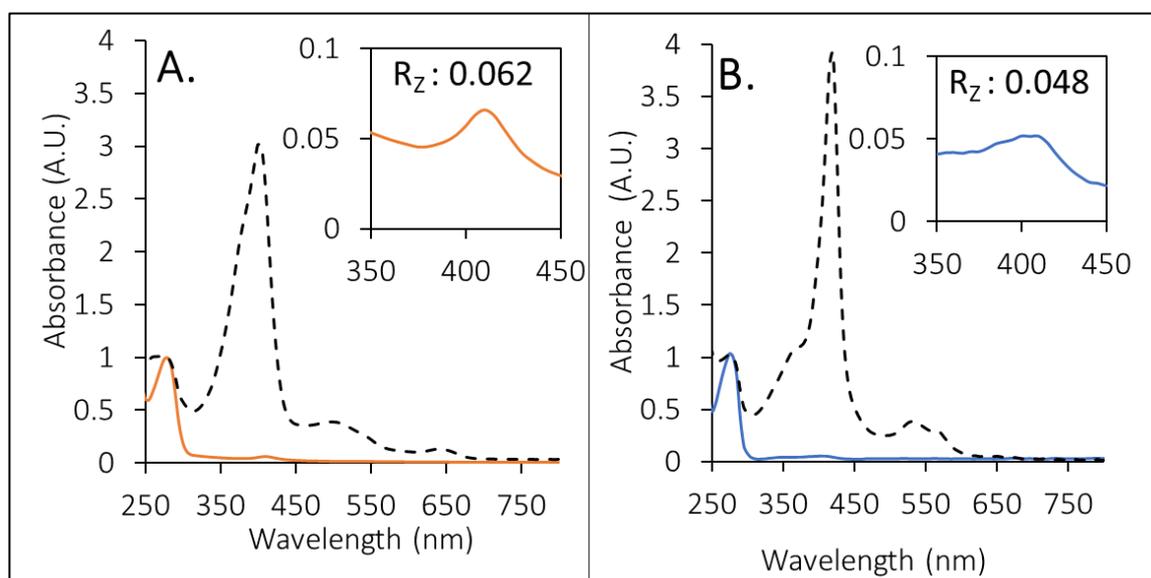
#### 4.2.4. Haem Extraction of Recombinant Protein.

The expression of recombinant protein was performed as previously described in Chapter 2 (see Sections 2.3 and 2.4) and protein purity was confirmed by SDS PAGE before extraction of bound haem from both HRP<sup>EC2</sup> and cytochrome *b*<sub>562</sub>. The absorbance spectrum of expressed HRP<sup>EC2</sup>, which had not been haem saturated, was observed to have an R<sub>z</sub> value of 0.14 A.U., equivalent to 5% of *holo*-HRP (R<sub>z</sub> 3.0 AU). Haem extraction of HRP<sup>EC2</sup> was performed

by buffer exchange of each protein into unfolding buffer (50 mM Tris, 6 M UREA, 1 mM DTT pH 8), and equilibration of protein and buffer overnight at 4 °C. Subsequently, the protein was buffer exchanged back into Tris 50 mM pH 8 by size exclusion, which separated protein (elution volume 35 mL) from free haem (Elution volume >100 mL) in the process. The resultant pure protein was observed to have an  $R_z$  value of 0.062 A.U., equivalent to 2% of *holo*-HRP ( $R_z$  3.0 AU) (Figure 4.13, A).

Whilst recombinant cytochrome  $b_{562}$  was shown to have a far higher  $R_z$  value of 3.9. upon purification, meaning a more drastic haem removal by butanone extraction was necessary (Chapter 2.4.7). The absorbance spectrum of haem extracted protein indicated the low-intensity presence of a Soret peak at 418 nm equivalent to an  $R_z$  value of 0.048 A.U., 1% of *holo*-Cytochrome  $b_{562}$  (Figure 4.13, B).

As both *apo*-HRP<sup>EC2</sup> and *apo*-Cytochrome  $b_{562}$  featured  $R_z$  values <2% of their *holo*-protein derivatives, each was determined to have a significant population of protein able to bind haem >98%. Furthermore, as the  $R_z$  value of SCF was previously calculated to be 1% of



**Figure 4.13. Absorbance Spectra of Recombinant *apo*-Protein used in Immunoassay Trial** **A**, the absorbance spectrum of *apo*-HRP<sup>EC2</sup> (Orange) with insert displaying the calculated  $R_z$  value and residual haem absorbance. The dashed line represents the absorbance spectrum of *holo*-HRP with an  $R_z$  value of 3 A.U. for context. **B**, the absorbance spectrum of *apo*-Cytochrome  $b_{562}$  (Blue) with insert displaying the calculated  $R_z$  value and residual haem absorbance. The dashed line represents the absorbance spectral of *holo*-Cytochrome  $b_{562}$  with an  $R_z$  value of 4 A.U. for context. All data corrected to 280 nm absorbance of 1 A.U. to reflect relative change in  $R_z$ .

---

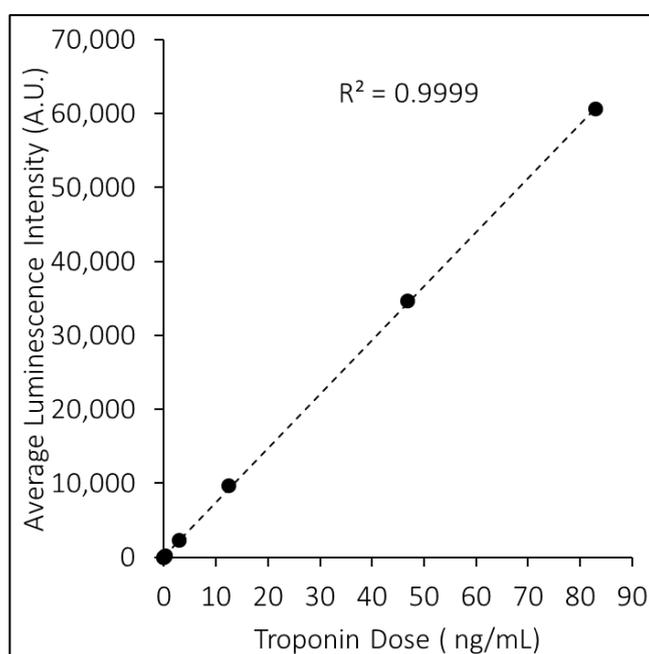
*holo*-HRP (section 4.2.2), it was reasoned that a low proportion of *holo*-protein in each sample would have a minimal effect on potential haem binding action.

#### 4.2.5. Assay Calibration

OCD's commercially available cardiac troponin assay was used as the basis for investigation of recombinant corrective function. The test is commonly performed in hospitals to assess the risk of an immediate heart attack in symptomatic patients, and it determines the urgent prescription of treatment <sup>276</sup>. The presence of cardiac troponin in the serum of the general population is not common, and the protein is generally only present in individuals who are experiencing a myocardial infarction. Consequently, as myocardial infarctions are rare, most positive results of serum from the general population within this assay are not due to the occurrence of a heart attack, but rather a false recognition event <sup>276,329</sup>. It is estimated approximately 5% of serum from a healthy population, not at risk of heart attack, falsely indicate troponin presence in the absence of SCF (Personal Communication, Dr Chris Thomas OCD, June 2019).

To obtain false-positive samples for investigation, serum from the general population was tested for cardiac troponin presence, in the absence of SCF as a correction factor. The patient serum which later demonstrated a decrease in generated signal when SCF was applied, were reasoned to be false-positive serum samples. These samples formed the basis of recombinant protein correction. It was conjectured if the application of recombinant protein, in place of SCF, could replicate the same decrease in signal, correction of rogue recognition has occurred.

Typically, the bioconjugate reagent used in the troponin assay featured SCF presence at a concentration of 0.5 mg/mL. However, for this investigation, a non-commercial sample of bioconjugate reagent was created by OCD which did not feature the presence of SCF. Instead, each recombinant corrective agent was spiked directly into serum samples for a working concentration equivalent to 10 x SCF. Freeze-dried protein was transported to OCD in Pencoed, UK, on dry ice and stored at -80 °C on arrival. When needed, the protein was rehydrated to 1 mg/mL concentrations and spiked into serum directly. All troponin test reactions were performed according to the diagnostic standard protocol (see Chapter 2.6.2.)<sup>276</sup>. As human sera is acellular no ethical statement was necessary for this investigation.

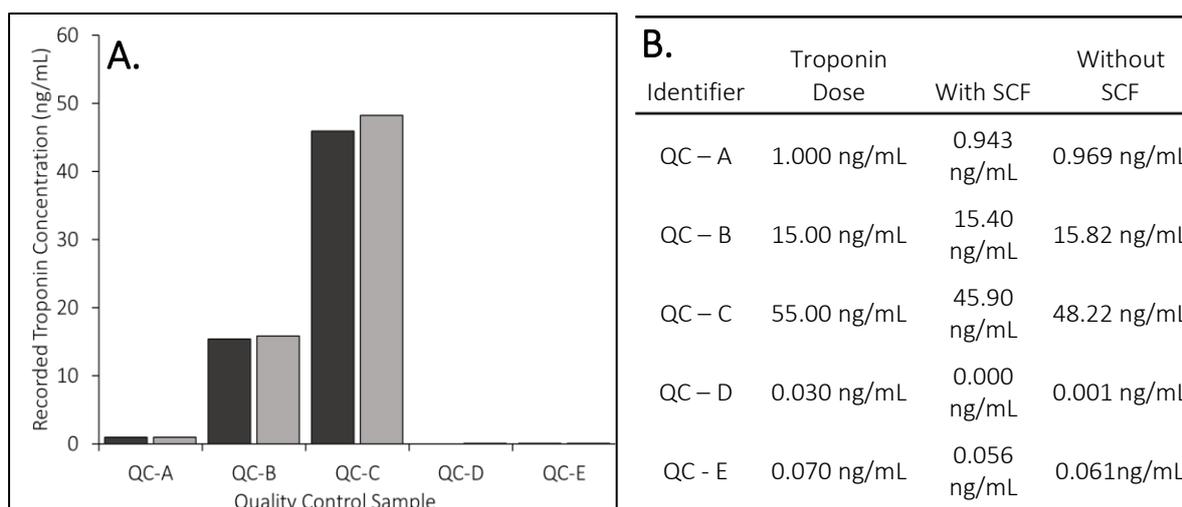


**Figure 4.14. Calibration of Troponin Assay in the Absence of SCF.**

Calibration of standard curve in the absence of SCF. Calibration performed using reference calibrators of known troponin concentration.

Before the investigation of recombinant protein correction, known troponin concentrations were used to establish a standard curve of luminescence against troponin concentration. This calibration was performed in the presence of SCF using reference standard controls of known troponin concentration (0.00 ng/mL, 0.027 ng/mL, 0.067 ng/mL, 0.122 ng/mL, 0.251 ng/mL, 2.87ng/mL, 12.4 ng/mL, 46.9 ng/mL, and 82.9 ng/mL)(Figure 4.14). A linear curve fit was then applied to each plotted data point, and the  $R^2$  value of that curve was calculated to be 0.9999, implying a clear dependency of observed luminescence on troponin concentration.

The calculated linear curve fit was then used to quantify the troponin present within several quality control reference samples, with and without SCF correction (Figure 4.15). Each reference sample was quantified for troponin as standard using the OCD Troponin protocol (see Chapter 2.6.2) and the previously established calibration curve (Figure 4.14) in either the absence or presence of SCF. Calculated troponin concentrations were then plotted (Figure 4.15, A) for comparison between conditions. A general consistency was observed between the quantified concentration of troponin in each sample and each known concentration (Figure 4.15, B). Furthermore, consistency was observed between samples and the presence or absence of SCF (Figure 4.15). This consistency of detected troponin concentration between conditions confirms that genuine troponin is still reported when samples are corrected with SCF. Any change in signal observed in the presence of SCF is, therefore, not reflective of quenching of genuine troponin recognition.



**Figure 4.15. Confirmation of Calibration Curve by Quantification of Troponin Present in Quality Control Reference Agents.**

**A**, the quantification of troponin present in each quality control reference sample as calculated in the presence of SCF (**Dark Grey**) using SCF calibration curve, and in the absence of SCF (**Light Grey**), using calibration curve calculated from standards in the absence of SCF. **B**, the absolute values of troponin concentration in each reference sample, and as calculated in the presence and absence of SCF.

#### 4.2.6. Corrective Action of Recombinant Protein

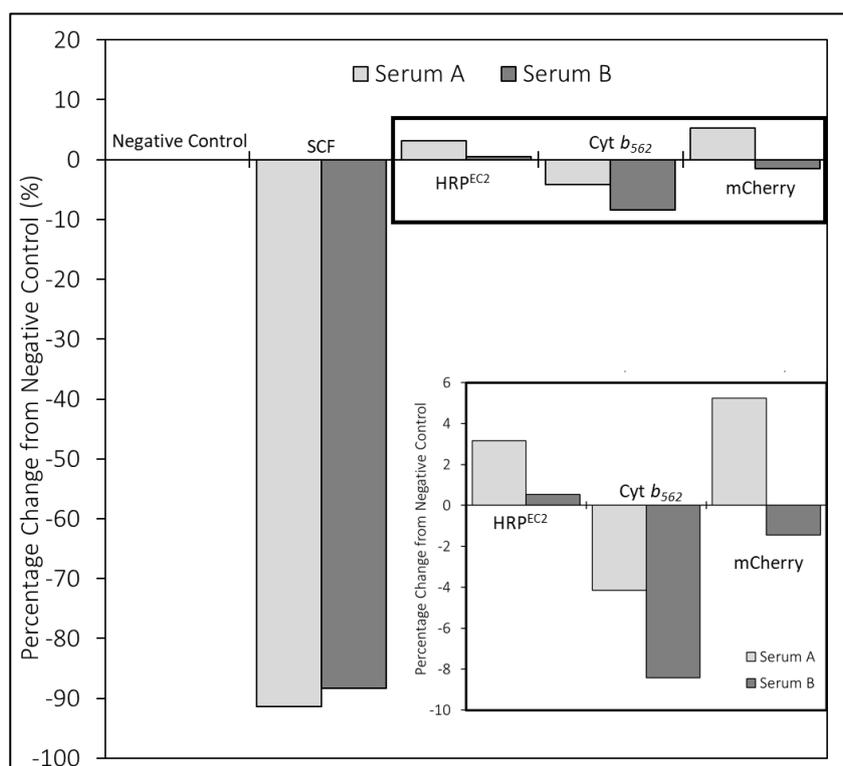
All test sera were first tested for troponin concentration in the absence of any correction factor, to determine a base level of false-positive reaction observed across sera. Then, the same sera were tested once again in the presence of SCF at standard commercial concentration 0.117 mg/mL to measure the decrease in signal observed by the addition of the corrective agent. Each trialled corrective agent (*apo*-HRP<sup>EC2</sup>, *apo*-Cytochrome *b*<sub>562</sub>, and mCherry), were then spiked into sera directly for a final working concentration 10 x SCF in the commercial assay (1.17 mg/mL). Detected values of luminescence from each sample were then converted to indicate perceived troponin presence with the established calibration curve (Figure 4.14). The detected troponin level in sera is presented in Table 4.3.

**Table 4.3. The Average Concentration of Troponin Detected in Sera with Each Corrective Agent of Investigation**

	Serum A	Serum B
Negative Control	0.123 ng/mL	0.086 ng/mL
With SCF	0.025 ng/mL	0.021 ng/mL
<i>apo</i> -HRP <sub>EC2</sub>	0.125 ng/mL	0.087 ng/mL
<i>apo</i> -Cytochrome <i>b</i> <sub>562</sub>	0.119 ng/mL	0.082 ng/mL
mCherry	0.127 ng/mL	0.086 ng/mL

An average of  $0.104 \pm 0.0091$  ng/mL troponin presence was detected in the absence of any corrective factor, whilst a reported concentration of  $0.023 \pm 0.0013$  ng/mL was reported in the presence of SCF (Table 4.3). This decrease of 0.081 ng/mL indicated a decrease of  $78 \pm 0.793\%$  in quantified troponin concentration by application of the correction factor SCF, as standard (Figure 4.16). This decrease indicated that SCF significantly reduced the false positive reporting of troponin within the sera of the general population.

However, reported concentrations of troponin calculated in sera spiked with each trial recombinant correction factor did not replicate this decrease (Table 4.3). Application of both *apo*—HRP<sup>EC2</sup> and mCherry was instead associated with a minor increase in reported troponin concentration from the negative control. An increase of 0.0013 ng/mL was observed upon application of *apo*-HRP<sup>EC2</sup> whilst an increase of 0.0016 ng/mL was calculated upon application of mCherry. This increase equalled a  $1 \pm 0.555\%$  rise in reported troponin concentration in HRP<sup>EC2</sup> spiked samples and  $2 \pm 1.612\%$  in mCherry spiked samples (Figure 4.16.). This increase indicated that neither recombinant *apo*-HRP<sup>EC2</sup> nor mCherry exerted any correction to the false



Percentage Change of Detected Troponin Concentration from Negative Control

	Serum A	Serum B
Negative Control	-	-
SCF	-91.4 %	-88.3 %
<i>apo</i> -HRP <sup>EC2</sup>	3.2 %	0.5 %
<i>apo</i> -Cytochrome <i>b</i> <sub>562</sub>	-4.2 %	-8.4 %
mCherry	5.2 %	-1.5 %

**Figure 4.16. Change in Troponin Concentration Observed by Addition of Corrective Agent of Investigation**

The calculated change in reported troponin concentration observed from false positive sera by addition of: SCF, *apo*-HRP<sup>EC2</sup>, *apo*-Cytochrome *b*<sub>562</sub>, and mCherry, in comparison to sera with no corrective agent (negative control). Insert displays and magnified view of the difference observed from addition of each recombinant protein. Table represents the raw values of percentage change of each corrective agent.

---

positive reporting of sera. Furthermore, as mCherry was included in the investigation as an experimental control, it can be assumed that *apo*-HRP<sup>EC2</sup> functions no better as a correction agent than any non-specific recombinant protein.

Application of *apo*-Cytochrome *b*<sub>562</sub> to sera was the only recombinant protein applied to yield a decrease in reported troponin concentration. A troponin concentration of  $0.101 \pm 0.0094$  ng/mL was representative of a decrease of 0.0038 ng/mL in reported concentration from negative control sera (Table 4.3). Therefore, it can be argued that the application of *apo*-Cytochrome *b*<sub>562</sub> did stimulate a decrease in false-positive reporting, which is possibly enabled by the haem binding function of *apo*-Cytochrome *b*<sub>562</sub>. However, this decrease was equal to a  $4 \pm 1.361\%$  drop reported troponin concentration (Figure 4.16), which was potentially within the range of possible error for investigation and is likely not significant.

Furthermore, the drop of  $4 \pm 1.361\%$  in reported concentration was far lower than the drop of 78% in concentration reported by application of SCF. As 10 x the concentration of *apo*-Cytochrome *b*<sub>562</sub> was applied to samples to achieve this change, a minimum *apo*-Cytochrome *b*<sub>562</sub> concentration ~200 x greater than SCF would be necessary to replicate the same efficiency of replication, and only if a linear dose-effect were observed. Therefore *apo*-Cytochrome *b*<sub>562</sub> is not a suitable recombinant protein for use as a correction factor.

---

## 4.3. Discussion

### 4.3.1. Identity of SCF

The initial hypothesis of this investigation was that SCF was formed of a heterogeneous assembly of HRP isoenzymes, any one of which could feature corrective properties. This array of isoenzymes was also anticipated to be heterogeneous between different production batches of SCF, due to their expression dependence on abiotic factors of plant growth<sup>54,55</sup>. However, results obtained by mass spectrometry indicated that SCF was formed of one singular isoenzyme, HRP C1A (Table 4.2). Furthermore, SDS PAGE of several batches of SCF indicated that the protein consistently sized at ~44 kDa and that no heterogeneity in isoform presence was observed between batches (Figure 4.5). Whilst calculation of the enzyme's  $R_z$  value confirmed that the agent was, as expected, 99% *holo*-HRP (Figure 4.7) and that butanone extraction (Figure 4.4), had almost eliminated all haem presence.

Mass spectrometry of SCF indicated further homogeneity with commercial *holo*-HRP as each was glycosylated (Table 4.2). This glycosylation pattern was, furthermore, consistent between both *holo*-HRP and SCF (Figure 4.8.) and indicated that glycans were not influenced by butanone extraction. Mass spectrometry further indicated that SCF was formed of HRP C1A isoenzyme, and the best means of recombinant replication would be the utilisation of the same sequence and haem extraction of the expressed protein. However, results indicate that glycosylation is essential for the replication of SCF's corrective function, as non-glycosylated *apo*-HRP did not yield corrective effect (Figure 4.16).

### 4.3.2. Replication of SCF by Recombinant Expression

The application of *apo*-Cytochrome *b*<sub>562</sub> was the only recombinant protein used in this investigation which exerted a decrease in the false signalling of troponin (Figure 4.16). A decrease of 4% of the false-positive signal, was encouraging for the potential link between haem binding and reduction in false-positive reporting. However, as the application of *apo*-cytochrome *b*<sub>562</sub> was performed at 10 x the working concentration of SCF, it was 20 x less efficient at reducing false positive reporting. Therefore, it is likely that haem binding is not a valid mechanism of corrective replication, and *apo*-Cytochrome *b*<sub>562</sub> is not applicable for widescale commercial correction.

---

Whilst the reverse effect was observed upon application of *apo*-HRP<sup>EC2</sup> to samples, whereby an increase of  $1 \pm 0.0555\%$  troponin concentration was reported (Figure 4.16). It is, therefore, reasoned that the critical biochemical stimulus of correction, is glycosylation, as the residues of glycosylation, and glycan presence, are the only difference between SCF and HRP<sup>EC2</sup>. It is reasoned that glycosylation may exert a more extensive influence within immunoassay than previously anticipated. Therefore, SCF's corrective properties are likely derived from protein-protein recognition rather than free haem binding, as free haem binding would have been replicated in the application of both *apo*-cytochrome *b*<sub>562</sub> and *apo*-HRP<sup>EC2</sup>.

### 4.3.3. The Importance of Glycosylation

Glycosylation can exert a significant influence on protein-protein recognition<sup>24,101,219,357</sup>, and therefore, may exert significant influence within test sera and consequently false reporting of immunoassay. Glycosylation is a PTM of significant importance to protein-protein recognition and has previously limited the therapeutic use of HRP<sup>54,358</sup>. The absence of glycans on protein is known to be linked to: reduced stability<sup>359</sup>, altered folding<sup>360</sup>, and protection against aggregation<sup>361</sup>, all of which would influence the recognition of SCF in test sera. Glycans themselves are also known to be key signalling agents in the occurrence of lethal congenital conditions such as hypogammaglobulinemia<sup>362,363</sup> and the onset of autoimmune conditions such as rheumatoid arthritis<sup>364</sup> and IgA Nephropathy<sup>365</sup>.

As glycans are demonstrated in this data to increase the protein's weight by 20%, it stands to reason these glycans may have a critical impact on the use of *apo*-HRP as a corrective factor<sup>54,74</sup>. Therefore, it is unlikely that the replication effects of SCF can be achieved by recombinant expression in *E.coli*, as cells cannot glycosylate protein<sup>75</sup>. It is hypothesised that a rogue recognition of *holo*-HRP glycans by immunoglobulins within patient sera stimulate subsequent false-positive signalling (see Figure 4.2). Therefore, glycosylation of SCF is critical to the removal of rogue signalling as the *apo*-enzyme can only competitively bind rogue recognition sites when glycosylated.

In future, de-glycosylation of SCF by alternative means, e.g. alkaline hydrolysis<sup>366</sup>, should be investigated, so that the protein-free glycans can be tested for corrective action, and further attempts of crystallisation of SCF can be trialled in the absence of glycosylation. It is reasoned that application of the glycans of either *holo*-HRP or SCF will function as competitive

---

binding agents and reduce the potential of HRP immobilisation within sera. Furthermore, if successful the application of glycans will eliminate the necessity for HRP haem extraction, which would reduce the carbon footprint of SCF.

#### 4.3.4. Conclusion

In summary, the application of recombinant *apo*-HRP<sup>EC2</sup> to sera did not reduce the intensity of false-positive reporting in immunoassay. However, results provide insight into the importance of the PTM glycosylation on the occurrence and removal of false-positive immune assay recognition. Future investigation should address the influence of glycosylation of SCF's function as a corrective factor, and if correction can be attained by the direct application of glycans to test sera. Chapter 5 will build on these results to determine the significance PTM exerts on the structural rigidity of both *apo* and *holo*-HRP.

**Chapter 5.** The Impact of Post Translational Modification on the Structure and Function of Horseradish Peroxidase.

## 5.1 Introduction

### 5.1.1. The Structural Influence of PTM

Post-translational modification is a process which has important implications for a protein's function and general biophysical characteristics (see Chapter 1.2). The change exerted by PTM is unique to each protein, and consistent effects are not observed across all proteins<sup>55,367</sup>. As PTM influence is varied; therefore, their associated modification to protein characteristics cannot be replicated reliably computationally, and the effect of modification cannot be predicted<sup>53</sup>.

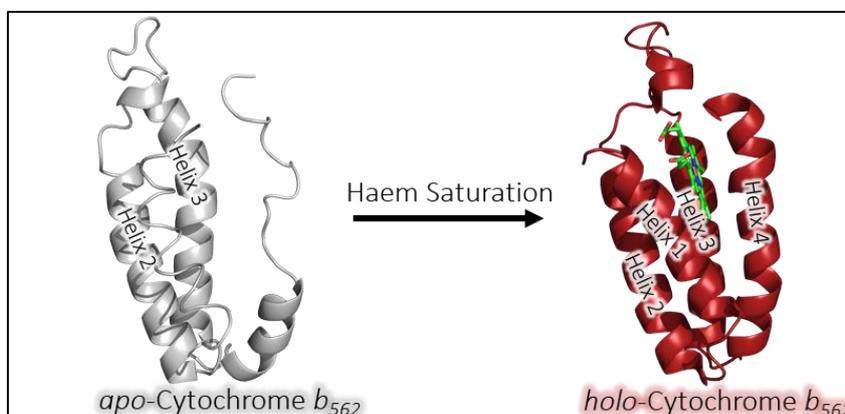
N-linked glycosylation of plant proteins is known to exert a significant influence upon protein folding and stability, while playing a key role in plant protein-protein interaction<sup>368</sup>. The exact role of glycans on plant peroxidases is the focus of numerous glycobiology studies with mixed results<sup>60,84,91,369</sup>. Removal of all glycans has been reported to reduce enzyme stability in the case of peanut peroxidase<sup>370</sup> and solubility in the case of HRP<sup>371</sup>. Whilst individual manipulation of HRP has been reported to decrease the thermostability, activity, and peroxide tolerance of HRP<sup>54,77,79,101,114</sup>.

Glycosylation increases both the weight and the hydrodynamic radius of a protein; however, glycan presence enhances protein stability. This improved stability is theorised to occur not as a result of stabilisation of the protein's folded state but rather the destabilisation of unfolded protein when glycosylated<sup>93</sup>. As glycosylation of HRP increases protein size by 20%, the associated change to structural dynamics is likely to be significant<sup>55</sup>.

Cofactor binding, in contrast, is a well-established process of enabling protein activity<sup>372</sup>. However, the structural implications of cofactor binding are not as well characterised. For example, haem binding has been previously observed to induce protein structure, in some proteins, such as cytochrome *b*<sub>562</sub> (Figure 5.1)<sup>373,374</sup>. The induction of structural folding observed by haem presence thus contributes to increased protein stability, and decreased flexibility.

However, the effects of haem binding, like glycosylation, are not conserved homogeneously in all proteins as haem binding methods differ between proteins and protein groups. For example, cytochrome *b*<sub>562</sub> and other *b* type cytochromes do not covalently bind

haem and are not catalytically active<sup>375</sup>, while cytochrome *c* type proteins do covalently bind haem and are typically catalytically active. As HRP is catalytically active but does not form a formal covalent bond haem, the effects of haem binding in HRP may not be consistent with that of the inactive cytochrome *b*<sub>562</sub>.



**Figure 5.1. Structural Rearrangement of Cytochrome *b*<sub>562</sub> Upon Binding of Haem.**

The structural change from a partially folded structure of *apo*-cytochrome *b*<sub>562</sub> to a fully folded 4-helical bundle structure of *holo*-cytochrome *b*<sub>562</sub> upon binding haem. This forms a helical shield arrangement around the porphyrin consisting of four  $\alpha$ -helices, which does not occur in the absence of haem.

### 5.1.2. The Significance HRP PTM

The only crystal structures of HRP in existence are that on non-glycosylated *holo*-HRP produced through recombinant means<sup>81,82</sup>. In the absence of crystal structures for both glycosylated and *apo*-HRP, the structural implications of both these forms of PTM on HRP are undetermined. As the heterologously expressed HRP can feature heterogeneous glycan assembly in *A.rusticana*<sup>376</sup>, non-native glycosylation in *P.pastoris*<sup>77</sup>, or no glycosylation<sup>24</sup> and predominantly *apo*-HRP formation in *E.coli*<sup>74</sup>, the effect of PTM is of considerable importance for recombinant replication.

Attempts to crystallise non-glycosylated *apo*-HRP were of limited success in Chapter 4, and no discernible diffraction data was obtained. Additionally, immunoassay trials indicated that the ability of recombinant *apo*-HRP to function as a corrective agent was eradicated in the absence of glycosylation. This is indicative that the non-glycosylated enzyme may be structurally heterogeneous to the glycosylated enzyme and unable to function in the hypothesised competitive binding of SCF. Results produced in Chapter 3 indicate that recombinant HRP remains soluble and active in the absence of glycosylation. However,

---

substitution of non-glycosylated surface asparagine residues for alanine was shown to increase protein stability in terms of its resistance to aggregation.

A greater understanding of the implications of both haem saturation and glycosylation will have significant implication for the interpretation of these results. As discussed in Chapter 4.3, crystallography will not be pursued for this endeavour due to the difficulty in generating glycosylated protein crystals. As such an alternate means of measurement was required for this analysis. In this investigation measurement of intrinsic protein, fluorescence will be used to determine and contrast the red edge excitation shift (REES) associated with several HRP PTM states.

### 5.1.3. Red Edge Excitation Shift

Fluorescence is the process by which absorption of a photon, amongst other stimulus, can promote an electron to an excited state <sup>377</sup>. This excited state is retained for several seconds, during which time vibrational relaxation of the electron to lower energy levels occurs. Then energy is released from the system as the excited electron returns to ground state by release of photon. Photons released are lower in energy than those absorbed which results in a stokes shift of light emission to lower wavelengths of energy <sup>378</sup>.

The transition of electron between ground and excited state in fluorescence stimulates a subsequent formation of a larger dipole moment in the excited fluorophore that was present within the relaxed ground-state fluorophore <sup>150</sup>. Thus upon excitation, solvent molecules surrounding the molecule are stimulated to reorganise by the dipole moment of the excited fluorophore <sup>379</sup>. The rate of solvent reorganisation is dependent on the number of solvent molecules which surround the fluorophore and can be slower than electronic redistribution which occurs between excited and ground state fluorophores <sup>268</sup>.

Typically, in a fully solvated environment, reordering of solvent occurs around the excited-state dipole within the lifetime of the excited-state before electrons return to their ground state. However, in a solvent occluded environment solvent reorganisation does not outpace the rate of emission and emission occurs first <sup>380</sup>. In solvent restricted environments, excitation occurs at progressively lower energy wavelengths and the excitation wavelength is gradually red shifted<sup>381</sup>. This phenomenon gradually selects for fluorophores which excite

---

towards the red end of the spectrum at a lower energy (longer) wavelength, with an inhomogeneous broadening of the spectrum and gives rise to red-edge excitation shift (REES)<sup>382</sup>.

#### 5.1.4. Using REES to infer Protein Structural Characteristics

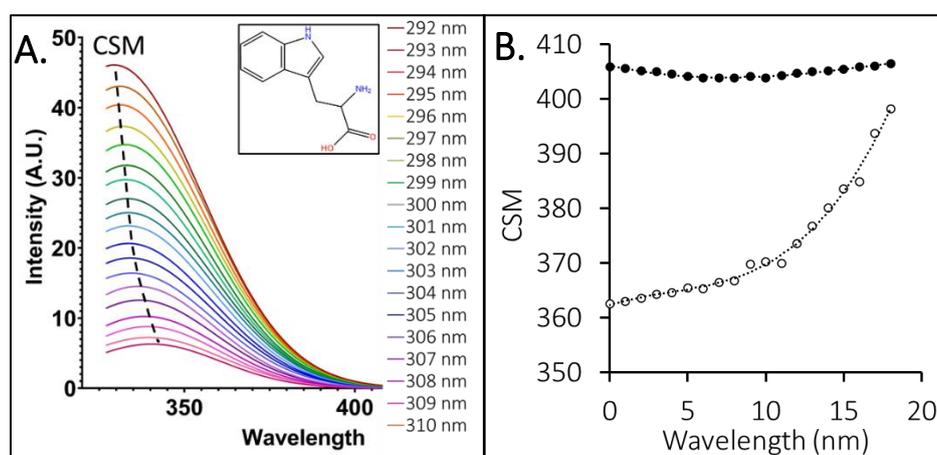
The REES effect is a unique measurement which can be used to infer fluorophore solvation, despite the optical silence of water molecules<sup>383</sup>. Understanding solvation variation is important to our understanding of a protein as water is critical for protein folding and is essential for internal dynamics, electron shuttling, and ligand rearrangement. Demchenko *et al.* first used REES as a means of probing the structural shift of human serum albumin upon isomerisation in 1981<sup>384</sup>. The group then later established that denatured protein exerts no REES effect as their fluorophores are fully solvent-exposed<sup>380</sup>. Since these initial experiments, the effect has been further used to track protein aggregation, ligand conformational change, and protein labelling<sup>152,383,385</sup>.

Investigation of a protein's REES can be used to indicate both localised and global protein flexibility by remote probing of fluorophore environments<sup>386</sup>. When a protein is folded, it has an intrinsic fluorescence capability which arises from the three canonical amino acids tryptophan, tyrosine, and phenylalanine<sup>387</sup>. Tryptophan is typically selected for REES measurement due to its unique spectral properties, its quantum yield of ~20 %, and its low insertion rate in protein<sup>379</sup>.

In a tightly folded protein, it is common for all internal tryptophan's present to be solvent restricted<sup>385</sup>. Whereas in a less rigid arrangement of protein structure, some tryptophan residues may no longer be isolated from the solvent. When excited, the solvent restricted amino acids will emit light before solvent reorganisation and dipole relaxation. Resultantly, this will produce an increased centre of spectral mass (CSM) of observed emission. The CSM is not the emission maximum but the overall average emission for each wavelength of excitation. In fluorophores of greatest solvent isolation, red shifting of excitation will gradually grow in intensity and eventually outpace an increase in the wavelength of excitation, leading to an observable REES effect (Figure 5.2)<sup>388</sup>. Therefore, fluorophores which are in a

tightly defined environment and which are solvent occluded will reduce in CSM and the spectrum will instead, blue shift <sup>389,390</sup>

Using the methodology as described by Catici *et al.* (2016) plotting the change in CSM against change in excitation wavelength gives a quantitative indication of the intensity of REES observed (Figure 5.2, B)<sup>385</sup>. Fitting the observed change in CSM with a non-linear exponential growth curve (Equation 2.7.) allows for the calculation of the amplitude of change ( $A$ ) and the curvature of the line ( $R$ ). Using both these values together can then generate an indication of the overall REES effect ( $A/R$ ) for each protein variant, which is reflective of the rate at which



**Figure 5.2. Red Shift of Tryptophan Emission in Folded and Unfolded HRP**

**A**, the shift in emission observed from a partially solvated fluorophore by increase of the wavelength of excitation. Spectra is used to calculate centre of spectral mass (CSM) of folded protein observed in **B**. Insert shows tryptophan with fluorescent indole group for reference. **B**, the shift in CSM of folded (**White**) and unfolded protein (**Black**) and wavelength of excitation is increased. Curve fitted to emission spectra is shown increase in gradient as wavelengths of excitation is increase, indicating REES effect. Data recorded from initial experimentation with F68C HRP and not used in this investigation.

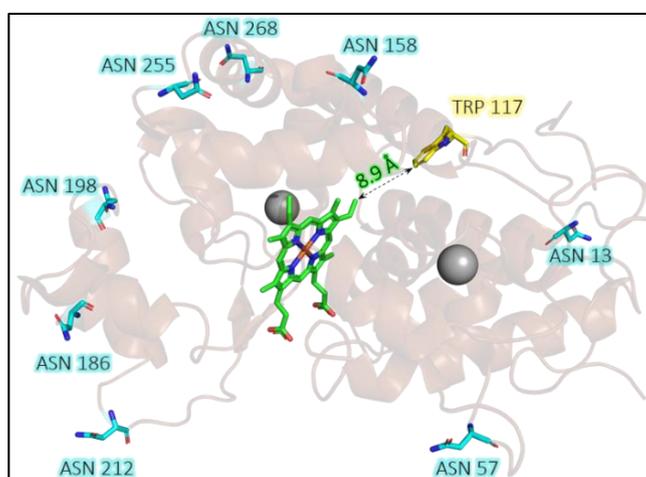
excitation shift occurs in respect to increase in excitation <sup>391,392</sup>. Contrasting the calculated REES intensity between the same protein in different conditions indicates the associated changes observed to fluorophore solvation and protein flexibility. The method can be applied to compare protein states, to ascertain which conditions yield the most tightly folded, or rigid, protein confirmation <sup>268</sup>.

#### 5.1.4. Chapter Aims and Objectives

This chapter aims to assert what structural information can be attained from REES measurement of several PTM variants of HRP. Measurement of CD spectra,  $T_m$ , and determined

catalytic efficiency will be used with to exude information on how protein structure dynamically changes in multiple PTM states, and to contrast with each observed REES effect. W117 being the only tryptophan present in HRP, is an ideal probe of protein flexibility as any REES effect observed will not be convoluted by the presence of multiple tryptophan emission signals at various states of solvation (Figure 5.3). Furthermore, the residue is located internal to the protein's structure and is within proximity, 8.9 Å to the bound haem, which makes the fluorophore an ideal reporter for modification to the functional centre of HRP. It is anticipated that W117 will be an effective reporter for any potential change to the active site of HRP which occurs upon haem binding and glycosylation.

This investigation will determine if variation in the REES effect in HRP can represent the dynamic structural influence exerted on the protein by both glycosylation, and haem saturation. It is hypothesised that there is a significant change in protein flexibility in the presence of each PTM which will result in a change to W117 solvation and a reported change in REES intensity will represent this and recorded CD spectra. This change is expected to manifest as a decreased structural flexibility in the presence of both PTMs, haem coordination, and glycosylation.



**Figure 5.3. Proximity of Tryptophan 117 to Haem and Glycosylation Sites in HRP.**

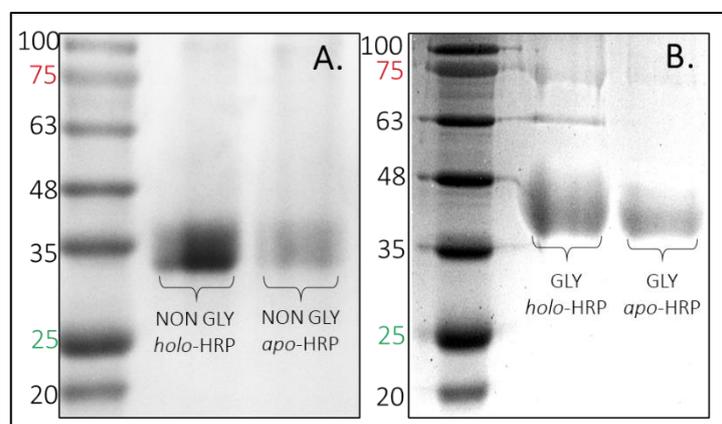
The measured distance between tryptophan 117 (Yellow) and haem porphyrin (Green). Surface glycosylation sites of HRP (Teal) show additionally to assess proximity to 117.

## 5.2. Results and Discussion

### 5.2.1. Protein Used in Investigation

As established, in Chapter 4 (see Figure 4.12), de-glycosylation of plant produced HRP was not possible. In place of de-glycosylated plant protein, recombinant non-glycosylated HRP produced in Chapter 3 was used. HRP<sup>EC1</sup> was initially trialled for use in this experiment; however, the variant was not stable for the length of time necessary for data collection. Instead, the second protein variant HRP<sup>EC2</sup> was used as non-glycosylated HRP.

HRP<sup>EC2</sup> was expressed and purified as described in Chapter 2.4.5. before buffer exchange into ddH<sub>2</sub>O and lyophilisation, so that both glycosylated and non-glycosylated protein was stored consistently. When necessary, protein was rehydrated in Tris buffer (50 mM Tris pH 8) to the desired protein concentration. Protein was gently agitated for 30 minutes upon rehydration before centrifugation at 150,000 xg for 5 minutes to remove the residual lyophilised powder, and concentration was confirmed *via* Bradford assay (see Chapter 2.5.2). The size of each protein was confirmed by SDS PAGE, to confirm that native HRP was glycosylated and approximately 12 kDa larger than pure HRP<sup>EC2</sup> (Figure 5.4)

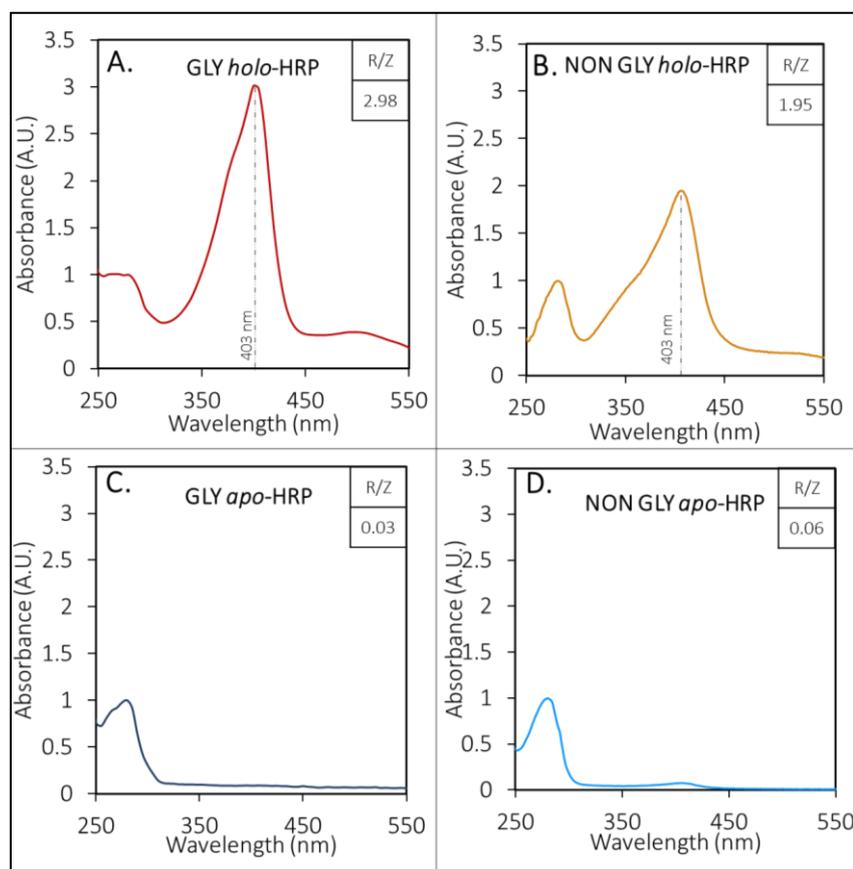


**Figure 5.4. Protein used in investigation.**

**A**, non-glycosylated recombinant protein. **B**, glycosylated natively produced plant protein.

Measurement of absorbance spectra then confirmed protein concentration and haem saturation before use (Figure 5.5). However, despite the excess of both cofactors, the  $R_z$  value of non-glycosylated *holo*-HRP did not match that of glycosylated *holo*-HRP being 1.95, and 2.98, respectively. The observed difference in  $R_z$  indicated that non-glycosylated HRP featured a Soret peak 2/3rds the size of glycosylated-HRP. Non-glycosylated *apo*-HRP featured an  $R_z$  value

of 0.06, equivalent to 2% haem saturation. Whilst glycosylated *apo*-HRP, featured an  $R_z$  value of 0.03, equivalent to 1.0% haem saturation.



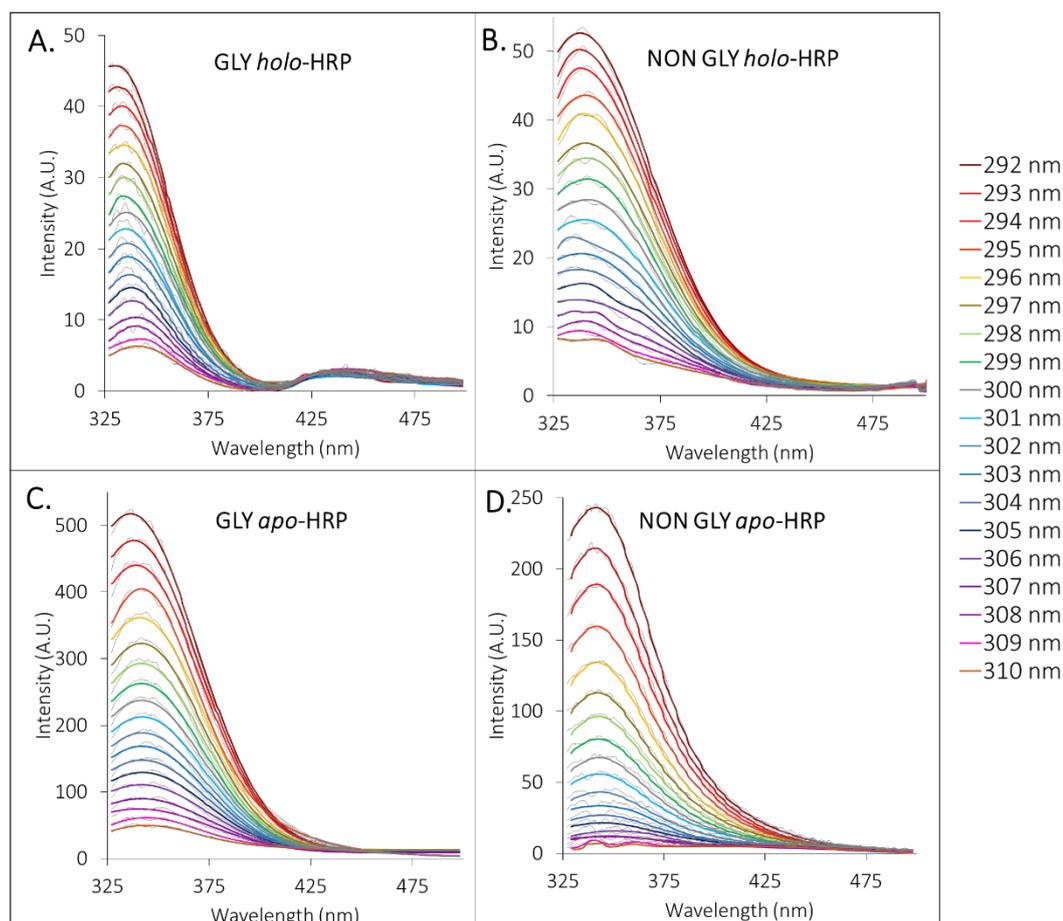
**Figure 5.5. Absorbance Spectra of Protein used in Investigation.**

**A**, GLY *holo*-HRP, **B**, NON GLY *holo*-HRP, **C**, GLY *apo*-HRP, **D**, NON GLY *apo*-HRP. With the  $R_z$  value of each protein shown in insert.

### 5.2.2. Tryptophan 117 Emission

Tryptophan emission of each protein variant was recorded at every 1 nm excitation ranging from 292 nm to 310 nm (Figure 5.6) as described previously (see Chapter 2.5.8.). Spectra were recorded 12 times for each variant, before averaging and subtracting of recorded spectra of Tris 50 mM pH 8 buffer. Data was typical of tryptophan fluorescence, except for one additional emission peak at  $\sim 440$  nm, in each *holo*-HRP sample, attributed to the presence of empty protoporphyrin IX, minus its central Fe core (Figure 5.6, A and B).

To account for the presence of two emission peaks spectra were fitted to the sum of two skewed Gaussians by way of equation 2.5 in Origin V 2020. The fitting of data to the sum of two skewed Gaussians allowed for the spectral component of tryptophan emission alone to



**Figure 5.6. Emission Spectra of 10  $\mu$ M HRP PTM Variants at Excitation 292-310 nm.**

Emission spectra of tryptophan 117 in GLY *holo*-HRP (A), NON GLY *holo*-HRP (B), GLY *apo*-HRP (C), NON GLY *apo*-HRP (D). Key (Right) shows the excitation wavelength of each spectrum. Raw data shown in grey with fitted plot shown overlain coloured by wavelength of excitation. Solid lines represent raw data fitted with Equation 2.5. Each graph's axis range is unique and chosen to best reflect the range in the individual variants and as such is not consistent between A-D.

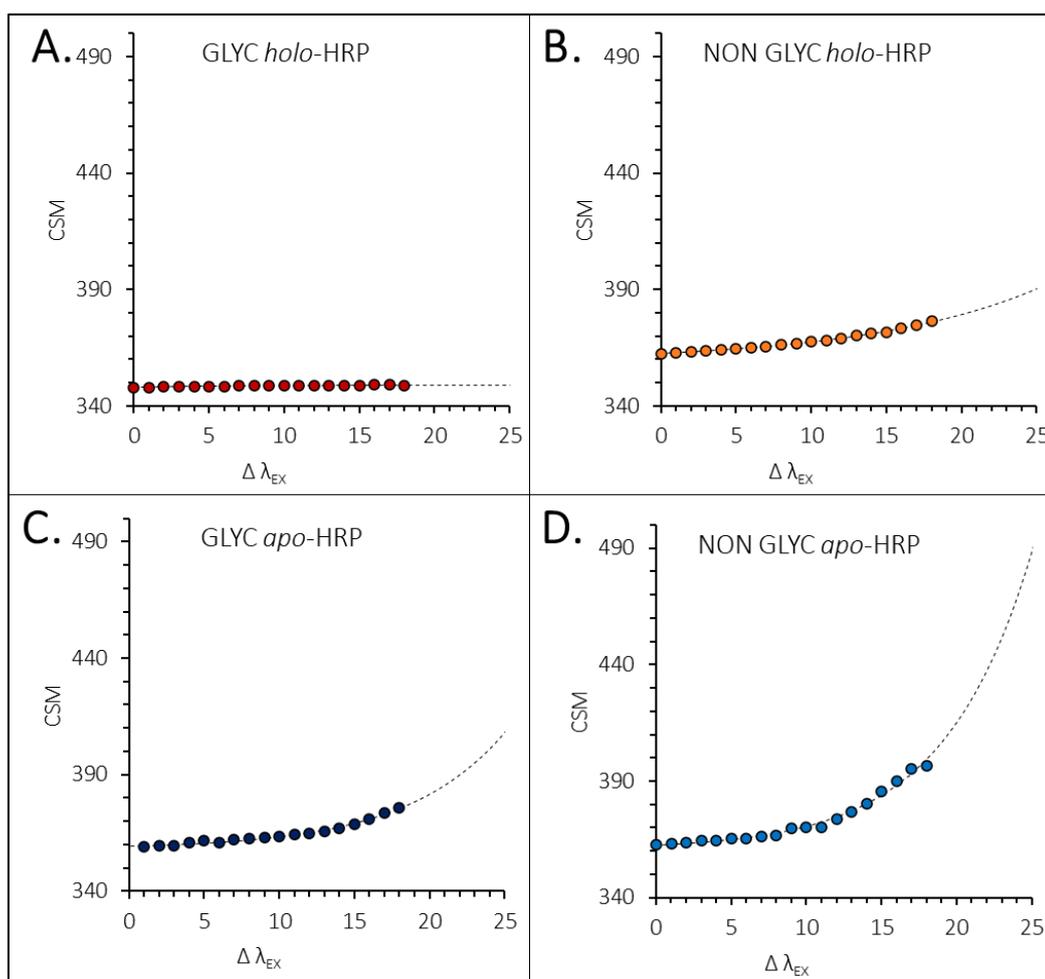
be extracted and prevented the secondary emission peak being included into the calculated CSM of tryptophan emission (Figure 5.6).

Fitted spectra were characteristic of tryptophan emission with an emission peak at  $\sim 335$  nm when excited at 292 nm. The largest detected intensity of tryptophan emission was emitted by glycosylated *apo*-HRP (Figure 5.6, C) with an emission intensity over 10 x larger than that observed in glycosylated *holo*-HRP (Figure 5.6, A). It was reasoned that this difference in emission intensity between *holo* and *apo* samples was likely due to fluorescence quenching by haem.

Protoporphyrin IX is a known agent of fluorescence quenching only when Fe is present and the porphyrin is fluorescent without the iron core. When iron is present the tetrapyrrole dissipates absorbed light energy across the conjugated double bond network in the vibrational relaxation of excitation energy without emission<sup>393</sup>. Tryptophan emission of haem binding proteins is known to be significantly reduced in the presence of bound haem<sup>394</sup>. Lower overall emission intensity was observed for both non-glycosylated *holo*-HRP and glycosylated HRP, indicating that quenching is happening regardless of HRP's glycosylation state.

### 5.2.3. REES Measurement

The CSM of each spectrum was calculated from each fitted Gaussian plot to recorded spectrum by way of Equation 2.6 (see Chapter 2.5.8) to quantify the change in wavelength from each 1 nm shift in excitation. Resulting CSMs were then plotted against the change in



**Figure 5.7. Red edge excitation shift of the Centre of Spectral Mass of Each HRP PTM Variant** Change in emission (CSM) of each PTM variant of HRP against change in excitation. GLY *holo*-HRP (A), NON GLY *holo*-HRP (B), GLY *apo*-HRP (C), NON GLY *apo*-HRP (D). Data fitted to Equation 2.7. to generate REES curve.

---

excitation wavelength and curve fitted in Origin V 2020 by way of Equation 2.7, to measure the change in emission associated with each change in excitation (Figure 5.7). The curvature of each fitted function is the first indication of REES as observed in Figure 5.7 B, C, and D. The larger the increase observed in the CSM at each 1 nm increase in excitation wavelength, the greater the red-shift observed in each spectrum. The straight line, observed in Figure 5.7, A, indicates that no relationship between the increase of excitation wavelength and CSM, which infers that no REES effect was observed in glycosylated *holo*-HRP.

Curvature observed from the change in CSM in respect to change in the excitation of non-glycosylated *holo*-HRP (Figure 5.7, B), glycosylated *apo*-HRP (Figure 5.7, C), and non-glycosylated *apo*-HRP (Figure 5.7, D) indicate the partial solvation of W117. Additionally, the initial CSM of each variant is consistent at 362 nm, indicates that each W117 fluorophore within these proteins is similarly solvated, suggesting homogeneity in protein structure.

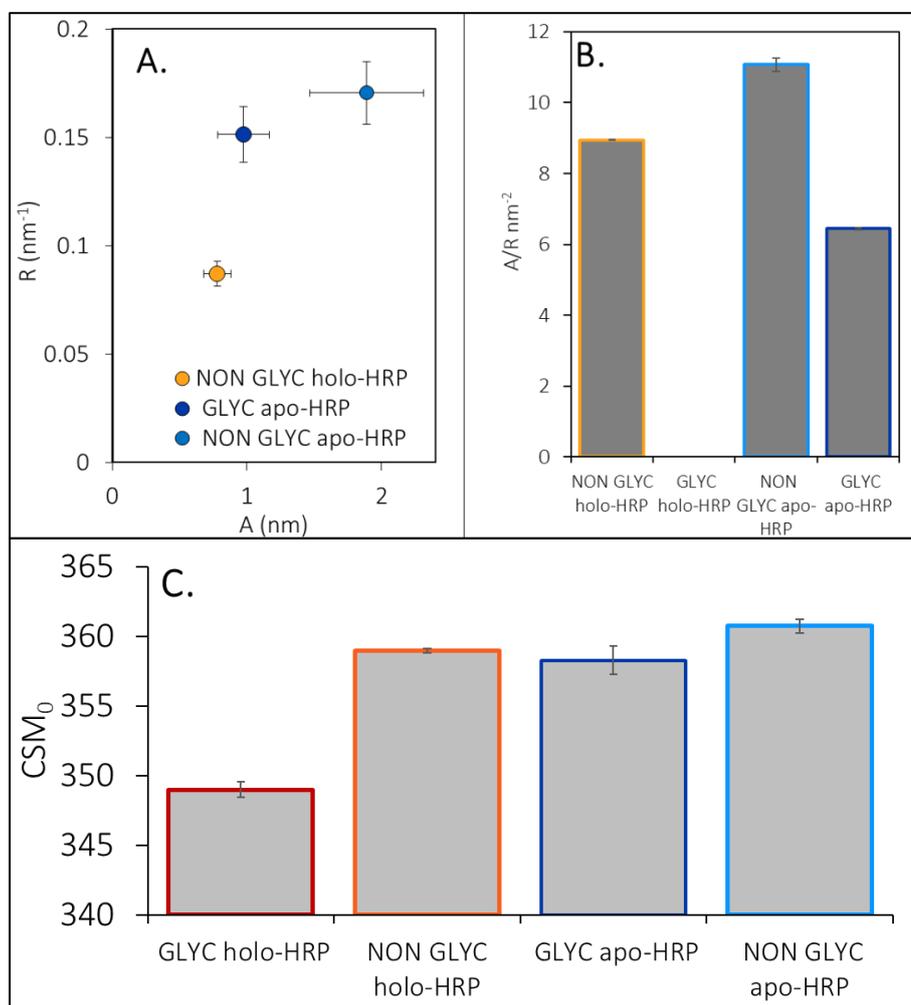
A clear distinction can be observed between the change in CSM of glycosylated and non-glycosylated proteins and additionally between both *holo* and *apo*-protein. The more extreme change in CSM observed in non-glycosylated protein infers that the environment of W117, in these variants is more variable. In turn this indicates protein structure is more dynamic, or less rigid, than in glycosylated protein. Whilst the increase in curvature observed as excitation is increased in *apo*-HRP in comparison to *holo*-HRP implies that the *apo* variant has a greater variation in W117 solvation states than *holo*-protein, implying a less rigidly defined secondary structure <sup>391</sup>.

The presence of a REES curvature displayed in B, C, and D, suggests that either localised or global HRP flexibility is influenced by PTM, as solvation is shown to change dependent on excitation energy applied. The upward curvature of spectra indicates that the emission CSM is red shifted and that the tryptophan fluorophore is partially solvated in these conditions. For each variant, the CSM<sub>0</sub> calculated is between 359 -361 nm (Figure 5.7, B-D), however, glycosylated *holo*-HRP is an exception to this with a CSM<sub>0</sub> of 349 nm (Figure 5.7, A). The observed consistency in CSM<sub>0</sub> indicates that the overall structure of each PTM variant except glycosylated *holo*-HRP is consistent, and it is only the flexibility of the protein which varies.

To quantify the REES effect observed in each variant, equation 2.7 was used wherein the amplitude ( $A$ ) of each change in CSM was divided by its curvature ( $R$ )(Figure 5.8). As

glycosylated *holo*-HRP exhibited no red-shift in emission, a curvature ( $R$ ) value of 0 nm was established, and thus no REES effect could be established for this variant, and CSM was consistent at approximately 349 nm throughout (Figure 5.7, A).

The observed REES effect of non-glycosylated *apo*-HRP was reflective of the most significant curvature of the fitted plot of CSM, which produced an amplitude of 1.88 nm and curvature of  $0.17 \text{ nm}^{-1}$  (Figure 5.8, A). Dividing  $A$  by  $R$  yielded a REES effect of 11 (Figure 5.8, B) which was reflective of the most significant shift in CSM per 1 nm excitation increase, observed from all PTM variants. The large shift in CSM observed indicated that W117 featured the greatest variation in solvation in non-glycosylated *apo*-HRP, and this was likely indicative of a more flexible protein structure.



**Figure 5.8. The Quantified Red Shift of W117 CSM of each HRP PTM Variant**

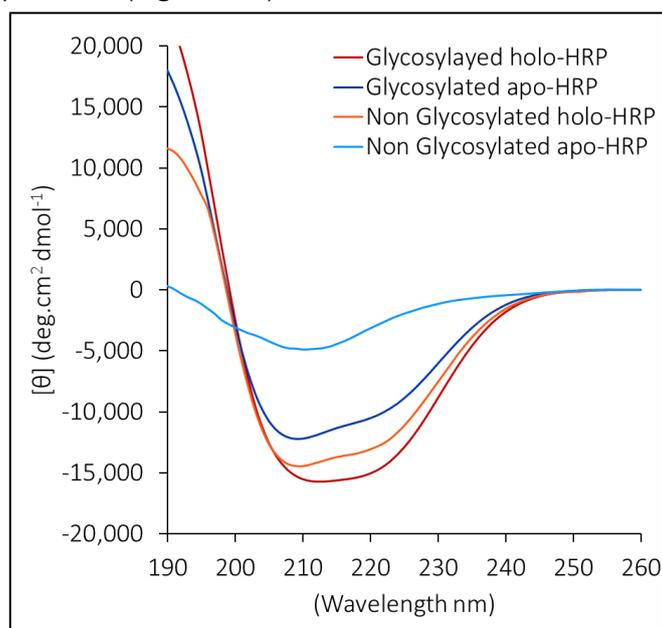
**A**, the plotted amplitude ( $A$ ) and curvature ( $R$ ) of each PTM variant's CSM. **B**, the REES effect ( $A/R$ ) of each variant. **C**, the CSM of each variant independent of excitation wavelength ( $\text{CSM}_0$ )

Comparison of both glycosylated *apo*-HRP and non-glycosylated *holo*-HRP with that of non-glycosylated *apo*-HRP indicated that both glycosylation and haem binding increased structural rigidity of HRP (Figure 5.8). This is argued as a decrease in REES effect observed from each form of PTM applied indicated the decrease in solvent exposure of W117, which is further indicative of reduced structural flexibility in each condition. The greatest decrease in flexibility was observed in glycosylated *apo*-HRP and is suggestive that glycosylation rigidifies protein to a greater extent than haem binding.

The decrease in flexibility inferred by both glycosylation and haem binding is indicative that glycosylated *holo*-HRP is highly rigidified and lacks flexibility which would induce a change in  $CSM_0$ . As W117 is not on the HRP surface or within the active site (Figure 5.3), the changes associated with the REES effect are not directly influenced by each form of PTM applied. Instead, the change in REES observed is indicative of a global change to protein stability in the presence of PTM, rather than a localised change of each coordinating ligand or glycosylation site.

#### 5.2.4. Circular Dichroism.

Analysis of the REES of PTM variants indicated that both glycosylation and haem binding reduced the flexibility of HRP (Figure 5.8). To assess if this increased rigidity influenced the



**Figure 5.9 .Far-UV Circular Dichroism Spectra of each PTM Variant at 293 K in Tris 5 mM pH 8.**

All spectra apart from non-glycosylated *apo*-HRP are characteristic of alpha helical secondary structure with minima at ~208 nm and ~ 222 nm and a maxima at ~192 nm.

---

secondary structure of HRP measurement of right-angled light scattering in the far UV spectral region was recorded for each variant at 293 K (Figure 5.9).

Far UV CD spectra indicated that three of the four PTM variants exhibited spectral minima at both 208 nm and 222 nm, and a maximum at 195 nm (Figure 5.9), characteristic of an  $\alpha$  helical arrangement of protein<sup>395</sup>. Glycosylated *holo*-HRP featured the spectrum with most negative molar ellipticity, with a minimum at 222 nm of  $-15,000 \text{ deg.cm}^2 \text{ dmol}^{-1}$ , while non-glycosylated *apo*-HRP, featured the least defined structural definition. The distinct secondary structure observed in glycosylated *holo*-HRP (Figure 5.9) is suggestive that the lack of a detected REES effect (Figure 5.7) was not due to the unfolding of the protein. Instead, the lack of detectable REES effect could possibly reflect a highly rigid protein structure.

The spectrum of non-glycosylated *apo*-HRP appeared to be mostly disordered as no clear hallmarks of secondary structure could be identified, such as the previously remarked minima at 208 nm and 222 nm (Figure 5.9, pale blue). The CD spectrum featured a molar ellipticity minimum of  $-2848 \text{ deg.cm}^2 \text{ dmol}^{-1}$  at 214 nm and which was not representative of an  $\alpha$ -helical structure. Instead, the observed spectrum featured characteristics of both  $\alpha$ -helical and  $\beta$  sheet structural arrangement, due to the presence of a spectral minimum between 212 nm and 215 nm which is characteristic of a  $\beta$  sheet secondary structure<sup>396</sup>, in combination with minimum at 208 nm. However, as no accompanying molar ellipticity maximum was observed at 198 nm, this PTM form is likely highly disordered protein population with no defined secondary structure.

All variants of HRP subjected to either glycosylation or haem coordination display the same basic characteristics which are indicative of an  $\alpha$ -helical secondary structure. However, the glycosylated *holo*-HRP population retains this folded spectral signature, yielding a CD spectrum more characteristic of an  $\alpha$  helical protein population. CD spectra indicate that the removal of one or both form of PTM resulted in protein population, which is less characteristically  $\alpha$ -helical. Consequently, this reflects the trend observed in REES spectra, whereby removal of each form of PTM, separately or in tandem, was associated with an increased curvature for the fitted plot of CSM (Figure 5.7). This curvature was reflective of increased protein flexibility, which is also represented in the decrease in structural definition observed in the CD spectra upon the removal of PTM (Figure 5.9).

---

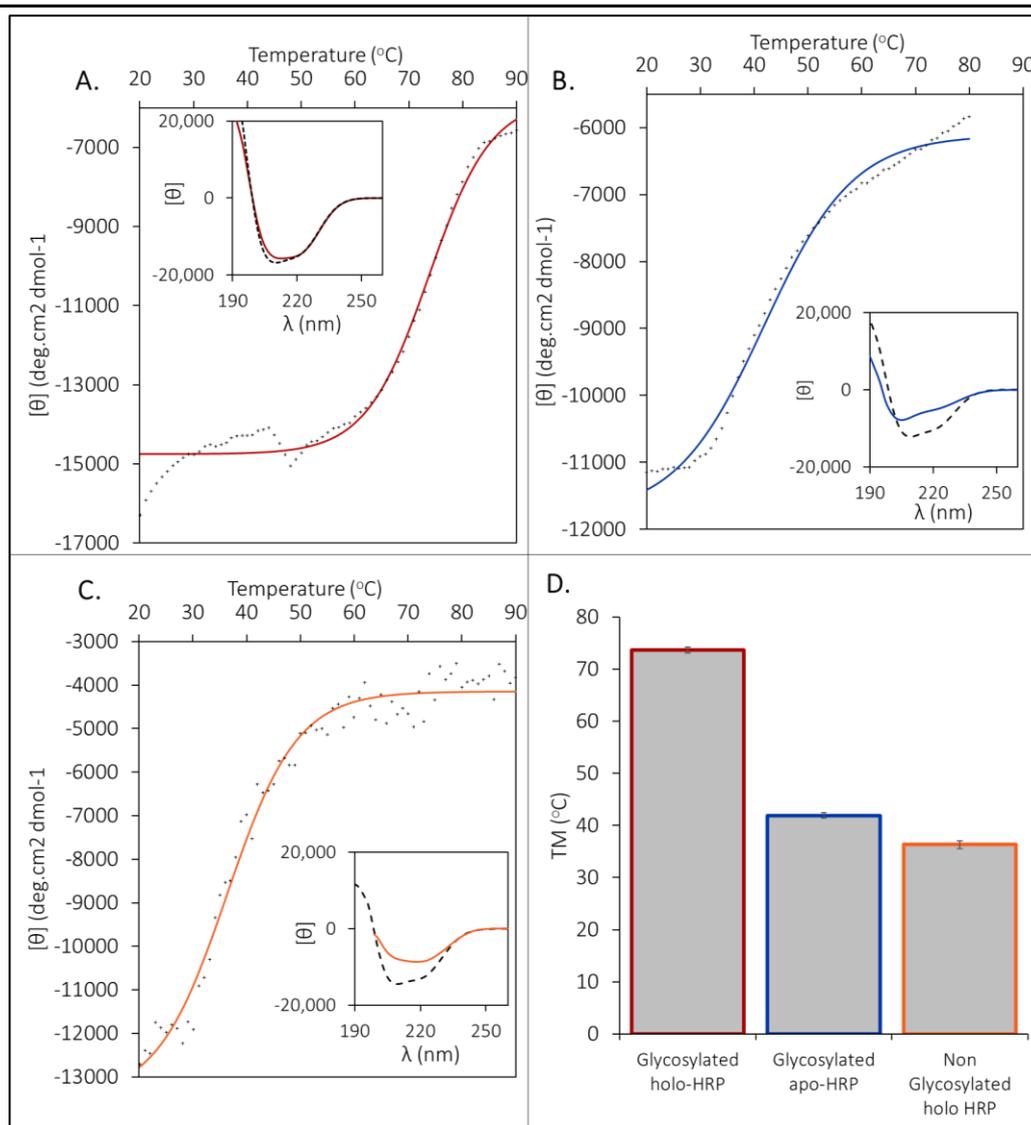
Results indicate that each form of PTM applied reduced flexibility to form a more rigid protein, and the presence of just one form of PTM can be sufficient to retain secondary structure (Figure 5.9). In contrast, the combination of PTMs exerts the greatest effect on protein stability. The absence of both PTMS results in a highly dynamic protein structure which is in constant flux and possesses no defined secondary structure. Thus, non-glycosylated *apo*-HRP was excluded from the analysis of thermal denaturation.

### 5.2.5. Thermal Induced Unfolding of HRP

As each form of “folded HRP” observed in Figure 5.9 featured a minimum at 222 nm, this wavelength was selected to monitor the midpoint of thermal denaturation ( $T_m$ ). Subsequently, each protein was subjected to a temperature ramp of 292-363 K at a ramp rate of 1 K/min whilst molar ellipticity at 222 nm was recorded. Resultant data were fitted to a Boltzmann distribution function in Origin to identify if a two-state transition (folded to unfolded protein) had occurred and identify the temperature at which transition occurred and protein became more unfolded than folded ( $T_m$ )(Figure 5.10).

A two-step transition of molar ellipticity at 222 nm was observed for each HRP PTM variant (Figure 5.10). Additionally, the dynode voltage recorded for each transition was observed to be with a range of 7.89 V to 8.16 V, indicating that all detector sensitivity did not need to be increased significantly to obtain signal (Table 5.1). This transition occurred at the highest temperature in glycosylated *holo*-HRP with a calculated transition,  $T_m$ , of  $347 \pm 0.57$  K (73 °C) (Figure 5.10, A). The transition from non-glycosylated *holo*-HRP was lower than that of glycosylated *holo*-HRP with a calculated difference 40 K lower at  $309 \text{ K} \pm 0.72$  (36 °C). The difference in  $T_m$  indicated that glycosylation dramatically increased thermal tolerance of HRP.

Comparison of glycosylated *apo*-HRP with that of glycosylated *holo*-HRP indicated that haem coordination, like glycosylation, also increased the thermal tolerance of HRP. A calculated  $T_m$  of  $315 \text{ K} \pm 0.54$  (42 °C) was observed for glycosylated *apo*-HRP, ~ 30 K lower than that of glycosylated *holo*-HRP. The decrease in  $T_m$  observed, indicated that haem binding, in addition to glycosylation increased the thermotolerance of HRP. While the observed  $T_m$  of both glycosylated *apo*-HRP and non-glycosylated *holo*-HRP differed by 6 K indicating the similarity in thermotolerance of HRP with just one form of PTM.



**Figure 5.10. Thermal Melt of HRP PRM Variants at a Temperature Range of 293 -363 K.**

Change in mean residue ellipticity at 222 nm of glycosylated *holo*-HRP (A), glycosylated *apo*-HRP (B), and non-glycosylated *holo*-HRP (C). The solid line represents fitted Boltzmann distribution function applied to date to measure the two-state transition from folded to unfolded protein. Inserts display the structure of each variant at 293 K (Dashed line) and after heating to 363 K (Coloured Line). D displays the  $T_m$  (point of transition between folded and unfolded protein) of each fitted Boltzmann distribution function.

It is hypothesised that the initially recorded increase in molar ellipticity of glycosylated *holo*-HRP at 318 K indicated the transition of protein to an intermediate secondary structure. Although the recorded spectrum of this variant, when held at 363 K (90 °C), indicated predominantly  $\alpha$ -helical and consistent secondary structure with that observed at 293 K (20 °C) as both minima at 208 nm and 222 nm were retained (Figure 5.10, A, Insert). The retained structural definition was suggestive that the population of glycosylated *holo*-HRP had not fully transitioned into unfolded protein, even at a temperature close to boiling.

Table 5.1. Dynode Voltage of Recorded 222 nm Molar Ellipticity for each HRP Variant.

	Glycosylated <i>holo</i> -HRP	Non-Glycosylated <i>holo</i> -HRP	Glycosylated <i>apo</i> -HRP
Minimum Dynode Voltage (V)	8.16	8.16	8.11
Maximum Dynode Voltage (V)	7.89	7.89	7.90
Average Dynode Voltage (V)	7.99 ± 0.05	7.99 ± 0.05	8.00 ± 0.05

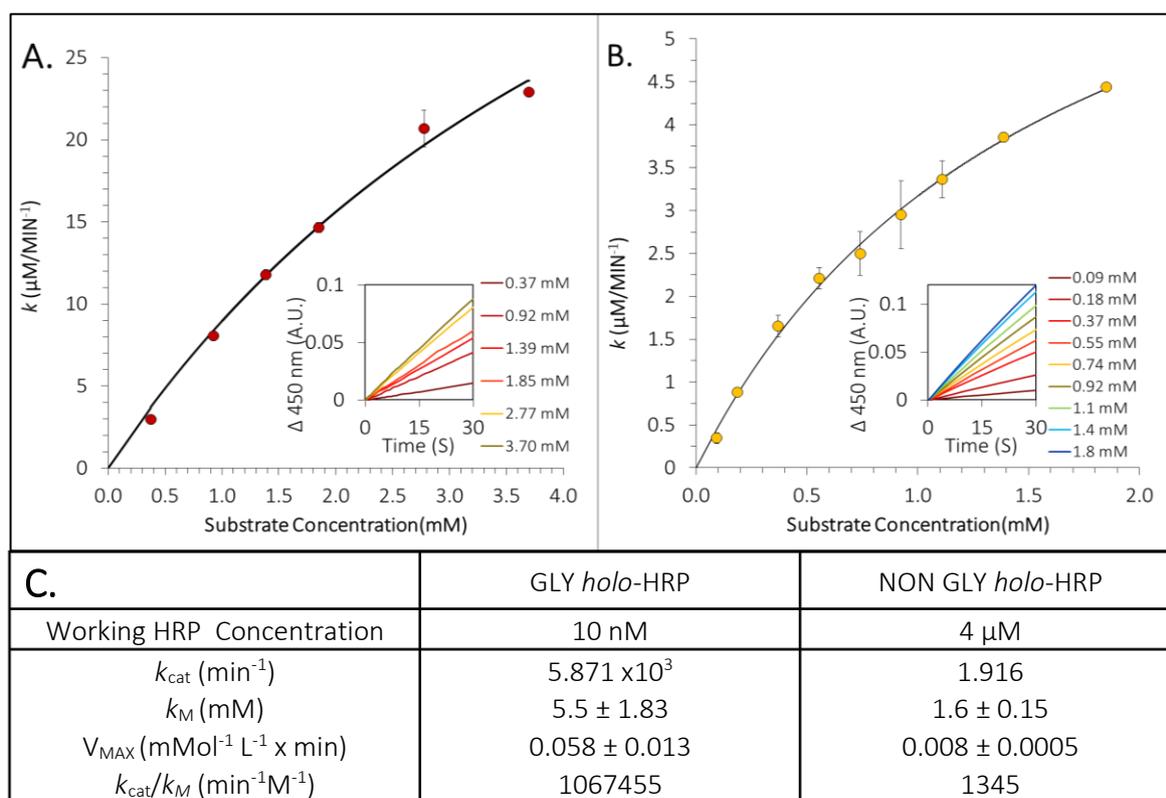
In contrast definition of the minima at 208 nm and 222 nm recorded for non-glycosylated *holo*-HRP and glycosylated *apo*-HRP at 363 K were less negative than observed previously at 292 K (Figure 5.10, B and C). This loss in mean residue molar ellipticity at both 208 nm and 222 nm indicates a loss of secondary structure across the protein population when heated to 363 K (90 °C). Furthermore, as the observed  $T_m$  of both glycosylated *apo*-HRP and non-glycosylated *holo*-HRP differed by 6 K, this indicated similarity in thermotolerance of protein when applied with each singular form of PTM.

Evidence of thermotolerance suggests that the presence of two forms of PTM produces a population of HRP which unfolds at a higher temperature than protein which has a singular form of PTM. This observation is consistent with the trend observed by both REES measurement and CD spectra (Figures 5.8 and 5.9). Each PTM individually provides a benefit to the thermotolerance of HRP, with glycosylation resulting in a higher  $T_m$  than that observed from haem coordination. Whereas the removal of both PTMs results in a protein structure which is not clearly defined and is possibly unfolded at 293 K (20 °C) (Figure 5.9).

#### 5.2.6. Steady-State Kinetics of *holo*-HRP

Investigation of *holo*-HRP catalysis was measured in the presence and absence of glycosylation. A range of OPD substrate concentrations from 0.9 mM to 4 mM was mixed with Tris 50 mM pH 8, and peroxide (working concentration of 3% (v/v)) before the addition of enzyme. Samples were mixed by pipette, held for 10 seconds and absorbance at 450 nm was recorded over 30 seconds at 20°C for 5 replicate readings for each concentration of substrate. Measurement of absorbance change at 450 nm was converted to substrate turnover per minute in Excel (Figure 5.11, A and B). Data were fitted to a Michaelis Menten curve in Origin V 2020 to establish both  $K_m$ , and  $k_{cat}$  (Figure 5.11, C).

Analysed data indicated that the rate at which glycosylated *holo*-HRP formed product was  $\sim 3,000$  x greater than non-glycosylated *holo*-HRP, with an observed  $k_{\text{cat}}$  of  $5.871 \times 10^3 \text{ min}^{-1}$  and  $1.196 \text{ min}^{-1}$  respectively. Samples of glycosylated HRP featured a slower rate of substrate saturation ( $V_{\text{MAX}}$ ) of  $0.058 \pm 0.013 \text{ mMolmin}^{-1}$  in comparison to the  $0.008 \pm 0.0005 \text{ mMolmin}^{-1}$  of non-glycosylated protein, despite the use of 400 x more non-glycosylated protein in measurement. This indicated a reduction of product formation by almost 4 orders of magnitude in the absence of glycosylation which far exceeds the change in haem saturation reported for each glycan form (Figure 5.5). Therefore, catalytic efficiency is different in each sample investigated which indicates that glycosylated-HRP has far superior catalytic efficiency to that of non-glycosylated-HRP.



**Figure 5.11. Michaelis Menten Plot of *holo*-HRP**

Substrate turnover of glycosylated *holo*-HRP (A) and non-glycosylated *holo*-HRP (B) at various substrate concentrations. Data fitted to Michaelis Menten equation in Origin V 2020 to obtain values for  $V_{\text{MAX}}$ ,  $k_M$  and  $k_{\text{cat}}$  (C). Inserts (A and B) display raw data for change in absorbance per second for each substrate concentration. Error bars display standard error of mean.

This is surprising as glycans themselves do not possess catalytic function. Instead, it is reasoned that the reduction in global protein structural flexibility in the presence of glycosylation, aids catalysis instead (Figure 5.8). The reported increase in rigidity of

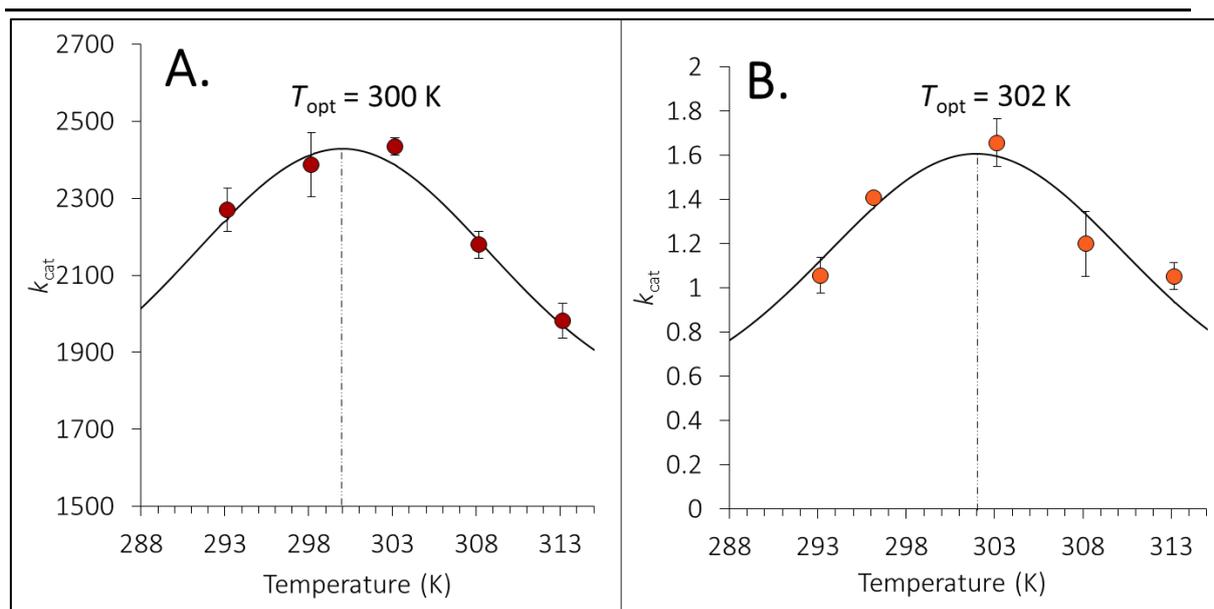
---

glycosylation observed in Figure 5.7 is reasoned to be a contributing factor to both the increase in  $T_m$  (Figure 5.10) and catalytic efficiency (Figure 5.11) observed in glycosylated HRP. It is hypothesised that the calculated increase to both  $k_{cat}$  and  $k_M$  is indicative of both greater uptake of  $H_2O_2$  from a more rigid active site alignment of the glycosylated HRP structure.

A similar observation was reported by Humer and Spadiut (2019), whereby the  $k_{cat}$  of glycosylated-HRP was almost 180 x greater than that of non-glycosylated HRP<sup>54</sup>. Additionally, this study found that the  $k_{cat}$  of non-glycosylated HRP could be increased with the replacement of surface non-glycosylated residues with aspartic acid. A quadruple non-glycosylated variant (N13D/N57S/N255D/N268D) produced in this study was reported to have a  $k_{cat}$  10x greater than benchmark non-glycosylated protein, although,  $k_{cat}$  was still 18 x less than that of glycosylated protein. As non-glycosylated HRP used in this experiment is removed of all eight surface asparagine residues in favour of alanine, aspartic acid may offer a catalytic advantage to HRP greater than that of alanine. However, all forms of non-glycosylated HRP have reduced catalytic efficiency in comparison to glycosylated-HRP. It is reasoned that regardless of the small differences observed between each form of non-glycosylated protein, HRP has a superior ability to transition through the catalytic cycle when glycosylated.

### 5.2.7. Temperature Dependence of Activity

As the measurement of substrate turnover was recorded at 293 K (Figure 5.11), it is unlikely, but possible, that the kinetic rate of turnover from non-glycosylated *holo*-HRP was limited by protein unfolding at 309 K (Figure 5.10). To confirm the rate of catalysis was not influenced by the temperature at which catalysis occurred the rate of product formation ( $k_{cat}$ ) was recorded over a range temperature (293-313 K) for the calculation of the optimum temperature of catalysis for each PTM form ( $T_{opt}$ ). To effectively record the influence of temperature on catalysis, independent any potential increase in  $k_M$ , the substrate concentration used in the investigation was 10 x that of the recorded  $k_M$  observed at 293 K (Figure 5.11, C). As  $k_M$  is not temperature dependent, any curve observed to  $T_{opt}$  plot is therefore not reflective of a change in  $k_M$  as the temperature is increased. Reactants were incubated for 5 minutes at the required temperature before mixing and measurement of absorbance change at 450 nm. Change in absorbance was then converted to the rate of substrate turnover and fitted to equation (Equation 2.8.) in Origin V 2020 to calculate the optimum temperature of turnover ( $T_{opt}$ ) (Figure 5.12).



**Figure 5.12. The Effect of Glycosylation on the Temperature Dependence on  $k_{cat}$  of *holo*-HRP**

The change in  $k_{cat}$  of both glycosylated (A) and non-glycosylated (B) *holo*-HRP fitted to equation of MMRT in Origin V 2020 to establish optimum temperature of turnover ( $T_{opt}$ ).

Error bars display standard error of mean.

The fitted temperature dependence curves showed a similarity in  $T_{opt}$  of both glycosylated and non-glycosylated *holo*-HRP, despite a rate of catalysis several orders of magnitude greater in glycosylated protein (Figure 5.12, A and B). Glycosylated protein was calculated to have  $T_{opt}$  of catalysis at 300 K, 2 K lower than non-glycosylated HRP. The observed  $T_{opt}$  of each protein was comparable between glycosylation states despite a difference of 40 K in each previously calculated  $T_m$  (Figure 5.10).

Furthermore, the calculated  $T_{opt}$  of glycosylated *holo*-HRP of 300 K was consistent, with previously reported  $T_{opt}$  of 303 K by both Wang *et al.* (2016) and Mohamed *et al.* (2014)<sup>397,398</sup>. It is possible that the difference in  $T_{opt}$  between variants of 2 K is within the margin of error of recorded data and that glycosylation has no effect on  $T_{opt}$  of catalysis. The  $T_{opt}$  of each protein is lower than that of each recorded  $T_m$ , indicating that turnover was reduced in each protein before unfolding occurred. Typically, the  $T_{opt}$  of catalysis for an enzyme is inherently limited by  $T_m$ , however as the  $T_m$  of glycosylated *holo*-HRP is 40 K greater than reported  $T_{opt}$ , this does not appear to be the case for HRP. The difference in  $T_m$  and  $T_{opt}$  signifies that the high stability supplied to HRP in the presence of glycosylation is not reflected in the thermotolerance of catalysis.

---

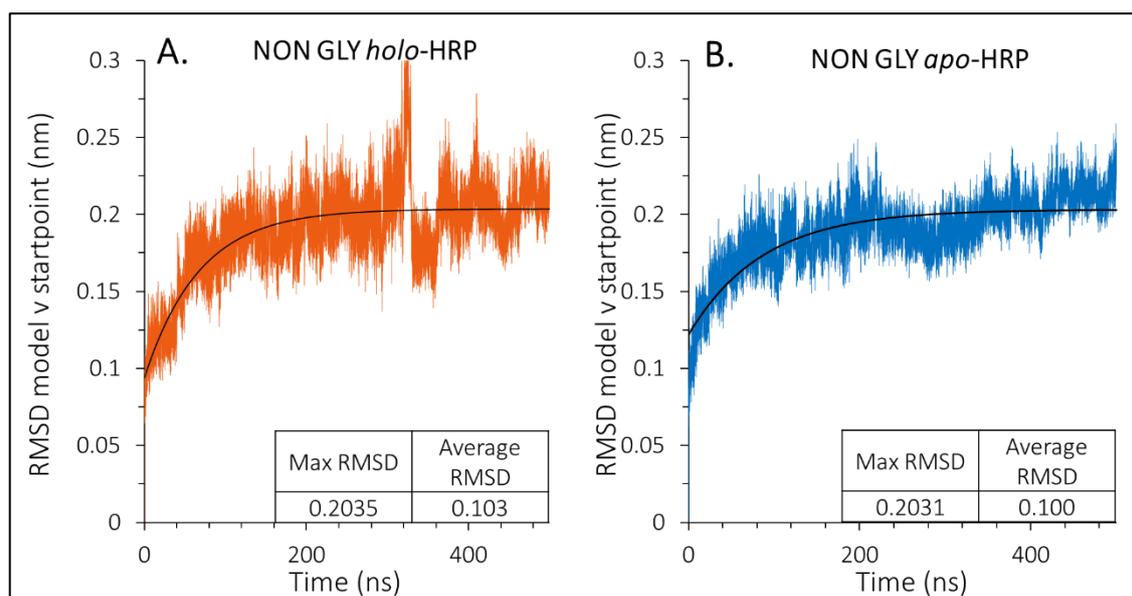
As the  $T_m$  of glycosylated *holo*-HRP is 347 K (Figure 5.10), it is unlikely that the decrease in catalytic efficiency above 300 K is reflective of protein unfolding. It is possible that if a structural transition occurs between 300-320 K as hypothesised, and this results in rearrangement within the active site, this will reduce the catalytic efficiency of the enzyme. This reduction in efficiency could possibly occur by blocking of H42 from binding peroxide or blocking of substrate binding in catalysis by a more flexible active site<sup>88,295,296</sup>. This change would be observed in both glycosylated and non-glycosylated protein equally, regardless of the difference in  $T_m$ . It is hypothesised that as HRP evolved for function in the plant *A. rusticana*, it was not necessary to evolve a peroxidase  $T_{opt}$  greater than the ambient temperature in which the plant could sustain life. As such, the catalytic efficiency of the enzyme is not as thermophilic as the observed  $T_m$  of glycosylated *holo*-HRP would suggest. The likely reason for this discrepancy is that in nature the enzyme does not need to function at high temperatures as the plant would not grow in conditions  $>50$  °C.

The  $k_{cat}$  of both glycosylated and non-glycosylated HRP initially increased between 293 K and 302 K (Figure 5.12, B), which indicated that the ambient temperature of the investigation did not reduce steady-state kinetics recorded at 292 K. Furthermore, the comparable  $T_{opt}$  between each variant indicates that glycosylation has minimal influence on the thermal tolerance of catalysis. Therefore, the presence of glycans on HRP does not increase the thermotolerance of catalysis, and the reaction is limited by temperatures, regardless of glycan presence.

### 5.2.8. Molecular Dynamics

In Figure 5.10, the  $T_m$  of glycosylated *holo*-HRP was calculated to be 30 K greater than that of glycosylated *apo*-HRP. To predict which residues of HRP were reduced in flexibility upon haem coordination, molecular dynamic (MD) simulation was performed. Unfortunately, the glycosylated HRP could not be analysed for comparison as no crystal structure of glycosylated HRP exists (the model feature in Figure 4.10. is only a representation of glycan identity). MD simulation was performed for structures with and without haem presence (see Chapter 2.7.2), for a total of 5 repeat simulations of 100 ns, wherein each successive simulation continued where the previous ended (Figure 5.13). A non-linear logarithmic curve was fitted to the average root mean square deviation (RMSD) of each model in Origin V 2020, to confirm the equilibration of each simulation.

The RMSD value obtained from each model indicated that the backbone of non-glycosylated *holo*-HRP was fractionally more flexible than non-glycosylated *apo*-HRP at 0.103 nm and 0.100 nm, respectively (Figure 5.13). This difference observed between the two protein forms was minor and was reasoned to not be a true reflection of the difference between *holo* and *apo*-HRP *in vivo*. Instead this analysis highlights the difficulty in predicting by simulation the difference between two protein saturation states when a PDB file for one version alone exists.



**Figure 5.13. RMSD of *holo* and *apo* non-Glycosylated -HRP.**

The RMSD of the model vs start point of non-glycosylated *holo*-HRP (A) and non-glycosylated *apo*-HRP (B) over a simulated total of 500 ns. Raw data generated by Dr Harley Worthy, Cardiff University School of BIOSI. Data is fitted with a non-linear log curve in Origin V 2020, to assess if RMSD had plateaued within the duration of simulation. Insert shows average and maximum RMSD of fitted curve.

The root mean square fluctuation (RMSF) of each residue was plotted to assert if individual residues were more flexible in *holo*-HRP than *apo*-HRP (Figure 5.14). The difference observed in the RMSF at each residue was then calculated to determine which were most flexible in the presence or absence of haem binding. No substantial difference was observed in the haem binding ligand H170 or those that were natively glycosylated (N13, N57, N158, N186, N198, N212, N255, and, N268) (Figure 5.14). The difference in RMSF was then assigned a colour in accordance to intensity (Chapter 2.7.2), and the colour was applied to each residue in 1H58 (Figure 5.14, B-E) to indicate regions of altered flexibility.

This colouration indicates that in general changes protein flexibility triggered by haem binding are not restricted to the active site of HRP (Figure 5.14, B and D). Instead, there are two regions of localised RMSF increase in *holo*-HRP which may indicate greater flexibility the opening to the active site of the *holo*-protein. Whilst the residues with the greatest flexibility in *apo*-HRP are mainly restricted to loop regions between helices, with one exception of a single helix on the posterior surface of the protein to the active site (Figure 5.14, E).

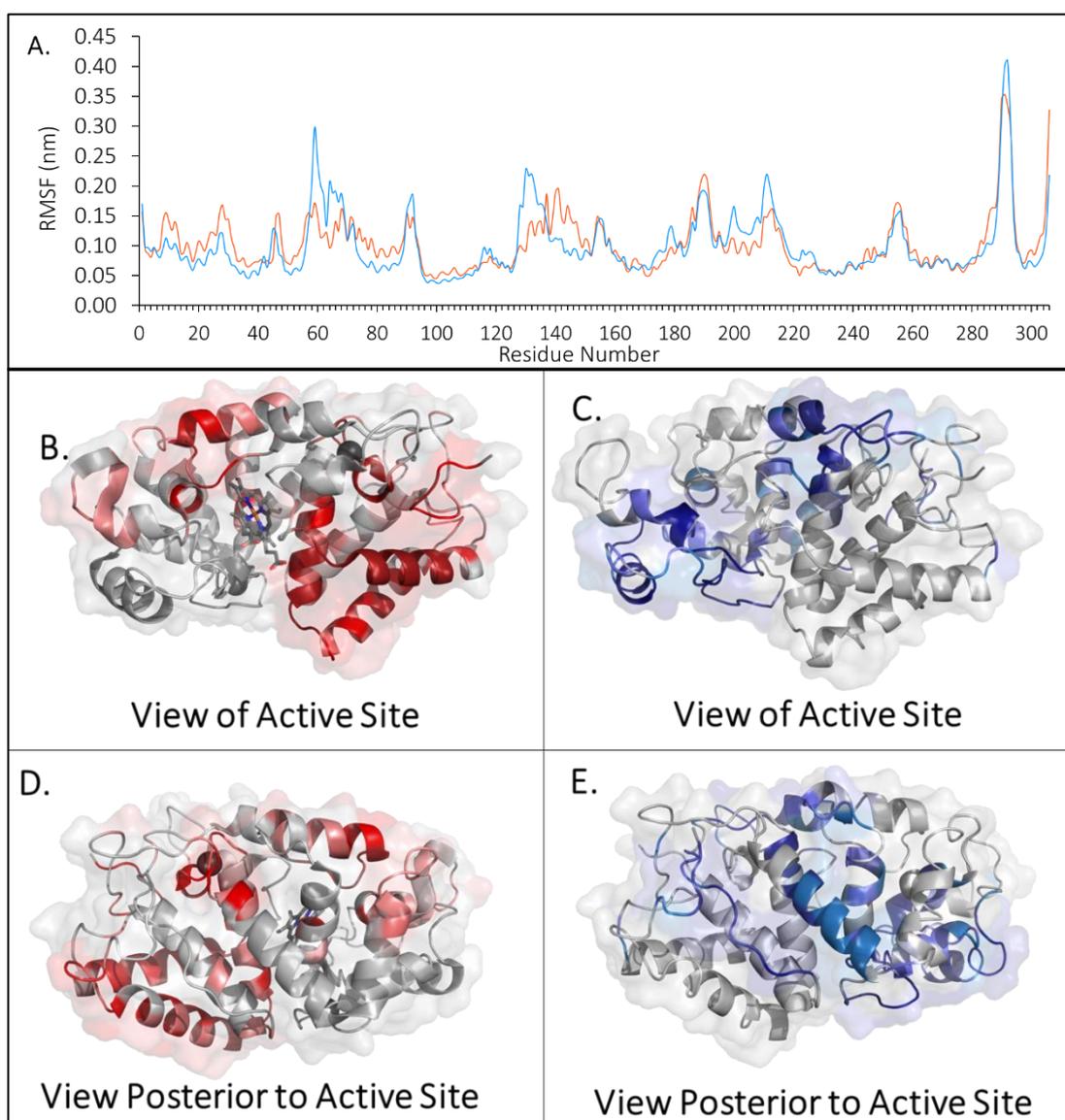


Figure 5.14. The Change in the RMSF of Residues of non-glycosylated HRP in the Presence and Absence of Haem Binding.

A, the RMSF of each residue of both non-glycosylated *apo*-HRP (Blue) and non-glycosylated *holo*-HRP (Orange). The RMSF of each used in B-E to colour residues in accordance with intensity of RMSF. Those with greatest flexibility in *holo*-HRP are shown as red (B and D) and those with greatest flexibility in *apo*-HRP shown as blue (C and E). A and C represent view of the protein's active site opening whilst D and E represent the protein rotated 180° so that the active site is pointed away.

## 5.3. Conclusion

### 5.3.1. The influence of PTM on HRP Flexibility

All measurement acquired in this investigation indicates that both glycosylation and haem coordination, reduce the flexibility of HRP (Table 5.2). Measurement of REES effect by W117 emission demonstrated that increased protein flexibility was observed upon the removal of each form of PTM. Significantly, HRP, which featured both haem coordination and glycosylation had no observed REES effect, while removal of both PTMS resulted in the largest REES effect (Figure 5.8). Removal of each form of PTM was shown to increase the solvent exposure of W117 with non-glycosylated *holo*-HRP, being marginally more flexible than glycosylated *apo*-HRP.

**Table 5.2. PTMs of HRP Ranked in Accordance to Measured Stability of each Data Set**

	REES Effect	CD Spectra	$T_m$	Enzyme Turnover *	$T_{opt}$
<b>Most Stable</b>	GLY <i>holo</i> -HRP	GLY <i>holo</i> -HRP	GLY <i>holo</i> -HRP	GLY <i>holo</i> -HRP	No Clear Difference
↑	GLY <i>apo</i> -HRP	GLY <i>apo</i> -HRP	GLY <i>apo</i> -HRP		
↓	NON-GLY <i>holo</i> -HRP	NON-GLY <i>holo</i> -HRP	NON-GLY <i>holo</i> -HRP	NON-GLY <i>holo</i> -HRP	
<b>Least Stable</b>	NON-GLY <i>apo</i> -HRP	NON-GLY <i>apo</i> -HRP	-		

\*Enzymes ranked by  $k_{cat}$  and not stability

CD spectra, meanwhile, demonstrated again that HRP which was not haem coordinated or glycosylated, featured the greatest variation in secondary structure (Figure 5.9). Furthermore, the correlation was observed between a lower REES effect of glycosylated *holo*-HRP (Figure 5.8) and an increased secondary structure definition (Figure 5.9), thermo-tolerance (Figure 5.10), and kinetic activity (Figure 5.11). The reverse trend was observed when protein was either non-glycosylated or non-haem bound. It is likely that the increased kinetic activity reported in the presence of glycosylation, was representative of reduced flexibility of HRP's active site, which is reflected in a decreased observed solvent exposure and localised flexibility of W117 in REES data (Figure 5.8).

The measured REES effect of each form of PTM was further proven to follow the same trend observed in each CD spectra,  $T_m$ , and  $k_{cat}$  (Table 5.2). This inferred structural rigidification

---

observed in REES correlated with substrate turnover ( $k_{cat}$ ), and thermal tolerance ( $T_m$ )(Figures 5.10 and 5.11). However, decreased structural flexibility was not reflected in the  $T_{opt}$  of catalysis, which was consistent regardless of glycosylation (Figure 5.12). The consistent  $T_{opt}$  indicates that increased rigidification of the active site increases catalysis in steady-state kinetics but does not increase the overall thermal tolerance of catalysis activity.

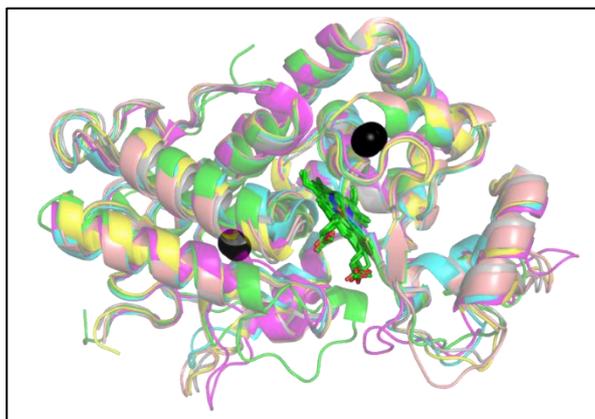
The absence of a REES effect of glycosylated *holo*-HRP was distinct to all other protein forms(Figure 5.8). It was observed that as the CSM of glycosylated *holo*-HRP did not increase with the increase in excitation wavelength which was indicative that the protein was folded. The combination of REES data with CD spectra further clarified that the folded protein was characteristically alpha-helical in structure (Figure 5.9). Furthermore, the  $T_m$  observed from glycosylated *holo*-HRP was the greatest of all PTMs of investigation.

While the interpretation of CD spectrum alone would indicate that the reverse effect was accurate of non-glycosylated *apo*-HRP, CD spectrum of this PTM variant indicated a protein with little apparent  $\alpha$ -helical content when compared with other PTM variants (Figure 5.9). The presence of a measurable REES effect indicated that whilst protein was highly flexible, the fluorophore was not completely solvent-exposed, and likely still within a form of folded protein structure. However, as this structure was not observed in the whole protein population, this organisation could not be characterised by the CD spectrum and is not sufficient for calculation of  $T_m$ , as it would not be apparent at which point transition to unfolded protein occurred. This ambiguity highlights the importance of the combination of REES data with the CD spectrum to ensure correct interpretation of protein organisation.

### 5.3.2. Influence of Glycosylation on Peroxidase Structure

HRP is characteristic of most class III plant peroxidases and is an extensively glycosylated protein. Structurally, class III peroxidases have a highly conserved  $\alpha$ -helical structure, two structurally stabilising calcium ions, one haem cofactor, and extensive glycosylation. Within the PDB database the recorded structures of alternative class III plant peroxidases, Soybean Peroxidase (1FHF)(38.4 % percentage similarity to HRP)<sup>399</sup>, Barley Grain Peroxidase (1BGP) (38.58 % percentage similarity to HRP)<sup>400</sup>, Switchgrass Peroxidase (5TWT) (51.31 % percentage similarity to HRP)<sup>401</sup>, Arabidopsis peroxidase (1QOP(89.38 % percentage similarity to HRP))<sup>402</sup>, and Royal Palm Peroxidase (3HDL) (43.89 % percentage similarity to HRP)

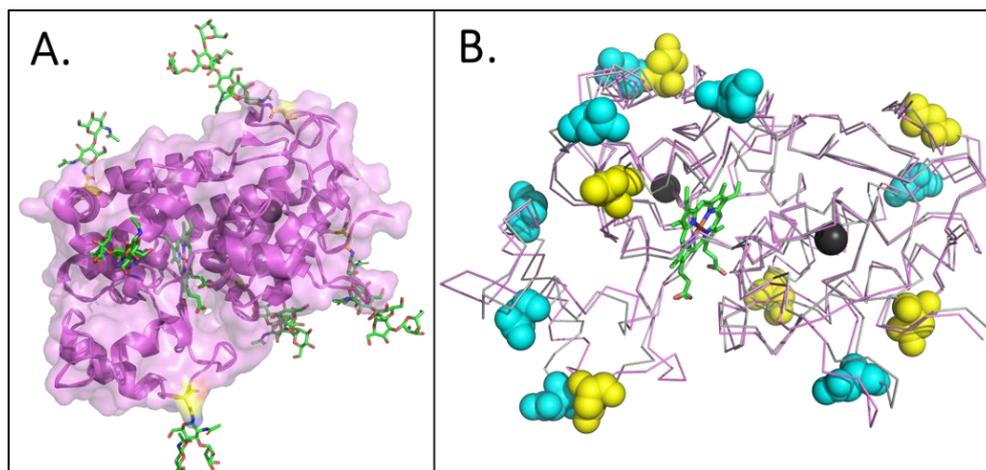
all share significant structural homology with that of HRP (1H58), and are also natively highly glycosylated (Figure 5.15). This is of note as the sequence identity of each alternative peroxidase is highly varied in comparison to HRP, with a percentage similarity ranging from 38.4 % for Soybean peroxidase to 89.38 % in Arabidopsis Peroxidase.



**Figure 5.15. Structural Homology Observed in Class III Plant Peroxidases**

Alignment of HRP (1H58) (**Grey**), Royal Palm Peroxidase (3HDL) (**Pink**), Soybean Peroxidase (1FHF)(**Peach**), Barley Grain Peroxidase (1BGP)(**Green**), Switchgrass Peroxidase (5TWT)(**Teal**), and Arabidopsis Peroxidase (1QOP)(**Yellow**). Coordinated haem shown as green sticks, and calcium  $^{2+}$  ions shown as black spheres.

Results obtained indicate that a consistent structural arrangement is observed between both glycosylated and non-glycosylated HRP(Figures 5.8, and 5.9). Comparison of the glycosylated structure of Royal Palm Peroxidase (Figure 5.16, A)and non-glycosylated HRP's structure indicates peroxidase homogeneity in structure regardless of glycosylation,(Figure



**Figure 5.16. Crystal Structure of Royal Palm Peroxidase**

**A**, crystal structure of RPP (3HDL) shown as cartoon (**Pink**) with attached glycans shown as sticks (**Green**). **B**, overlay of the ribbon structure of glycosylated RPP (Pink) and the crystal structure of non-glycosylated HRP (1H58) (**Grey**). With the glycosylation sites of each are shown as spheres for reference (**yellow** for RPP, **teal** for HRP)

---

5.16, B)<sup>82,403</sup>. The observed homogeneity of the structure reflects that while the functional effects of glycosylation are divergent; the structure of the protein is homologous. This homogeneity indicates that the glycosylation state does not induce variation in the structure of each peroxidase, as each has a consistent tertiary structure. This validates data obtained in this investigation which indicate that HRP is consistently structured regardless of PTM, but PTM does significantly reduce the flexibility which is observed within each structure.

### 5.3.3. Influence of Haem Binding on Peroxidase Structure

HRP coordinated with haem, featured both an decrease in REES intensity (Figure 5.8) and an increase in  $T_m$  (Figure 5.10). The reduction in REES magnitude of glycosylated protein indicated that a decrease in W117 flexibility occurred when HRP was coordinated with haem (Figure 5.8). This decrease in flexibility occurred regardless of glycosylation, and the CD spectrum indicated that haem coordination in the absence of glycosylation was sufficient to maintain HRP's secondary structure (Figure 5.9). Furthermore, the extent to which flexibility of HRP was reduced upon haem coordination was consistent with that of glycosylation, which suggests that haem coordination, has a comparable effect on protein stabilisation. It is hypothesised that haem binding aids HRP stability by decreasing active site flexibility when coordinated with H170. It is anticipated that haem coordination reduces the flexibility of peripheral aromatics F41, F152, F172, F179, and D221 (see Chapter 1, Figure 1.4) within the active site via the formation of pi-pi stacking interactions with haem porphyrin, which in turn stabilise and anchor the protein's secondary structure<sup>404</sup>.

CD spectra (Figure 5.9), demonstrated that both forms of *apo*-HRP were observed as less characteristically  $\alpha$ -helical in the absence of haem binding. This less defined structure of *apo*-HRP indicated that the population *apo*-HRP was more varied than that of *holo*-HRP. This confirmed the previous expectation that haem binding would trigger a modification to the secondary structure of HRP (Section 5.1.4). However, instead of triggering a change to protein confirmation, as predicted, haem binding instead reduced the flexibility of the existing secondary structure of the protein. This is the case in several other haem binding proteins such as haemoglobin, wherein the removal of the haem ligand increased protein flexibility and reduced structural definition, complicating efforts of structural determination<sup>404</sup>.

The observed similarity of both RMSD and RMSF observed from MD simulation (Figure 5.14) is suggestive that coordination of haem exerts minimal influence on both localised and global protein flexibility. However, results observed from the calculated REES effect (Figure 5.8) and  $T_m$  (Figure 5.10) indicate that haem binding exerts significant change to flexibility and protein is more rigid when coordinated with haem. Thus, it is hypothesised that REES measurement is a more effective means of predicting the structural effects exerted by PTM than that of MD simulation.

#### 5.3.4. Concluding Statement

Evidence compiled indicates that when coordinated with haem and glycosylated HRP is an extremely stable protein. The solvation of W117 in glycosylated *holo*-HRP signifies this as protein appears to have no-flexibility to allow for the change in solvation around the fluorophore. While both forms of PTM individually reduce the flexibility of HRP, maintaining a consistent structure within both non-glycosylated *holo*-HRP and glycosylated *apo*-HRP.

When glycosylation is absent, both *apo* and *holo*-HRP, have increased flexibility (Figure 5.8) which is believed to trigger a decrease in thermostability (Figure 5.10). Meanwhile, the elimination of both PTMs together results in a highly flexible protein with no clearly defined secondary structure. The lack of structural definition in the absence of both PTMs may explain why non-glycosylated *apo*-HRP failed to function as a corrective agent in commercial immunoassay in Chapter 4. As, if non-glycosylated *apo*-HRP is highly heterogeneous in structure the same protein-protein interactions which are critical to corrective function are not achievable.

It is proposed that when non-glycosylated, recombinant HRP, should be haem saturated to increase structural rigidity. However, alternative means of protein stabilisation to glycosylation should also be investigated. Particularly as de-glycosylation of HRP has significance to the use of the protein as a potential bio-therapeutic<sup>24,60,64</sup>. In future, the REES effects of artificial glycosylation should be investigated to assess the potential of increasing HRP stability. It has been previously reported that artificial glycosylation increases the stability of the non-glycosylated recombinant protein<sup>405</sup>. Quantification of this change by REES measurement would allow for comparison of artificial glycosylation benefit to that of native

glycosylation. This would in turn determine if artificial glycosylation is suitable means of reducing non-glycosylated protein structural flexibility.

## **Chapter 6.** The Effect of Oligomerisation on The Function of Disparate Fluorescent Proteins.

---

## 6.1. Introduction

### 6.1.1. The Importance of Protein Oligomerisation

Protein oligomers, or what is commonly referred to as quaternary structure, are arrangements of monomeric proteins which combine *via* specific intermolecular interaction to form a single multimeric structure<sup>208</sup>, or polymeric protein chains. Protein oligomers are more common in nature than that of protein monomers<sup>189</sup>, and homo-dimers are the most frequently oligomeric formation of protein<sup>201,406</sup>. It has been demonstrated previously that the association of fluorescent proteins can be promoted by the use of linker sequences<sup>407</sup>, domain insertion<sup>176</sup>, and the forming of oligomers from individual monomers<sup>202</sup>. This post-translational interaction of subunits has several benefits, which compensates for the increased metabolic expense of expression<sup>207</sup>. Benefits are increased control of surface-exposed residues, a lower surface area to volume ratio, and, control of active site orientation between molecules<sup>201,208,408</sup>. The outcome of which is that paired proteins have a greater resilience to environmental degradation and enhanced cooperation between catalytic or fluorescent subunits.

### 6.1.2. Previous Synthetic Oligomerisation of Monomeric Proteins.

Modern approaches to protein engineering have begun to utilise oligomerisation as a means of synthetically replicating the observed benefits of association<sup>409</sup>. There are several means of generating multimeric proteins through protein engineering<sup>189</sup>. For example, one simple method is to join molecules *via* a disulphide bridge by cysteine mutagenesis to form covalent linkage<sup>189</sup>. However, this form linkage does not specify how monomeric proteins align and frequently results in the for formation between homo-dimeric entities in addition to hetero-dimeric protein<sup>410</sup>(see Chapter 1.3.10). This method can only be utilised in oxidising conditions, which prevents the use of the PTM in the cell cytoplasm<sup>410</sup>.

Alternatively, domain insertion is a means of functional linkage, which is not dependent on the environment. This linkage involves the incorporation of two active protein regions into one continuous peptide chain<sup>411</sup>. Previous research by the DDJ lab demonstrated the possibility of generating functional communication between chromophores *via* domain insertion. Arpino *et al.* (2012) combined the genetic sequence of the fluorescent EGFP with the sequence of the redox-active Cytochrome *b*<sub>562</sub> (see Figure 7.31). In this study the gene

sequence of one protein was inserted into several possible sites of the second protein to form one dual functioning protein <sup>176</sup>.

The subsequent chimeric protein was proven to feature communication between chromophores, and quenching of EGFP emission was demonstrated to be dependent on both the proximity and redox sensitivity of the haem chromophore in Cytochrome *b*<sub>562</sub>. The observation of quenching in this domain-linked protein highlights the biosensing potential of chromophore communication, whilst practically demonstrating that partnerships can yield almost complete energy transfer. However, there are limitations to domain insertion as the method is highly disruptive and frequently results in both structural and functional disturbance <sup>411</sup>.

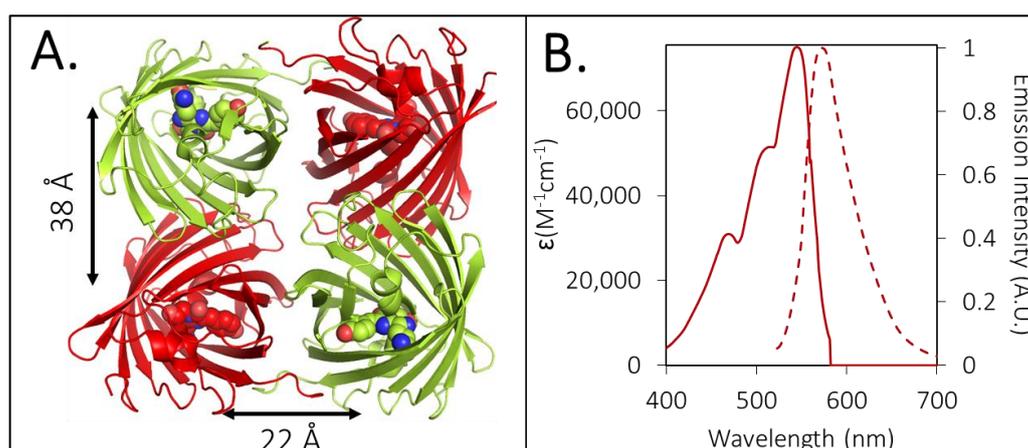
Subsequent work by Worthy *et al.* (2019), investigated if functional coupling could be achieved between two SPAAC (Strain Promoted Azide Alkyne Cycloaddition) linked (see Chapter 1.4.4) homodimeric proteins <sup>202</sup>. This study utilised interface prediction for defined nnAA incorporation, and subsequent molecular attachment, to anchor the monomeric pair. Functional analysis of the formed homodimeric protein then demonstrated that increasing the chromophore proximity of two sfGFP monomers, by the formation of the triazole linkage, produced a functionally enhanced sfGFP dimer. Wherein Worthy *et al.* (2019) demonstrated how identical fluorescent monomers could be linked by Click Chemistry to improve function, Arpino *et al.* (2012) demonstrated by domain insertion that communication between two structurally diverse proteins can be achieved by enforced proximity of chromophores. In this chapter, and Chapter 7, it will be investigated if the linking of two naturally disparate proteins, by click chemistry, can generate synergistic function, and replicate the diverse functional coupling achieved by Arpino *et al.* (2012).

### 6.1.3. Oligomeric History of Fluorescent Proteins

As discussed in Chapter 1.3.5. GFP<sup>WT</sup> can form homodimeric assemblies at high concentration <sup>167</sup>. Mutagenesis of residue A206V, in the formation of sfGFP, reduces this homodimeric tendency <sup>162</sup>, although weak intermolecular alignment is still noted when protein is stored at high concentration. This study aims to utilise this effect to link sfGFP with an additional, structurally homologous, protein.

The fluorescent protein DsRed, like GFP<sup>WT</sup>, naturally features a higher-order quaternary structure and is natively tetrameric<sup>154,412</sup>. GFP and DsRed are both autofluorescent  $\beta$ -barrel proteins from the metazoan super family<sup>413</sup>. DsRed is derived from the Bilateral coral genus, *Discosoma*, and GFP obtained from the Cnidarian species, *Aequorea Victoria*<sup>154,414</sup>. To date, engineering of each FP has primarily focused upon reducing both GFP and DsRed's tendency to oligomerise<sup>154,209,415</sup>. This has been performed to enhance each protein's use as a molecular imaging probe<sup>416</sup>, and reduce their influence upon the system of investigation<sup>412</sup>.

In nature, DsRed is composed of four identical subunits which are held together by hydrophobic interaction, hydrogen bonding, and, salt bridges (Figure 6.1, A)<sup>412</sup>. The tetrameric arrangement of DsRed features both immature (green) and mature chromophore forms (red) (Figure 6.1, A.)<sup>412</sup>. The combination of both chromophore forms allows for a wide range of excitation from 370 nm up to the  $\lambda_{\max}$  of 545 nm (Figure 6.1, B).



**Figure 6.1. Characteristics of Oligomeric DsRed.**

**A**, the two green fluorescent and two red fluorescent monomeric subunits of DsRed (PDB 1G7K), shown with the measured distance between each chromophore (Spheres). **B**, excitation (solid line) and emission (dashed line) of oligomeric DsRed.

Examination of the recorded spectral characteristics of DsRed on FPBase (<https://www.fpbases.org/protein/dsred/>) indicates that there is a significant functional difference associated with the oligomeric state of DsRed (Figure 6.1, B). The separation of DsRed into either dimeric or monomeric derivatives is accompanied by a reduced maturation time and an increased wavelength of both absorbance and emission (Table 6.1). However, in the case of monomeric DsRed, there was a 63% decrease in the reported quantum yield of protein, which was not observed when dimeric.

It is hypothesised that the loss of fluorescence efficiency in monomeric DsRed is stimulated by oligomeric separation. Reduced quantum yield is observed across almost all monomeric DsRed derivatives<sup>180,417–421</sup>, whilst dimeric derivatives retain, if not improve, quantum yield in comparison to tetrameric protein<sup>180,407</sup>. It is reasoned that the greater surface exposure which occurs upon monomeric separation increases the surface area of the protein which in turn decreases protein quantum yield.

For this investigation mCherry, a derivative of monomeric mRFP1 (Table 6.1) with a quantum yield of 0.22 was selected for the dimeric partnership with sfGFP. It is reasoned that the combination of green (sfGFP) and red chromophores (mCherry) in the investigation will mimic the combination of mature and immature chromophores within tetrameric DsRed and allow for the gradual stepping down of energy from the green to red region of the visible light spectrum by way of FRET (see Chapter 1.3.8.).

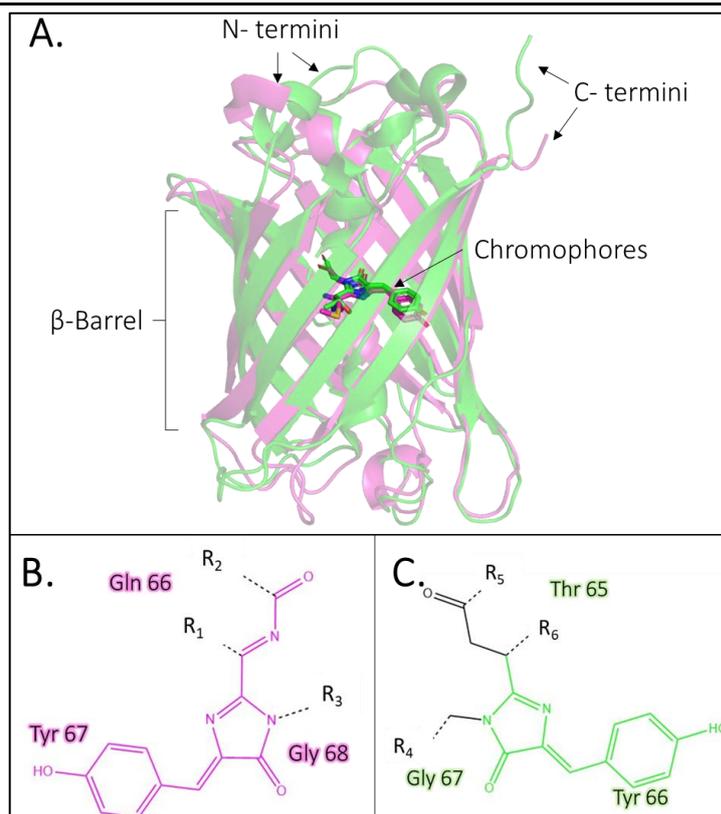
**Table 6.1. Spectral Properties of DsRed, obtained from FPBase**  
(<https://www.fpbases.org/protein/dsred/>)

	Tetrameric DsRed (DsRed)	Dimeric DsRed (Dimer 2)	Monomeric DsRed (mRFP1)
$\lambda_{Abs}$ (nm)	546	552	584
$\lambda_{Emis}$ (nm)	574	579	607
Extinction Coefficient ( $M^{-1}cm^{-1}$ )	72,500	69,000	50,000
Quantum Yield $\Phi$	0.68	0.69	0.25
Maturation Time (min)	1600	120	60

#### 6.1.4. Structural and Functional Comparison of mCherry and sfGFP

Structurally mCherry and sfGFP share the same overall  $\beta$ -can structure, but just 28% amino acid homology (Figure 6.2, A). Each protein is formed predominantly of  $\beta$ -sheets which fold to form a  $\beta$ -barrel around a central  $\alpha$ -helical core that contains a fluorescent chromophore (Figure 6.2, A)<sup>422</sup>. The mature chromophore of each protein features three residues which undergo covalent rearrangement(see Chapter 1.3.4.), which takes 13.6 minutes in sfGFP and 15 minutes in mCherry<sup>422</sup>.

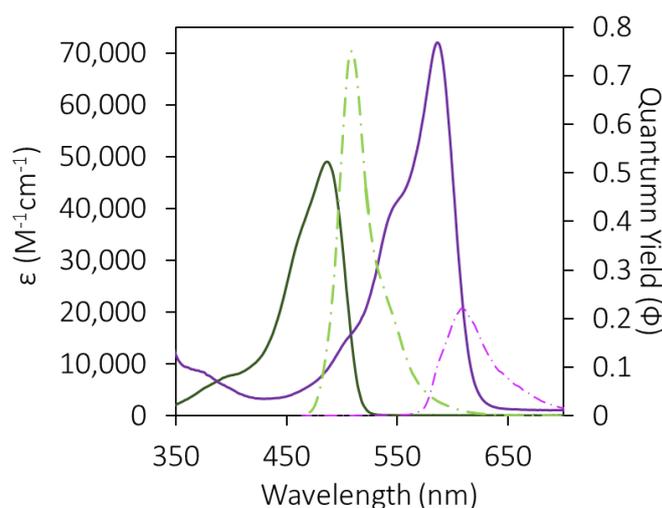
The increased wavelengths of mCherry excitation and emission are a result of the mCherry chromophore's extended conjugated double bond network, in comparison to sfGFP, which extends from Tyrosine 67 into Glutamine 66 (Figure 6.2, B)<sup>412,423</sup>. Emission of mCherry occurs at 587 nm is approximately 100 nm greater than that of sfGFP at  $\lambda_{Max}$  485 nm. However, the quantum yield of mCherry is low, approximately 3 x less than that of sfGFP at 0.22  $\Phi$  (587



**Figure 6.2. The Structural Properties of both sfGFP and mCherry.**

**A**, structural alignment of mCherry (PDB code 2h5q) and sfGFP (PDB code 2b3p). Molecular configuration of mCherry chromophore (**B**) and sfGFP chromophore (**C**).

nm), and  $0.75 \Phi$  485nm respectively (Table 6.2). Although the molar extinction coefficient of mCherry is 47% greater than that of sfGFP at  $72,000 \text{ M}^{-1} \text{ cm}^{-1}$  (587 nm) and  $49,000 \text{ M}^{-1} \text{ cm}^{-1}$  (485 nm) respectively (Figure 6.3).



**Figure 6.3. The Absorbance and Emission Spectra of sfGFP and mCherry.**

Absorbance presented as molar extinction coefficient (**Solid Line**) and emission presented as quantum yield (**Dashed Line**) of sfGFP (**Green**) and mCherry (**Purple**).

Table 6.2. The Fluorescence Properties of sfGFP and mCherry obtained from FPBase.org

	sfGFP	mCherry
$\lambda_{\text{MAX}}$ (nm)	485	587
$\lambda_{\text{EM}}$ (nm)	510	610
Molar Extinction Coefficient ( $\text{M}^{-1} \text{cm}^{-1}$ )	49,000	72,000
Quantum Yield $\Phi$	0.75	0.22
Maturation time (min)	15	13.6

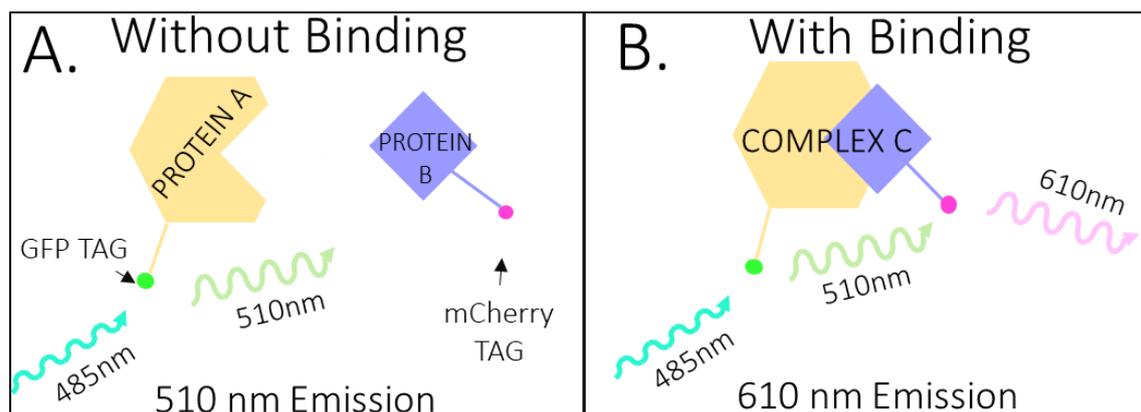
## 6.1.5. Functional Communication between sfGFP and mCherry

Energy transfer between the FPs is typically observed by way of FRET<sup>424</sup> and is described in Chapter 1.3.8. FRET is the non-radiative energy transfer which occurs between chromophores which share significant spectral overlap and is dependent on both distance (~8- 10 nm) and orientation between chromophores<sup>425</sup>(see Chapter 1, Figure 1.11). The distance at which a 50% transfer of energy between chromophores arises is known as the Förster radius ( $R_0$ ), and in the case of sfGFP and mCherry, is stated on FPbase.org to be 52.89 Å (Table 6.3). When chromophores are aligned at a distance greater than the FRET radius, energy transfer tails off rapidly as energy transfer is inversely proportional to the sixth power of distance<sup>426</sup> (Figure 6.4, A). As a result, the distance between each protein can be calculated, and the process can be utilised for the measurement of protein interaction in real-time within cells<sup>424,426</sup> (Figure 6.4).

It is anticipated that the triazole linkage between the two FPs will align the chromophores within the predicted  $R_0$  of 52.89 Å (Table 6.3), and an energy transfer > 50% will be observed. In this case, sfGFP will function as the donor with  $\lambda_{\text{Em}}$  510 nm, which will stimulate mCherry absorbance at  $\lambda_{\text{Ex}}$  of 587 nm, and subsequent emission at 610 nm.

Table 6.3. The FRET Properties of sfGFP and mCherry.

	sfGFP + mCherry
J Coupling Constant ( $\lambda$ ) ( $*1^{15}\text{M}^{-1} \text{cm}^{-1} \text{nm}^4$ )	1.8
$R_0$ (Å)	52.89
$\kappa^2$	0.6667



**Figure 6.4. The Initiation of FRET Between Fluorescent Proteins as a Means of Inferring Protein Interaction.**

**A**, In the absence of Protein A and Protein B association no energy is transferred from sfGFP to mCherry and no emission at 610 nm occurs upon excitation at 485 nm. **B**, upon the formation of Complex C, sfGFP and mCherry Tags align within FRET radius allowing for FRET between chromophores and the emission of light at 610 nm after excitation at 485 nm.

#### 6.1.6. Aims and Objectives.

This chapter aims to determine: 1, if the linkage of two fluorescent proteins by click chemistry is possible; 2, if the fluorescence properties (e.g. quantum yield) of both sfGFP and mCherry are altered upon dimerisation; and 3, if direct communication between the chromophore pair is initiated. To achieve this, interface prediction will be utilised to identify sites of compatible interfaces between sfGFP and mCherry. Then, site-directed mutagenesis will be used to incorporate nnAAs AZF and SCO-K (see Chapter 1.4.4) for protein linkage *via* click chemistry. Wherein, FRET between the chromophore pair is expected and reasoned to be enhanced, consistent with the previous linkage of GFP-type proteins<sup>202</sup>. It is anticipated that linking diverse proteins by way of interface prediction, as described by Worthy *et al.* (2019), will replicate the formation of novel synergistic function as observed by Arpino *et al.* (2012), whilst minimising the risk of structural disruption.

## 6.2. Results and Discussion

### 6.2.1. Selection of an sfGFP and mCherry<sup>WT</sup> Oligomerisation Interface

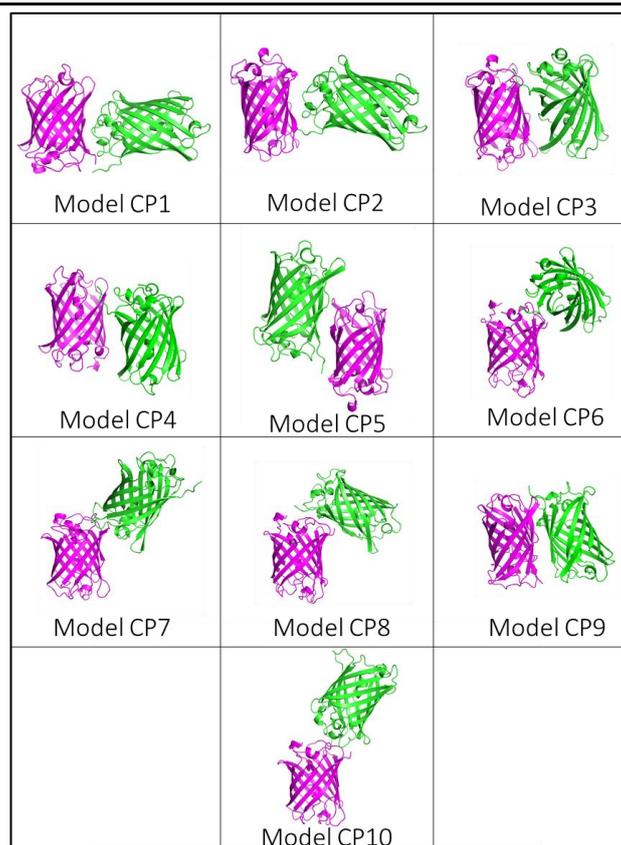
As demonstrated by Reddington *et al.* (2012), residue choice for nnAA incorporation is critical for SPAAC to occur at high efficiency<sup>265</sup>. It is known that the surface microenvironment surrounding each nnAA has significance for the ability to form molecular attachment<sup>265</sup>, and not all surface-exposed residues on sfGFP are compatible with SPAAC. Consequently, the incorporation of nnAA must occur within two protein regions which are compatible for interfacing, or else, cross-linking between molecules will not occur. It is rationalised that nnAA incorporation, within the predicted region of the interface *in vivo*, enhances the alignment of nnAAs and increases the affinity of residues to form a molecular attachment. Molecular modelling and interface ranking allow for the selection of the regions which are most compatible for cross-linking. This approach simulates the alignment between proteins *in silico* and predicts the weak interactions which occur between residues.

ClusPro is a server which clusters the predicted alignment between two proteins to determine which surface interactions are most likely to occur. The ClusPro server rotates one ligand-protein around a secondary receptor protein on three planes of axis sampling 1000s of possible interface orientations before ranking<sup>427</sup>. The ranking then utilises the Fourier correlation to eliminate non-compatible interfaces based on electrostatic repulsion, hydrophobicity, and the number of van der Waals contacts<sup>280</sup>. This prediction allows for subsequent clustering of interfaces to determine the most compatible regions of within a 9 Å radius<sup>428</sup>. The simulation assumes that the greater the frequency of interfaces are predicted to occur at a given location, the greater chance of native residue interaction<sup>427</sup>.

ClusPro docking (<https://cluspro.bu.edu>) (MA, USA)<sup>427</sup> was used for the naïve docking of sfGFP and mCherry. The “Dock” simulation on the ClusPro server was performed with sfGFP (PDB: 2B3P)<sup>168</sup> and mCherry (PDB: 2H5Q)<sup>429</sup> (see Chapter 2.7.3.). The top 30 most common interface sites predicted, based upon a balance of all intermolecular forces, were downloaded

**Table 6.4. Number of Members Featured in Each of The Top 10 Models Sampled by Alignment Between sfGFP a mCherry in ClusPro.**

Model Number	1	2	3	4	5	6	7	8	9	10
Number of Members	154	111	84	71	50	45	42	39	34	26



**Figure 6.5. Top 10 Models Generated by ClusPro server.**

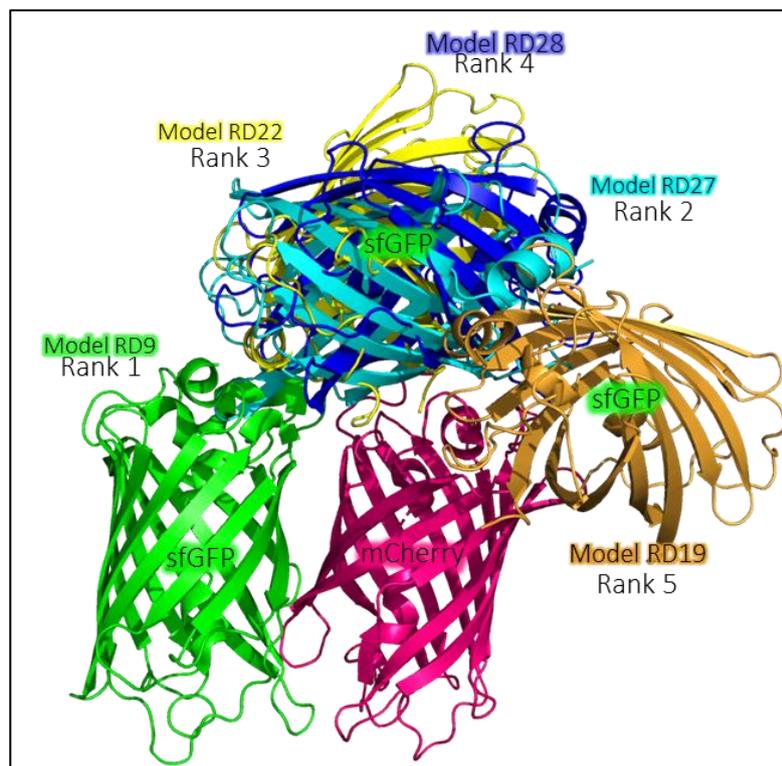
The top 10 ranked interfaces of sfGFP, PDB 2b3p, (**Green**) and mCherry, PDB 2h5q, (**Pink**). The top clusters ranked by the number of members for each cluster (See Table 6.4.) for assessment, of which the top 10 are displayed in Figure 6.5. From the ClusPro modelling, model CP1 featured the greatest number of clustered members (Table 6.4). However, one problem face with ClusPro simulation is that it samples rigid copies of each PDB file and does not account for the natural flexibility of each protein<sup>430,431</sup>. To address this issue and to provide a more accurate energetic-based ranking of binding interfaces, further ranking of models was performed using the RosettaDock software package<sup>432</sup>(see Chapter 2.7.4).

RosettaDock simulation predicted protein-protein interaction by the modelling of individual side chain environments and, later, several subsequent rounds of backbone

**Table 6.5. Ranked Placement of ClusPro Acquired Models According to the Total Energy between Molecules.**

Ranked Placement	Model ID	Total Energy (kJ/mole)	Interface Energy (kJ/mole)
1	RD9	-426	-13.6
2	RD27	-412.4	-14.6
3	RD22	-379.2	-11.6
4	RD28	-369.8	-12.2
5	RD19	-369.6	-10.6

relaxation<sup>433</sup>. Then the simulation arranges the molecules in short independent segments before energy minimisation<sup>430</sup>. RosettaDock simulation returns structures which are then modified per the individual residue environments and the van de Waals attractions of interfaces<sup>281,430</sup>. These simulations are then ranked based on the total energy which exists between proteins at a given alignment, and the localised energy which exists between protein at a given interface (Table 6.5.)<sup>282</sup>.



**Figure 6.6. The 5 Lowest Energy Interfaces as Predicted by Rosetta Dock.**

The top 5 alignments of sfGFP (Various Colours) and mCherry (Magenta) ranked according to total energy of interaction by Rosetta Dock. Model CP9 featured the smallest distance between chromophores of 18 Å.

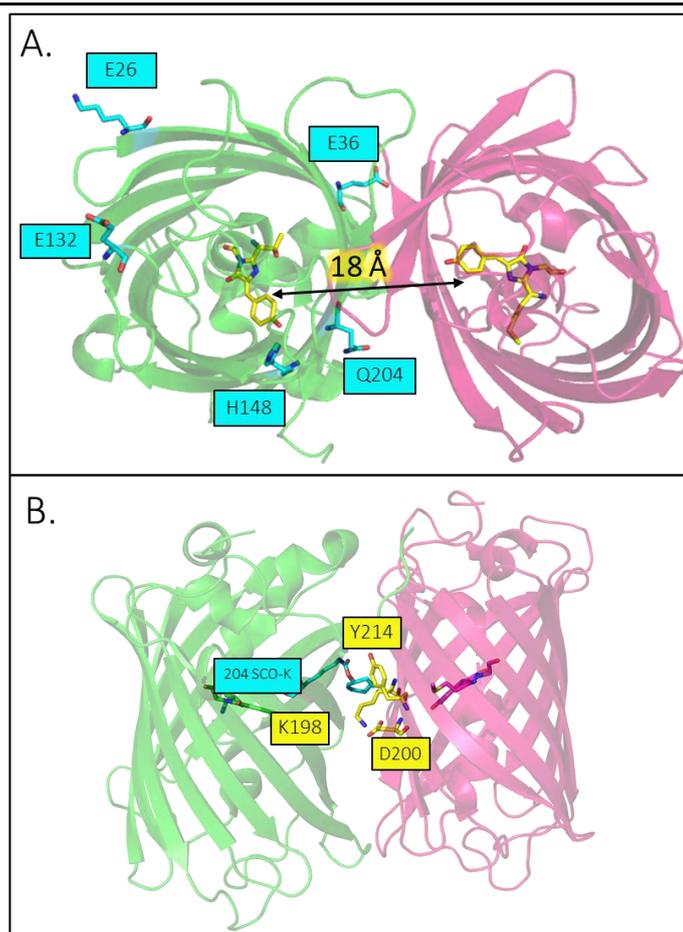
The lowest five energy state models obtained from ClusPro are ranked according to RosettaDock output (Table 6.5). Results highlight the limitation of using ClusPro interface prediction alone as the regions which are sampled the most frequently by ClusPro (Figure 6.5, Models 1-8) are not those which feature the lowest energy requirements for interaction and, are not featured in the top 5 simulations ranked in RosettaDock (Figure 6.6). Model RD9 is shown to have the lowest total energy score of -426 kJ/mole, and model RD27 features the lowest interface energy between molecules.

Model RD9 was selected in favour of model RD27 for further analysis for several reasons which were as follows. Firstly, the total energy between molecules was determined to

be a more feasible prediction of protein interaction than interface energy, as it was calculated based upon all residues in each model, not just those at the region of the potential interface. Secondly, Model RD9 was the only structure which sampled a parallel alignment of  $\beta$  barrels (Figure 6.6), and so model RD9 features the closest alignment of the two chromophores with a measured distance of 18 Å between the chromophores. As the FRET is inversely proportional to the 6<sup>th</sup> power of the distance between chromophores, the closer chromophore alignment at this region of the interface was reasoned to be more conducive to FRET.

### 6.2.2. Residue selection for AZF Incorporation

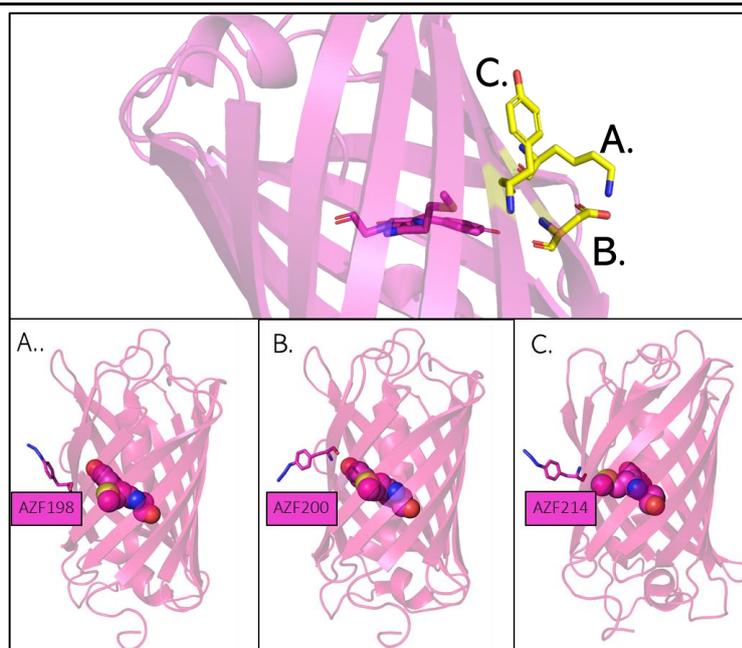
The first stage of residue selection was to identify the residues which were present within the interface of model RD9. As there is a decreased yield when incorporating SCO-K into protein, it was decided that it would be beneficial to see if any known sfGFP TAG variants previously utilised for SCO-K expression were located within this interface (Personal communication, Dr Dafydd Jones, January 2018). Accordingly, all existing surface-exposed TAG mutants of sfGFP were modelled into Model 9 to see if any were present within the interface (Figure 6.7, A). Both residues E36 and Q204 were observed within the interface between the two models. Q204 was selected in favour of E36, as both homo and heterodimers have formed between sfGFP monomers previously at this residue<sup>202,265</sup>. This residue was mutated in PyMOL to SCO-K, and mCherry residues within a measured 5 Å distance that were surface-exposed were identified (Figure 6.7, B).



**Figure 6.7. Selection of Residues for nnAA Incorporation in Model RD9.**

Both A and B feature model 9 with sfGFP shown as **Green**, mCherry as **Pink**, with chromophores shown internally as sticks. **A**, Existing sites of TAG mutation shown as sticks (**Teal**) in sfGFP which could be utilised for SCO-K incorporation. Q204 is shown as within the region of interface between the two molecules. **B**, selection of surface exposed residues on mCherry expansion of 204 (Teal) by 5 Å shown as sticks (**Yellow**).

The three surface-exposed mCherry residues, in 5 Å proximity to residue 204 SCO-K in sfGFP were K198, D200, and Y214 (Figure 6.8). Residues were mutated to AZF in PyMOL to ensure that no steric clashes would occur upon the incorporation of nnAA (Figure 6.8, A-C). From this analysis, no steric clashes were observed, and all three residues were determined to be suitable targets for nnAA incorporation by SDM (see Chapter 2.2.3). Residue K198 in mCherry is the structural homologue of sfGFP Q204. As sfGFP incorporation of both AZF and SCO-K to residue 204 were previously reported to exert minimal influence to emission<sup>202</sup>, mCherry K198 was selected as the first target residue for AZF incorporation. It was reasoned that should mCherry 198 AZF be unsuitable for oligomerisation D200 and Q214 would subsequently be mutated for AZF incorporation.



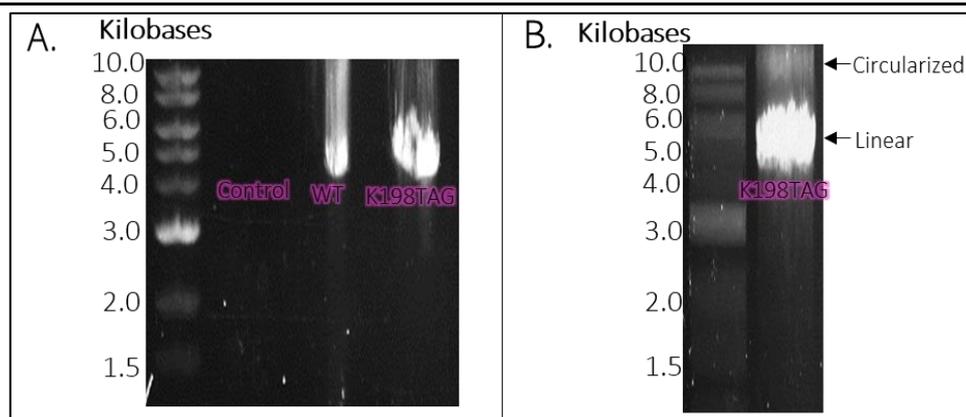
**Figure 6.8. Potential Residues of nnAA Incorporation in mCherry.**

Three residues of potential nnAA incorporation shown as sticks (Yellow) and mutated to AZF in PyMOL (Pink). Letters represent the amino acid of residue of incorporation. **A**, K198 AZF, **B**, D200AZF, and **C**, Y214AZF.

### 6.2.3. Site-Directed Mutagenesis of mCherry<sup>WT</sup> to mCherry<sup>198TAG</sup>

As described in Chapter 1 (see Chapter 1.3.2.), the incorporation of nnAA utilises the amber stop codon “TAG”. To incorporate AZF to residue 198, SDM was first necessary to substitute 198 AAG with 198 TAG (Figure 6.9). Modification of the genetic sequence of mCherry was performed by SDM (see Chapter 2.2.3.1.) using mCherry<sup>198TAG</sup> mutagenesis oligonucleotides (see Chapter 2.2.2.). Agarose gel electrophoresis of amplified pBAD - mCherry<sup>198AZF</sup> indicated that amplified plasmid was of the same size as pBAD- mCherry<sup>WT</sup> at approximately 4.8 kB (Figure 6.9, A). No amplification was observed in the control sample without the template. The pBAD- mCherry<sup>198TAG</sup> plasmid was subsequently phosphorylated and ligated (see Chapter 2.2.3.4) before final check of plasmid ligation by gel electrophoresis. The mutated plasmid was shown to have circularised by the presence of a newly formed larger band (Figure 6.9, B) and was transformed into DNA propagation NEB 5- $\alpha$  *E. coli* cells and grown overnight at 37°C on carbenicillin supplemented LB agar.

The sequence of pBAD- mCherry<sup>198TAG</sup> was confirmed by Eurofins Genomics after DNA extraction from cells (see Chapter 2.2.5). Once the sequence was confirmed plasmid was transformed along with the pDULE plasmid (AzF incorporation apparatus, see Chapter 2.3.3.) into Top10<sup>TM</sup> *E. coli* cells for expression.



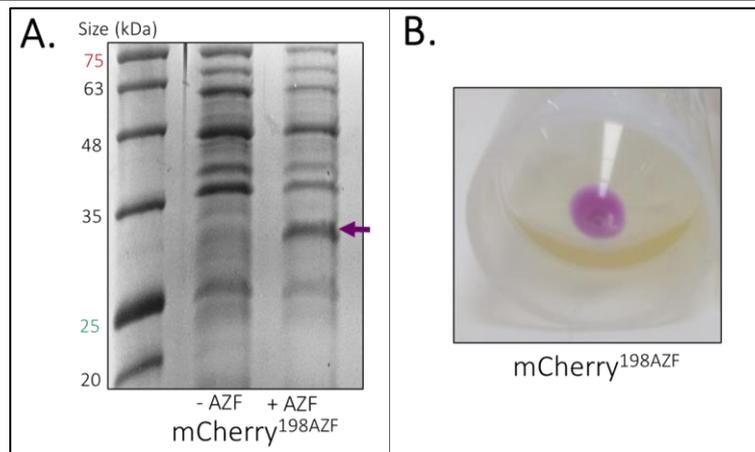
**Figure 6.9. Generation of pBAD-mCherry<sup>198AZF</sup>.**

**A**, Agarose gel electrophoresis of bands formed from amplification of pBAD-mCherry<sup>WT</sup> with mCherry<sup>198TAG</sup> mutagenesis oligonucleotides. Control well represents amplification observed without pBAD-mCherry<sup>WT</sup> template and WT well represents template along. **B**, Results of blunt end ligation of amplified pBAD-mCherry<sup>198AZF</sup> showing the presence of both linear and circularised plasmid.

#### 6.2.4. Expression and Purification of mCherry<sup>198AZF</sup>

Cells transformed with both pDULE and pBAD - mCherry<sup>198TAG</sup> were induced as described in Chapter 2.3.3. Confirmation that protein expression was dependent on AZF presence was achieved by growth of a control cell culture which was supplied with nnAA. This control step was incorporated to ensure that cells were transformed with pBAD-mCherry<sup>198TAG</sup> and not pBAD-mCherry<sup>WT</sup>. In the absence of nnAA cells the amber stop codon (TAG) at position 198 would prematurely terminate translation generating a truncated, non-functional protein if cells were transformed with pBAD - mCherry<sup>198TAG</sup>, however, cells transformed with pBAD - mCherry<sup>WT</sup> would continue to express functional protein.

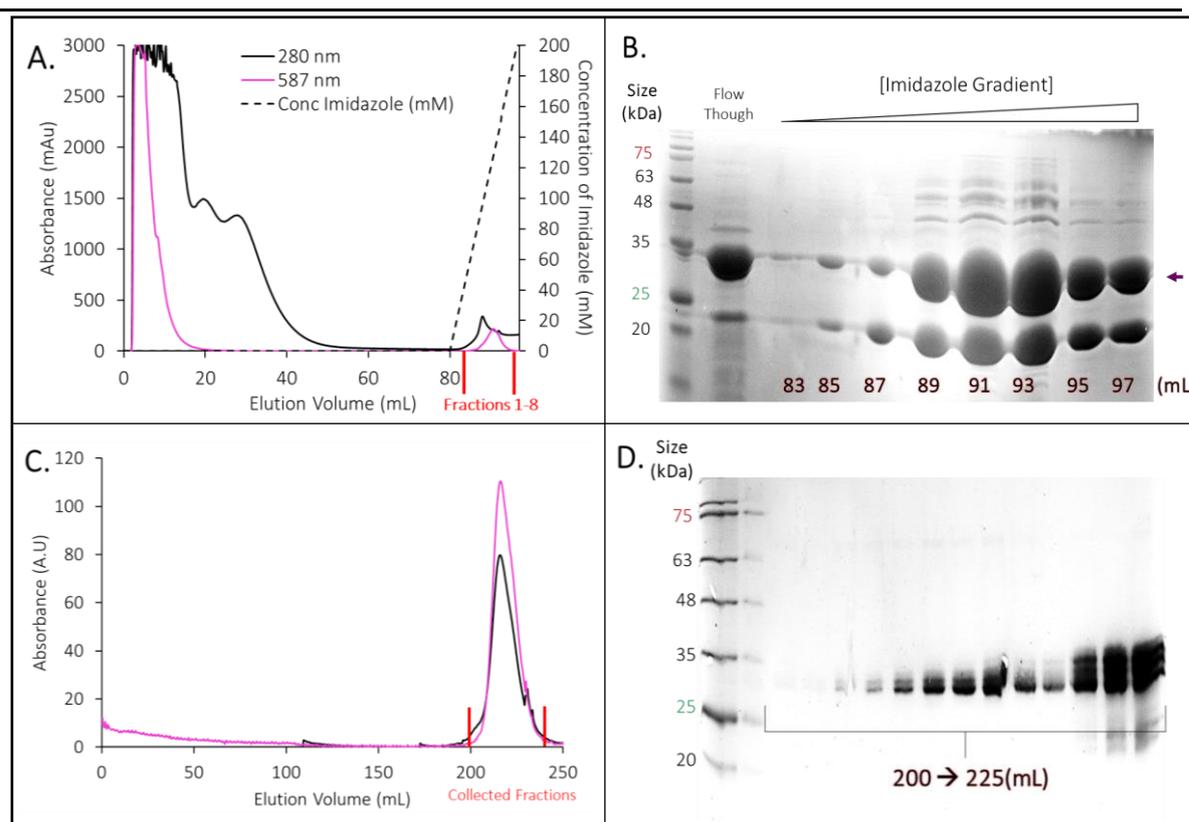
SDS PAGE (Figure 6.10, A) indicated that a protein consistent with mCherry expression (~27 kDa) was expressed only in the presence of supplied AZF (Figure 6.10, A). Furthermore, cells which were supplied with AZF were observed to change colour overnight and turn bright magenta in colour (Figure 6.10), once more indicative of mCherry expression. Based on these results, large scale expression of mCherry<sup>198AZF</sup> was achieved by the addition of nnAA, and arabinose to 1L cell culture (see Chapter 2.3.3), in overnight incubation at 37 °C in the dark. Induced cells were lysed by French pressure cell lysis (see Chapter 2.3.5.) and purification of expressed protein was achieved in two steps: the first of which was IMAC (see Chapter 2.4.1.), and the second of which was SEC (see Chapter 2.4.2.).



**Figure 6.10. Expression of mCherry<sup>198AZF</sup>.**

**A**, SDS PAGE of cells transformed with pDULE and pBAD-mCherry<sup>198TAG</sup> after induction with 1% arabinose. Only supplied with AZF show no band consistent with expression of mCherry at ~27 kDa, indicated by arrow. **B**, colouration of cells supplied with AZF consistent with expression of mCherry.

Protein lysate was applied to Ni<sup>2+</sup> resin column for affinity purification and fractions were collected after application of an imidazole gradient (20-200 mM) and the increase of absorbance at both 280 nm and 587 nm (Figure 6.11, A). SDS PAGE of fractions indicated the presence of protein sized at ~27 kDa consistent with mCherry and several additional contaminate proteins (Figure 6.11, B). Removal of non-specific protein was achieved by SEC (Figure 6.11, C). Again, absorbance at both 280 nm and 587 nm was observed, and fractions were collected as both increased in intensity at 200 mL elution volume. SDS PAGE of collected fractions indicated the presence of two proteins, one sized at ~27 kDa and the other ~23 kDa (Figure 6.11, D). Banding at ~27 kDa was consistent with the presence of full-length mCherry, whilst banding at 23 kDa was consistent with mCherry which had cleaved when boiled in SDS. This secondary band is standard in both mCherry<sup>WT</sup> and DsRed expression and occurs due to cleavage of the protein at Q66 when protein is boiled in SDS<sup>414</sup>. This cleavage occurs due to the presence of an unstable acylimine within the chromophore of mCherry which is prone to nucleophilic attack upon denaturation<sup>414</sup>.



**Figure 6.11. Purification of mCherry<sup>198AZF</sup> by IMAC and SEC.**

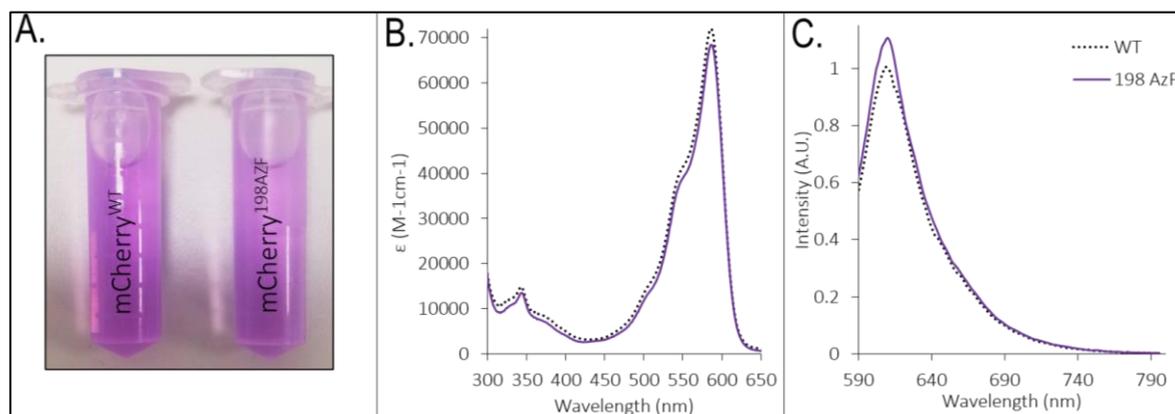
**A**, chromatograph of affinity purification of mCherry<sup>198AZF</sup> from lysate using 5 mL Ni<sup>2+</sup> on AKTA Prime. Absorbance at 587 nm (**Magenta**) and 280 nm (**Black**) indicates the presence of mCherry<sup>198AZF</sup> in column flow through (elution volume 0-50 mL) and after washing of column in gradient of imidazole (Elution Volume 80-100 mL). **B**, SDS PAGE of collected fractions from, indicating the presence of mCherry<sup>198AZF</sup> due to presence of band at 27 kDa. **C**, SEC chromatograph of mCherry<sup>198AZF</sup>. Absorbance at both 280 ml (**Black**) and 587 ml (**Magenta**) is shown to increase at volume of 200 mL. **D**, SDS PAGE of fractions obtained at elution volume 200-225 mL from SEC indicating the presence a band at ~27 kDa.

### 6.2.5. Functional Characterisation of mCherry<sup>198AZF</sup>

It has been previously observed that altering the chromophore environment of fluorescent proteins can lead to functional changes, such as photo-switching<sup>275</sup> and alteration of spectral characteristics<sup>252</sup>(see Chapter 1.3.6.). As the side chain of K198 does not directly interact with the chromophore, no spectral change was predicted from AZF incorporation. However, to determine the potential effect of dimerisation between sfGFP and mCherry, it was first necessary to confirm this prediction so that a baseline of mCherry<sup>198AZF</sup> function could be established for later comparison.

The colour of mCherry<sup>198AZF</sup> indicated very little spectroscopic change upon the incorporation of nnAA (Figure 6.12, A). The molar extinction coefficient of mCherry<sup>198AZF</sup>, was consistent to mCherry<sup>WT</sup> with a minor decrease from 72,000 M<sup>-1</sup> cm<sup>-1</sup> in mCherry<sup>WT</sup> to

68,400  $M^{-1} cm^{-1}$  in mCherry<sup>198AZF</sup> at a  $\lambda_{max}$  of 587 nm (Figure 6.12, B). The fluorescent activity of mCherry<sup>198AZF</sup> was maintained upon the incorporation of nnAA with excitation at  $\lambda_{EX}$  587 nm and emission at  $\lambda_{Em}$  610 nm (Figure 6.12, C). A small increase in the quantum yield of mCherry<sup>198AZF</sup> was observed in comparison to mCherry<sup>WT</sup>. Thus, no substantial change in mCherry<sup>198AZF</sup>'s functional characteristics was observed by AZF incorporation.



	$\lambda_{Max}$	$\lambda_{Emis}$	$\epsilon$ ( $M^{-1}cm^{-1}$ )	Quantum Yield $\Phi$
mCherry <sup>WT</sup>	587 nm	610 nm	72,000	0.22
mCherry <sup>198AZF</sup>	587 nm	610 nm	68,400	0.24

**Figure 6.12. Characteristics of mCherry<sup>198AZF</sup> in comparison to mCherry<sup>WT</sup>.**

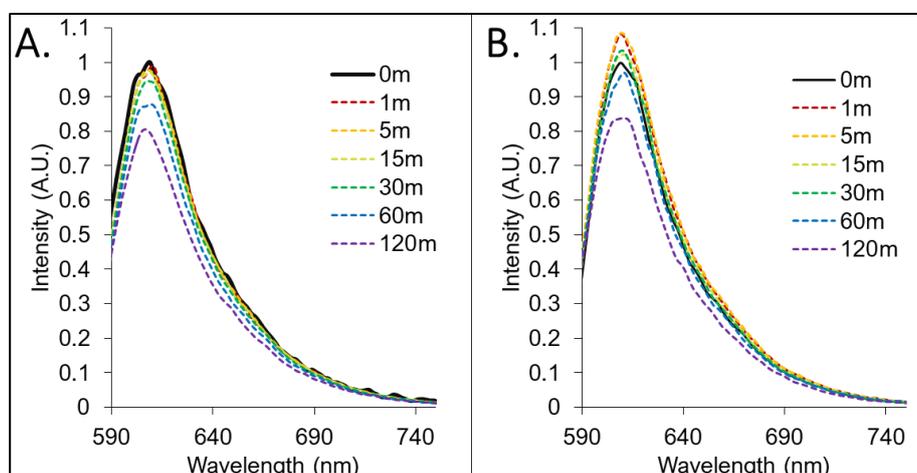
**A**, 10  $\mu M$  concentrations of both mCherry<sup>WT</sup> and mCherry<sup>198AZF</sup>, colouration of mCherry<sup>198AZF</sup> shown to be slightly less intense than that of mCherry<sup>WT</sup>. Absorbance (**A**) and Emission (**B**) of mCherry<sup>198AZF</sup> (Purple solid line) and mCherry<sup>WT</sup> (Black Dotted Line). Exact values of each shown in table.

#### 6.2.6. UV Irradiation of mCherry<sup>198AZF</sup>

Under the influence of UV light, AZF reacts to form a nitrene radical<sup>299,431</sup>. The previous incorporation of AZF into mCherry exerted photocontrol over its fluorescent properties through UV light exposure<sup>252,265</sup>. While residue 198 is not expected to modulate photo control of fluorescence, the effect of UV light irradiation was investigated to confirm no photo-mediated change occurred (Figure 6.13, A).

Dark state mCherry<sup>198AZF</sup> that had been expressed purified, and stored, in the dark, was exposed to UV irradiation, and excited at 587 nm several times over 2 hours (with readings at, 0 m, 1 m, 5 m, 15 m, 30 m, 60 m, and 120 m). When excited at 587 nm emission at 610 nm did increase 9% within the first 5 minutes of UV exposure (Figure 6.13, B). However, this increase was not sustained, and emission dropped back to pre-exposure levels after 30 further minutes

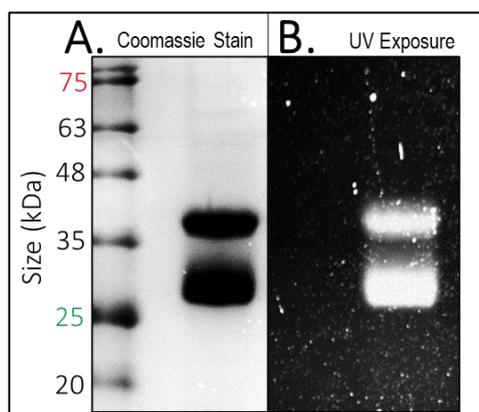
of UV exposure. Over 2 hours of UV exposure of both mCherry<sup>WT</sup> and mCherry<sup>198AZF</sup> emission, decreased to 20% and 16% of pre-irradiation intensity, respectively. Therefore, results demonstrate that UV irradiation of mCherry<sup>198AZF</sup> stimulated no greater change to function than observed in mCherry<sup>WT</sup>.



**Figure 6.13. The Effect of UV Exposure on Emission of mCherry<sup>WT</sup> and mCherry<sup>198AZF</sup>.** A, mCherry<sup>WT</sup>. B, mCherry<sup>198AZF</sup>. Both samples exposed to UV for a total of 120 minutes (Data normalised to an initial intensity of 1).

#### 6.2.7. Molecular Attachment of mCherry<sup>198AZF</sup> with Cy3 DBCO

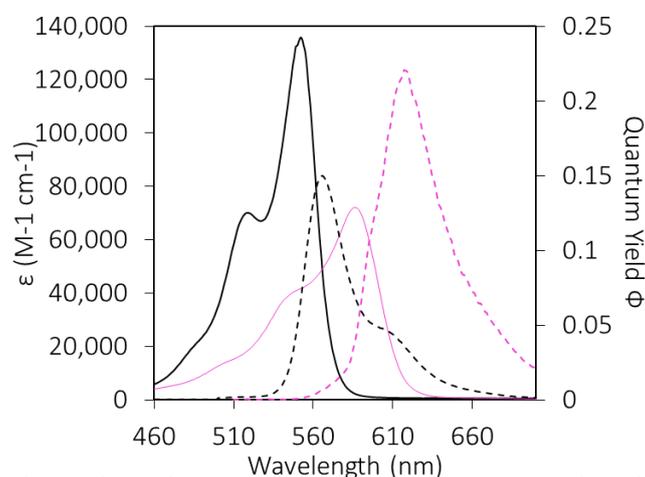
Cy3 DBCO is a reactive azide probe, which can tag azide protein by copper-free click reaction<sup>434</sup>. Purified mCherry<sup>198AZF</sup> was mixed with Cyanine3 dibenzocyclooctyne (Cy3 DBCO), to confirm if the AZF residue was accessible for click chemistry. Protein and probe were mixed at a working concentration of 10  $\mu$ M respectively and incubated overnight at room temperature. SDS PAGE of mCherry<sup>198AZF</sup>: Cy3 DBCO mix the next day indicated the presence



**Figure 6.14. SDS PAGE of mCherry<sup>198AZF</sup> Incubated with Cy3 DBCO.**

SDS PAGE as seen by Coomassie Stain (A), and by UV light exposure (B) of 10  $\mu$ M mCherry<sup>198AZF</sup> mixed with 10  $\mu$ M CY3 DBCO. A bands consistent mCherry<sup>198AZF</sup> at ~27 kDa is observed by coomassie stain in A, and under UV light in B.

of mCherry<sup>198AZF</sup> by banding at 27 kDa (Figure 6.14, A), which was subsequently observed to fluoresce in UV light (Figure 6.14, B). The fluorescence of SDS PAGE banding at 27 kDa, consistent with mCherry<sup>198AZF</sup>, indicated that incorporated AZF had formed a molecular attachment with Cy3 DBCO.



**Figure 6.15. Spectral Overlap Observed Between Cy3 DBCO and mCherry<sup>198AZF</sup>.**

Absorbance corrected to a Molar Extinction Coefficient (**Solid line**) and emission corrected to a maximum intensity of Quantum Yield (**Dotted Line**) of both Cy3 (**Black**) and mCherry (**Magenta**).

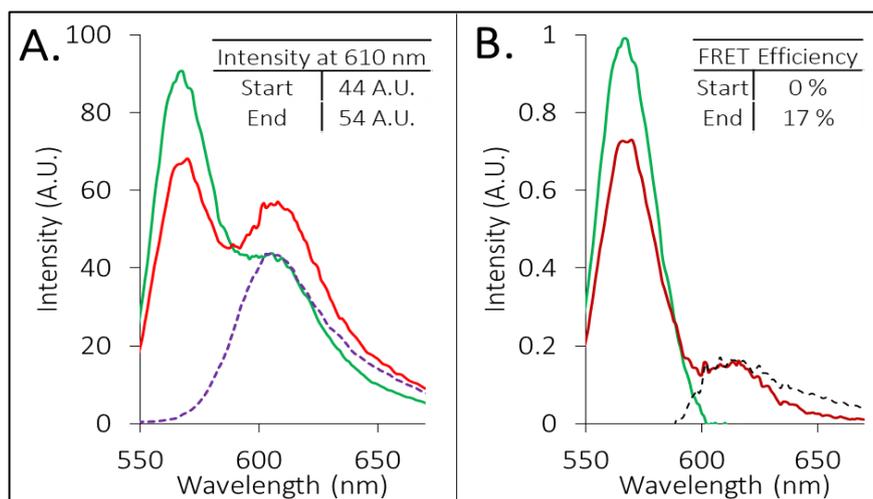
Furthermore, as Cy3 is excited at 555 nm with an emission of 565 nm (Figure 6.15), therefore a significant spectral overlap exists between Cy3 DBCO and mCherry, with a J coupling constant of  $4.83 \cdot 10^{15} \text{ M}^{-1} \text{ cm}^{-1} \text{ nm}^4$  (Table 6.6) Thus, Cy3 DBCO emission at 565 nm is expected to stimulate enhanced excitation of the mCherry fluorophore if the pair are aligned within  $48.85 \text{ \AA}$  (Table 6.6). Thus, when linked by molecular attachment (as observed in Figure 6.14) the excitation of the protein: probe mix at 515 nm is expected to stimulate the emission of mCherry at 610 nm by way of FRET.

**Table 6.6. Spectral Characteristics of mCherry<sup>198AZF</sup> and Cy3 DBCO.**

	Cy3 DBCO	mCherry <sup>198AZF</sup>
$\lambda_{\text{Abs}}$ (nm)	555	587
$\lambda_{\text{Emis}}$ (nm)	565	610
Extinction Coefficient ( $\text{M}^{-1}\text{cm}^{-1}$ )	135,850	68,400
Quantum Yield $\Phi$	0.15	0.24
J Coupling Constant ( $\lambda$ ) ( $\cdot 10^{15} \text{ M}^{-1} \text{ cm}^{-1} \text{ nm}^4$ )	4.83	
$R_0$ ( $\text{\AA}$ )	48.85	

To establish if chromophore communication was enhanced by molecular attachment, a fresh protein: probe mix was incubated overnight at equal concentration. Emission stimulated by excitation at 515 nm was recorded immediately upon mixing and then repeated

after overnight incubation in dark conditions. Following incubation, for 18 hours, the sample was again excited at 515 nm, and the emission spectrum recorded (Figure 6.16, A). Subsequently, the donor emission intensity of Cy3 at 565 nm was shown to decrease from 90 A.U. to 65 A.U., whilst the acceptor emission intensity of mCherry at 610 nm was observed to increase from 42 A.U. to 56 A.U. (Figure 6.16, A), after overnight incubation.



**Figure 6.16. Emission of Cy3 DBCO and mCherry<sup>198AZF</sup>.**

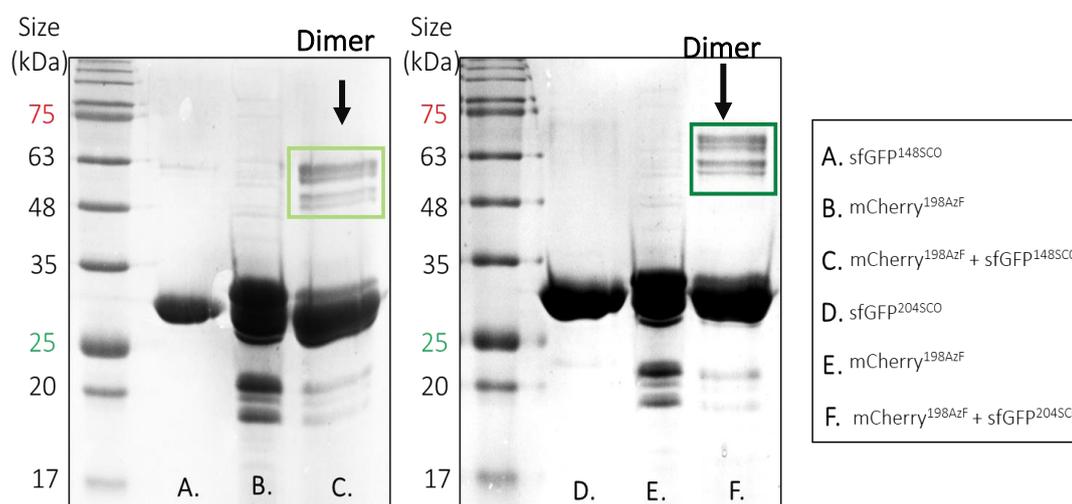
**A**, raw emission spectra of 1:1 mix of Cy3 DBCO and mCherry<sup>198AZF</sup> at  $\lambda_{EX}$  515 nm immediately after mixing (**Green**) and after 18 hours (**Red**) with emission of mCherry at 515 nm presented for reference (**Purple Dashed**). Insert displays raw intensity of emission at 610 nm. **B**, the corrected emission spectrum of mix at  $\lambda_{EX}$  515 nm; with initial fluorescence (**Green**), and after 18 hours (**Red**) after subtraction mCherry<sup>198AZF</sup> emission. The calculated difference between start and finish emission spectra presented (**Black Dashed**), with calculated FRET efficiency of start and end presented in inset.

Spectra were then corrected to an initial 565 nm emission intensity of 1 A.U. in Microsoft Excel 2019, and the emission spectrum of mCherry<sup>198AZF</sup> (excited at 515 nm) was subtracted from each (Figure 6.16, B). Calculation of FRET efficiency (see Chapter 2.5.7.) of corrected spectra indicated a FRET efficiency of 0% at the start of incubation and 17% after incubation for 18 hours. This increase in FRET efficiency indicates that more of the chromophores of Cy3 DBCO and mCherry are aligned within 48.85 Å after incubation. Therefore, it can be concluded that mCherry<sup>198AZF</sup> is capable of both cycloaddition (Figure 6.14), and FRET (Figure 6.16) with Cy3 DBCO.

#### 6.2.8. SPAAC between mCherry<sup>198AZF</sup> and sfGFP SCO-K

The reaction of mCherry<sup>198AZF</sup> with sfGFP containing SCO-K incorporated at either residues 148 or 204 was performed by mixing of 100  $\mu$ M of each for 24 hours at room

temperature (See Chapter 2.4.6, for sfGFP purification). SDS PAGE of the sample, after incubation, indicated the presence of monomeric mCherry and sfGFP at ~27 kDa, in addition to two new protein forms sized at ~50 kDa and ~60 kDa (Figure 6.17, C and F). Neither of these protein forms was observed in monomeric samples and were only present where azide and alkyne had been mixed. It was concluded that the additional protein banding observed was indicative of dimerised mCherry and sfGFP, which had occurred at both residues 148 (Figure 6.17, C) and 204 (Figure 6.17, F) of sfGFP. Unfortunately, mass spectrometry of the sample was not considered at this stage; this analysis would have conclusively verified the triazole linkage of sfGFP and mCherry. The presence of two newly formed protein bands was reasoned to be an artefact of boiling of mCherry in SDS PAGE and was representative of dimeric protein, again cleaved at the mCherry chromophore <sup>414</sup>.



**Figure 6.17.** The Formation of Dimeric Protein by incubation of mCherry<sup>19AZF</sup> with sfGFP SCO.

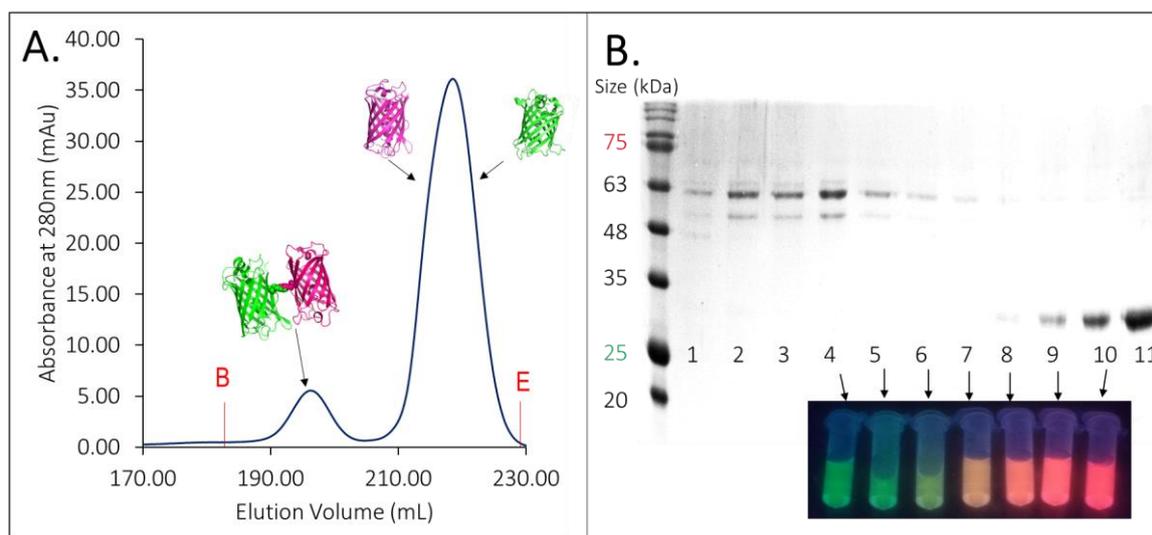
Combination of mCherry<sup>198AZF</sup> (B and D) with both sfGFP<sup>148SCO</sup> (A) and sfGFP<sup>204SCO</sup> (D) results in dimer formation (C and F). Dimeric protein is indicated at approximately 63 kDa by green boxes was once again formed of two bands implying cleavage of mCherry component during boiling in SDS. Analysis by ImageJ indicates percentage of dimeric protein formed from combination with sfGFP<sup>148SCO</sup> (C) accounts for 20% of total protein in lane; whilst dimer formed by combination with sfGFP<sup>204SCO</sup> accounts for 28% of protein in lane.

Dimeric protein observed, formed a smaller double band (~54 kDa and 63 kDa) when formed from sfGFP<sup>148SCO</sup> in comparison to dimer formed from sfGFP<sup>204SCO</sup> (~63 kDa and 68 kDa). This observation is consistent with Worthy (2018), whereby protein linkage at residue 148 would form a smaller band on SDS PAGE in comparison to protein which was linked at residue 204 <sup>431</sup>. This difference observed size was reasoned to be an effect in the modification of interface between the two molecules, with dimer formed at residue 148 being marginally more

compact than dimer formed at residue 204. As sfGFP<sup>204SCO</sup> has spectral characteristics more like sfGFP<sup>WT</sup> than sfGFP<sup>148SCO</sup> (see Chapter 1.3.6.), dimer formed from mCherry<sup>198AZF</sup> and sfGFP<sup>204SCO</sup> (GFPCH<sup>204-198</sup>) was used in favour of dimer formed from sfGFP<sup>148SCO</sup> (GFPCH<sup>148-198</sup>) in the subsequent analysis.

### 6.2.9. Separation of Monomeric Protein from GFPCH<sup>204-198</sup>

To characterise the functional properties of GFPCH<sup>204-198</sup>, the separation of dimer from monomer was first necessary. Separation of dimeric protein was achieved by SEC (see Chapter 2.4.2) in the replication of methodology applied by Worthy *et al.*, (2019)<sup>202</sup>. The sample was concentrated to a total volume of 2 mL and applied to Superdex 16/600 S200 size exclusion column at a flow rate of 1 mL/min. Two absorbance peaks eluted from the column at 190 mL and 210 mL elution volume (Figure 6.18, A). SDS PAGE indicated that the first elution peak was that of dimeric GFPCH<sup>204-198</sup> whilst the second peak was monomeric protein (Figure 6.18, B). Examination of fractions under UV light (Figure 6.18, B), while imprecise, provided an opportunity to assess the nature of the fractions quickly. Fractions which were shown to contain dimeric protein by SDS PAGE appeared to be green in colour and emitted only green light under UV light. In contrast, fractions which showed a mix of both dimer and monomer appeared to be coloured more orange.



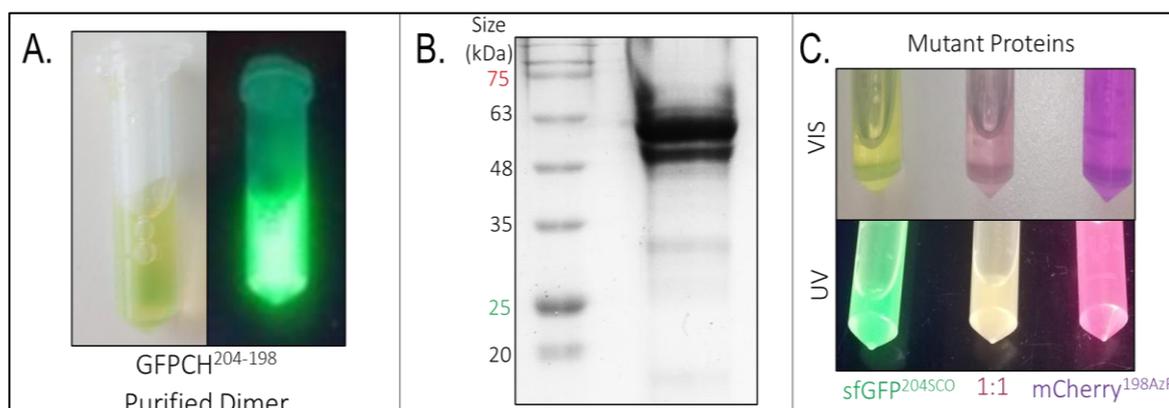
**Figure 6.18. Separation of GFPCH<sup>204-198</sup> from Monomeric Protein by SEC.**

**A**, UV Absorbance of two protein peaks eluted by size exclusion, one at 190 mL indicating dimeric elution and a second at 210 mL indicating monomeric elution. **B**, SDS PAGE of eluted fractions, with the same samples shown under UV light in insert.

### 6.2.10. Absorbance of GFPCH<sup>204-198</sup>

The observation that dimeric fractions were green was unexpected due to the presence of both green and red chromophores within GFPCH<sup>204-198</sup> (Figure 6.18, B). However, observation of concentrated GFPCH<sup>204-198</sup> confirmed only a slight red hue was present in purified GFPCH<sup>204-198</sup> (Figure 6.19, A). It was speculated that this might be due to the presence of free sfGFP monomer within the sample, yet, SDS PAGE and ImageJ analysis of GFPCH<sup>204-198</sup> indicated that the presence of monomer accounted for just 2% of total protein presence (Figure 6.19, B). Furthermore, comparison of fractions from the SDS PAGE indicated that mCherry monomer eluted before sfGFP (Figure 6.18, B) indicating that this 2% monomer presence is more likely to be mCherry monomer than sfGFP. Comparing GFPCH<sup>204-198</sup> (Figure 6.19, A) colouration to that of monomeric protein (Figure 6.19, C), indicates an evident change in spectroscopic properties of protein when dimerised. Dimeric protein resembled the colouration observed in sfGFP<sup>204SCO</sup> and did not match that observed from an equal concentration of mCherry with sfGFP (Figure 6.19, C).

For confirmation and quantification of the spectroscopic change observed, the absorbance spectrum of GFPCH<sup>204-198</sup> were measured (Figure 6.20). Measurement indicated the molar extinction coefficient of GFPCH<sup>204-198</sup> featured an absorbance peak at 587, this was later confirmed the presence of the mCherry chromophore (Figure 6.20, A). However, the spectrum further demonstrated that the absorbance properties of each chromophore present had been significantly altered upon dimerisation.



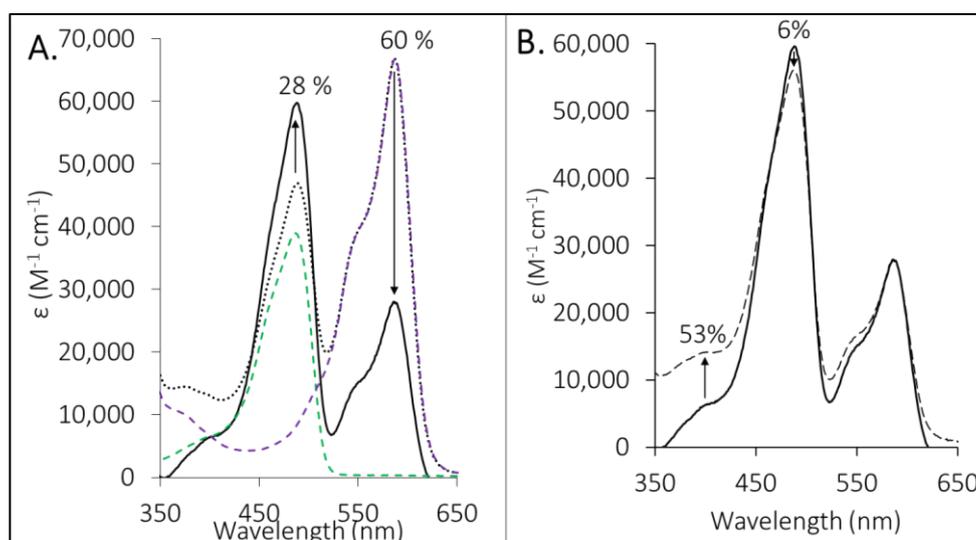
**Figure 6.19. Colour of Pure Dimer GFPCH<sup>204-198</sup> and Mixed Monomeric Proteins.**

**A**, appearance of dimer in both visible and UV light. **B**, SDS PAGE of GFPCH<sup>204-198</sup> Dimer, quantified to contain 2% monomeric presence. **C**, appearance of each monomer when separated and mixed at 1:1 ratio both in visible and UV light.

Comparison of dimeric absorbance to the sum absorbance of mCherry<sup>198AZF</sup> and sfGFP<sup>204SCO</sup> molar extinction coefficient indicated significant spectral change upon

dimerisation. The absorbance at 485 nm of dimeric GFPCH<sup>204-198</sup> increased 28% in comparison to the sum of both monomeric molar extinction coefficients, from 45,900 M<sup>-1</sup>cm<sup>-1</sup> to 57,300 M<sup>-1</sup>cm<sup>-1</sup>. Whilst the reverse was observed in absorbance at 587 nm whereby a decrease of 60% from 68,400 M<sup>-1</sup>cm<sup>-1</sup> to 29,923 M<sup>-1</sup>cm<sup>-1</sup> occurred on dimerisation. This change confirms the visual observations of GFPCH<sup>204-198</sup>; however, it does not indicate the reason for this change.

To explore if the change was stimulated by the fluorescence induced photobleaching of GFPCH<sup>204-198</sup>, the absorbance spectrum of a freshly purified dimer sample was compared before and after extended excitation measurement (20 repeat spectral readings) (Figure 6.20, B). Although a 6% decrease was observed in absorbance at 485 nm after fluorescence, there was overall very little change to absorbance associated with either chromophore (485 nm and 587 nm) after extended excitation. Instead, an increase in absorbance at 410 nm was observed (Figure 6.20, A). This increase is reasoned to be due to background noise, as it correlated with neither chromophore's absorbance.



**Figure 6.20. GFPCH<sup>204-198</sup> Molar Extinction Coefficient.**

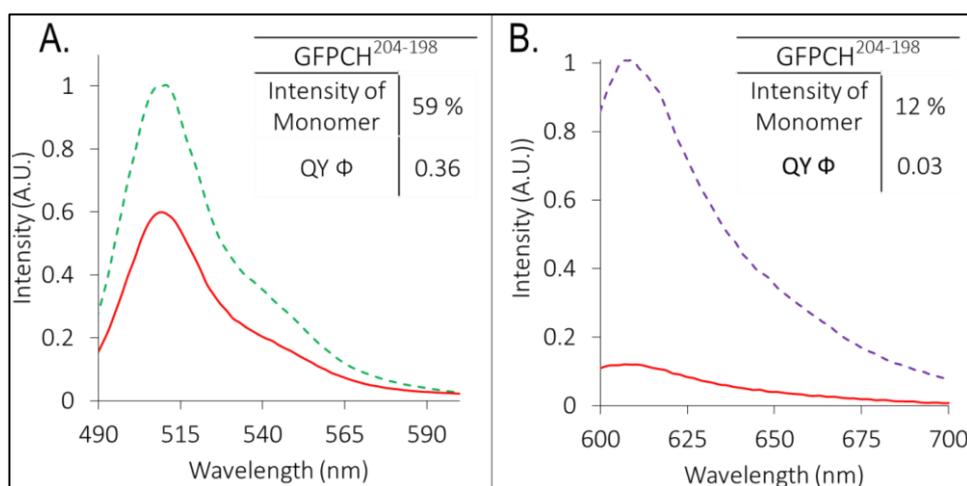
**A**, molar extinction coefficient of dimeric GFPCH<sup>204-198</sup> (**Black**), and both monomeric sfGFP<sup>204SCO</sup> (**Green Dashed**) and monomeric mCherry<sup>198AZF</sup> (**Purple Dashed**) and the sum of both (**Black Dots**). **B**, comparison of dimeric GFPCH<sup>204-198</sup> before (**Black**) and after extended excitation of chromophores (**Black Dashed Line**).

### 6.2.11. Fluorescence of GFPCH<sup>204-198</sup>

To ascertain if the changes observed in absorbance of GFPCH<sup>204-198</sup> were reflected in fluorescence, the dimeric protein was excited at both 485 nm and 587 nm and compared with monomeric protein (Figure 6.21). This comparison was performed as FRET was not expected

between monomeric protein as the distance between chromophores was expected to exceed the 10 nm radius in which FRET can occur. Excitation of 10  $\mu\text{M}$  GFPCH<sup>204-198</sup> at 485 nm yielded emission of light at 510 nm (Figure 6.21, A). The intensity of emission from GFPCH<sup>204-198</sup> was observed to be 60% the intensity of 10  $\mu\text{M}$  sfGFP<sup>204SCO</sup> monomer after excitation at 485 nm. However, the number of sfGFP chromophores are halved in GFPCH<sup>204-198</sup> as for a 10  $\mu\text{M}$  sample of dimer only 5  $\mu\text{M}$  of sfGFP is present. Therefore, the emission intensity of GFPCH<sup>204-198</sup> at 510 nm is 20% greater than expected when chromophore concentration is considered.

The opposite effect is observed in the mCherry chromophore (Figure 6.21, B). After excitation at 587 nm emission at 610 nm of 10  $\mu\text{M}$ , GFPCH<sup>204-198</sup> is equivalent to 12% of the recorded emission intensity of 10  $\mu\text{M}$  mCherry<sup>198AZF</sup>. Furthermore, when this value is doubled to account for chromophore intensity, emission intensity just 24%, that of monomeric mCherry<sup>198AZF</sup> is calculated. Consequently, the emission intensity of the mCherry chromophore is  $\frac{1}{4}$  the intensity of monomeric mCherry<sup>198AZF</sup>. Whilst the calculated quantum yield of GFPCH<sup>204-198</sup>'s mCherry chromophore at 610 nm (Figure 6.21, B) is less than 10% the value of the sfGFP chromophore at 510 nm at 0.03  $\Phi$  and 0.36  $\Phi$  respectively (Figure 6.21, A).



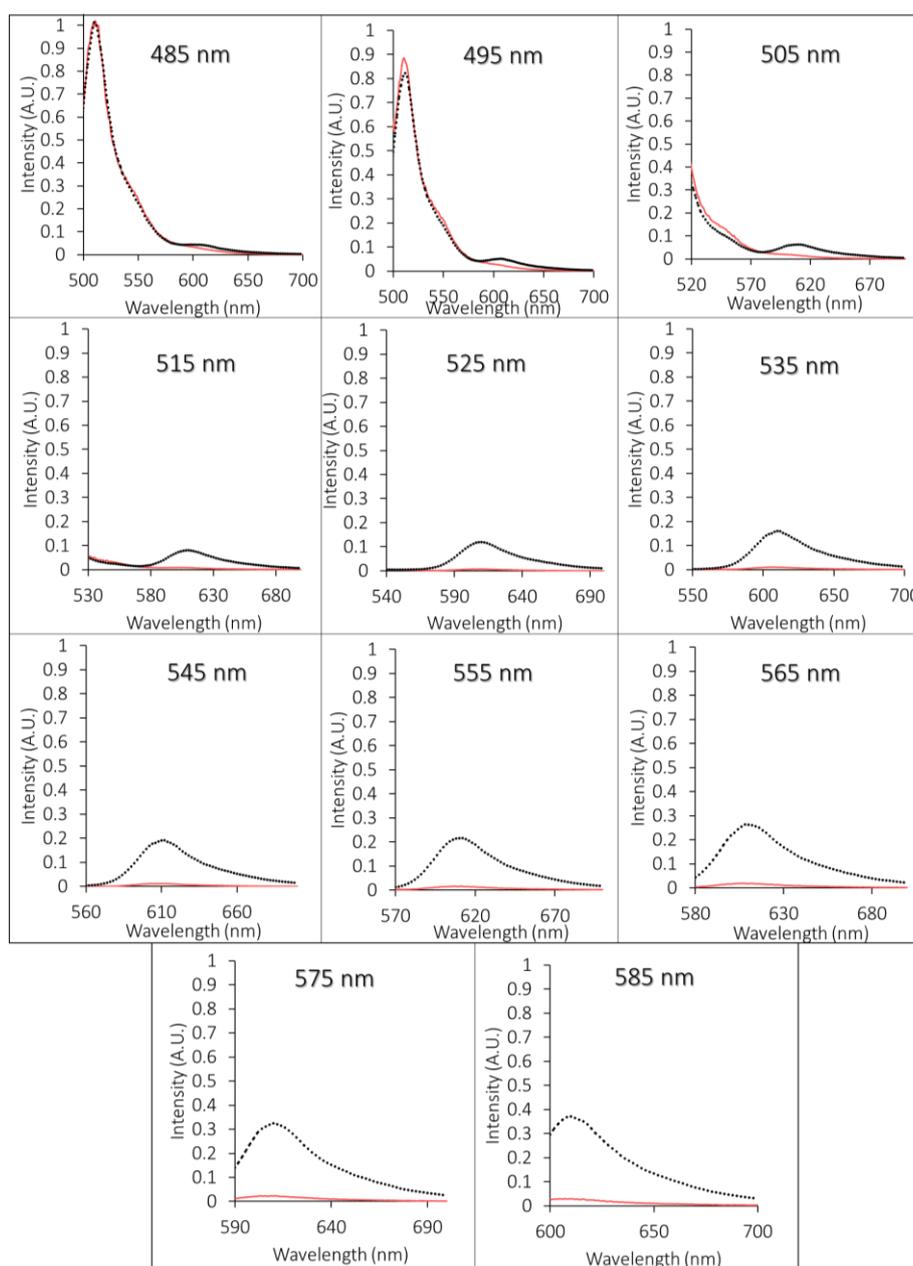
**Figure 6.21.** Emission of GFPCH<sup>204-198</sup> in Comparison to Monomeric sfGFP<sup>204SCO</sup> and mCherry<sup>198AZF</sup>.

GFPCH<sup>204-198</sup> (red) excited at 485 nm (A) and 587 nm (B). The emission of monomeric sfGFP<sup>204SCO</sup> (Green) and mCherry<sup>198AZF</sup> (Purple) shown as dashes. The calculated quantum yield of the dimer at each excitation wavelength shown as insert. Datasets corrected to monomeric intensity of 1.

To assess if the decrease of emission at 610 nm was conserved across all wavelengths of excitation, dimeric GFPCH<sup>204-198</sup> emission was recorded at 10 nm intervals between 485 - 585 nm and compared with the emission of a non-linked mix of monomeric protein. Protein was excited at 10 nm intervals, and emission spectra were compared after normalisation to an

initial emission intensity of 1 A.U. after excitation at 485 nm (Figure 6.22). It was observed that GFPCH<sup>204-198</sup> was able to emit light at 485 nm upon excitation at 485 nm, 495 nm, and 505 nm. However, a severely reduced emission at 610 nm was observed across all wavelengths of excitation, and only an emission intensity of 0.02 A.U. was observed upon excitation at 587 nm, in comparison to the emission of 0.26 of monomeric protein.

The lack of emission at 610 nm is consistent with both the observed colour and absorbance of GFPCH<sup>204-198</sup> (Figures 6.19, and 6.20). The 88% decrease in emission intensity of



**Figure 6.22. Emission of GFPCH<sup>204-198</sup> and Monomer Mix** GFPCH<sup>204-198</sup> (Red) and combined sample of *sfGFP<sup>204SCO</sup>* and mCherry<sup>198AZF</sup> (Dotted Line) after excitation at 10 nm intervals between 485 nm and 585 nm. Data corrected to a 485 nm intensity of 1.

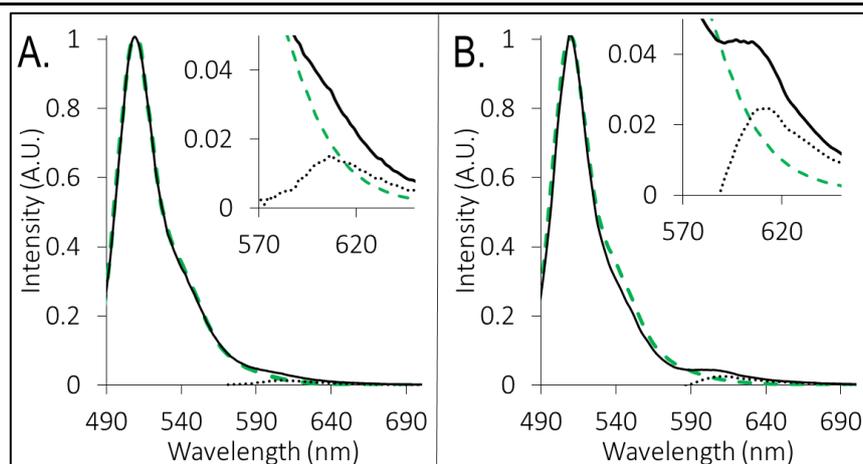
the mCherry chromophore in GFPCH<sup>204-198</sup> after excitation at 587 nm (Figure 6.22) is greater than the decrease in absorbance of 60% observed previously (Figure 6.20). This contrasts the 20% increase in emission at 510 nm and the 28% increase in absorbance at 485 nm of the sfGFP chromophore in GFPCH<sup>204-198</sup>.

The exact stimulus for each chromophore's functional change is unknown. However, results indicate that following dimerisation, the green chromophore of GFPCH<sup>204-198</sup> has a greater capacity to interact with light whilst the red chromophore has a reduced capacity to interact with light. Although the difference likely observed is exaggerated by the initial quantum yield of mCherry<sup>198AZF</sup> being just 32% that of sfGFP<sup>204SCO</sup>.

The decreased emission at 610 nm may be a result of internal quenching (see Chapter 1.3.8.) of light within dimeric GFPCH<sup>204-198</sup>. A form of quenching, will be explored in detail in Chapter 7, wherein excitation energy during fluorescence is transduced from the form of light energy to chemical, and dissipated within the protein<sup>435</sup>. Fluorescence quenching is a frequent observation in many haem binding proteins, wherein the intrinsic fluorescence of amino acids is lost as electronic excitation is dissipated across the haem porphyrin ring and resulting in the quenching of light as the p orbital is sparsely populated with electrons<sup>150</sup>. However, there is no known record of energy quenching by mCherry; and this explanation does not account for the characteristics which are retained in the dimer, such as, emission at 510 nm (Figure 6.21).

#### 6.2.12. FRET of GFPCH<sup>204-198</sup>

This investigation aimed to stimulate an increase in FRET between dimeric fluorescent proteins. Therefore, the efficiency of FRET from the sfGFP to the mCherry chromophore in GFPCH<sup>204-198</sup> was quantified. Excitation of GFPCH<sup>204-198</sup> at 485 nm was expected to stimulate energy transfer between the chromophore pair, which would result in emission at 610 nm if the pair were aligned within 52.89 Å (Table 6.3). As the predicted interface between protein was expected to align chromophores within 18 Å, it was anticipated that when dimerised energy transfer would be observed, if reduced in intensity, between chromophores; especially as FRET was previously observed in the combination of Cy3 DBCO and the mCherry chromophore when linked at 198 AZF (Figure 6.16).



**Figure 6.23. Comparison of Emission Spectra of Dimeric GFPCH<sup>204-198</sup> and Monomeric sfGFP<sup>204SCO</sup> and mCherry<sup>198AZF</sup>.**

**A**, emission of GFPCH<sup>204-198</sup> (**Black**), in comparison to sfGFP<sup>204SCO</sup> (**Green Dashed**), and mCherry with difference between the spectra shown in insert (**Black Dots**). **B**, emission of sfGFP<sup>204SCO</sup> and mCherry<sup>198AZF</sup> mix (**Black**) at excitation 485 nm in comparison to just monomeric sfGFP<sup>204SCO</sup> (**Green Dashed**) with difference between the spectra shown in insert (**Black Dots**).

To assess if energy transfer observed between chromophores was enhanced by proximity emission of 10  $\mu\text{M}$  of both GFPCH<sup>204-198</sup> and a mixed monomeric protein were compared. Emission of each sample was recorded after excitation at 485 nm, and spectra were corrected to a 510 nm emission intensity of 1 A.U. for comparison (Figure 6.23). Subtraction of the emission profile of monomeric, non-mixed, sfGFP<sup>204SCO</sup>, was then performed to calculate the proportion of signal which was generated from the mCherry chromophore alone.

A minor emission peak was observed at 610 nm in both GFPCH<sup>204-198</sup> (Figure 6.23, A), and the mixed monomeric protein sample (Figure 6.23, B). This peak was most distinct in the mixed monomer sample with a reported intensity of 0.024 A.U. upon sfGFP<sup>204SCO</sup> subtraction. Whilst emission at 610 nm from dimeric GFPCH<sup>204-198</sup> was 0.014 A.U., when sfGFP<sup>204SCO</sup> spectrum was subtracted. Emission at 610 nm indicates that mCherry is stimulated to a small degree by sfGFP, when both mixed in solution and when dimerised. However, the increased proximity of chromophores in GFPCH<sup>204-198</sup> did not increase energy transfer. This difference indicates that there is greater functional communication between chromophores when protein is free in solution, and not within the dimeric protein.

The observation that decreased distance did not increase mCherry emission contradicts initial expectation as enhanced chromophore proximity was not expected to decrease energy transfer. Without an increase in emission at 610 nm, it is not likely that FRET

has occurred, despite the forced proximity of chromophores. Therefore, FRET efficiency between sfGFP and mCherry is not enhanced upon dimerisation. The principles of FRET indicate that energy transfer is inversely proportional to the chromophore distance, orientation, and their spectral overlap<sup>436</sup>. As the excitation and emission of each protein are consistent when protein is free in solution and when dimerised, it is argued that a lack of spectral overlap does not inhibit FRET. Furthermore, as the average distance between the chromophores is certain to decrease with dimerisation, it is additionally unlikely that FRET is hindered by chromophore distance. It is feasible, however, that the orientation at which chromophores are aligned within GFPCH<sup>204-198</sup> is not conducive to FRET, but this does not account for the lack of emission at 610 nm when the mCherry chromophore is excited directly.

The significant disruption to mCherry excitation observed previously (Figure 6.21) may indicate why no increase in emission at 610 nm was observed in dimeric GFPCH<sup>204-198</sup> (Figure 6.23, A). It is reasoned that as the mCherry chromophore is a less efficient at excitation when dimerised (Figures 6.20. and 6.21.), the enhanced distance between chromophores will not result in a detectable increase in energy transfer. As mCherry is frequently used as a fluorescent protein tag in FRET studies<sup>437-439</sup>, the internal changes which result in a decrease in chromophore function, that occurred upon oligomerisation, is significant and must be explored further.

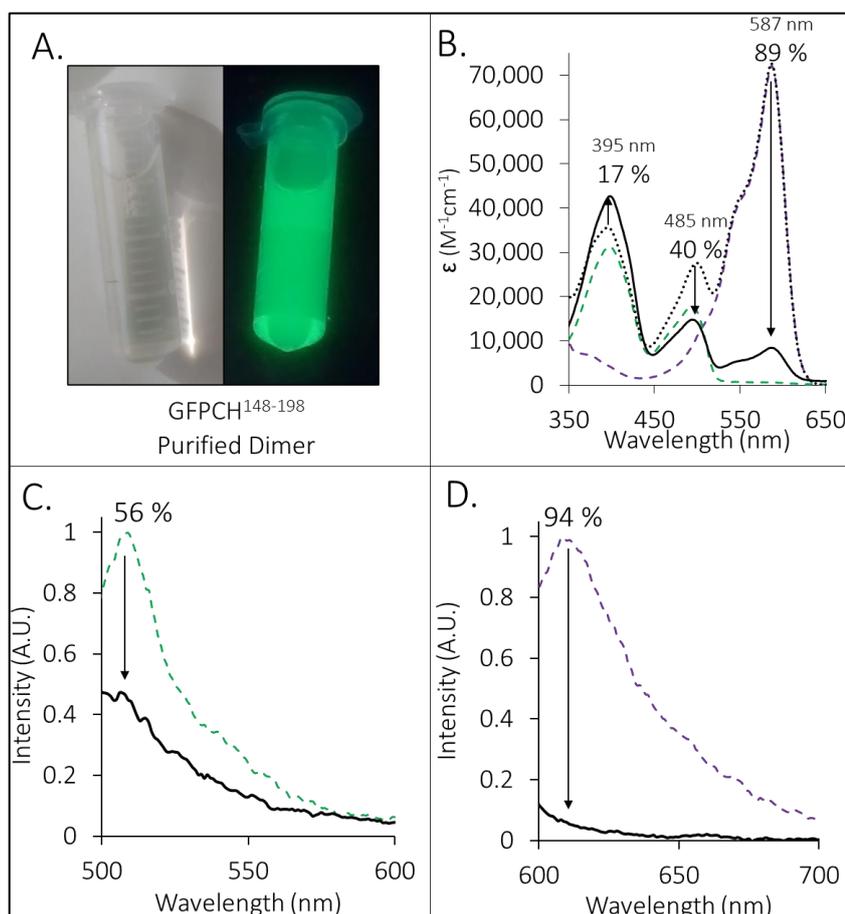
### 6.2.13. Functional Characterisation of GFPCH<sup>148-198</sup>

As cross-linking of mCherry<sup>198AZF</sup> to sfGFP<sup>204SCO</sup> reduced the capacity of the mCherry chromophore to interact with light (Figures 6.20 and 6.21), it was later assessed if these observations were replicated when mCherry was cross-linked with sfGFP<sup>148SCO</sup>. For this, the dimer formed in 6.17, C was purified as described previously (see Chapter 6.2.9) and both the absorbance and emission spectra of GFPCH<sup>148-198</sup> were recorded.

As sfGFP<sup>148SCO</sup> unlike sfGFP<sup>204SCO</sup>, favours protonation of Y66 within the chromophore (see Chapter 1.3.6.) two distinct excitation maxima (395 nm and 492 nm) exist in sfGFP<sup>148SCO</sup>. Previous study of sfGFP dimerisation at residue 148 by Worthy *et al* (2019), indicated that the formation of the dimeric triazole link at residue 148 produced conditions favourable to deprotonation of Y66, which increased excitation observed at 485 nm, whilst decreasing excitation at 395 nm<sup>202</sup>. It was anticipated that dimerisation of sfGFP<sup>148SCO</sup> with mCherry<sup>198AZF</sup>

would stimulate this same effect and de-protonation of the sfGFP chromophore, would additionally be favoured when sfGFP<sup>148SCO</sup> was cross-linked with mCherry<sup>198AZF</sup>.

The colouration of purified dimeric GFPCH<sup>148-198</sup> (Figure 6.24, A) was consistent with colour observed in GFPCH<sup>204-198</sup> (Figure 6.19), although less saturated. Pure protein was once again shown to have lost all red colouration associated with the presence of mCherry and did not visually appear to emit red light (Figure 6.24, A). The colouration of pure GFPCH<sup>148-198</sup> was less intense than that of GFPCH<sup>204-198</sup>. This decrease in observed colouration was expected due to the lower molar absorption coefficient of sfGFP<sup>148SCO</sup> (17,300 M<sup>-1</sup>cm<sup>-1</sup> at 485 nm) in comparison to sfGFP<sup>204SCO</sup> (39,800 M<sup>-1</sup>cm<sup>-1</sup> at 485 nm).



**Figure 6.24. Characteristics of GFPCH<sup>148-198</sup>.**

**A**, visual appearance of GFPCH<sup>148-198</sup> in both visible and UV light. **B**, molar extinction coefficient changes of GFPCH<sup>148-198</sup> (**Black**) in comparison to sfGFP<sup>148SCO</sup> (Green Dashed), mCherry<sup>198AZF</sup> (Purple Dashed), and the sum of both monomers (**Black Dots**). **C**, emission of GFPCH<sup>148-198</sup> (**Black**) after excitation at 485 nm in comparison to monomeric sfGFP<sup>148SCO</sup> (**Green Dashed Line**). **D**, emission of GFPCH<sup>148-198</sup> (**Black**) after excitation at 587 nm in comparison to monomeric mCherry<sup>198AZF</sup> (**Purple Dashed Line**).

The molar extinction coefficient of GFPCH<sup>148-198</sup> decreased by 89% at 587 nm from 71,700 M<sup>-1</sup>cm<sup>-1</sup> to 8,200 M<sup>-1</sup>cm<sup>-1</sup> at 587nm upon dimerisation (Figure 6.24, B), and was far greater than the 60% loss in absorbance observed at 587 nm in GFPCH<sup>204-198</sup> (Figure 6.20). In GFPCH<sup>148-198</sup>, however, unlike GFPCH<sup>204-198</sup> an additional absorbance loss of 40% at 485 nm was observed from 22,600 M<sup>-1</sup>cm<sup>-1</sup> to 13,500 M<sup>-1</sup>cm<sup>-1</sup>. While an increased molar extinction coefficient of 17% from 35,200 M<sup>-1</sup>cm<sup>-1</sup> to 41,200 M<sup>-1</sup>cm<sup>-1</sup> was observed at 395 nm.

The decrease in absorbance at 485 nm contrasts observation by Worthy *et al.* (2019) which reported an increase in absorbance at 485 nm when sfGFP forms a dimeric linkage at residue 148. This increase in function is rationalised to be a result of a conformational change in residue 148 upon dimerisation. This change in residue 148 then allows for the formation of a long-range symmetrical water channel to form between monomers, functionally connecting each chromophore and promoting de-protonation switching the absorbance spectrum from a  $\lambda_{Max}$  at 395 nm to 485 nm<sup>202</sup>. However, in the instance of dimeric GFPCH<sup>148-198</sup> the increase by 17% at 395 nm was unexpected as it goes against the trend observed by Worthy *et al.* (2019), but it could be indicative that the protonated state of sfGFP is favoured when linked to mCherry<sup>198AZF</sup>.

Excitation of 10  $\mu$ M GFPCH<sup>148-198</sup> at 485 nm (Figure 6.24, C), indicates a decrease of 56% emission intensity of GFPCH<sup>148-198</sup> in comparison to monomeric sfGFP<sup>148SCO</sup>. An emission intensity of 0.46 A.U. was reported for GFPCH<sup>148-198</sup>, in comparison to the 1 A.U. observed from sfGFP<sup>148SCO</sup> at equal concentration. Doubling of GFPCH<sup>148-198</sup> emission at 510 nm to account for chromophore concentration, yields an 8% decrease in emission intensity to 0.92 A.U. when dimerised. This decrease further supports the hypothesis that protonation of sfGFP<sup>148SCO</sup> chromophore is favoured when dimerised with mCherry<sup>198AZF</sup>, as fewer molecules of sfGFP would be excited at 485 nm when protonation is favoured.

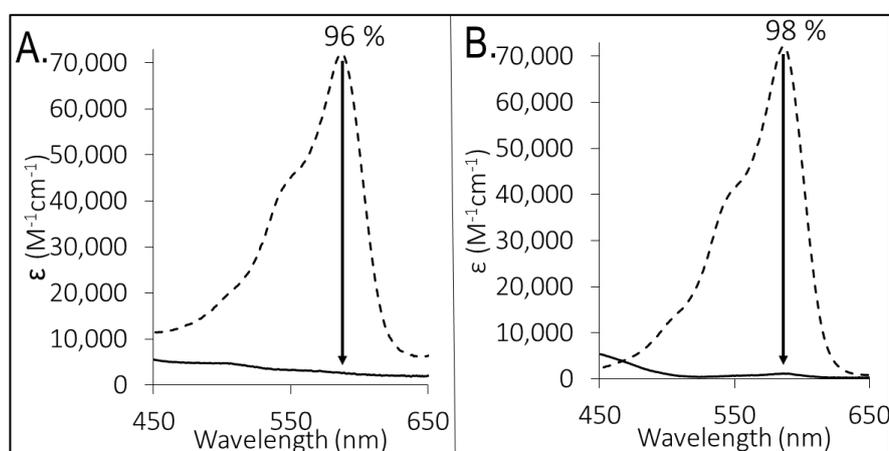
Furthermore, GFPCH<sup>148-198</sup> emission at 610 nm appeared to have been close to eliminated after excitation at 587 nm (Figure 6.24, D). A decrease of 94% of emission intensity was observed upon excitation at 587 nm; however, due to the indistinct 610 nm emission peak, emission may not be stimulated by the mCherry chromophore, and may instead be a result of background noise from sfGFP excitation. The drastic reduction of emission at 610 nm of GFPCH<sup>148-198</sup> is consistent with observations of GFPCH<sup>204-198</sup> and is further evidence of

significant change to the mCherry chromophore when dimerised. Consistent with GFPCH<sup>204-198</sup>, GFPCH<sup>148-198</sup> appears green in colour, has an 89% decrease in molar extinction coefficient at 587 nm, and a 94% decrease in emission at 610 nm after excitation at 587 nm (Figure 6.24).

#### 6.2.14. Unfolding of mCherry<sup>198AZF</sup>

One possible explanation for the loss of both excitation and emission of the mCherry chromophore is unfolding of the mCherry  $\beta$ -barrel upon the formation of dimeric protein. The unfolding of the  $\beta$ -barrel would expose the internal chromophore to the environment and thus would be fluorescently quenched by the solvent. To test this mCherry<sup>WT</sup> was unfolded by both chemical and thermal denaturation for comparison to the loss of function observed in both GFPCH<sup>204-198</sup> and GFPCH<sup>148-198</sup> (Figure 6.25).

Firstly 10  $\mu$ M mCherry<sup>WT</sup> was prepared fresh in Tris 50 mM pH 8 was unfolded by heating to 100 °C for 10 minutes (Figure 6.25, A). Absorbance spectra were recorded before and after the application of heat and corrected to the molar extinction coefficient of mCherry<sup>WT</sup> (Figure 6.25, A). A loss of absorbance was observed upon heating, and the thermally denatured protein exhibited a 96% decrease in absorbance intensity at 587 nm, decreasing from 72,000 M<sup>-1</sup>cm<sup>-1</sup> to 2,600 M<sup>-1</sup>cm<sup>-1</sup>. Furthermore, as no distinct absorbance maximum was observed at 587 nm, absorbance observed was likely due to background from the unfolded protein and is not indicative of the mCherry chromophore presence.



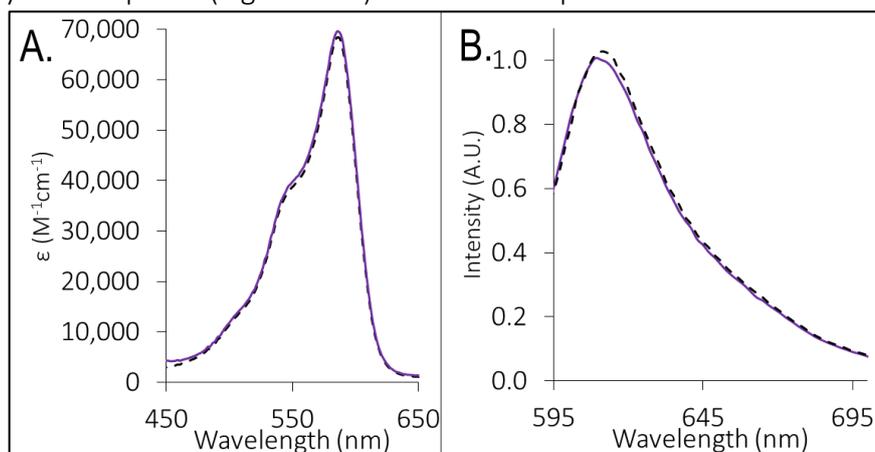
**Figure 6.25. Absorbance Spectra of Unfolded mCherry<sup>WT</sup>.**

Protein unfolded by boiling (A), and protein which was denatured by buffer exchanging into 6M guanidinium for 24 hours (B). Sample before denaturation (Dashed Line) and after denaturation (Black) with the loss of absorbance at 587 nm labelled as a percentage.

Next, 10  $\mu\text{M}$  mCherry<sup>WT</sup> was buffer exchanged into 6 M Guanidinium Hydrochloride and incubated at room temperature for 24 hours, for chemical denaturation. Absorbance at the end of 24 hours was again representative of a large-scale loss of mCherry chromophore. Absorbance at 587 nm was observed to decrease 98% from 72,000  $\text{M}^{-1}\text{cm}^{-1}$  to 1,200  $\text{M}^{-1}\text{cm}^{-1}$ . The decrease in absorbance of 98% did indicate that  $\sim 2\%$  of mCherry<sup>WT</sup> had not unfolded in 6M Guanidinium Hydrochloride.

Each method of protein unfolding, both thermal and chemical, applied to mCherry<sup>WT</sup> stimulated almost complete absorbance loss at 587 nm. It was reasoned that as dimeric mCherry still featured absorbance at 587 nm (Figures 6.20, and 6.24), mCherry was not denatured in the formation of a triazole linkage with sfGFP. As the loss of absorbance at 587 nm was not due to unfolding, it was next investigated if the loss of absorbance was triggered by reaction between 198 AZF and SCO-K to form a triazole linkage.

Fresh mCherry<sup>198AZF</sup> was prepared to a 10  $\mu\text{M}$  concentration and incubated with a 20 x molar excess of SCO-K at room temperature for 24 hours, to replicate dimerisation conditions applied previously (Chapter 6.2.8). The absorbance of mCherry<sup>198AZF</sup> after incubation showed an increase in absorbance of 1.5% at 587 nm (Figure 6.26, A), and an emission decrease at 610 nm of 2 % (Figure 6.26, B). Therefore, the reaction between 198 AZF and SCO-K does not initiate the functional loss of the mCherry chromophore, and functional loss observed in dimeric protein must be specific to the combination of protein. These findings match previous observation that cycloaddition between 198 AZF and DBCO resulted in no functional change of the mCherry chromophore (Figure 6.16). As 198 AZF is proven to react with both DBCO (Figure



**Figure 6.26. Reaction of mCherry<sup>198AZF</sup> with a 20 x Molar Excess of SCO-K.** Absorbance (A) and emission (B) of mCherry<sup>198AZF</sup> before (Black Dashed Line) and after (Purple) 24- hour incubation with 20 x excess of SCO at room temperature.

6.16) and SCO-K (Figure 6.26) without functional change, this indicates that formation of a triazole linkage itself does not prevent the mCherry chromophore from interacting with light.

---

## 6.3.1. Discussion

### 6.3.1. Explanation of mCherry Chromophore Functional Loss

Observation consistent in the analysis of both GFPCH<sup>204-198</sup> and GFPCH<sup>148-198</sup> was the significantly reduced absorbance at 587 nm and emission at 610 nm (Figures 6.20, 6.21 and 6.24). Absorbance at 587 nm was observed to decrease by 60% in GFPCH<sup>204-198</sup> and 89% in GFPCH<sup>148-198</sup>. At the same time emission at 610 nm was reduced 88% in GFPCH<sup>204-198</sup> and 94% in GFPCH<sup>148-198</sup>. Changes in the emission of each dimer are shown to be greater than changes to absorbance, which is indicative that the mCherry chromophore is more capable of absorbing light than emitting it when dimerised. These results contradict expectation as it was initially hypothesised that linking of both FPs would increase functional communication *via* processes such as FRET and result in a more significant emission of protein at 610 nm on excitation at a wavelength optimal for sfGFP.

These changes are consistent with each generated dimer as the mCherry chromophores in both GFPCH<sup>204-198</sup> and GFPCH<sup>148-198</sup> have compromised function and show no internal communication between chromophores. Whereas the sfGFP chromophore shows enhanced absorbance at 485 nm (Figure 6.20) and emission at 510 (Figure 6.21) in GFPCH<sup>204-198</sup>, but enhanced absorbance at 395 nm in GFPCH<sup>148-198</sup> (Figure 6.14). These results in combination, contrast the previously observed functional enhancement by SPAAC linkage of sfGFP homodimers<sup>202</sup>, as the mCherry chromophore appears to be functionally weakened upon dimerisation

The exact cause of functional impairment to mCherry is unclear, although the change appears to be specific to sfGFP dimerisation. The function of the mCherry<sup>198AZF</sup> chromophore is consistent with mCherry<sup>WT</sup> and is retained after: nnAA incorporation (Figure 6.12), UV irradiation (Figure 6.13), reaction with DBCO (Figure 6.16), and SPAAC with SCO-K (Figure 6.26). Additionally, functional communication between mCherry<sup>198AZF</sup> and Cy3 DBCO is enhanced after cycloaddition by way of FRET (Figure 6.16). It is only after the forced proximity of the mCherry<sup>198AZF</sup> chromophore to sfGFP at both residues 148 and 204 SCO that functional loss is observed in mCherry (Figures 6.20, 6.21, and 6.24).

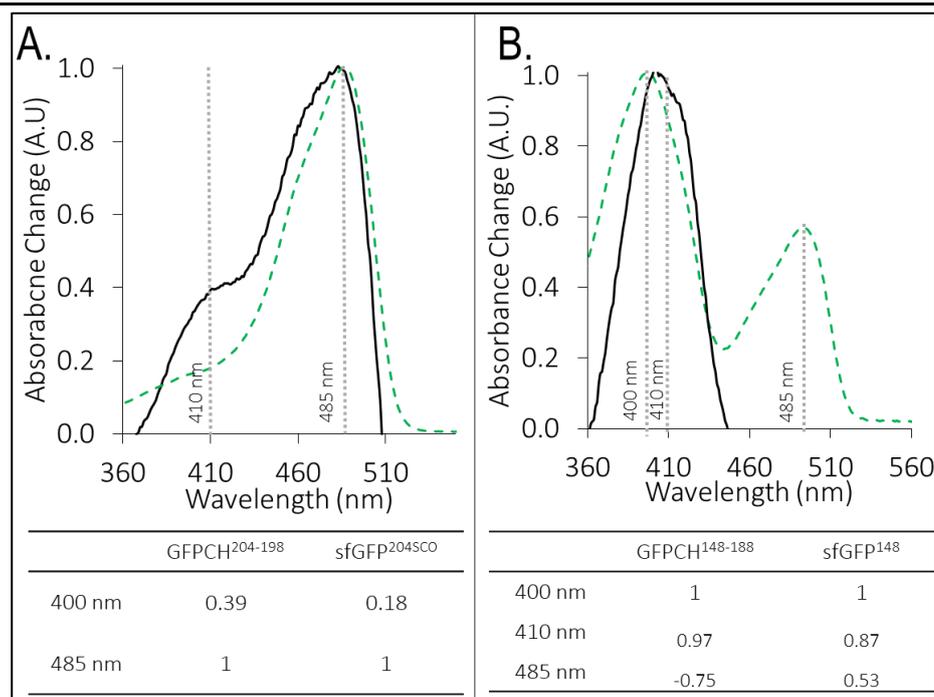
It is possible that the unfolding of mCherry may stimulate some decrease in absorbance observed. However, it is unclear why the decrease is not consistent in each dimer, and why the

function of the green chromophore within each protein appears to have been enhanced in tandem (Figures 6.20, 6.24 and 6.26). Absorbance at 485 nm was increased in GFPCH<sup>204-198</sup> (Figure 6.20), whilst emission at 510 nm was also shown to increase by 20% (Figure 6.21) indicating no impairment to function of sfGFP<sup>204SCO</sup> upon dimerisation. Whereas absorbance of GFPCH<sup>148-198</sup> was shown to increase by 22% at 382 nm, and emission 510 nm of GFPCH<sup>148-198</sup> was consistent with monomeric sfGFP<sup>148SCO</sup> indicating protonation of the sfGFP chromophore rather than a loss in function.

Although it is impossible to conclusively establish the exact cause of functional loss without further investigation, results from several alternative studies may go some way to an explanation. It has been observed previously by Wiens *et al.*, 2016, that the forced proximity of green and red tdTomato variants resulted in the loss in function of either one or both chromophores<sup>407</sup>. The source of this function was reasoned in this paper to be the failure of chromophore formation within the DsRed subunit. As both mCherry and tdTomato are variants of DsRed, the mechanism of this switch could be a potential cause for results observed in this chapter<sup>180</sup>. However, this study recorded a loss of both green and red chromophore functions, which not observed during this experiment.

An alternative explanation is that the inherent fluorescent function of the red chromophore is the limiting factor here rather than any GFP specific interaction. Prangma *et al.* (2020) postulated that the decreased quantum yields of the Red FPs are a result of the longer time spent in a dark state of the red chromophore<sup>440</sup>. The longer dark state of each chromophore arises from the increased size in the conjugated double bond network which results in longer relaxation time and a reduced fluorescence lifetime of each chromophore. If the mCherry chromophore is switched to a more prolonged dark state by dimerisation this could enable the emission reduction observed. However, this does not explain the additional change in absorbance observed and why this effect is not present mCherry is linked with Cy3 DBCO.

Subtraction of monomeric molar extinction coefficient from the absorbance of each dimeric protein offers the most consistent explanation for change observed (Figure 6.27). Absorbance at 410 nm was shown to increase in GFPCH<sup>204-198</sup> (Figure 6.27, A) and GFPCH<sup>148-198</sup> (Figure 6.27, B) and resultantly constituted a more significant proportion of signal in each

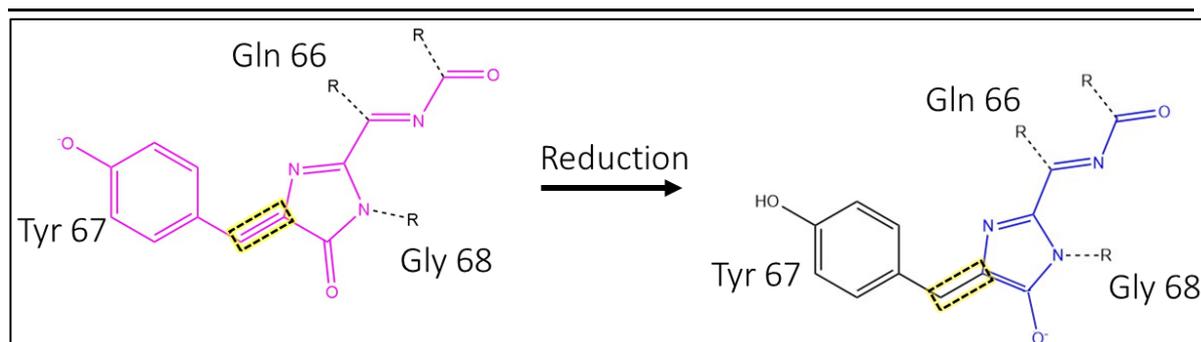


**Figure 6.27. Absorbance of Dimeric Protein at Green Wavelengths in Comparison To Expected Characteristics Of Each Monomeric sfGFP.**

Absorbance spectra of GFPCH<sup>204-198</sup> (A) and GFPCH<sup>148-198</sup> (B) after subtraction of the molar extinction coefficients of each monomeric counterpart and correction of  $\lambda_{MAX}$  to 1 A.U. Absorbance of GFPCH<sup>204-198</sup> (A) indicates a greater absorbance at 410 nm in dimeric protein (Black) in comparison to absorbance of sfGFP<sup>204SCO</sup> (Green Dashed Line). Whilst absorbance of GFPCH<sup>148-198</sup> (B) indicates that dimeric protein (Black) shifts in absorbance from  $\lambda_{MAX}$  of 400 nm towards longer wavelengths of light in comparison to monomeric sfGFP<sup>148SCO</sup> (Green Dashed Line) with the absorbance at 410 nm shown to increase in proportion to absorbance at 400 nm.

dimeric protein in comparison to monomeric components. This effect was unexpected and has not been observed previously in the work of SPAAC induced fluorescent protein dimerisation<sup>202,275,431</sup>. However, the effect has been observed in the previous study of DsRed protein in reducing conditions and is reminiscent of the green chromophore of tetrameric DsRed<sup>420,441</sup>.

Work by Cloin *et al.* (2017), demonstrated that the oxidation of the chromophore of mCherry is also influential to both absorbance and emission<sup>441</sup>. This study highlighted that when the chromophore of mCherry is reduced, it can lead to a photo-switching event which was dubbed “greening.” Greening of mCherry occurs by the reduction of the double bond between C $\alpha$  and C $\beta$  of the tyrosine of Y67 (Figure 6.28) and mediates a decreased absorbance at 587 nm and an increased absorbance at 410 nm, as was observed in dimeric protein. This effect is observed to an extent in Figure 6.27, wherein absorbance at 410 nm in both dimers is measured to be greater than that predicted of the sfGFP chromophore alone.



**Figure 6.28. Reduction of mCherry Chromophore Observed in by Cloin *et al* (2017).**

Reduction of the double bond between the C $\alpha$  and C $\beta$  of Y67 and indicated by green and yellow box. Image created in ChemBiodraw.

Cloin *et al.* generated this switch by incubation of protein with  $\beta$ -mercaptoethanol, however, it is feasible that direct molecular attachment of sfGFP to mCherry, induces a comparable reduction in the mCherry chromophore. This effect is also observed in a series of alternative red FPs derived from DsRed, mKate and HcRED in a study by Subach *et al.* (2012), wherein chromophores were reported to have lower propensity to absorb light at 587 nm and increased emission at 460 nm<sup>420</sup>.

Photo-switching is a well-established process across many fluorescent proteins; for example, the red-switching of GFP has been previously reported<sup>442–444</sup>. It is reasoned that the formation of dimeric protein may trigger similar switching within the mCherry by reduction of the chromophore upon dimerisation. Figure 6.27 indicates an increase in absorbance at 410 nm observed from both dimeric GFPCH<sup>148-198</sup> and GFPCH<sup>204-198</sup>. This increase in absorbance is minimal in comparison to the reduction in absorbance at 587 nm. However, this is consistent with the findings by Cloin *et al.* (2017) which indicated the loss of absorbance at 587 nm was disproportionately more extensive than the gained absorbance at 410 nm.

This effect is a possible explanation for the change observed in the dimerisation of mCherry with sfGFP (Figure 6.20 and 6.24), wherein, dimerisation yields an increase in absorbance at 410 nm (Figure 6.27). It is possible that if the mCherry chromophore were reduced this would result in the apparent loss of function due to the masking of signal from “greened” mCherry chromophore by the more intense spectral properties of sfGFP at overlapping wavelengths. Although this explanation cannot be proven based upon evidence obtained so far it does go the furthest in explaining all results observed. Thus this explanation

forms the most likely hypothesis moving forward to future investigation with more protein variants.

### 6.3.2. Functional Communication between Chromophores

This chapter aimed to develop a novel SPAAC induced partnership of sfGFP and mCherry. SPAAC linkage of mCherry and sfGFP generated both GFPCH<sup>204-198</sup> and GFPCH<sup>148-198</sup>; although chromophore communication within each was not as achieved due to compromised function observed in the mCherry component. As the utility of a fluorescent protein is dependent on both brightness and overall photostability<sup>445</sup> it is not feasible to develop this dimeric system further as a fluorescent probe. However, this work does raise several important questions as to the effect proximity induces on the function of fluorescent proteins.

Initial expectations were that the covalent linkage of two fluorescent proteins would enhance functional communication. However, findings suggest that FRET cannot be achieved due to the impaired function observed in the mCherry upon dimerisation (Figure 6.21, and 6.24). The most plausible explanation for this loss of function is dimerisation dependent photo-switching of the mCherry chromophore by reduction (Figure 6.28). Consequently, green shifting would reveal characteristics reminiscent of tetrameric DsRED which comprises of two mature red chromophores and two immature green chromophores and may indicate a reduction, and photo-switching, is triggered by oligomerisation. Thus, it would suggest that the green chromophore present in DsRED is conserved in mCherry and is likely present in other DsRed derived FPs, in the right conditions<sup>407,423,437,440,441</sup>. This could be investigated further in future by the attachment of mCherry<sup>198AZF</sup> several different proteins to determine if the effect is sfGFP specific and if a redox sensing protein such as Cytochrome *b*<sub>562</sub> is also reduced in the process of dimerisation.

This study together with that of Worthy *et al.*<sup>202</sup> may indicate that how we interpret the emission of various FPs in FRET analysis, is not reflective of all interactions which can arise. FRET is generally considered to be a passive process between two FPs physically which are separated, however, as both GFP and DsRed have a strong tendency to dimerise at high concentration, FP self-association may have significance to interpretation of FRET data. When fused to partner proteins that have a natural tendency to interact, this could potentially

increase the local concentration of the FPs and bring them closer together in space, thus promote their interaction.

Thus, some effects observed in FRET measurement may be reflective of the underlying artefacts of the FPs interaction, as observed in this study, and not reflective of the proximity or orientation between chromophores. Furthermore, if the change in mCherry chromophore function observed in this study is induced by oligomerisation then it is highly significant due to the frequent use of mCherry as a fluorescent partner probe <sup>423</sup>. The next step of this investigation will be to crystallise protein and obtain definitive evidence of the chromophore environment of mCherry<sup>198AZF</sup> when both monomeric and dimeric.

### 6.3.3. Conclusion

Oligomerisation is currently an emerging field in protein engineering and *de novo* protein design. Results presented offer an indication that the formation of an oligomeric interface is both feasible and predictable between the two fluorescent proteins. Oligomerisation demonstrated a consistent approach for the generation and purification of more structurally complex dimers and will be applied once more in Chapter 7. However, results also indicate that the resultant function of a dimeric protein is a complex issue and relies upon how the internal biochemistry of each protein is altered upon dimerisation. Lastly, these results contradict the assumption that the function of two proteins is greater than one. This finally indicates that caution must be taken when assuming the properties of an oligomeric protein will be consistent with that of monomeric protein <sup>189,206,406,446</sup>.

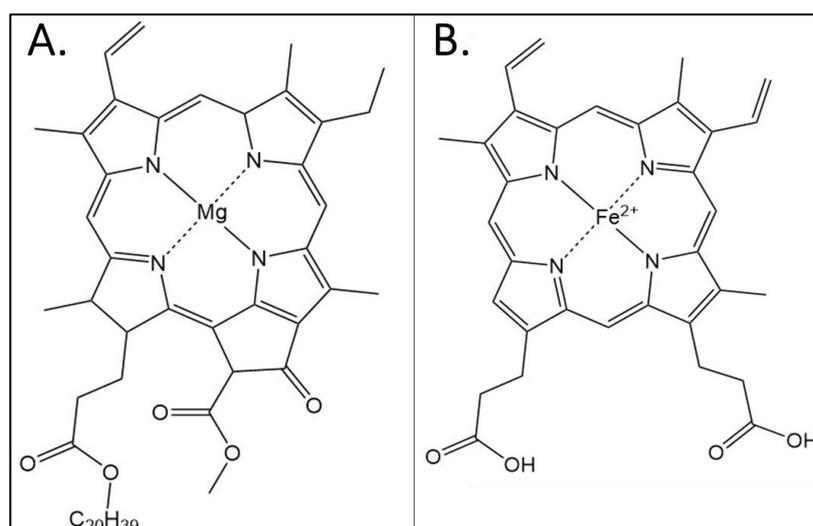
## **Chapter 7.** The Generation of Structurally Diverse Protein Dimers and Oligomers

## 7.1 Introduction

### 7.1.1. The Importance of Chromophore Communication

The synthetic combination of varied biological functions is a core principle of biotechnology<sup>176</sup>; for example, the conjugation of antibody with enzyme is critical in the application of many immunoassays<sup>447</sup>. Chapter 6, demonstrates a practical approach for the construction of functionally linked heterodimers *via* interface prediction, and underscores the feasibility of linking two structurally similar proteins, GFP and mCherry (Figure 6.17), by SPAAC mediated molecular attachment. This chapter aims to use the same methodology applied in Chapter 6 to link two functionally diverse proteins, to establish a means of chromophore communication, which is not observed in nature.

Chromophore communication is a central feature of many biological processes, such as photosynthesis. One significant group of protein chromophores are tetrapyrroles<sup>448</sup>, which are a highly important group of biological molecules that mediate many of the reactions on which life is dependent (9), e.g. light harvesting, electron transport, and oxygen transfer (10). Tetrapyrroles are often referred to as the pigments of life as they exhibit strong absorbance of visible light, with Chlorophyll A possessing the strongest molar extinction coefficient of any known biological compound<sup>6</sup>. The number of double bonds present in tetrapyrroles (Figure



**Figure 7.1. The Structure of Two Common Tetrapyrroles Chlorophyll A and Haem Protoporphyrin IX.**

Chlorophyll A (A) is constituted with a central magnesium ion for electron transfer whilst haem protoporphyrin IX (B) is constituted with a central iron atom. Both central atoms are responsible for the reactivity of each tetrapyrrole and feature in critical steps of respiration and photosynthesis. Image created using ChemBiodraw.

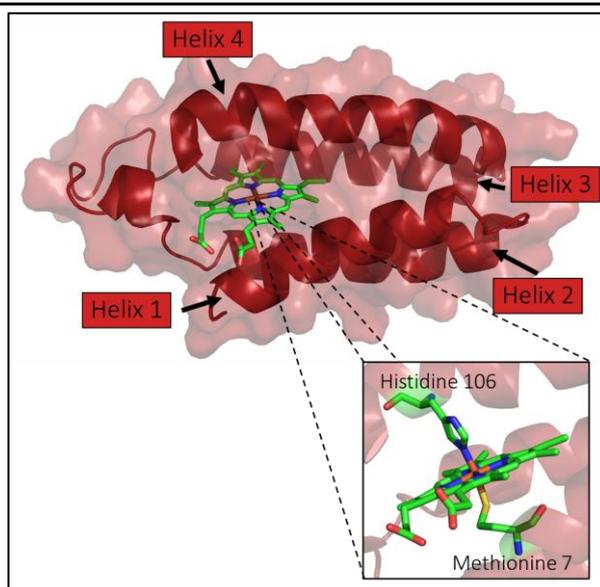
7.1) far exceeds that within a general fluorescent protein chromophore (Figure 6.2). This provides tetrapyrroles with an enhanced ability to absorb light as each different double bond length is able to absorb a different wavelength of light <sup>6</sup>. Structurally tetrapyrroles are organic cofactors which contain a metal ion held in an organic complex by four peripheral nitrogen atoms <sup>449</sup> (Figure 7.1). A methylene bridge links the four nitrogen atoms to the centrally coordinated metal ion, which then facilitates function <sup>450</sup>.

Tetrapyrroles can be either fluorescent, in the case of Zinc porphyrin, or non-fluorescent, in the case of haem porphyrin. In fluorescent porphyrins, absorbed light energy promotes the excitation of electrons to higher energy states, which induce the emission of light from the chromophore. However, in non-fluorescent tetrapyrroles, such as haem porphyrin, the d-orbital of the metal ion initiates quenching of fluorescence, due to the decrease in electron density within the orbital <sup>451</sup>. In this orbital electron-energy transfer is believed to result in the reabsorption/quenching of fluorescence as the partly filled orbital can undergo ligand field d- d transition <sup>452</sup>. This effect was observed previously in Chapter 5, whereby the presence of haem porphyrin resulted in the reduced emission of W117, due to fluorescence quenching in *holo*-HRP (Figure 5.6.).

This transfer of energy is consistent with that observed between fluorescent chromophores (FRET) and exhibits the same principles as previously described in Chapter 1.3.8. Therefore, energy transfer is dependent-upon: distance, orientation, and spectral overlap between chromophores <sup>453</sup>. Consequently, haem porphyrin can function as an energy acceptor during resonance energy transfer, making haem binding proteins model partners for attachment to a fluorescent energy donor, such as sfGFP.

### 7.1.2. Cytochrome *b*<sub>562</sub>

Cytochrome *b*<sub>562</sub> is one of the simplest haem-binding protein scaffolds. Structurally the protein consists of four  $\alpha$ -helical chains with a centrally co-ordinated iron protoporphyrin IX (Figure 7.2). The 12 kDa protein consists of 106 amino acids which form a narrow structure 25 Å in diameter <sup>374,454</sup>. Cytochrome *b*<sub>562</sub> has a high haem binding affinity and will readily uptake cofactor from the local microenvironment (15). Upon haem binding, the coordinating residues of cytochrome *b*<sub>562</sub> rigidifies the protein to form the  $\alpha$ -helical bundle arrangement observed in Figure 7.2, improving stability, and reducing flexibility (17) (see Chapter 5, Figure 5.1).



**Figure 7.2. Structure of Cytochrome  $b_{562}$  (1QPU).**

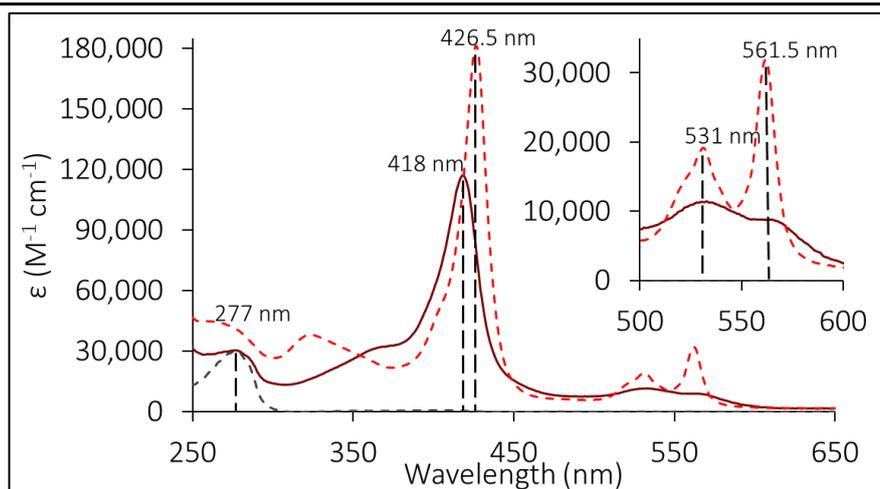
*Holo*-Cytochrome  $b_{562}$  shown as cartoon (Red) with coordinated haem porphyrin (Sticks). Coordination of central iron atom (Orange) by Methionine 7 and Histidine 106 (Also Sticks coloured by atom) is shown in insert.

Within cytochrome  $b_{562}$  haem is co-ordinated to Histidine 106 and Methionine 7 (Figure 7.2, Insert). Although unlike HRP (see Chapter 1.2.6), cytochrome  $b_{562}$  exhibits no enzymatic activity, and as a *b* type cytochrome, does not covalently bind haem<sup>455</sup>. Instead, the native role of cytochrome  $b_{562}$  within the periplasm of *E. coli* is unverified, but the protein is believed to be involved in electron transport<sup>456</sup>.

When saturated a haem Soret peak at 418 nm is observed in the absorbance spectrum of cytochrome  $b_{562}$  in addition to two small  $\alpha$  and  $\beta$  peaks at 531 nm and 561.5 nm (Figure 7.3.). When cytochrome  $b_{562}$  is reduced the absorbance redshifts and both  $\alpha$  and  $\beta$  peaks increase in the intensity (Table 7.1). When reduced the wavelengths at which absorbance is increased overlap with that of sfGFP emission (510 nm) (Figure 7.3), and this indicates that the potential of energy transfer between the haem chromophore and sfGFP.

**Table 7.1. Spectral Properties of Cytochrome  $b_{562}$**

	APO $\lambda_{\text{Max}}$ 277 nm	Oxidized $\lambda_{\text{Max}}$ 418 nm	Reduced $\lambda_{\text{Max}}$ 426.5 nm
$\epsilon$ ( $M^{-1} \text{ cm}^{-1}$ )	30,000	117,000	181,000
$\alpha$ Band (531 nm)	-	11,300	17,600
$\beta$ band (561.5 nm)	-	8,800	32,000



**Figure 7.3. Absorbance Spectra of apo, Oxidized, and Reduced, Cytochrome  $b_{562}$ .**

Molar Extinction Coefficient of Apo (Grey Dashed Line), oxidised (dark red), and reduced (red dashed line) cytochrome  $b_{562}$ . Insert shows both  $\alpha$ -Band (533 nm) and  $\beta$ -Band (563 nm).

### 7.1.3. Communication between Cytochrome $b_{562}$ and GFP

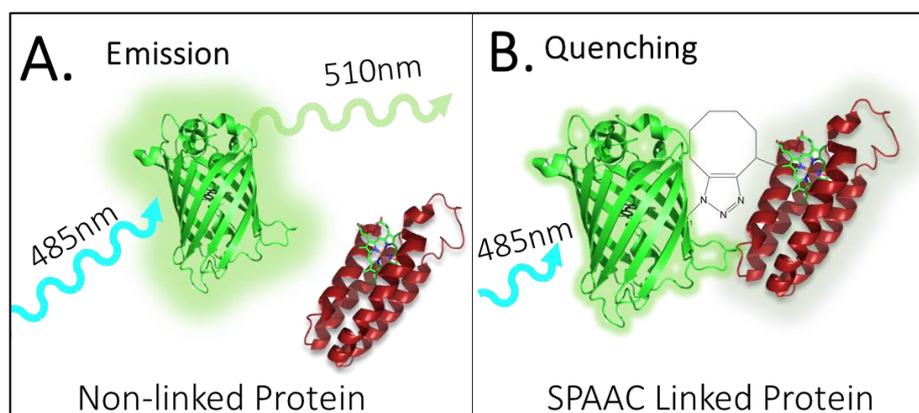
The partnership between cytochrome  $b_{562}$  and EGFP has previously been employed for functional chromophore communication when protein is linked by both linker sequence and domain insertion<sup>176,372,457–460</sup>. The joining of proteins *via* linker sequence was reported by Takeda *et al.* (2001) and demonstrated that upon haem binding fluorescence of EGFP decreased *via* FRET mediated quenching<sup>458</sup>. Furthermore, when the same group later substituted EGFP and haem bound cytochrome  $b_{562}$  with a blue fluorescent protein and coordinated cytochrome  $b_{562}$  with zinc porphyrin, fluorescence energy transfer was demonstrated between the pair<sup>457</sup>.

This work was later furthered by Arpino *et al.* (2012), who demonstrated that the distance between EGFP with cytochrome  $b_{562}$  chromophores could be reduced by domain insertion<sup>176</sup>. This study generated several hybrid EGFP-cytochrome  $b_{562}$  proteins by insertion of the genetic sequence of cytochrome  $b_{562}$ , within the genetic sequence of EGFP. This approach firstly utilised domain insertion for the combination of protein and then later directed evolution to enhance the function of the produced protein. The decreased distance between chromophores of expressed protein resulted in a greater transfer of energy than sampled previously by Takeda *et al.* (2001) and resulted in a near 100% signal quenching in one variant (CG6). The resultant emission of CG6 at 510 nm was subsequently found to be responsive to both haem binding and hydrogen peroxide. However, despite increased communication between chromophores, one consequence of this approach was that many generated variants

had a reduced quantum yield with a lower haem binding capacity in comparison to monomeric sfGFP and cytochrome  $b_{562}$ .

### 7.1.4. Aims

This chapter utilises the same methodology applied in Chapter 6 to build a functionally linked heterodimer. It is aimed that linkage of two functionally disparate and structurally unrelated proteins sfGFP and cytochrome  $b_{562}$  will result in a communication between chromophores upon SPAAC assembly, which will result in the quenching of sfGFP emission by way of FRET in dimeric protein. It is expected that SPAAC induced dimerisation will replicate chromophore proximity observed by Arpino *et al.* (2012), without the risk of impaired function from domain insertion<sup>461</sup>. Energy transfer between chromophores is expected to be further influenced by both chromophore distance (Figure 7.4) and enhanced with increased haem absorbance in reducing conditions<sup>462</sup>.



**Figure 7.4. Distance Dependent FRET Predicted Between GFP And Cytochrome  $b_{562}$ .**

**A**, predicted emission of sfGFP at 510 nm when sfGFP and cytochrome  $b_{562}$  are not dimeric.

**B**, predicted FRET of sfGFP emission to haem porphyrin when cytochrome  $b_{562}$  protein is linked by way of SPAAC attachment.

Lastly, it will be investigated if the nnPTM SPAAC can be combined with natural PTM to form higher-order oligomers (trimers and tetramers), and further increase the complexity of oligomeric assembly. Oligomerisation will be achieved by the secondary mutation of cytochrome  $b_{562}$  for the presence of a surface-exposed cysteine. Cysteine mutagenesis will utilise residues previously established in the formation of homo-dimeric cytochrome  $b_{562}$  to link two heterodimeric assemblies of sfGFP and cytochrome  $b_{562}$ <sup>299</sup>(Figure 7.5). If successful, the generation of a higher order, tetrameric protein, will indicate the potential of greater protein functional diversity for future work. This tetramer will be redox sensitive due to the presence of disulphide bonds and enable the transition of non-natural oligomeric formation

from the combination of two individually functioning protein monomers to three. This would enable the combination of two different metal porphyrins, into the same system in future with no modification to the oligomeric interfaces.

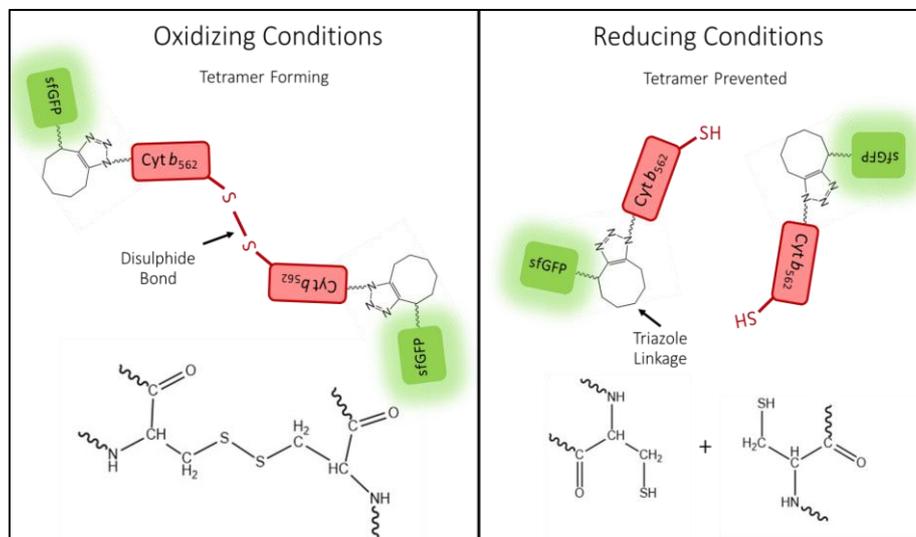


Figure 7.5. Proposal for the Formation of Tetrameric sfGFP and Cytochrome *b*<sub>562</sub>.

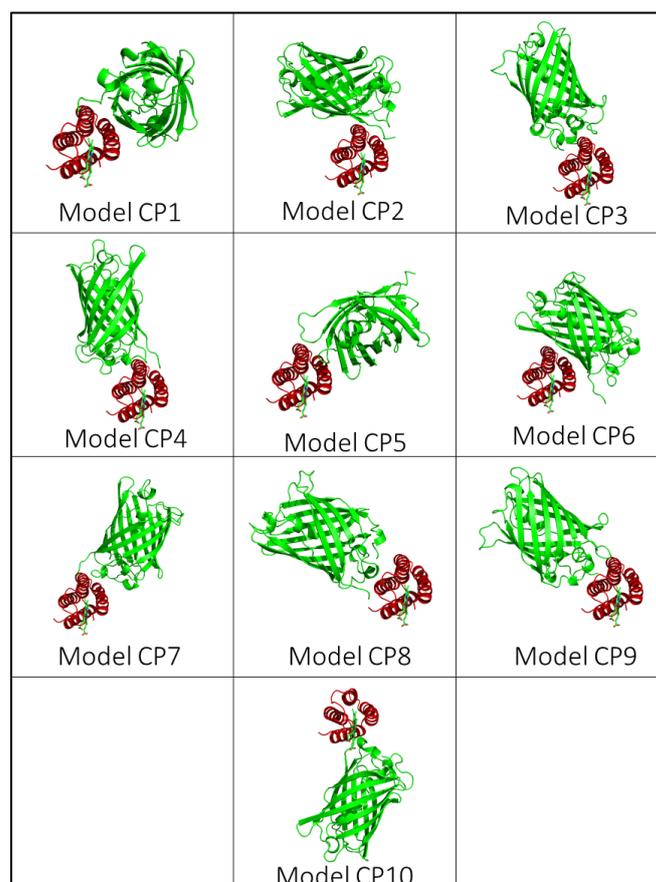
Incubation of protein in oxidizing conditions is conducive to tetramer formation by disulphide formation of surface cysteine thiol groups. Whilst in reducing conditions disulphide bonds are broken and prevented from reforming ensuring dimeric protein.

## 7.2. Results and Discussion

### 7.2.1. *In Silico* Simulation of Interface between sfGFP and Cytochrome $b_{562}$

As stated previously (Chapter 6.2.1.) interface selection for nnAA incorporation is highly influential for the potential reactivity of residues <sup>265</sup>. Consequently, the chance of successful dimer generation is increased when nnAA is located within a compatible region of protein interface. Thus, interface prediction between sfGFP and cytochrome  $b_{562}$  was performed as described in Chapter 6.2.1.

Firstly, ClusPro “Dock” Simulation (<https://cluspro.bu.edu>) (MA, USA) was used to align sfGFP (PDB 2B3P) and cytochrome  $b_{562}$  (PDB 1QPU)(See Section 2.7.3.)<sup>168,374</sup>. Of the 30 alignments generated by ClusPro, the top 10 models ranked according to cluster number were downloaded for assessment (Figure 7.6). The most frequently sampled interface between the two models was represented in model CP1 (Table 7.2.) This model featured a predicted 204



**Figure 7.6. The Top 10 interface predictions as ranked by ClusPro.**

sfGFP, PDB 2B3P, (Green) and  $b_{562}$ , PDB 1QPU, (Red), with haem incorporated back into structure in PyMOL (Green Sticks). The top clusters displayed according to ClusPro interface prediction and ranked by the number of members for each cluster (See Table 7.2).

**Table 7.2. The Number of Clustered Models in Each of the Top 10 interface Predictions Obtained by ClusPro Docking between sfGFP (2B3P) and Cytochrome *b*<sub>562</sub> (1QPU).**

	Model Rank									
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Cluster Members	204	98	89	72	58	53	43	43	39	31

simulated members and was sampled more than twice as frequently as the second most common alignment represented by Model CP2 (Table 7.2).

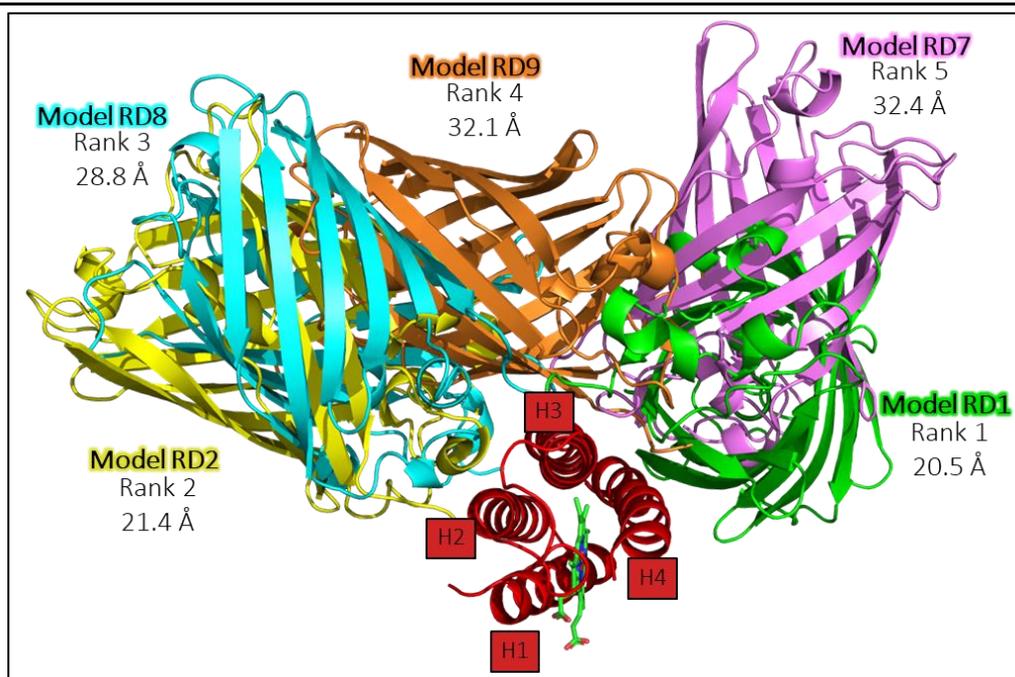
As described in Chapter 6, however, ClusPro simulation only predicts the rigid alignments of protein, as observed within the PDB depository for each structure<sup>430,431</sup>. To simulate backbone relaxation in each predicted interface models were then further assessed and ranked using Rosetta Dock<sup>282</sup>. Models were ranked by the calculation of individual residue environments and the van de Waals attractions between interfaces, to generate an estimation of the amount of energy between interfaces and total energy between molecules (Table 7.3).

The top 5 models from this simulation were subsequently represented in PyMOL, and the distance between chromophores was measured in each predicted alignment. To calculate chromophore distance each model was aligned with both the PDB entries for sfGFP (2B3P) and cytochrome *b*<sub>562</sub> (1QPU), and the distance between the central iron of haem and the hydroxyl group of Tyr 66 of the chromophore was measured (Figure 7.7).

Whilst the lowest Interface Energy of -10.024 (kJ/mole) was reported between monomers in model RD9, model RD1 was reported to have the second-lowest interface energy of -9.632 (kJ/mole)(Table 7.3). However, based on high clustering number (Table 7.2) and favourable energies from RosettaDock (Table 7.3) (Figure 7.7), model RD1 was selected as an interface for designing SPAAC-based linkage of the two proteins. The calculated interface score of -358.2(kJ/mole) for model 1 was notably less negative than the -426 (kJ/mole) reported between molecules in model RD9 of sfGFP-mCherry alignment (Table 6.5). This difference is

**Table 7.3. Ranked Placement of ClusPro Acquired Models According to the Total Energy between Molecules.**

Ranked Placement	Model ID	Total Energy (kJ/mole)	Interface Energy (kJ/mole)
1	RD1	-358.2	-9.632
2	RD2	-354.0	-9.091
3	RD8	-353.4	-7.078
4	RD9	-352.9	-10.024
5	RD7	-350.5	-8.83



**Figure 7.7. The 5 Lowest Energy Interfaces as Calculated by Rosetta Dock.**

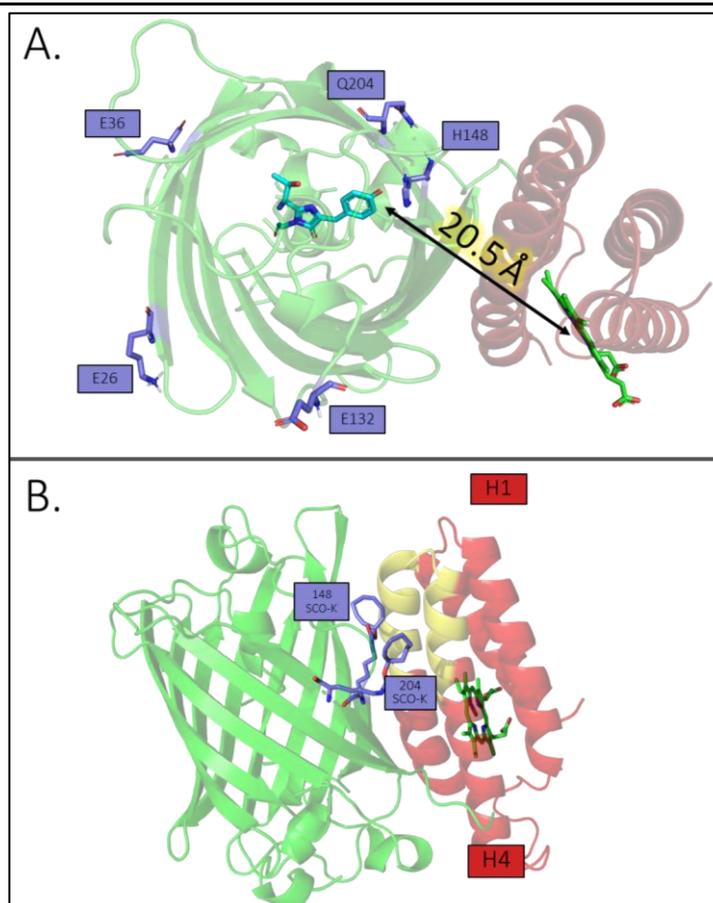
The top 5 alignments of sfGFP (Various Colours) and  $b_{562}$  (Red) ranked according to Total energy of interaction by Rosetta Dock. Distances measured between each model chromophore pair is shown with each model. Model RD1 is measured to align with closest chromophore proximity.

potentially indicative that cytochrome  $b_{562}$  might not interface as readily with sfGFP as mCherry; however, as the limits of SPAAC have yet to be established, it is not known if this difference is significant.

### 7.2.2. Residue Selection for nnAA Incorporation in Cytochrome $b_{562}$

Residues within the interface displayed in Model RD1 which were surface exposed were reasoned to be the best residues of nnAA incorporation in both sfGFP and cytochrome  $b_{562}$  (Figure 7.8.). Inspection of existing residues of previous SCO-K incorporation in sfGFP (E26, E36, E132, H148, and Q204) was first performed to determine if any overlapped with the region of the interface in Model RD1 (Figure 7.8, A). Both residues sfGFP 148 and 204 were within the region of the interface of sfGFP and cytochrome  $b_{562}$  in model RD1(Figure 7.8, A).

Residues 148 and 204 in sfGFP were mutated to SCO-K residues in PyMOL (see Chapter 2.7.5.), before selection of all surface-exposed residues in cytochrome  $b_{562}$ , which were within a 6 Å radius (Figure 7.8, B). Residues which were highlighted by this selection were collectively located upon either helix 3 or 4 of cytochrome  $b_{562}$ . This localisation was considered of benefit

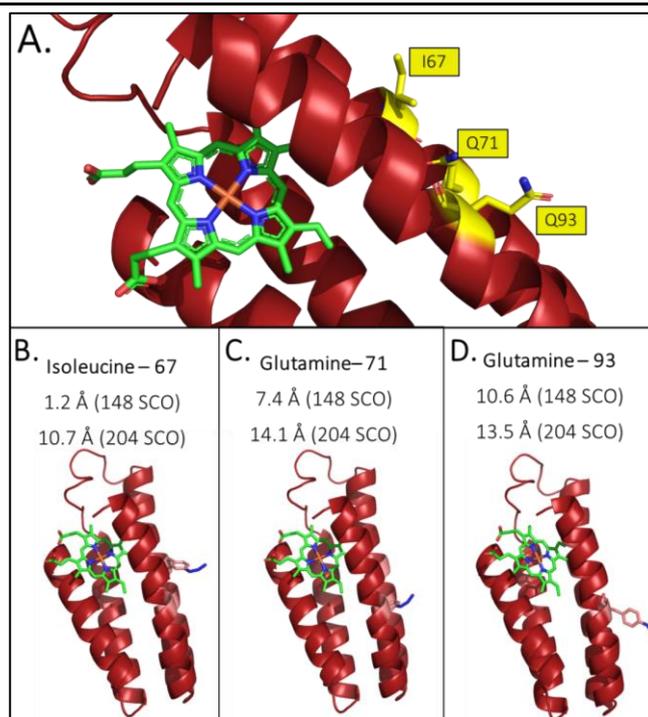


**Figure 7.8. The Alignment of sfGFP and Cytochrome  $b_{562}$  Observed in Model RD1.**

Alignment of sfGFP (2B3P) and  $b_{562}$  (1QPU) with model RD1 with chromophores shown as sticks internally. **A**, residues of previous sfGFP SCO incorporation E26, E36, E132, H148, and Q204 shown as sticks (**Purple**). **B**, residues 148 and 204 mutated to SCO-K, and all residues within a 6 Å radius of these two residues on cytochrome  $b_{562}$  are highlighted (**Yellow**).

as this was anterior to the opening of the haem binding pocket, and it was anticipated that the nnAA incorporation within this region would have a minimal influence upon the haem binding of cytochrome  $b_{562}$ .

Three surface-exposed residues, I67, Q71, and Q93 (Figure 7.9) were identified by this analysis. Residues I67 and Q93 were both located on Helix 3 whilst, Q71 was located on Helix 4. As no clear preferential residue of nnAA incorporation could be established between the three potential residues; all three were selected for site-directed mutagenesis. It was reasoned that selection of multiple residues would increase the chance of project success and subsequently oligonucleotides were designed to mutate codons 67, 71, and 93 to TAG in site-directed mutagenesis (Chapter 2, Table 2.4).



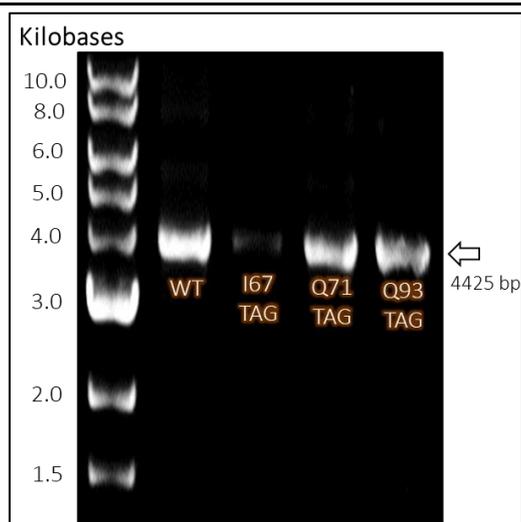
**Figure 7.9. Potential Sites of nnAA Incorporation in Cytochrome  $b_{562}$ .**

**A**, Residues 6 Å proximity to residues 148 and 204 on sfGFP highlighted (**Yellow**), and those which are surface exposed shown as sticks. All residues were mutated to AZF in PyMOL by *in-silico* mutagenesis and measured for proximity to SCO-K in sfGFP. **B**, Isoleucine 67, **C**, Glutamine 71, and, **D**, Glutamine 93.

### 7.2.3. TAG Mutagenesis of Cytochrome $b_{562}$

Following the identification of three target residues for nnAA incorporation, site-directed mutagenesis was performed with TAG mutagenesis oligonucleotides (See Section 2.2.3). Analysis of the PCR product by agarose gel electrophoresis (Figure 7.10), indicates amplification of plasmid to a size consistent with pBAD-cyt  $b_{562}^{\text{WT}}$  (~4400 bp) in all samples. All amplified bands were phosphorylated and ligated before transformation into DNA propagating NEB® 5- $\alpha$  *E. coli* cells (see Chapter 2.1.3.). Cells that were transformed with pBAD-cyt  $b_{562}^{67\text{TAG}}$  and pBAD-cyt  $b_{562}^{93\text{TAG}}$  did not yield colonies in multiple (>10) transformation attempts, and these mutants were not pursued further. Colonies which did grow after transformation of cells with pBAD-cyt  $b_{562}^{71\text{TAG}}$  were subsequently used to inoculate 10 mL media for DNA propagation. DNA was extracted and sent for sequencing (Eurofins Genomics, UK) which confirmed the sequence identity of the plasmid as pBAD-cyt  $b_{562}^{71\text{TAG}}$ .

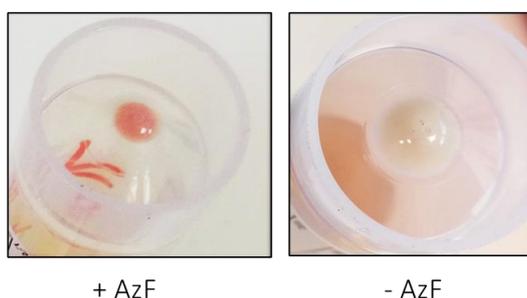
At this stage, it was decided that work would continue with the generation of protein from the newly formed pBAD-cyt  $b_{562}^{71\text{TAG}}$  plasmid and mutations I67 TAG and Q93 TAG would not be pursued further. Sequenced pBAD-cyt  $b_{562}^{71\text{TAG}}$  plasmid was transformed into TOP10™



### 7.10. Mutagenesis of Cytochrome $b_{562}$ for nnAA Incorporation.

SDM of the pBAD- $b_{562}^{WT}$  plasmid using I67, Q71, and Q93 TAG mutagenesis oligonucleotides. Amplified product formed banding at approximately 4425 bp in size, consistent with WT plasmid.

*E. coli* cells in a double transformation with pDULE. Resultant colonies were tested for expression of cytochrome both with and without nnAA. Expression of cytochrome  $b_{562}^{71AZF}$  was observed by the pink colouration of cells and was achieved only in cells which had been supplied with AZF indicating expression was dependent on nnAA (Figure 7.11).



**Figure 7.11. Expression of Cytochrome  $b_{562}^{71AZF}$ .**

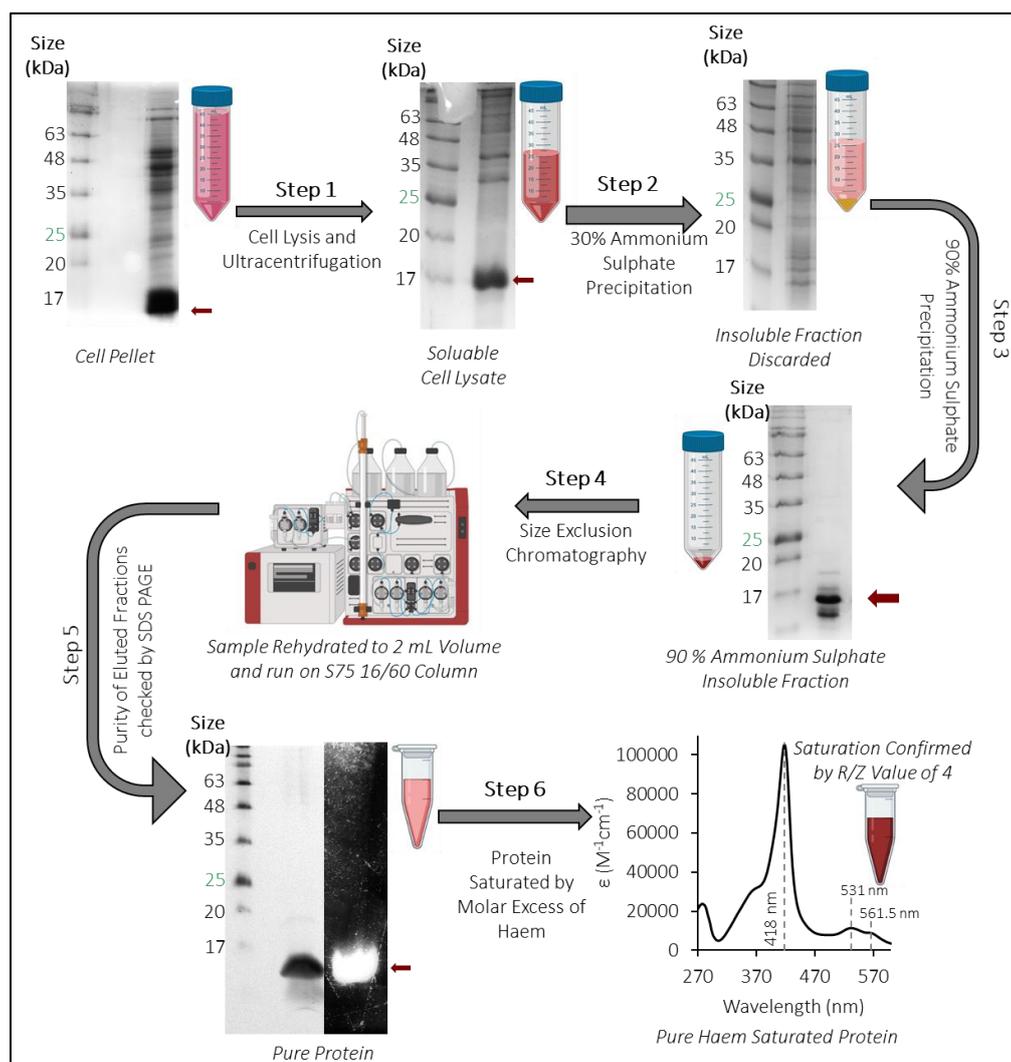
Pink coloration indicative of cytochrome  $b_{562}$  expression observed in cell pellet of culture which had been supplied with 1 mM AZF. Pink colour is not observed in control cells which were not supplied with nnAA. Both cultures grown overnight at 37 °C after induction of cells with 1% arabinose.

### 7.2.4. Purification of Cytochrome $b_{562}^{71AZF}$

The purification of cytochrome  $b_{562}^{71AZF}$  was performed as described previously (see Chapter 2.4.7.), with the exception that growth and purification were performed in dark conditions<sup>299,461</sup>. Two cycles of ammonium sulphate precipitation were applied to the lysate to separate less soluble protein contaminants (Schematic 7.1, Step 2 and 3)(see Chapter 2.3.6.). Contaminate protein of lower solubility than cytochrome  $b_{562}$  was removed by a 30% (w/v)

## Chapter 7. Functional Fluorescence Quenching by Resonance Energy Transfer in Dimeric Protein.

ammonium sulphate precipitation before precipitation of cytochrome  $b_{562}^{71AZF}$  by 90% (w/v) ammonium sulphate precipitation. This precipitant was resuspended to 2 mL volume in Tris 50 mM pH 8 and applied to an S75 16/600 column for size exclusion chromatography (Schematic 7.1, Step 4)(see Chapter 2.4.2.). Eluted protein was subsequently checked for purity by SDS PAGE before pure fractions were pooled and concentrated (Schematic 7.1, Step 5). To ensure AZF incorporation pure protein was mixed with Cy3 DBCO and incubated overnight (see



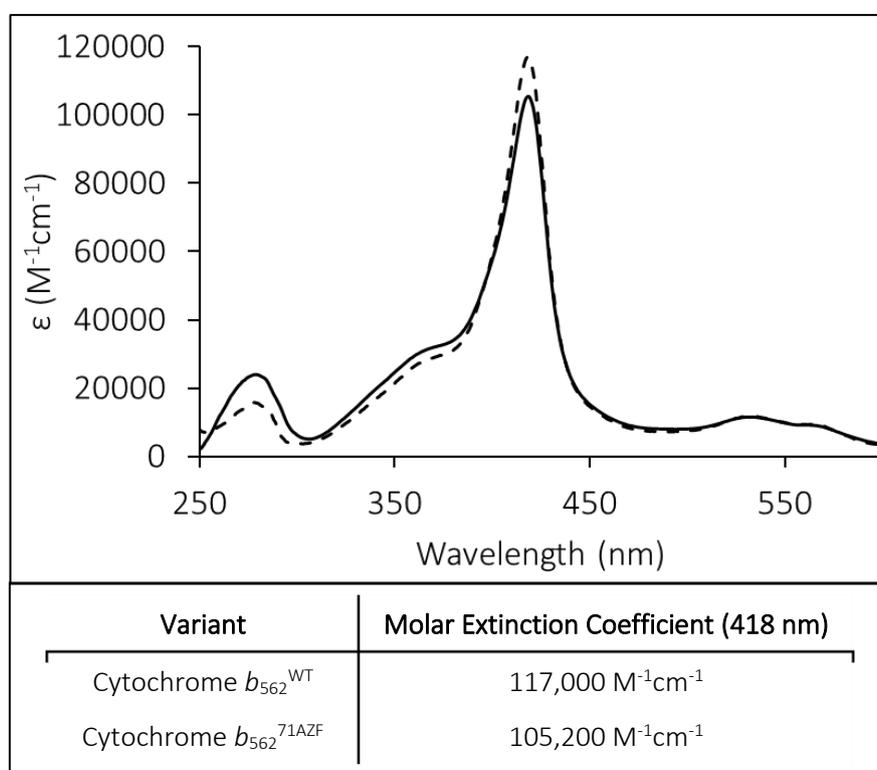
**Schematic 7.1. Purification of Cytochrome  $b_{562}^{71AZF}$ .**

Total protein presence at each step shown by SDS PAGE and  $b_{562}$  indicated by red arrow. **Step 1**, cell lysis and centrifugation to remove insoluble protein. **Step 2**, a 30% ammonium sulphate precipitation of protein from the lysate. **Step 3**, a 90% ammonium sulphate precipitation from lysate. **Step 4**, rehydration of precipitate and size exclusion chromatography. **Step 5**, SDS PAGE of pooled pure fractions of  $b_{562}$  purification after incubation with 10 x molar excess of Cy3 DBCO. Fluorescence of protein band confirms feasibility of AZF cycloaddition. **Step 6**, haem saturation of pure protein and removal of free haem by desalting before saturation check by  $R_Z$  calculation. The absorbance spectrum of saturated protein displayed corrected to molar extinction coefficient. Images created on biorender.com

Chapter 2.8.1). SDS PAGE of protein/probe mix indicated the presence of a fluorescent band at 12 kDa consistent with that of cytochrome  $b_{562}^{71AZF}$  confirming AZF incorporation (Schematic 7.1, Step 5). Finally, the pure protein was saturated with haem by mixing with a 10x molar excess of free haem and subsequent desalting to acquire an  $R_z$  value of 4 (Schematic 7.1, Step 6)(see Chapter 2.5.4.).

### 7.2.5. Absorbance of Cytochrome $b_{562}^{71AZF}$

As this investigation aims to generate a haem sensitive fluorescent probe, the maintained binding affinity of cytochrome  $b_{562}$  is vital. Upon haem saturation, the absorbance spectrum of cytochrome  $b_{562}^{71AZF}$  was recorded, and calculation of the variant's molar extinction coefficient performed (see Chapter 2.5.3.)(Figure 7.12). The absorbance spectrum of cytochrome  $b_{562}^{71AZF}$  indicated the presence of a Soret band at 418 nm, and both  $\alpha$  and  $\beta$  bands at 531.5 nm and 561.5 nm respectively, consistent with the absorbance spectrum of cytochrome  $b_{562}^{WT}$  (Figure 7.12.). The molar extinction coefficient of cytochrome  $b_{562}^{71AZF}$  was calculated to be  $105,200 \text{ M}^{-1}\text{cm}^{-1}$  at 418 nm. This change represented a decrease of 10% from the  $117,000 \text{ M}^{-1}\text{cm}^{-1}$  at 418 nm of cytochrome  $b_{562}^{WT}$  and indicated that the nAA



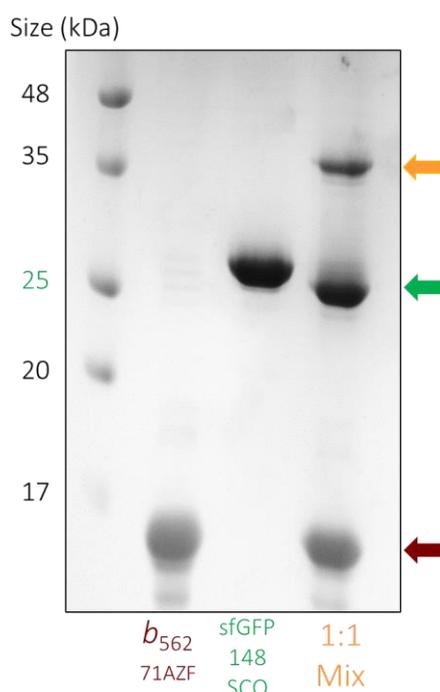
**Figure 7.12. Comparison of Molar Extinction Coefficient of Cytochrome  $b_{562}^{71AZF}$ .** The absorbance spectra of both cytochrome  $b_{562}^{71AZF}$  (Solid Line) and cytochrome  $b_{562}^{WT}$  (Dashed Line) Table displays molar extinction coefficient of each at 418 nm.

incorporation had a limited influence on absorbance. However, a difference of 10 % is small as possibly within the margin of error for the investigation.

### 7.2.6. Cycloaddition of sfGFP<sup>148SCO</sup> and Cytochrome *b*<sub>562</sub><sup>71AZF</sup>

The formation of dimeric protein, between cytochrome *b*<sub>562</sub><sup>71AZF</sup> and sfGFP<sup>148SCO</sup>, was achieved by mixing of 100 μM of each protein at a 1:1 molar ratio, and incubation overnight at 20 °C (see Chapter 2.8.3.). The next day, the sample was analysed by SDS PAGE, and the formation of a faint novel protein band was observed at an approximate weight of 39 kDa. This band was consistent with that expected of the combination of sfGFP and cytochrome *b*<sub>562</sub> (27 kDa and 12 kDa). However, the band was too low in concentration for visualisation after UV exposure.

It was reasoned that the rate of interaction between sfGFP and cytochrome *b*<sub>562</sub>, was, therefore, limited by sfGFP self-association within the sample. As SPAAC can occur at temperatures ranging from 0-160 °C<sup>463</sup>, it was reasoned that increasing the incubation temperature would not impede the reaction between azide and alkyne, but may decrease weak association between molecules. To investigate if SPAAC efficiency could be increased,



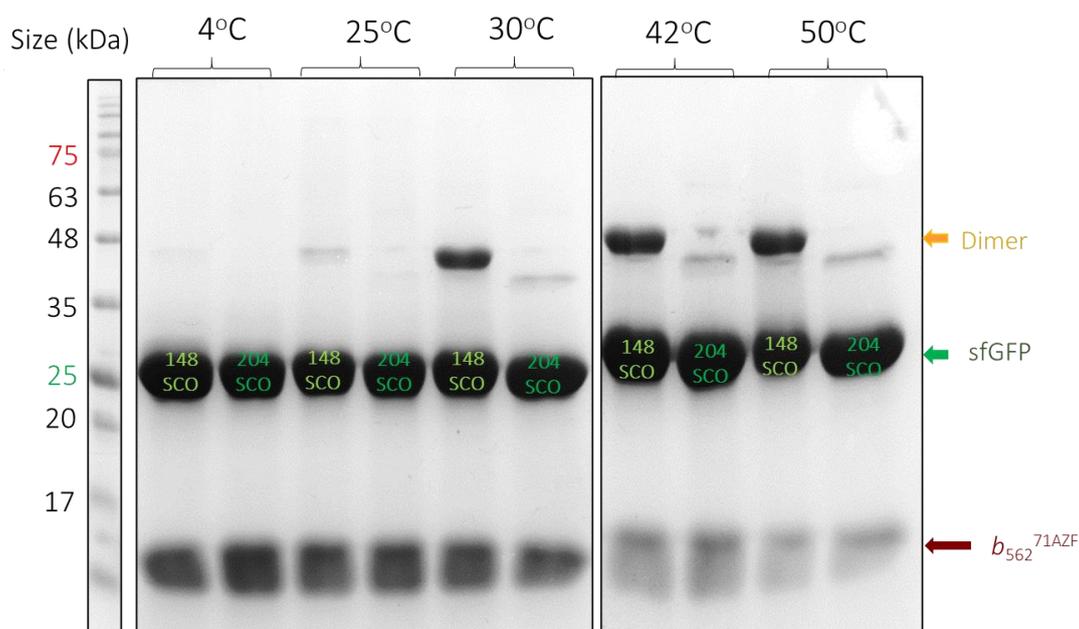
**Figure 7.13. Formation of Cytochrome *b*<sub>562</sub> sfGFP Dimer at 37°C**

SDS PAGE of separate 100 μM samples of *b*<sub>562</sub><sup>71AZF</sup> (Red) and sfGFP<sup>148SCO</sup> (Green) (Lanes 1 and 2) and sample of 1:1 molar ratio of each (50 μM) (Lane 3). All samples incubated overnight at 37 °C prior to SDS PAGE. When both proteins are mixed a third band (Orange), is observed sized at approximately 39 kDa is observed within Lane 3.

cytochrome  $b_{562}^{71AZF}$  and sfGFP<sup>148SCO</sup> were mixed once more overnight at 37 °C. SDS PAGE of incubated protein demonstrated the intensity of the newly formed band (~39 kDa) had increased with the increased incubation temperature (Figure 7.13).

### 7.2.7. The Influence of Temperature on Dimer Formation

As demonstrated in Figure 7.13., increasing incubation temperature at which SPAAC occurs from 20 °C to 37 °C, stimulated an increase in dimeric protein formed. To identify the optimum temperature for cycloaddition five different temperatures (4 °C, 25 °C, 30 °C, 42 °C and 50 °C) were used to incubate mixed samples of sfGFP and cytochrome  $b_{562}^{71AZF}$  overnight. In this instance no temperature higher than the  $T_m$  of apo-cytochrome  $b_{562}$  (55 °C) was investigated, to minimise the potential effects of thermal denaturation<sup>464</sup>.



**Figure 7.14. Influence of Temperature on the Intensity of Dimer Formation.**

SDS PAGE analysis of dimer formed (Orange) sfGFP (Green) and  $b_{562}^{71AZF}$  (Red) across temperature range from 4 °C to 50 °C. All lanes from the same Gel but cropped to remove lane breaks.

SDS PAGE of incubated protein indicated the association between sfGFP and cytochrome  $b_{562}$  was increased with temperature (Figure 7.14). The presence of a protein band consistent with dimeric protein at ~39 kDa was shown to be greatest at the highest temperatures of investigation (42 °C and 50 °C), and lowest at 4 °C (Figure 7.14). While both sfGFP<sup>148SCO</sup> and sfGFP<sup>204SCO</sup> formed a band consistent with dimeric protein (~39 kDa) when mixed with cytochrome  $b_{562}^{71AZF}$ .

## Chapter 7. Functional Fluorescence Quenching by Resonance Energy Transfer in Dimeric Protein.

The ~39 kDa band, consistent with dimeric protein, varied between each sfGFP variant and was most intense in the combination of sfGFP<sup>148SCO</sup> and cytochrome *b*<sub>562</sub><sup>71AZF</sup>, rather than sfGFP<sup>204SCO</sup> and cytochrome *b*<sub>562</sub><sup>71AZF</sup>. The band formed at ~39kDa was far weaker in intensity after incubation at both 4 °C and 25 °C, in comparison to incubation at ≥30 °C, indicating a low intensity of dimer formation without application of heat.

ImageJ analysis of dimer formed Figure 7.14 (see Chapter 2.5.1.), quantified the percentage of total protein which had formed dimer in each reaction condition (Table 7.4). The temperature which yielded the highest percentage of dimeric protein was 42 °C in sfGFP<sup>148SCO</sup> samples and 50 °C in sfGFP<sup>204SCO</sup> samples. Whilst the band corresponding with dimeric protein (~39 kDa) was observed to be 8 x greater in samples formed with sfGFP<sup>148SCO</sup> (25% at 42 °C) than with sfGFP<sup>204SCO</sup> (3% at 42 °C). Herein, dimers are referred to as GFP*b*<sup>148-71</sup> (the combination of sfGFP<sup>148</sup> and cytochrome *b*<sub>562</sub><sup>71AZF</sup>) and GFP*b*<sup>204-71</sup> (the combination of sfGFP<sup>204</sup> and cytochrome *b*<sub>562</sub><sup>71AZF</sup>).

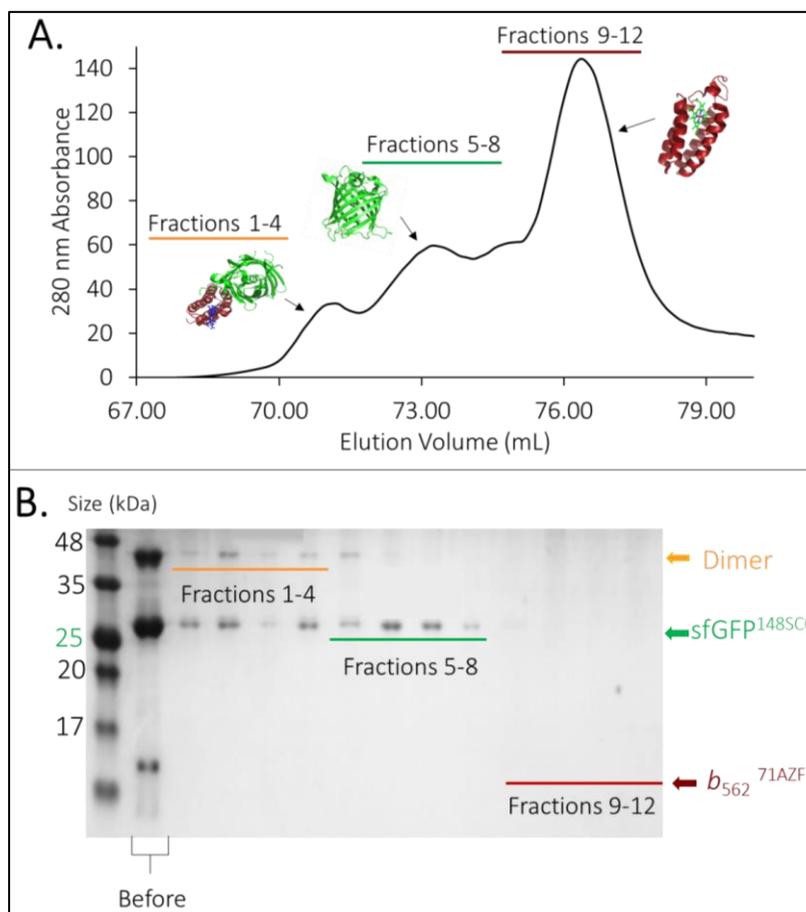
**Table 7.4. Percentage Formed Dimer at Each Investigated Temperature**

Temperature of Incubation	Percentage of Dimer Formed	
	sfGFP <sup>148SCO</sup>	sfGFP <sup>204SCO</sup>
4 °C	1%	-
25 °C	3%	-
30 °C	20%	0.5%
42 °C	25%	3%
50 °C	21%	4%

### 7.2.8. Separation of Heterodimer from Monomer

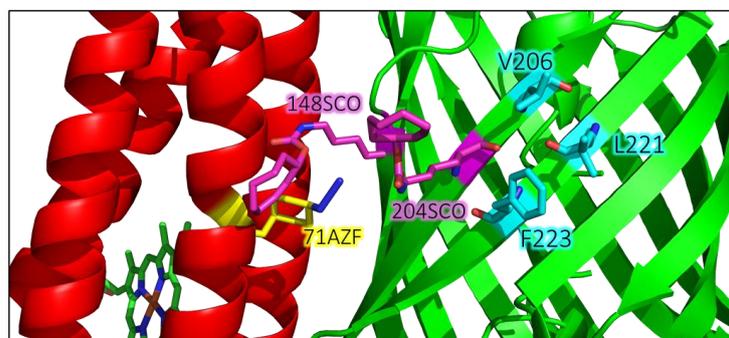
Trial separation of GFP*b*<sup>148-71</sup> from monomeric protein was performed by size exclusion chromatography, as it would be difficult to separate sfGFP monomer from dimer via nickel affinity chromatography alone (see Chapter 2.4.2)(Figure 7.15). The resultant chromatograph indicated that protein elution began at 65 mL and featured several merged peaks of absorbance at 280 nm, which lacked discrete definition (Figure 7.15, A). Subsequently, SDS PAGE of fractions, revealed that monomeric sfGFP<sup>148SCO</sup> and GFP*b*<sup>148-71</sup> were not separated by SEC (Figure 7.15, B). In this instance, no sample of purified GFP*b*<sup>148-71</sup> was obtained, and only the separation of monomeric cytochrome *b*<sub>562</sub><sup>71AZF</sup> was achieved. Several buffer conditions (2M NaCl and pH 6) were trialled to aid in protein separation, however no condition enabled successful monomer separation. It was hypothesised, that the failure to separate dimeric protein was due to interaction between sfGFP<sup>148SCO</sup> (monomer) and GFP*b*<sup>148-71</sup> (dimer), as the

elution volume of each was shown to overlap, despite the known size difference of each protein form.



**Figure 7.15.** The Separation of dimeric Protein from Monomeric by SEC at 23 °C.

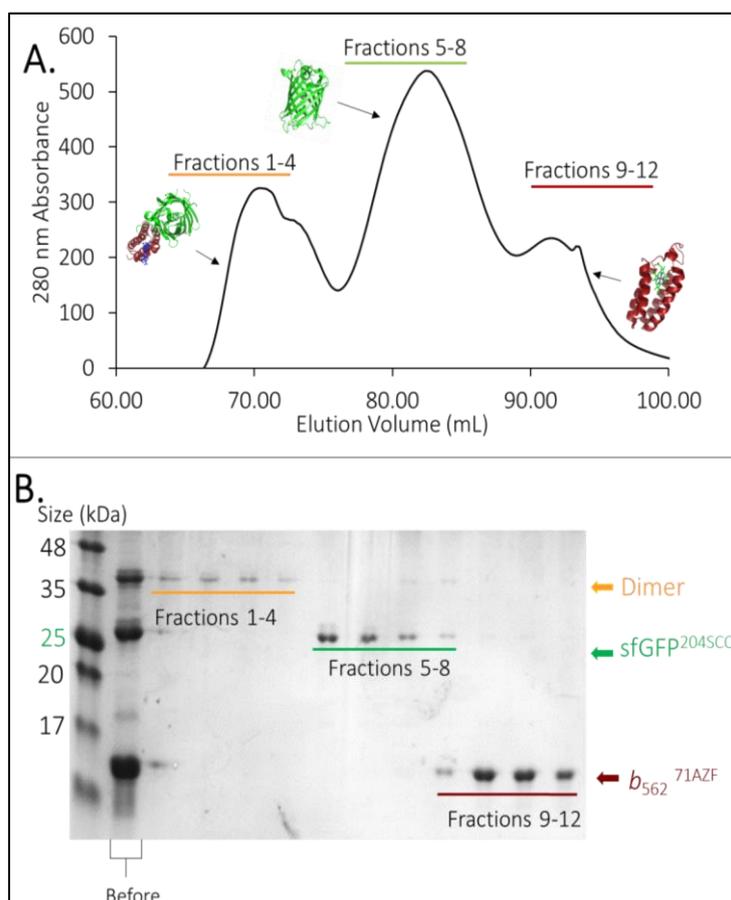
**A**, Separation of  $\text{GFP}_{b^{148-71}}$  from monomeric protein by size exclusion chromatography. Chromatograph represents absorbance at 280 nm (**black line**) as protein eluted from Superdex S200 16/600 column. **B**, SDS PAGE of fractions obtained, indication no separation of  $\text{GFP}_{b^{148-71}}$  from  $\text{sfGFP}^{148\text{SCO}}$ . Lane titled “Before” displays sample prior SEC for comparison.



**Figure 7.16.** Hydrophobic Residues Responsible for Self-Association in sfGFP.

Alignment of sfGFP (**Green**) and  $b_{562}$  (**Red**) as obtained in Model RD1. Three hydrophobic residues V106, L221, and F223, known to contribute to sfGFP self-association displayed as sticks (**Teal**).

Self-association of sfGFP is known to arise spontaneously at high concentration<sup>169,465</sup> *via* interaction at the surface exposed hydrophobic residues V206, L221, and F223<sup>167</sup> (Figure 7.16). As these three residues are located adjacent to the molecular interface represented in Model RD1 it was reasoned that they would still be accessible for self-association, even after SPAAC with cytochrome *b*<sub>562</sub>. Furthermore, as residue 204 is closer to these hydrophobic residues, self-association between sfGFP molecules may cause reduced accessibility to residue 204 for SPAAC, in comparison to residue 148. As increased incubation temperature was previously observed to increase the rate of SPAAC between monomers (Figure 7.14), it was hypothesised this may have been achieved, by surmounting of the intermolecular attraction at residues 206, 221 and 223, breaking apart monomeric sfGFP. Consequently, it was investigated if an increase of temperature during purification would enhance dimer separation for analysis.



**Figure 7.17. Analysis of the Separation of Dimeric Protein by SEC at 42 °C.**

**A**, Separation of GFP<sup>b148-71</sup> from monomeric protein by size exclusion chromatography at 42 °C. Chromatograph represents absorbance at 280 nm (**black line**) as protein eluted from Superdex S200 16/600 column. The first peak is representative of dimer, whilst the second is representative of sfGFP monomer, and the third *b*<sub>562</sub> monomer. **B**, SDS PAGE of fractions obtained, indication no separation of GFP<sup>b148-71</sup> from sfGFP<sup>148SCO</sup>. Lane titled “Before” displays sample prior to SEC for comparison. Pure dimer is observed in Fractions 1-4.

## Chapter 7. Functional Fluorescence Quenching by Resonance Energy Transfer in Dimeric Protein.

Separation of proteins was subsequently trialled with buffer incubated in a 42 °C water bath. Elution of protein was observed, again, after 65 mL, and in this instance, a greater separation of elution peaks was observed (Figure 7.17, A). Increased separation of major protein forms was reflected in SDS PAGE analysis of fractions which indicated that those which eluted first (Fractions 1-4) comprised of pure dimeric protein (Figure 7.17, B). It was concluded that the association between sfGFP was the stimulus for dimeric protein association as, upon the application of heat, enhanced protein separation was detected.

### 7.2.9. Absorbance Spectrum of Oxidised GFP $b^{148-71}$ and GFP $b^{204-71}$

As observed in Chapter 6 (Figure 6.23) the increased proximity between fluorescent chromophores resulted in a decrease in function in the mCherry chromophore. Therefore, the effect of SPAAC mediated proximity between the chromophore of sfGFP and cytochrome's haem porphyrin is uncertain. Thus far, nnAA incorporation to cytochrome  $b_{562}$  has stimulated minor change to absorbance (Figure 7.12); however, this might not be conserved in protein linked by SPAAC assembly. Effects observed by Arpino *et al.* (2012) demonstrate that the direct linking of EGFP and cytochrome  $b_{562}$  by domain insertion can yield associated spectral change to the  $\lambda_{MAX}$  of haem's Soret peak<sup>176</sup>. Furthermore, sfGFP absorbance change can be indicative of the loss of fluorophore function (Figure 6.20) whilst reduced Cytochrome  $b_{562}$  absorbance loss can be indicative of a reduced availability of cofactor (more *apo* protein) (Figure 7.3); both of which would severely impede FRET within constructed dimeric protein.

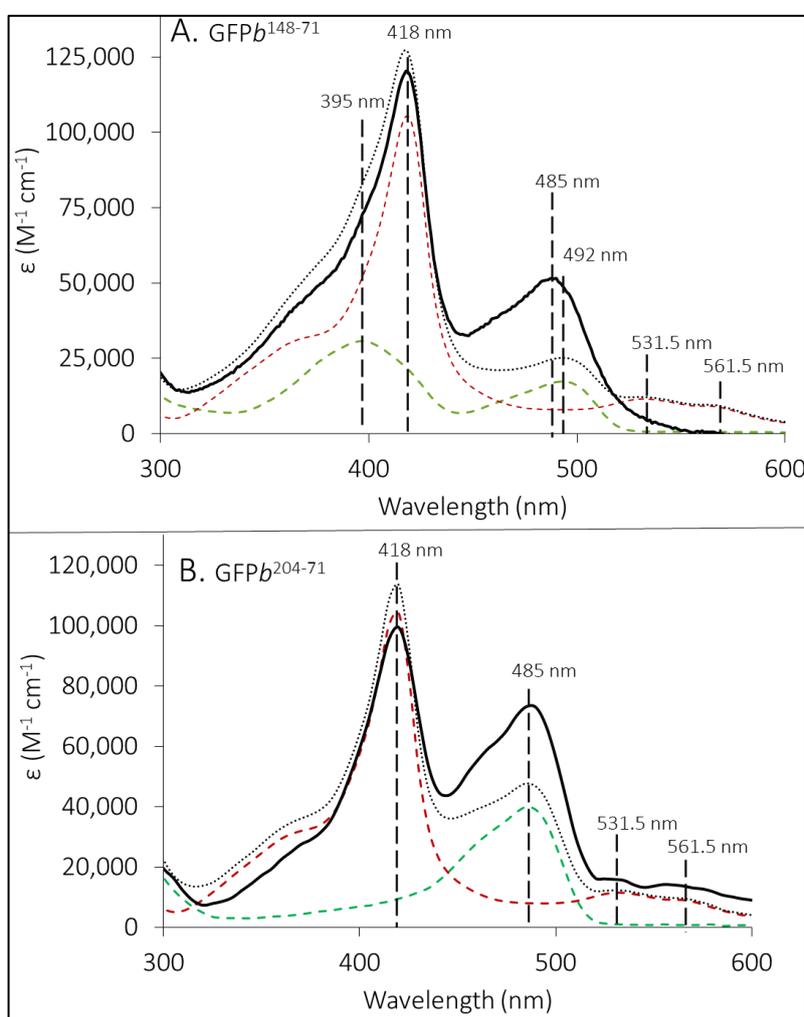
The absorbance spectra of both dimers indicate the retained presence of both haem porphyrin and sfGFP chromophore by absorbance peaks at both 418 nm and 485 nm (Figure 7.18). Addition of each monomeric protein's molar extinction coefficient indicates the

**Table 7.5. Molar Extinction Coefficient of Each Chromophore**

Molar Extinction Coefficient (M <sup>-1</sup> cm <sup>-1</sup> )	Monomer			Dimer	
	sfGFP <sup>148SCO</sup>	Cytochrome $b_{562}^{71AZF}$	sfGFP <sup>204SCO</sup>	GFP $b^{148-71}$	GFP $b^{204-71}$
Haem Soret Peak	-	105,200 (418 nm)	-	120,300 (418 nm)	99,300 (418 nm)
sfGFP Chromophore	31,000/17,300 (395 nm/ 492nm)	-	39,800 (485 nm)	50,900 (485 nm)	73,100 (485 nm)

absorbance change expected from the mixing of each monomer. The absorbance of each dimer was shown to be consistent in intensity with that predicted of monomeric cytochrome  $b_{562}^{71AZF}$ , and the  $\lambda_{MAX}$  of dimeric protein was retained at 418 nm, consistent with cytochrome  $b_{562}^{WT}$  (Figure 7.18). Only a minimal absorbance change at 418 nm occurred upon dimerisation, from  $105,200 \text{ M}^{-1} \text{ cm}^{-1}$  to  $120,300 \text{ M}^{-1} \text{ cm}^{-1}$  (15% increase) in  $\text{GFP}b^{148-71}$ , and from  $105,200 \text{ M}^{-1} \text{ cm}^{-1}$  to  $99,300 \text{ M}^{-1} \text{ cm}^{-1}$  (6% decrease) in  $\text{GFP}b^{204-71}$  (Table 7.5).

Whilst sfGFP chromophore absorbance was observed to increase from  $39,800 \text{ M}^{-1} \text{ cm}^{-1}$  to  $73,100 \text{ M}^{-1} \text{ cm}^{-1}$  at 485 nm in  $\text{GFP}b^{204-71}$  and in  $\text{GFP}b^{148-71}$  from  $17,300 \text{ M}^{-1} \text{ cm}^{-1}$  to  $50,900 \text{ M}^{-1} \text{ cm}^{-1}$ . A predicted increase of 47% from  $17,200 \text{ M}^{-1} \text{ cm}^{-1}$  to  $25,200 \text{ M}^{-1} \text{ cm}^{-1}$  was



**Figure 7.18. Spectral Characteristics of  $\text{GFP}b^{148-71}$  and  $\text{GFP}b^{204-71}$  in Oxidising Conditions.**

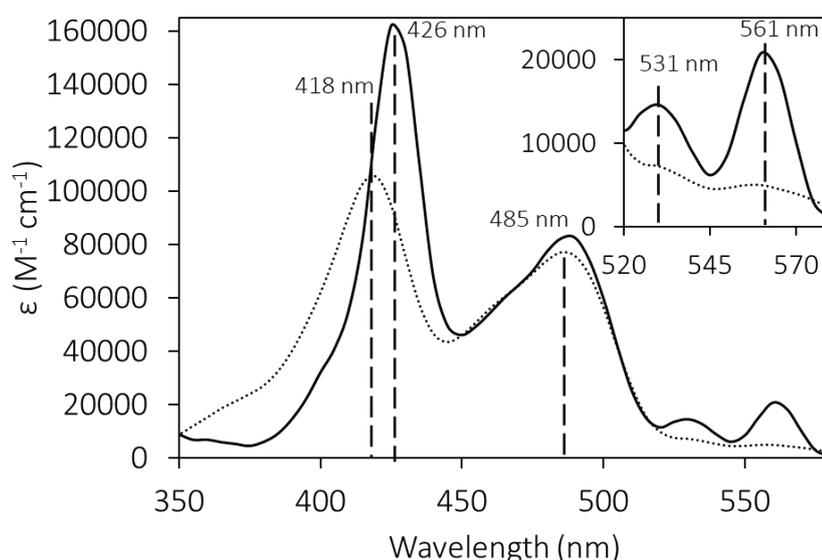
**A**,  $\text{GFP}b^{148-71}$ , and constituent monomers sfGFP<sup>148SCO</sup> and  $b_{562}$ . **B**,  $\text{GFP}b^{148-71}$ , and constituent monomers sfGFP<sup>148SCO</sup> and cytochrome  $b_{562}^{71AZF}$ . Relative absorbance of dimer (**Black**),  $b_{562}^{71AZF}$  (**Red Dashed**), and sfGFP (**Green**). The combined sum of each monomeric absorbance shown for comparison (**Grey Dotted**). Data shown as molar extinction coefficients. The wavelengths associated with each absorbance peak are shown for reference of the  $\lambda_{MAX}$  of each chromophore.

expected to occur at 492 nm in GFPb<sup>148-71</sup>, however, a greater ~200% was observed from 17,200 M<sup>-1</sup> cm<sup>-1</sup> at 492 nm to 50,900 M<sup>-1</sup> cm<sup>-1</sup> at the new  $\lambda_{MAX}$  485 nm. Whilst an increase of 20% was predicted to occur in GFPb<sup>204-71</sup> from 39,800 M<sup>-1</sup> cm<sup>-1</sup> to 47,700 M<sup>-1</sup> cm<sup>-1</sup> at 485 nm but an increase of 84% was instead observed from 39,800 M<sup>-1</sup> cm<sup>-1</sup> to 73,200 M<sup>-1</sup> cm<sup>-1</sup>, again 4 x greater than predicted.

The increase in absorbance intensity of >84 % at 485 nm was far greater than that expected in each dimer indicates that the sfGFP chromophore has an increased ability to absorb light after SPAAC. This observation cannot be explained by either concentration error as the absorbance of the haem porphyrin at 403 nm remains consistent to monomeric Cytochrome b<sub>562</sub>. While the absorbance increase was most notable in GFPb<sup>148-71</sup>, where the spectral shift in chromophore absorbance from 492 nm to 485 nm was recorded. This change is consistent with de-protonation of Y66 of the chromophore as described previously (see Chapter 1.3.6)(33) and, this indicates that the chromophore environment of sfGFP in GFPb<sup>148-71</sup> has been altered by cycloaddition <sup>202</sup>.

#### 7.2.10. Absorbance of GFPb<sup>204-71</sup> in Reducing Conditions

A target of this investigation was to establish if the dimeric protein was sensitive to oxidation. To determine if dimeric protein retained oxidation sensitivity GFPb<sup>204-71</sup> was



**Figure 7.19. Relative Spectral Change of GFPb<sup>204-71</sup> Associated with Reducing Conditions.** Oxidised (Dotted Line) and reduced (Solid Line) GFPb<sup>204-71</sup>. Reducing conditions stimulated with 10-fold molar excess of DTT. Data shown as molar extinction coefficients and insert displays an enhanced view of wavelengths 520-600 nm. The wavelengths associated with each absorbance peak are shown for reference.

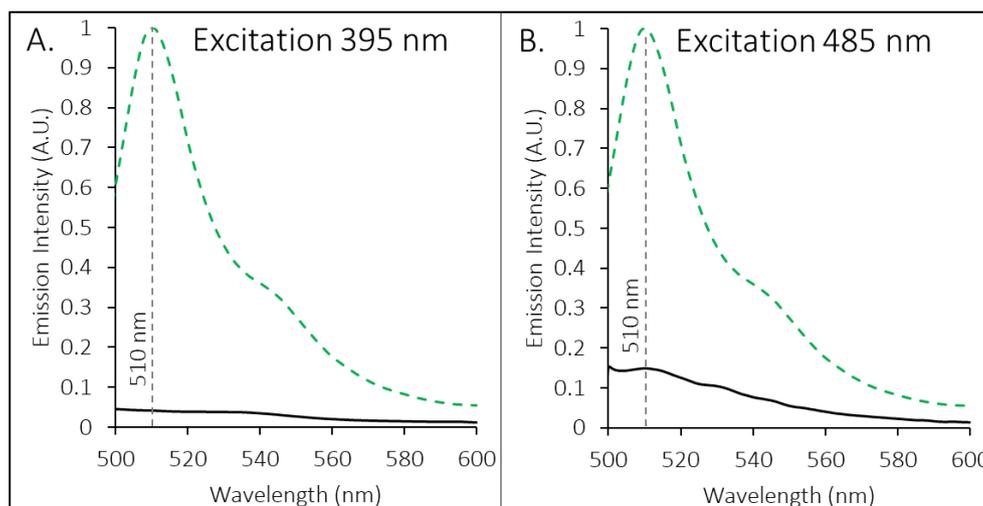
incubated with a 10-fold molar excess of DTT. The recorded absorbance spectrum of reduced GFPb<sup>204-71</sup> (Figure 7.19), indicates protein is similarly sensitive to redox conditions as cytochrome *b*<sub>562</sub><sup>WT</sup> (Figures 7.3). The Soret peak of GFPb<sup>204-71</sup> was recorded to shift in reducing conditions from a  $\lambda_{\text{MAX}}$  418 nm to  $\lambda_{\text{MAX}}$  426.5 nm, whilst also increasing in intensity, from 105,200 M<sup>-1</sup> cm<sup>-1</sup> in oxidised GFPb<sup>204-71</sup> to 162,000 M<sup>-1</sup> cm<sup>-1</sup>. Additionally, both  $\alpha$  and  $\beta$  peaks indicated an increase in relative absorbance intensity, with the  $\beta$  peak at 561.5 nm displaying the most significant increase, consistent with that expected with cytochrome *b*<sub>562</sub><sup>WT</sup>.

The increase of absorbance at 561 nm of both GFPb<sup>148-71</sup> and GFPb<sup>204-71</sup> is noteworthy as this overlaps sfGFP emission at 510 nm. It is expected that if communication exists between chromophores when dimerised, emission of sfGFP at 510 nm will be decreased in reducing conditions due to increased absorbance of GFPb<sup>204-71</sup> in this range. Furthermore, as the sfGFP chromophore is isolated from the solvent by the  $\beta$  barrel, any change observed in sfGFP emission would, therefore, be reflective of the change in haem absorbance<sup>466</sup>.

### 7.2.11. Excitation and Emission of GFPb<sup>148-71</sup> and GFPb<sup>204-71</sup>

Results obtained in Chapter 6, (Figure 6.23), indicated that energy transfer between the chromophores sfGFP and mCherry did not increase with proximity, however, emission of GFPCH<sup>204-198</sup> at 510 nm was enhanced upon dimerisation (Figure 6.21). As demonstrated in Figure 7.18, the molar extinction coefficient of each dimer at 485 nm increased. Therefore, it was anticipated that an increase in the emission of each sfGFP chromophore would be observed upon excitation at 485 nm, as the dimer had a greater ability to absorb light than monomeric protein, consistent with that observed from excitation of GFPCH<sup>204-198</sup> (Figure 6.21).

To assess emission change produced upon dimerisation of sfGFP<sup>148SCO</sup>, both monomer and dimer were excited at both 395 nm and 485 nm and corrected to a 510 nm emission intensity of 1 A.U. (Figure 7.20). Upon the excitation of GFP<sup>b148-71</sup> at 395 nm a drastically reduced emission spectrum was observed with an intensity of 0.04 A.U. at 510 nm and no clear emission maximum (Figure 7.20, A). As there was no clear peak of emission, it is reasoned that emission was quenched at 395 nm excitation, and emission observed is from naturally fluorescent amino acids, and not from the chromophore itself.

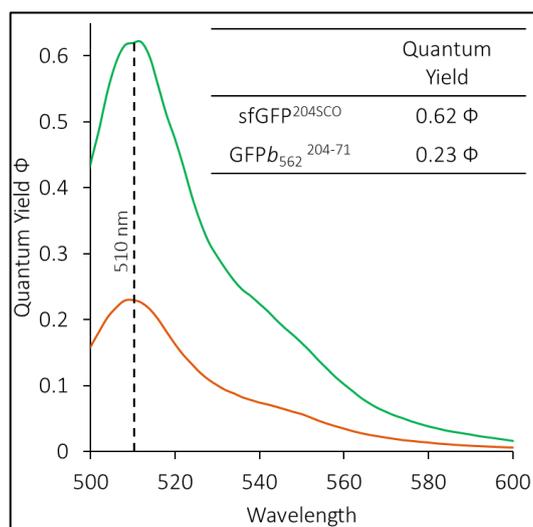


**Figure 7.20. Emission of GFP<sup>b148-71</sup> After Excitation at Both 395 nm and 485 nm.**

Emission of 1  $\mu$ M of sfGFP<sup>148SCO</sup> (Green) and GFP<sup>b148-71</sup> (Black) excited at both 400 nm (A) and 485 nm (B). Spectra corrected to a sfGFP<sup>148SCO</sup> 510 nm emission intensity of 1 AU

In comparison excitation of GFP<sup>b148-71</sup> at 485 nm did initiate a clear emission maximum at 510 nm with an emission intensity of 0.15 A.U., reduced 85% in comparison with monomeric sfGFP<sup>148SCO</sup>. Again, this decrease in emission is reasoned to indicate quenching of sfGFP by the proximity of haem in cytochrome *b*<sub>562</sub>, as dimerisation was the only change applied to protein. Both observations are indicative that the fluorescence of GFP<sup>b148-71</sup> is considerably diminished in comparison to sfGFP<sup>148SCO</sup>, despite the increased molar extinction coefficient of the sfGFP chromophore observed previous (Figure 7.18).

In contrast, the excitation of GFPb<sup>204-71</sup> at 485 nm did stimulate a clear emission of dimer at 510 nm. Therefore, the quantum yield of GFPb<sup>204-71</sup> was calculated to quantify fluorescence observed (see Chapter 2.5.6) (Figure 7.21). The obtained quantum yield of GFPb<sup>204-71</sup> was 0.23  $\Phi$  and was 63% reduced from that of sfGFP<sup>204SCO</sup>. As emission was expected to decrease 50% in intensity (0.31  $\Phi$ ), this value is lower than that expected of a chromophore population which was halved in concentration. Furthermore, this decrease contrasts the increased molar extinction coefficient of GFPb<sup>204-71</sup> at 485 nm and is likely indicative of quenching of sfGFP when protein is dimerised.



**Figure 7.21. Quantum Yield of GFPb<sup>204-71</sup> at 510 nm.**

Emission of 1 $\mu$ M GFPb<sup>204-71</sup> (Orange) and sfGFP<sup>204SCO</sup> (Green) after excitation at 485 nm. Insert displays the quantum yield of both proteins at 510 nm emission.

To some extent, both GFPb<sup>148-71</sup> and GFPb<sup>204-71</sup> were quenched upon dimerisation. This effect was greatest when GFPb<sup>148-71</sup> was excited at 400 nm, where a reduction of 85% emission intensity was observed at 510 nm. Neither dimeric protein stimulated the same decrease in emission intensity at 485 nm as the 100% emission quenching reported by Arpino *et al.* (2012). Instead, each sfGFP chromophore was still observed to emit light at 510 nm when dimerised, albeit to a lesser extent. Comparison of both dimeric emissions after 485 nm, furthermore, indicates that the quenching observed is not equal between the two sites of sfGFP interface as emission at 510 nm was quenched 85% within GFPb<sup>148-71</sup> and 63% within GFPb<sup>204-71</sup>. As the residue at which molecules are linked influences chromophore distance, and orientation, it is reasoned that the different intensities of emission observed are indicative of FRET quenching, due to the varied dipole-dipole alignment of chromophores. This effect is not likely due to the distance between chromophores, as chromophore distance would remain relatively consistent

when protein is linked at either residue 148 or residue 204 as the residue pair are on parallel  $\beta$  strands.

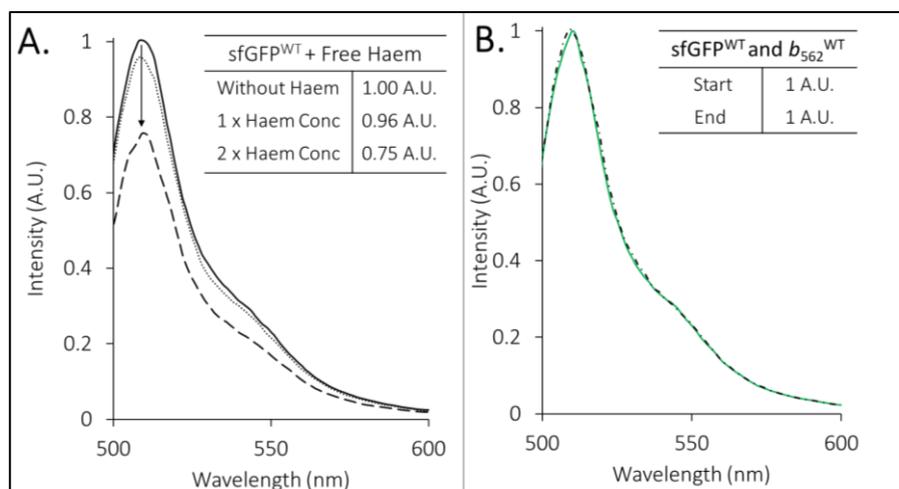
### 7.2.12. The Sensitivity of sfGFP<sup>WT</sup> Emission to Haem

To establish if the quenching observed in Figures 7.20 and 7.21 is specific to the dimeric attachment of sfGFP and cytochrome  $b_{562}$ , two control experiments were established. The first control was to assess if sfGFP<sup>WT</sup> emission was influenced by the presence of haem porphyrin in solution. While the second control assessed if the combination of sfGFP<sup>WT</sup> and cytochrome  $b_{562}$ <sup>WT</sup> at an equal concentration in solution, decreased the emission of sfGFP.

A decrease in emission of 4% in sfGFP emission was observed upon the overnight incubation of sfGFP with an equal molar excess of free haem, and 25% upon incubation with a two-fold molar excess of free haem (Figure 7.22, A). At each concentration, the emission decrease was not as intense as that observed in GFP $b^{148-71}$  when excited at both 395 nm (Figure 7.20), and complete emission quenching was not replicated. As haem porphyrin is not capable of emitting light it does not contribute positively to the signal recorded within the spectrum. Therefore the decrease in sfGFP emission is indicative of fluorescence quenching rather than FRET, which in this instance is triggered upon the application of haem. It is theorised that quenching of emission is observed due to the association between free haem porphyrin and the poly-histidine tag of sfGFP, which has previously been reported<sup>311</sup>. However, as cytochrome  $b_{562}$ <sup>71AZF</sup> is haem saturated before SEC and combination with sfGFP<sup>148SCO</sup>, it is reasoned that there is minimal free haem present within the dimeric protein to associate in this way.

The combination of sfGFP<sup>WT</sup> with cytochrome  $b_{562}$ <sup>WT</sup> at a 1:1 molar ratio stimulated no change in sfGFP emission (Figure 7.22, B). Therefore, it can be assumed that there is no native association between wild type proteins and emission is uninfluenced by the presence of haem when it is coordinated with cytochrome  $b_{562}$ . This indicates that quenching observed in both GFP $b^{148-71}$  and GFP $b^{204-71}$  is not induced by the combination of cytochrome  $b_{562}$  with sfGFP in solution, but rather is induced by the proximity of haem porphyrin. Thus, when protein is free in solution chromophores are distant, however, in dimeric protein haem is coordinated with cytochrome  $b_{562}$  and chromophores are within an estimated distance of 20.5 Å (Figure 7.8).

This then mediates the transfer of emission energy from sfGFP to haem by way of FRET, which is inversely proportional to the 6<sup>th</sup> power of chromophore distance <sup>467</sup>.



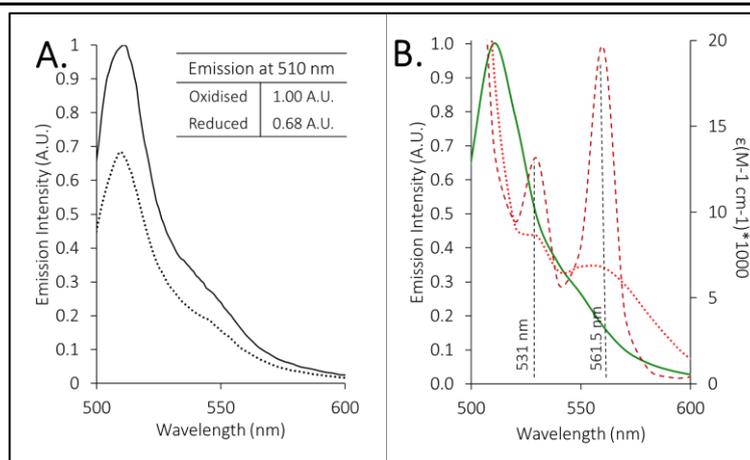
**Figure 7.22. Effect of Haem binding on Emission of sfGFP.**

**A**, emission spectra of 1 μM sfGFP (**Black**) in comparison to sfGFP incubated with 1 μM (**Dotted**) and 2 μM of free haem (**Dashed**) overnight. **B**, Equal concentration of 1 μM cytochrome *b*<sub>562</sub><sup>WT</sup> and 1 μM sfGFP<sup>WT</sup> before (**Green**) and after (**Dashed**) overnight incubation together. Data corrected to an emission intensity of sfGFP of 1 A.U. at 485 nm. Insert displays absolute value of emission intensity of emission at 510 nm.

### 7.2.13. Redox Sensitivity of GFPb<sup>204-71</sup> Emission

The sensitivity of GFPb<sup>204-71</sup> emission to reducing conditions was investigated by measurement of emission at 510 nm after excitation at 485 nm in both oxidising and reducing conditions. Emission of the dimer was first recorded in oxidising conditions before incubation of the dimer with a 10-fold molar excess of DTT to stimulate reducing conditions. As anticipated, GFPb<sup>204-71</sup> emission at 510 nm decreased a further 32 % in the reduced protein when compared to the oxidised form, (see Chapter 2.5.5) (Figure 7.23, A).

The decrease in emission supported the expectation that an increased absorbance of reduced haem between 530 nm and 570 nm increased the efficiency of energy transfer of energy between chromophores (Figure 7.23, B). This furthermore is evidence that the sfGFP chromophore, in dimeric GFPb<sup>204-71</sup>, is sensitive to redox conditions, whereas, monomeric sfGFP<sup>WT</sup> is not <sup>374,468</sup>. Thus, change in emission is indicative of further functional communication between chromophores as haem porphyrin coordinated to cytochrome *b*<sub>562</sub> is redox-sensitive.



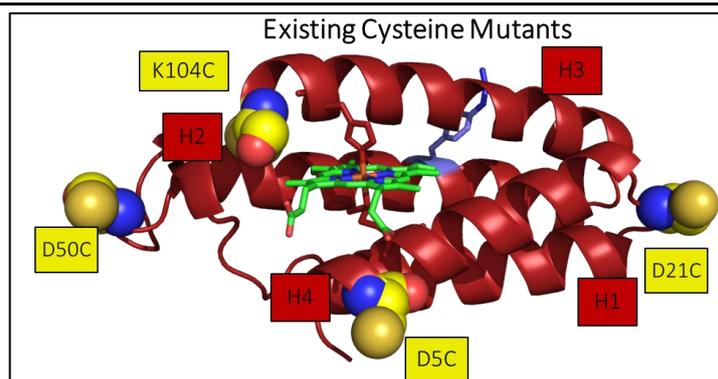
**Figure 7.23. Effect of Reducing Conditions on the Emission of GFPb<sup>204-71</sup>**

**A**, emission of 1  $\mu$ M GFPb<sup>204-71</sup> in oxidising conditions (**Black**) and after 1-hour incubation with 10-fold molar excess of DTT (**Dots**). **B**, overlay of cytochrome *b*<sub>562</sub><sup>71AZF</sup> absorbance in oxidising (**Dots**) and reducing conditions (**Dashes**) against emission spectra of sfGFP (**Green**). Data corrected to a 510 nm emission of 1 A.U. after excitation at 485 nm in oxidising conditions.

#### 7.2.14. The Rationale for the Double Mutagenesis of Cytochrome *b*<sub>562</sub><sup>71AZF</sup>

Thus far, this chapter has focused on the application of SPAAC in the formation of a functionally linked dimeric protein. Therefore, as the emission of GFPb<sup>204-71</sup> has been established to be sensitive to both haem proximity (Figures 7.20 and 7.21) and redox conditions (Figure 7.23), the following work will investigate the formation of higher-order oligomers, through the combination of both SPAAC and disulphide bridging, in the formation of one tetrameric protein.

Disulphide bond formation, in principle, is simple; however, utilisation of such a bond would allow for the formation of a tetrameric protein that is redox-sensitive. It is anticipated that any functional change observed upon the generation of tetramer will be reversible by separation of tetrameric protein to dimeric in reducing conditions, and this will provide further functional sensitivity to environmental reduction<sup>410</sup>. Cytochrome *b*<sub>562</sub> has no surface cysteines<sup>275</sup> but has been previously mutated to form four cytochrome *b*<sub>562</sub> cysteine mutants (D5C, D21C, D50C, and K104C) within the DDJ LAB (Figure 7.24). These mutations were initially established as a means of protein attachment to electrodes<sup>299,469,470</sup>, although, self dimerisation was reported in oxidative conditions<sup>299</sup>. Thus, work commenced upon the selection of one such residue, which would stimulate homo-dimerisation while exerting neutral influence on the interface between 71 AZF and sfGFP.



**Figure 7.24. Possible Residues for Cysteine Mutagenesis for Formation of Homodimers.** Cytochrome  $b_{562}^{\text{WT}}$  (PDB 1QPU) shown as cartoon (Red) with coordinated haem (Green sticks) shown as representation of haem binding pocket. The 4 existing locations of possible cysteine mutagenesis represented are mutated cysteines (Yellow Spheres). Helices are numbered for reference (Red Boxes)

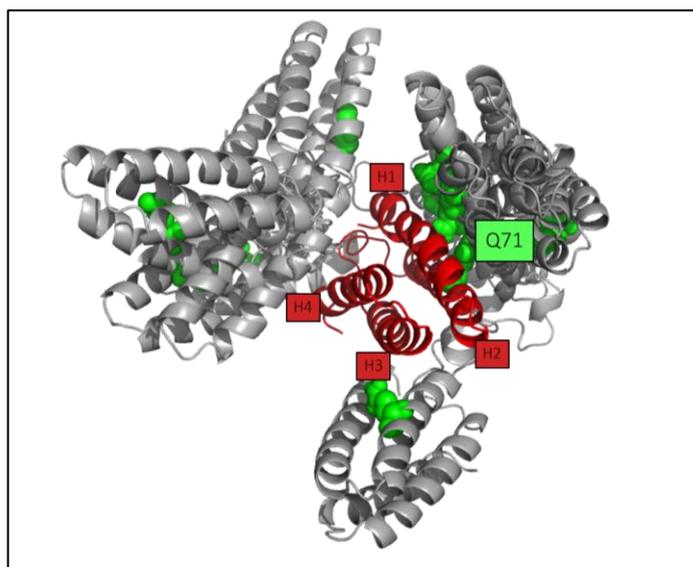
The formation of tetrameric protein was dependent on compatibility between the two protein interfaces to assemble with separate protein appendages and exert no interference upon one another. Consequently, it was necessary to carefully select a surface residue for cysteine mutagenesis upon cytochrome  $b_{562}$  that would be congruent to SPAAC at 71 AZF. The following criteria rationalised residue selection: 1, The residue must be in a potential region of self-association in ClusPro alignment. 2, It must not conflict with the established interface between sfGFP and cytochrome  $b_{562}$ , and 3, The residue must not block the haem binding pocket. Therefore, if an existing cysteine mutant fulfilled all criteria (Figure 7.24), this mutant would be used in the generation of a double cysteine mutant.

To determine which regions of cytochrome  $b_{562}$  were most likely to self-align ClusPro “Dock” simulation was performed with cytochrome  $b_{562}$  (PDB 1QPU) acting as both receptor and ligand. It was reasoned regions of cytochrome  $b_{562}$  which were most compatible for self-association would be highlighted by ClusPro simulation, and regions that were least compatible would not be represented. ClusPro simulation was used once more for this alignment, rather than analysis of the literature, as the methodological approach had been proven successful twice previously in this thesis (Figures 6.5, and 7.6).

Arrangement of each of the top 10 “balanced” models in PyMOL demonstrated that alignment between molecules was most likely to occur in one of two orientations (Figure 7.25). These orientations were located firstly between helices 1 and 4, and secondly between helices 1 and 2. However, all clusters which interface at the surface between helices 1 and 2 directly

overlapped with 71 AZF (Figure 7.25). This strategy was used over the identification of potential crystal contact sites as it allowed for the removal of haem porphyrin within simulation and thus predicted potential *apo* protein contacts. As residue 71 was intended for SPAAC attachment to sfGFP, these alignments were therefore eliminated from the investigation. Subsequently, the best region of cysteine incorporation is rationalised to be on either helix 1 or 4 (Figure 7.25).

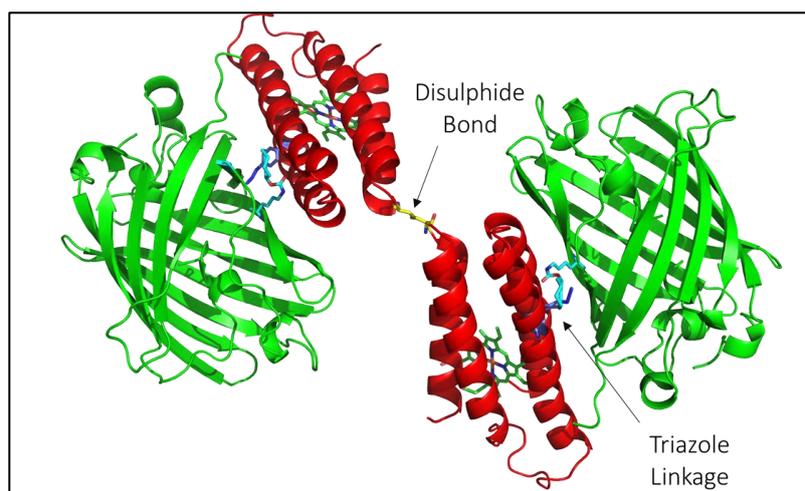
Examination of existing cysteine mutants in PyMOL displays that all four cysteine mutations are located on either helix 1 or 4 and are anterior to residue 71 (Figure 7.24).



**Figure 7.25. The Top 10 “Balanced” Clusters of Cytochrome  $b_{562}$  Alignment**

Obtained interfaces observed within the top 10 models as ranked by ClusPro alignment of cytochrome  $b_{562}^{\text{WT}}$ , PDB 1QPU, (Red) as both ligand and receptor, with 71 AZF incorporated back into receptor structure in PyMOL (Green Spheres).

However, all residues were located on the same plane as the opening to the haem binding site. Residue Asp 21 was selected as the most favourable site for cross-reactivity as it was situated at the greatest distance from the haem binding site (Figure 7.24). It was argued that attachment *via* residue 21 would allow for the association between cytochrome  $b_{562}$  monomers without impeding haem sensitivity to redox conditions. Simulation of the proposed tetramer alignment was generated using two copies of Model RD1 (Figure 7.8) aligned at residue 21 in PyMOL (Figure 7.26). As all previously discussed criteria were met with this alignment, it was reasoned that this was the best mutation to begin tetrameric formation.

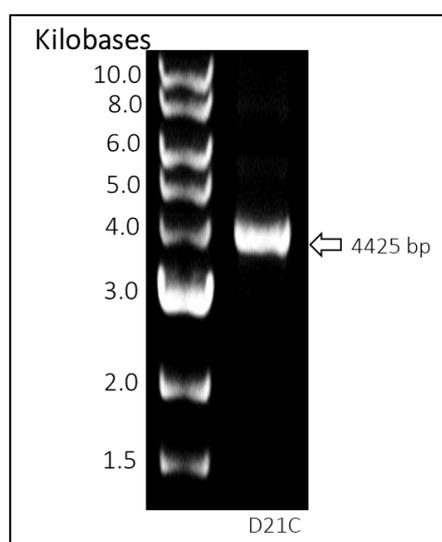


**Figure 7.26. The Proposed Alignment of Protein in Tetramer Formation.**

Representation of proposed tetramer formed when two copies of  $\text{GFPb}^{148-71}$  (Model RD1) are linked by disulphide bond at residue 21. Cytochrome  $b_{562}^{\text{WT}}$  (PDB 1QPU) shown as cartoon (Red) with conjugated haem shown as sticks coloured by element. Mutations 21 Cysteine (**Yellow Sticks**) and 71 AZF (**Purple Sticks**) shown on cytochrome  $b_{562}$ , whilst 148 SCO-K shown on sfGFP (**Teal Sticks**).

### 7.2.15. Mutagenesis, Expression and Purification of Cytochrome $b_{562}^{\text{DM}}$

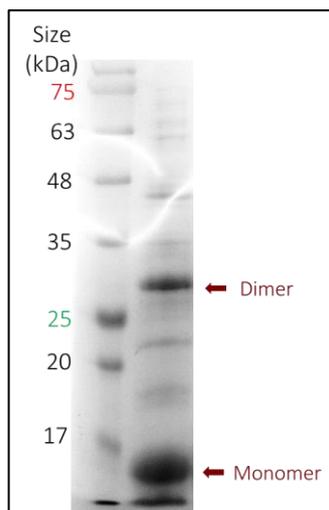
The generation of cytochrome  $b_{562}^{71\text{AZF-21CYS}}$  was performed by SDM of  $\text{pBAD-}b_{562}^{71\text{AZF}}$  plasmid, using mutagenesis oligonucleotides as first described by Reddington (2013) (see Chapter 2.2.2)<sup>299</sup>(Figure 7.27, A). DNA of a size consistent with  $\text{pBAD-}b_{562}^{71\text{AZF}}$  (4425bp) was observed in the sample amplified with D21C mutagenesis oligonucleotides, and the sample was subsequently phosphorylated, ligated, and transformed into *E. coli* DH5 $\alpha$  cells for DNA



**Figure 7.27. Cysteine Mutagenesis of  $\text{pBAD-Cytochrome } b_{562}^{71\text{azf}}$ .**

Amplified DNA observed from amplification of cytochrome  $b_{562}^{71\text{AZF}}$  with D21C mutagenesis oligonucleotides, sized at approximately 4,425 bp, consistent with cytochrome  $b_{562}^{\text{WT}}$ .

propagation. Plasmid sequence was confirmed by Eurofins Genomics (see Chapter 2.2.5) and the double mutant will herein be referred to as cytochrome  $b_{562}^{DM}$ .



**Figure 7.28. Non-Reducing Gel of Cytochrome  $b_{562}^{DM}$  Expressing Cell Lysate.**

Gel shows presence of two bands of over expression within cell lysate indicated by red arrows: one consistent with monomeric cytochrome  $b_{562}^{DM}$  at 12 kDa and the second consistent with dimeric cytochrome  $b_{562}^{DM}$  at 24 kDa.

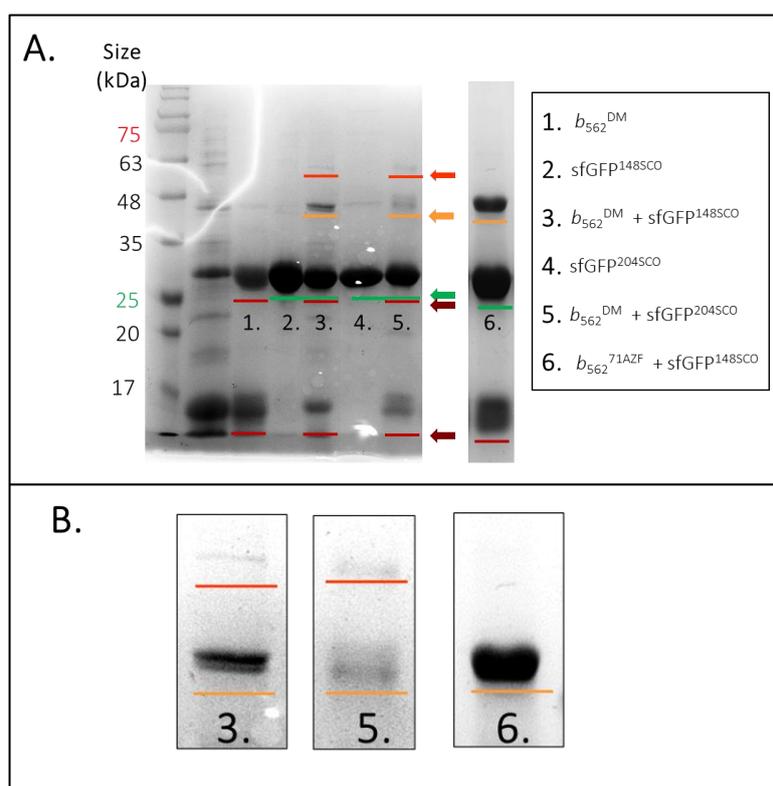
Once generated cytochrome  $b_{562}^{DM}$  was co-transformed with pDULE plasmid into TOP10™ *E.coli* cells for expression of the protein. The homo-dimerisation of expressed protein was confirmed by non-reducing SDS PAGE analysis (Figure 7.28). Two proteins bands consistent with both monomeric (12 kDa) and dimeric cytochrome  $b_{562}$  (24 kDa) were observed in the cell lysate. Protein was precipitated by 90% ammonium sulphate cut and was resuspended in 2 mL of  $\text{CuSO}_4$  5 mM Tris 50 mM pH 8 to promote further disulphide formation<sup>471</sup> and purified as described previously (see Section 7.2.4.).

#### 7.2.16. Oligomeric Assembly of sfGFP and Cytochrome $b_{562}^{DM}$

Trial oligomerisation of purified cytochrome  $b_{562}^{DM}$  was performed by mixing 100  $\mu\text{M}$  of cytochrome  $b_{562}^{DM}$  with either sfGFP<sup>148SCO</sup> or sfGFP<sup>204SCO</sup>. Protein was incubated overnight at 42 °C, before non-reducing SDS PAGE analysis (Figure 7.29, A). SDS PAGE of the sample following incubation indicated the formation of an additional band sized at approximately 51 kDa, which was larger than that of dimeric protein (Figure 7.29, Lanes 3 and 6). This band is consistent with trimeric protein formed from two molecules of cytochrome  $b_{562}^{DM}$  and one of sfGFP (12 +12 +27). No band is observed, in any lane, which was consistent with that of tetrameric protein at ~78 kDa. The formation of trimeric protein (Figure 7.29, Lanes 3 and 6) is

evidence, however, that SPAAC induced cycloaddition can be combined with disulphide bond formation for the generation of higher-order oligomers.

As protein oligomers in nature commonly form by the association between multiple identical subunits, the addition of each subsequent monomer is as likely to occur as the first formation of a homo-dimer<sup>406,472</sup>. However, the combination of multiple heterogeneous monomers to form one tetramer is not equally likely at each stage of assembly, as each contributing reaction is governed by a unique rate. Therefore, it is likely that cycloaddition between both azide (cytochrome  $b_{562}$ ) and alkyne (sfGFP) occurs at a different rate to that of disulphide bond formation (between cytochrome  $b_{562}$ ). Consequently, when monomers are mixed one protein product is more likely to form than the other, and the secondary reaction may not occur as favourably as the initial reaction.



**Figure 7.29. Non-Reducing Gel of Cytochrome  $b_{562}^{DM}$  Mixed With sfGFP.**

**A**, non-reducing SDS PAGE of tetrameric assembly between sfGFP and cytochrome  $b_{562}^{DM}$ . Formation of novel protein band sized at approximately 51 kDa highlighted (**Dark Orange**) in wells 3 and 5. Dimeric GFP $b^{148-71}$  and GFP $b^{204-71}$  highlighted also at 39 kDa (**Light Orange**). Identity of protein within each well shown in key. Well 6 is presented for comparison and displays formation of dimer from Figure 7.15. **B**, close-up view at range of 40-63 kDa from wells 3, 4, and 6, with newly formed bands highlighted.

## Chapter 7. Functional Fluorescence Quenching by Resonance Energy Transfer in Dimeric Protein.

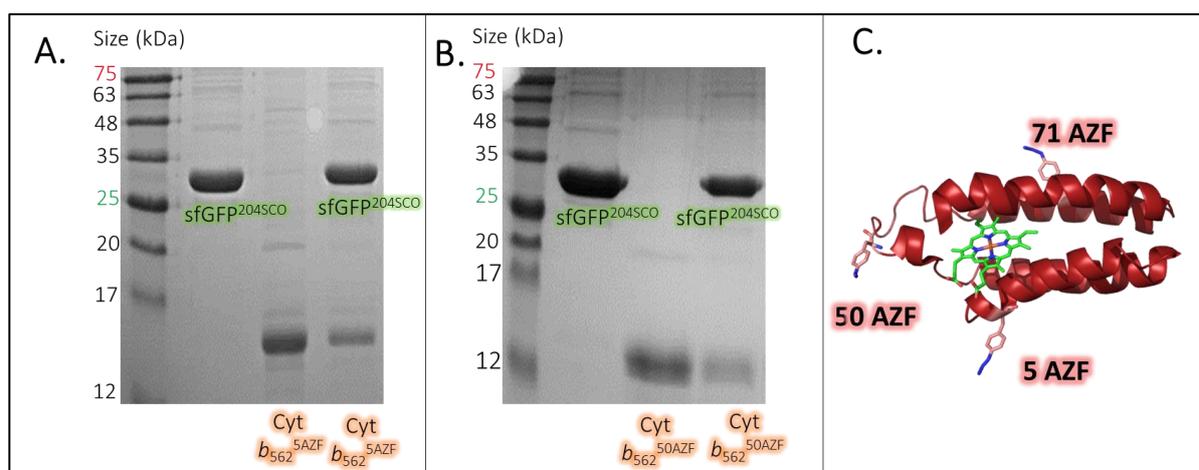
---

Therefore, it is reasoned that the separation of protein assembly into two different reaction steps is necessary to increase the yield of oligomeric protein formed<sup>473</sup>. To achieve this in the future formation of tetramer will be split into multiple steps. The first step will be for the formation and purification of a single SPAAC linked dimer as SPAAC is the rate limiting step of this reaction and dependant AZF not being exposed to light. Then secondly, the pure dimer will be incubated with CuSO<sub>4</sub> to stimulate disulphide formation and form tetramers without the potential of intermediate trimer formation. As the creation of covalent triazole linkage cannot be reverse this will allow for the formation of tetramer from dimer in a single reaction step.

## 7.3. Conclusion

### 7.3.1. Discussion of Oligomerisation of Cytochrome $b_{562}$ and sfGFP

The initial aim of this investigation was to establish if a more complex dimeric assembly of protein was viable by SPAAC attachment. Subsequently, by the application of interface prediction and nAA chemistry two non-symmetrical dimers were formed,  $\text{GFP}b^{148-71}$ , and  $\text{GFP}b^{204-71}$  (Figure 7.14). Further to this, the combination of both cycloaddition and disulphide bond formation allowed for the formation of non-natural mediated trimeric protein (Figure 7.29). The formation of dimer observed in the combination of cytochrome  $b_{562}^{71\text{AZF}}$  with  $\text{sfGFP}^{204\text{SCO}}$  (Figure 7.14) contrasts previous observation within the DDJ lab, wherein, the combination of the cytochrome mutants, 5 AZF and 50 AZF, with  $\text{sfGFP}^{204\text{SCO}}$ , did not yield dimeric protein (Figure 7.30). The ability to form dimer at residue 71 AZF in cytochrome  $b_{562}$ , therefore demonstrates, the importance of interface prediction when selecting possible residues for nAA incorporation, and subsequent SPAAC linkage.



**Figure 7.30. SDS PAGE of  $\text{sfGFP}^{204\text{SCO}}$  mixed with Alternative Cytochrome AZF Mutants.**

**A**, combination of 100  $\mu\text{M}$  cytochrome  $b_{562}^{5\text{AZF}}$  and 100  $\mu\text{M}$   $\text{sfGFP}^{204\text{SCO}}$ . **B**, combination of 100  $\mu\text{M}$  cytochrome  $b_{562}^{50\text{AZF}}$  and 100  $\mu\text{M}$   $\text{sfGFP}^{204\text{SCO}}$ . The presence of a dimeric protein band at approximately 39 kDa is not observed in either combination. **C**, the relative positions of both residues 5 AZF and 50 AZF in comparison to residue 71 AZF on cytochrome  $b_{562}$ . Gels by Jacob Pope (2016).

One limitation of this investigation, however, was the inability to form tetrameric protein (Figure 7.29). Although, the tetrameric assembly of protein is reasoned to be possible in the future as both forms of protein linkage were compatible in the formation of hetero and homodimers. It is likely that the inability to form a tetrameric protein is due to the time

constraints of this project and with future refinement of methodology, tetrameric protein formation is feasible.

The formation of dimeric protein between sfGFP and cytochrome *b*<sub>562</sub> yielded a click efficiency of up to 25% of monomeric protein (Table 7.4). This was of note as it is comparable to the efficiency of 28% dimeric protein yield reached between sfGFP and mCherry (GFPCH<sup>198-204</sup>) (Figure 6.17). This illustrates the importance of interface prediction, as even in structurally diverse protein, linkage is possible, but only if nnAA incorporation is situated within a compatible interface (Figure 7.30).

One significant influence of dimer formation was that of incubation temperature (Figure 7.14). This was unexpected as in previously published dimeric study incubation at room temperature was sufficient for substantial dimer formation<sup>260,265</sup>. It is probable temperature is not directly an influencing factor in azide/alkyne reactivity but influences the self-association of sfGFP<sup>169,465</sup>. Self-association of sfGFP's hydrophobic residues V206, L221, and F223 (Figure 7.16) indicate that homodimeric association of sfGFP is most likely at this interface. This explains why the higher concentration of dimer is formed at sfGFP residue 148 rather than 204 (Figure 7.14.), and additionally, why it is difficult to separate monomer from dimer at room temperature (Figure 7.15). It is possible that this region of self-association may be utilised in future to aid oligomeric assembly.

### 7.3.2 Functional Communication between Chromophores

Combination of two chromophores, haem porphyrin and the sfGFP chromophore in both GFPb<sup>148-71</sup> and GFPb<sup>204-71</sup> was confirmed by absorbance at both 418 nm and 485 nm (Figure 7.18). This contrasts findings by Arpino *et al* (2012) which recorded a shift in Soret peak absorbance to 422 nm in constructed protein but is consistent with results observed by Takeda *et al.* (2001).

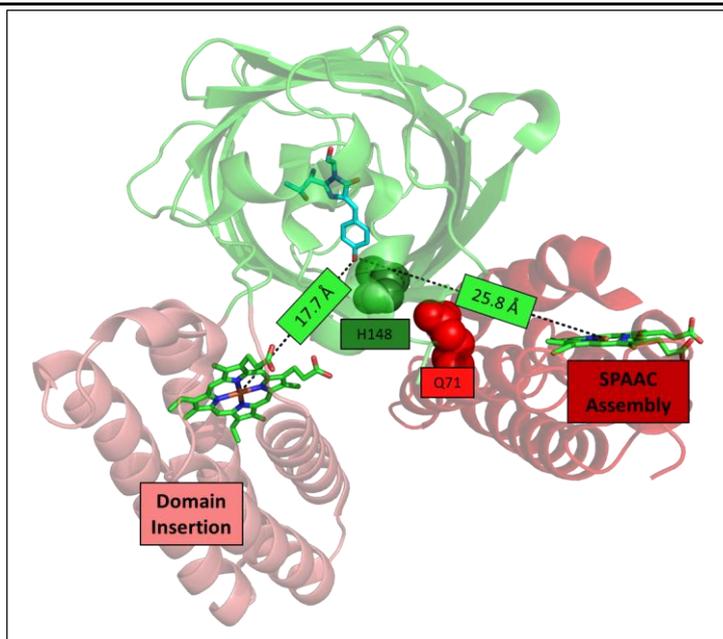
One observed difference in SPAAC linked proteins is the increase in relative absorbance at 485 nm of the sfGFP chromophore. This mirrors results observed in cycloaddition of sfGFP<sup>204SCO</sup> in Chapter 6 (Figure 6.20) and observation by Worthy *et al* (2019) whereby cycloaddition of residues 148 and 204 resulted in a net increase in chromophore absorbance. The effect is reasoned to be due to the modified mobility of molecular water at the site of the interface between the two molecules. This, in turn, induces water channel re-arrangement,

which modifies the solvation around the chromophore of sfGFP, increasing the permanence of interaction with water (see Chapter 1 Figure 1.10). Although the effect was previously rationalised to be specific to shared water channels formed *via* homo-dimeric linkage of sfGFP<sup>202</sup>. This effect can be explored further in future by the dimerisation of Cytochrome *b*<sub>562</sub><sup>71AZF</sup> to sfGFP at an alternative site of interface, for example, 180° from residue 148 on the  $\beta$  barrel.

Both dimers formed were fluorescently active and capable of emission at 510 nm after excitation at 485 nm (Figures 7.20 and 7.21). Whilst the emission of GFP*b*<sup>204-71</sup> was subsequently proven to be influenced by reducing conditions (Figure 7.23). The decrease in emission observed in each dimer (Figures 7.20 and 7.21) is believed to be a direct result of FRET emission quenching by haem porphyrin<sup>453</sup>, as each featured a higher molar extinction coefficient at 485 nm in comparison to monomeric sfGFP (Figure 7.18). Quenching of emission is achieved by the presence of the iron ion ligand within the haem porphyrin which is only partially filled with electrons which undergo ligand field d-d transition and result in the strong absorbance of fluorescence<sup>452</sup>.

Complete signal quenching of 100%, as previously observed by Arpino *et al* (2012), was not recorded in this investigation. Instead, emission quenching ranged from 63% in GFP*b*<sup>204-71</sup> when excited at 485 nm to 85% in GFP*b*<sup>148-71</sup> when excited at 400 nm. This difference in quenching efficiency can be explained when chromophore distance is contrasted (Figure 7.31). As quenching is known to obey the principles of FRET, energy transfer is inversely proportional to the 6<sup>th</sup> power of chromophore distance, orientation, and dipole alignment<sup>453</sup>. Thus, as the protein generated by Arpino *et al* (2012) (PDB 3U8P) features a closer alignment of chromophores of 17.7 Å in comparison to the predicted distance of 25.8 Å by SPAAC it is expected that distance between chromophores enhances FRET efficiency in the Arpino assembly.

The increased proximity between chromophores observed by Arpino *et al* (2012) is reasoned to be an effect of both domain insertion, in addition to the alternative interface sampled between cytochrome *b*<sub>562</sub> and sfGFP (Figure 7.31). As no ClusPro predicted model of protein alignment (Figure 7.6) featured proximity between the haem binding pocket and the  $\beta$ -barrel of sfGFP, this site of interface does not appear to be suitable for replication *via* SPAAC.



**Figure 7.31. Comparison of Chromophore Alignment of GFP and Cytochrome  $b_{562}$  Achieved by Domain Insertion and SPAAC Assembly.**

Alignment of crystal structure (3u8p) of EGFP and cytochrome  $b_{562}$  linked by domain insertion by Arpino *et al* (2012) with predicted alignment between sfGFP (2b3p) and cytochrome  $b_{562}$  (1QPu) obtained by RosettaDock (Figure 7.7). GFP (**Green**) used as centre of alignment with chromophore shown centrally (**Teal Sticks**). The measured distance between Tyr 66 hydroxyl group and central iron of each haem chromophore overlain on top (**Green Boxes**). Residues mutated for non-natural incorporation shown as spheres and labelled. Representation of cytochrome  $b_{562}$  aligned according to domain insertion shown as **Light Red** and according to computerised modelling for SPAAC assembly shown as **Dark Red**.

### 7.3.3. Sensitivity of GFP $b^{204-71}$ to Reducing Conditions

In addition to being FRET capable, GFP $b^{204-71}$  was also observed to be sensitive to reducing conditions (Figure 7.23). The incubation of GFP $b^{204-71}$  in reducing conditions stimulated a shift in haem Soret peak absorbance from 418 nm to 426.5 nm (Figure 7.19), while additionally, emission was also reduced by 32% (Figure 7.23). The sensitivity of emission to environmental oxidation is significant as it offers the potential to measure oxidation by fluorescence and enables GFP $b^{204-71}$  to function as an environmental biosensor.

This is significant, as the ability to measure intracellular oxidation is limited, and many research groups are focused on the formation of biologically compatible redox probes<sup>185,474–476</sup>. Furthermore, the accumulation of reactive oxygen species, such as H<sub>2</sub>O<sub>2</sub>, is critically important for intracellular processes and can lead to important cellular signalling events, e.g. programmed cell death<sup>477</sup>. Thus sensing of these species within cells is desirable<sup>474</sup>. As SPAAC

click chemistry is biorthogonal this enables the possibility of future investigation to utilise GFP<sup>b204-71</sup> as a biological reporter intracellularly for the detection of oxidative conditions by way of fluorescence emission.

### 7.3.4. Concluding Statement

Modern biology is heavily reliant upon light signalling for investigation and monitoring of the internal cell environment. This chapter presents a SPAAC induced energy transfer system which can be used for the remote probing of oxidising conditions, whilst highlighting the diversity of monomers and functionalities which can be combined in molecular assembly. The incorporation of SPAAC compatible residues in cytochrome *b*<sub>562</sub> and sfGFP demonstrates that both functionally and structurally distinct proteins can be linked *via* SPAAC. It is feasible that in future this methodology could be applied in the linkage of more functionally or structurally diverse protein combinations for the generation of more biorthogonal systems of chromophore communication<sup>470</sup>.

## Chapter 8. Discussion

---

## 8.1. General Overview

Post-translational modification is a powerful system by which protein diversity is enhanced, and the alteration of both the structure and function of a protein can be achieved. In this thesis, the influence of both natural PTM (Chapters 3, 4 and 5), and non-natural PTM (Chapters 6 and 7) have been investigated, forming two key themes of research. Insight has been gained as to the influence of PTM exerts on the: recombinant expression (Chapter 3), interaction (Chapter 4), and flexibility (Chapter 5), of the enzyme HRP. This research has indicated that significant control of a protein's structural rigidity is induced by both glycosylation and haem binding.

Chapters 6 and 7, demonstrated that nnPTM could be used to induce oligomerisation between two sequentially (Chapter 6), and structurally (Chapter 7), diverse protein monomers. Chapter 6, demonstrated that the combination of two protein functions does not always induce a predictable synergy between functional centres. While, in contrast, Chapter 7 did provide evidence of functional communication between chromophores, which was influenced by proximity.

## 8.2. The importance of native PTM

Modern approaches to protein production, frequently utilise recombinant protein expression for many commercial <sup>478</sup>, industrial <sup>479</sup>, and pharmaceutical <sup>24</sup> purposes. As previously discussed in Chapter 1.2.4, recombinant expression of HRP would be a potentially rewarding venture for the acquisition of a commercially highly significant enzyme. The importance of the native PTM on HRP was therefore explored in Chapters 3-5 of this thesis. In these sections, it was investigated how the absence of native PTM influenced the expression, function, and rigidity of HRP.

The first section of work presented in Chapter 3 explored how the presence of both pre- and post-peptide sequences influenced the expression of HRP in the recombinant expression host *E. coli*. Results demonstrated that expression of the full-length proto-HRP sequence in *E. coli* enabled the soluble expression of HRP<sup>EC2</sup> which could be both purified and stored for use. This chapter demonstrated that while HRP<sup>EC1</sup> was soluble, without native glycans, the protein was observed to frequently precipitate from solution. Mutagenesis of

---

these surface N-linked glycan-sites in HRP<sup>EC2</sup> enabled increased protein stability and allowed for the formation of higher concentration samples without precipitation. The increased stability observed upon the mutagenesis of non-glycated asparagine residues demonstrates that the absence of PTM can itself be a stimulus of impaired protein function. Most importantly, the results presented in this chapter enabled the development of a methodology by which recombinant, non-glycosylated HRP could be expressed and purified for work in future chapters.

In future investigation, further advancement to both recombinant HRP stability and purification should be investigated. As presented by the formation of HRP<sup>EC2</sup> in Chapter 3, the removal of surface asparagine residues goes some way to tackling this issue, however, it is anticipated that further advancement is possible. It is anticipated that chemical glycosylation of recombinant HRP using nnPTM will improve protein function for future use, whilst inferring an enhanced structural rigidity to recombinant protein which was not previously feasible. Although the activity of recombinant HRP is not currently suitable for widespread application, Chapter 5, hints that glycosylation of the recombinant protein either by the transference of expression host, or synthetic glycan attachment may reduce structural flexibility and increase catalytic efficiency.

The function of *apo*-HRP as a corrective factor was constructed in the presence and absence of glycosylation in Chapter 4. Results presented in this chapter demonstrated that not only was the corrective function of *apo*-HRP decreased in the absence of glycosylation, but non-glycosylated protein also functioned equally as poorly as a corrective agent as mCherry. This provides clear evidence that the interaction of HRP, with other molecules in an immunoassay, is dependent upon the presence of glycosylation. In the absence of glycosylation, the agent is not recognised in immunoassay and does not reduce any rogue activity observed within test sera.

Chapter 5, explored the possible functional change observed from the application of PTM to assess the possible stimulus for results reported in Chapter 4. Findings presented in Chapter 5 demonstrated that HRP which is both haem saturated and glycosylated is by far the most stable form of HRP, and both haem binding and glycosylation individually, induce a degree of protein rigidity. The extent by which both function and stability of HRP were aided by the

---

presence of both forms of PTM was of surprise. Glycosylated *holo*-HRP had a catalytic turnover several orders of magnitude greater than that of non-glycosylated *holo*-HRP. Whilst non-glycosylated *apo*-HRP was disproportionately less stable than all other HRP PTM variants. The lack of stability observed in non-glycosylated *apo*-HRP is likely a contributing factor in the protein's lack of corrective function within an immunoassay. It is reasoned that in the absence of PTM, the structural composition of HRP is naturally highly flexible and unable to bind to the same protein-protein recognition sites as glycosylated *apo*-HRP. Therefore, glycosylation is an essential feature for the replication of SCF.

One additional set of experiments which would be a useful addition to this study is the comparison of multiple different forms of plant peroxidase enzymes, or a prokaryotic non-glycosylated peroxidase enzyme by REES analysis. As observed in Figure 5.16 the structures of many plant peroxidases are homologous, however, their individual glycosylation patterns are heterogeneous. The contrast of several glycosylated states and locations, will therefore indicate the degree of flexibility which is afforded to each peroxidase by specific regions of glycosylation. It is reasoned that certain residues of glycosylation are more beneficial to enforcing structural rigidity than others and maybe conserved between species. Thus, identifying these regions of conserved peroxidase glycosylation will indicate the regions which are most benefitted by the application of glycosylation. Consequently, this knowledge will identify the best target regions for the non-natural attachment of chemically synthesised glycans to HRP or other recombinant peroxidase enzymes.

### 8.3. The Use of Non-Natural PTM for Protein Attachment

In this thesis, the use of non-natural chemistry as a means of functional attachment for two pairs of naturally disparate proteins was investigated. This approach was crucially spontaneous, upon the mixing of nnAA incorporated protein, and occurred without the need for a catalyst. Interface simulation was employed to enhance the residue selection for nnAA incorporation so the site of attachment between the two molecules could be complementary to natural protein chemistry. The utilisation of non-natural PTM in this way has an advantage over the application of natural PTM as it allows for specified orientation between

---

chromophores based upon interface prediction. This technique was beneficial to the formation of the dimeric interface between sfGFP and both mCherry and cytochrome *b*<sub>562</sub>.

The application of azide and alkyne in this instance was a successful means of defining specific molecular anchoring of protein at a defined interface. Using each dimeric protein out of the context of natural systems allowed for the exact effects of interaction alone to be determined. Chapter 7 demonstrates an integrated protein function within the dimeric protein formed between sfGFP and cytochrome *b*<sub>562</sub>. This cohesive function is observed in the quenching of sfGFP emission by way of FRET from the sfGFP chromophore to the haem porphyrin of cytochrome. Furthermore, Chapter 7 demonstrates the linkage of protein by both natural and non-natural PTM within one system. Together these results highlight the potential of both interface prediction and molecular anchoring, to the generation of novel synergistic function between disparate proteins. Effectively, the mCherry chromophore function is switched off upon the association of protein to sfGFP, as observed in Chapter 6. This effect has significant implications for mCherry's application as a molecular probe *in situ*, especially, if the effect is induced, as hypothesised, by the reduction of the mCherry chromophore.

Consequently, the energy transfer between fluorescent chromophores in dimeric mCherry and sfGFP was decreased, in comparison to that observed between non-linked monomers. Thus, if the induced proximity between chromophores was not guaranteed by SPAAC, results would be interpreted to suggest that chromophores are not within proximity. This has implications for use of sfGFP and mCherry in combination as fluorescent protein probes. Commonly the fluorescent protein pair are used in unison as protein probes and are present in cells at a high localised concentration. It is noted, that whilst the pair are not expected to form a close association, as was induced by SPAAC, fluorescent proteins do tend to interact and interface prediction presented in Chapter 6 may suggest that interaction between the two FPs is feasible at this site of interface.

Thus, the increased proximity of chromophores may have some degree of influence on fluorescence function and convolute the spectra which are obtained from the sample. Therefore, this may be interpreted as a reduced FRET efficiency between chromophores which are aligned in proximity. This occurrence would then have significant implications for the calculation of both the orientation and distance between chromophores which may not be

---

reflective of the true association. Consequently, small changes to a protein's localised environment can have large and disproportionate effects on function, as observed in dimeric mCherry. This highlights that while the compatible interfaces can be predicted between molecules, the functional output from dimeric protein is more complex, and not as easy to predict.

## 8.4. Future Work

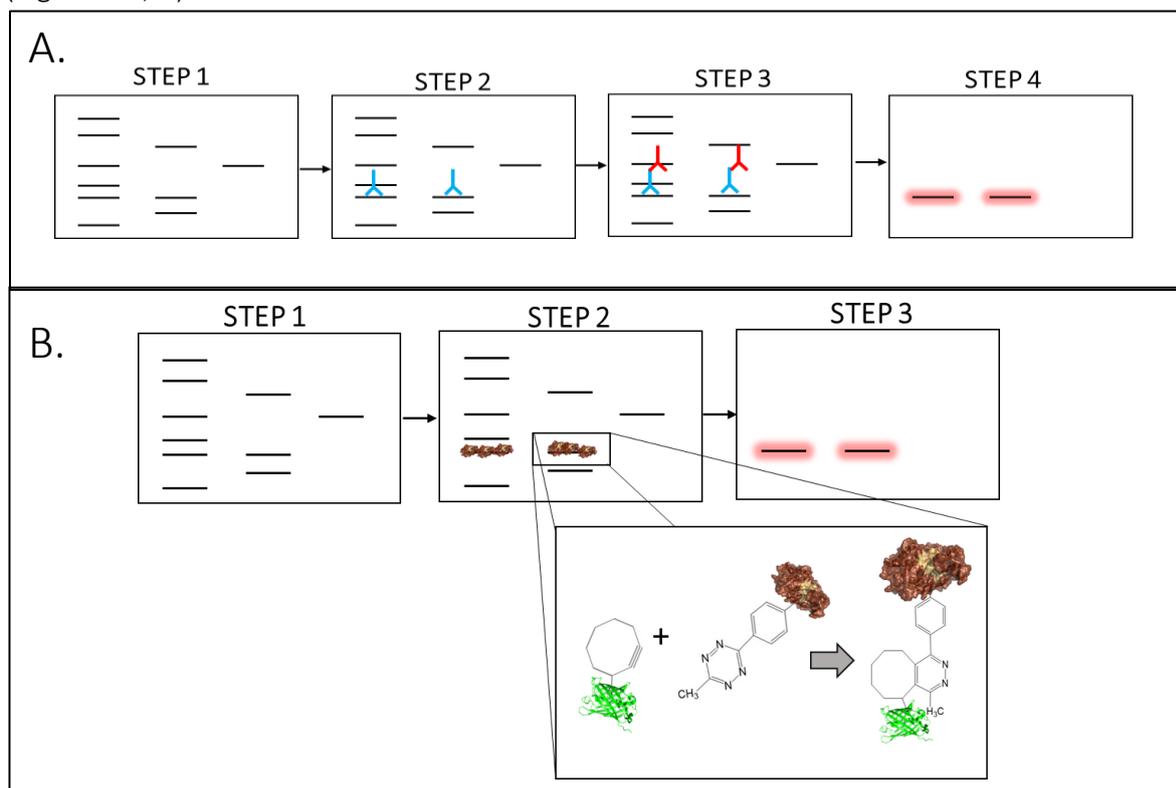
### 8.4.1. Formation of a Novel Biomolecule Detection Assay Using nnPTM of HRP

It is proposed that in future the work of this thesis will be expanded upon for the generation of a series of novel molecular biosensors. The first biosensors which are proposed could be generated as a short-term investigation using existing molecular linkers which are already commercially available. This system would utilise nnPTM, and plant produced HRP, as a molecular probe of nAA incorporation.

It is reasoned that plant produced HRP is an ideal enzyme for the generation of a series of molecular detection enzymes. This is reasoned due to the proteins extensive previous use in many existing detection systems, including both ELISAs and western blots<sup>321,480</sup> (Figure 8.1, A). It is hypothesised that the attachment of HRP to one end of a specific molecular linker *via* nnPTM, would allow for the direct attachment of HRP to specified chemistry of an immobilised protein upon a nitrocellulose membrane. Subsequently, washing of the membrane and application of substrate would indicate and identify the specified protein due to the catalytic turnover of immobilised HRP (Figure 8.1, B).

Initial attachment of the probe to HRP would utilise the native chemistry readily available upon the plant produced protein. The attachment could utilise any chemistry available in numerous amino acid side chains, for the conjugation of the probe to protein. There are numerous systems which can be used to link probe to protein, for example, the primary amines present within lysine residues, of which HRP has 6, for the molecular attachment of linkers; e.g. NHS esters<sup>222</sup>, carbonates<sup>481</sup>, and, imidoesters<sup>482</sup>, to name a few. The first proposed attachment will be the application of an NHS ester to surface lysines as this form of attachment is typically used in the conjugation of HRP to the antibody<sup>222</sup>.

Whilst the proximal end of the attached probe could be supplemented with any chemistry for the detection of a specified protein. It is initially proposed that a molecule such as 6-Methyl-Tetrazine-PEG5-NHS should be used in the investigation. This molecule is suggested as it would allow HRP to be used as a potential detection of SCO-K incorporation, as tetrazine is a highly reactive agent capable of cycloaddition with carbon-carbon triple bonds (Figure 8.1, B).



**Figure 8.1. Proposed Mechanism for the Detection of Protein Using HRP as Probe of nnAA Incorporation.**

**A**, the current mechanism of protein detection using a western blot. In this reaction protein is separated by SDS PAGE and transferred to a nitrocellulose membrane by western blotting (**Step 1**), before application of a specific protein binding antibody which associates with a specific feature of the desired protein, e.g. polyhistidine tag (**Step 2**). The membrane is then washed before the application of a secondary antibody, typically conjugated with HRP (**Step 3**) for the colorimetric, or luminescent detection of protein presence by substrate turn over (**Step 4**).

**B**, represents the proposed mechanism for the detection of specific protein features, in this instance incorporation of the nnAA SCO-K. Initially detection is consistent with the existing methodology and features protein separation by SDS PAGE before transference to nitrocellulose membrane by western blot (**Step 1**). In **Step 2** HRP which has been conjugated 6-Methyl-Tetrazine-PEG5-NHS to feature a surface exposed tetrazine residue is used to probe the presence of SCO-K incorporating protein. If protein is present cycloaddition occurs between the tetrazine and the cyclooctyne of SCO-K immobilising HRP. Finally, the membrane is washed to remove non-bound protein before application of substrate for the detection of immobilised HRP (**Step 3**).

---

Although recombinant HRP, which is generated in Chapter 3 would be an ideal protein probe for use in a later investigation, the lower catalytic ability of the non-glycosylated enzyme would reduce the efficiency of the proposed system. However, the ability to modify the protein by SDM would provide an alternative advantage to the use of HRP without the need of an intermediary molecular linker, e.g. *via* disulphide attachment. As Chapter 3 detailed a general methodology for the recombinant expression and purification of active HRP within *E. coli*, it is reasoned that both SDM and nnAA can now be performed upon HRP which would be the basis of a long-term investigation into this concept. For example, if either SCO-K or AZF were incorporated into one, or several residues upon the surface of HRP, the enzyme could be used to directly react with a target protein by SPAAC. It is reasoned that this system would be advantageous, as it would eliminate the necessity of intermediate antibodies whilst also allowing for the turnover of many substrate molecules from molecular recognition event.

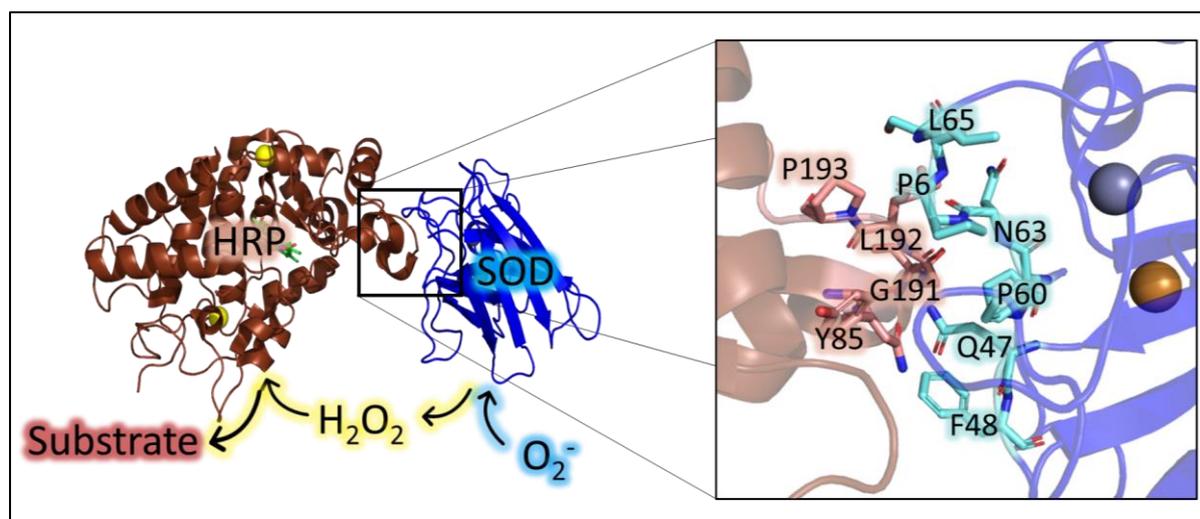
#### 8.4.2. Oligomerisation of HRP to Form a Novel Biosensing Dimer

A long-term goal of future experimentation would be to further investigate the generation of novel protein biosensors *via* oligomerisation. The analysis featured in Chapters 6 and 7 into the use of nnAA to generate novel protein oligomers can generally be expanded on to sample of more functionally disparate protein targets for investigation. However, it is proposed that the greatest benefit to oligomerisation will be to combine two functionally complementary enzymes for one synergistic function. It is anticipated that as interface prediction and nnPTM incorporation can be used as a reliable and consistent means of attachment of two functionally and structurally disparate proteins, the approach can be applied to a wider field of functional communication. It is proposed that this, in future, will then allow for the development of more complex functional combinations and expand research out of the field of FRET-based chromophore communication.

As the catalysis of HRP is peroxide sensitive, this makes the enzyme an ideal protein target for combination with several peroxide forming enzymes in the formation of one biosensing enzyme dimer. Peroxide forming enzymes such as superoxide dismutase (SOD) and glucose oxidase, would both be potential targets of investigation. Initially, SOD would be selected for a partnership with HRP, as the enzyme has an existing history of recombinant expression within the DDJ lab. The molecular attachment of HRP with SOD would then enable the formation of a biosensing dimer, which could be applied in the detection of superoxide

radicals (Figure 8.2). However, glucose oxidase would additionally be a good target for combination, as it produces peroxide in the presence of glucose which in turn could be combined with HRP for the detection of glucose present.

Using the methodology, which was previously applied in Chapter 6, the PDB structures of both HRP and SOD were investigated for a potential region of interface using the ClusPro server. The top-ranked structure was then analysed in PyMOL for possible residue selection for nnAA and molecular anchoring between the two enzymes (Figure 8.2). Several residues in each protein are therefore identified which can be used for future nnAA incorporation within the enzyme interface. The dimerisation of protein is reasoned to be beneficial as increased proximity between active sites would reduce the distance over which peroxide must diffuse. This is expected to increase the sensitivity of HRP to peroxide formed by SOD and reduce the concentration of formed peroxide, which is necessary for the stimulation of HRP, allowing for greater sensitivity of the system to superoxide radicals.



**Figure 8.2. ClusPro Predicted Interface Between HRP and Superoxide Dismutase**

The top ranked interface of interaction between HRP (1H58) and SOD(2SOD). Arrows display the basic reaction pathway between the pair in which superoxide is converted to peroxide within SOD. Hydrogen peroxide is then used in the turnover of substrate by HRP to produce a colorimetric indication of peroxide presence. Insert displays potential residues of nnAA incorporation for cycloaddition between the molecular pair.

This thesis focused on the application of the nnAAs, AZF and SCO-K, for click chemistry, however alternative, more reactive nnAA functional groups are available for the covalent attachment of protein. One such mechanism is inverse electron demand Diels–Alder (IEDDA) which utilises both trans-cyclooctyne (TCO) and tetrazine for a cycloaddition reaction which is 7 orders of magnitude more reactive than that of SCO<sup>483</sup>. If the formation of biosensing

---

dimeric enzymes is feasible, these alternative nnAA will be investigated as a potential means of increasing dimeric efficiency.

## 8.5 Concluding Statement.

This thesis demonstrates that PTM has several important implications for protein expression, interaction, activity, and flexibility. Furthermore, results demonstrate that nnPTM is an excellent means of controlling protein reactivity and communication in a biorthogonal fashion. Chapter 4, demonstrates that without the application of PTM protein function can be lost. Whilst Chapter 5 details the functionally significant alterations which can be induced upon the application of PTM. Chapter 6 and 7 together highlight the ability to form novel protein dimers by the application of nnPTM. Which, in the results observed in Chapter 7, stimulate communication between chromophores. However, the results obtained in Chapter 6 demonstrate that combination of two protein functions does not produce an equal dimeric protein function, and that complex intermolecular interaction can have a magnified impact on function when protein is dimeric. Finally, Chapter 7 combined both natural PTM and nnPTM in the formation of a trimeric protein.

Chapter 3, meanwhile, presents a methodology for the expression and purification of soluble and active recombinant HRP within *E. coli*. Therefore, it is now feasible that nnAA mutants of HRP can be generated for future investigations of protein dimerisation. It is lastly reasoned that the combination of the techniques applied in Chapters 6 and 7 with the means of HRP production detailed in Chapter 3 allows for the formation of many further biosensing dimeric protein complexes. This finally opens the potential of future investigations to link more complex protein combinations for the formation of novel, more functionally significant, biosensors.

---

## 8.6. Publication Associated with This Thesis

Worthy, H. L., Auhim, H. S., Jamieson, W. D., Pope, J. R., Wall, A., Batchelor, R., **Johnson, R.**, Watkins, D, Rizkallah, P, Castell, O, Jones, D. D. (2019). Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry. *Nature Communications Chemistry*. <https://doi.org/10.1038/s42004-019-0185-5>.

Pope, J., **Johnson, R.**, Jamieson, D, Worthy, H, Kailasam, S., Auhim, H., Watkins, D., Rizkallah, P., Castell, O., Jones D., (2020) The effect of proximity on the function and energy transfer capability of fluorescent protein pairs. *Advanced Science*. <https://doi.org/10.1002/advs.202003167>

## Chapter 9. Bibliography

1. Gutiérrez-Preciado, A., Romero, H. & Peimbert, M. An Evolutionary Perspective on Amino Acids. *Nat. Educ.* **3**, 29 (2010).
2. Cobb, M. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.* **15(9)**, (2017).
3. Crick, F. Central dogma of molecular biology. *Nature* **227**, pages561–563 (1970).
4. Berg, J., Tymoczko, J. & Stryer, L. Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point. *W. H. Free. Co.* **5**, 12–20 (2002).
5. Gronemeyer, P., Ditz, R. & Strube, J. Trends in upstream and downstream process development for antibody manufacturing. *Bioengineering* **1**, 188–212 (2014).
6. Berg, J., Tymoczko, J. & Stryer, L. *Amino Acids Are Encoded by Groups of Three Bases Starting from a Fixed Point*. New York: *W H Freeman* (2002).
7. Rehman, I. & Botelho, S. *Biochemistry, Tertiary Structure, Protein*. *StatPearls* (2018).
8. Gaston, M. A., Jiang, R. & Krzycki, J. A. Functional context, biosynthesis, and genetic encoding of pyrrolysine. *Curr. Opin. Microbiol.* **14**, 342–9 (2011).
9. Chen, G. *et al.* A review of enhancers for chemiluminescence enzyme immunoassay. *Food Agric. Immunol.* **28**, 315–327 (2017).
10. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
11. Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).
12. Kia-Ki, H. & Martinage, A. Post-translational chemical modification(S) of proteins. *Int. J. Biochem.* **24**, 19–28 (1992).
13. Barber, K. W. & Rinehart, J. The ABCs of PTMs. *Nat. Chem. Biol.* **14**, 188–192 (2018).
14. Guo, C., Cheng, M. & Gross, M. L. Protein-Metal-Ion Interactions Studied by Mass Spectrometry-Based Footprinting with Isotope-Encoded Benzhydrazide. *Anal. Chem.* **91**, **2**, 141, (2019).
15. Chuh, K. N., Batt, A. R. & Pratt, M. R. Chemical Methods for Encoding and Decoding of Posttranslational Modifications. *Cell Chem. Biol.* **23**, 86–107. (2016).
16. Li, B. L., Langer, J. A., Schwartz, B. & Pestka, S. Creation of phosphorylation sites in proteins: Construction of a phosphorylatable human interferon  $\alpha$ . *Proc. Natl. Acad. Sci. U. S. A.* **86**, 558–590 (1989).
17. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett.* **588**, 206–212 (2014).
18. Raran-Kurussi, S., Cherry, S., Zhang, D. & Waugh, D. S. Removal of affinity tags with TEV protease. *Methods Mol. Biol.* **1586**, 221–230 (2017).
19. Hatlem, D., Trunk, T., Linke, D. & Leo, J. C. Catching a SPY: Using the SpyCatcher-SpyTag and related systems for labeling and localizing bacterial proteins. *Int. J. Mol. Sci.* **20**, (2019).
20. Los, G. V. *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
21. Winston, S. E., Fuller, S. A., Eveleigh, M. J. & Hurrell, J. G. R. Conjugation of Enzymes to Antibodies. *Curr. Protoc. Mol. Biol.* **11**, 1 (2000).
22. Wisdom, G. B. Alkaline Phosphatase Labeling of Antibody Using Glutaraldehyde. *Protein Protoc. Handb.* **1**, 397–409 (1996).
23. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and

- 
- challenges. *Front. Microbiol.* **5**, 172 (2014).
24. Schmidt, F. R. Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.* **65**, 363–72 (2004).
  25. Siloto, R. M. P. & Weselake, R. J. Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatal. Agric. Biotechnol.* **1**, 181–189 (2012).
  26. Melnikov, S. V. & Söll, D. Aminoacyl-TRNA synthetases and tRNAs for an expanded genetic code: What makes them orthogonal? *Int. J. Mol. Sci.* **20**, 19–29 (2019).
  27. Ulmer, K. M. Protein engineering. *Science (80-. )*. **219**, 666–670 (1983).
  28. Itakura, K. *et al.* Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science (80-. )*. **198**, 1056–63 (1977).
  29. Buckel, P. Recombinant proteins for therapy. *Trends Pharmacol. Sci.* **17**, 450–456 (1996).
  30. Sauna, Z. E. *et al.* Recent advances in (therapeutic protein) drug development. *F1000Research* **7**, 113–120 (2017).
  31. Berg, P., Baltimore, D., Brenner, S., Roblin, R. O. & Singer, M. F. Asilomar conference on recombinant DNA molecules. *Science (80-. )*. **455**, 290–1 (1975).
  32. Buyel, J. F., Twyman, R. M. & Fischer, R. Extraction and downstream processing of plant-derived recombinant proteins. *Biotechnol. Adv.* **1**, 902–13 (2015).
  33. Jia, B. & Jeon, C. O. High-throughput recombinant protein expression in *Escherichia coli*: Current status and future perspectives. *Open Biol.* **6**, 160–196 (2016).
  34. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749 (2007).
  35. Wanner, B. L., Kodaira, R. & Neidhardt, F. C. Regulation of lac operon expression: Reappraisal of the theory of catabolite repression. *J. Bacteriol.* **136**, 947–54 (1978).
  36. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411–21 (1999).
  37. Consortium, S. G. *et al.* Protein production and purification. *Nat. Methods* **5**, 135–146 (2008).
  38. Gileadi, O. Recombinant protein expression in *E. coli* : A historical perspective. *Methods Mol. Biol.* **1586**, 978–989 (2017).
  39. Froger, A. & Hall, J. E. Transformation of Plasmid DNA into *E. Coli* using the heat shock method. *J. Vis. Exp.* **6**, 253 (2007).
  40. Solioz, M. & Bienz, D. Bacterial genetics by electric shock. *Trends Biochem. Sci.* **15**, 175–7 (1990).
  41. Green, E. R. & Meccas, J. Bacterial Secretion Systems: An Overview. *Microbiol. Spectr.* **4**, 1128 (2016).
  42. Sjöström, M., Wold, S., Wieslander, A. & Rilfors, L. Signal peptide amino acid sequences in *Escherichia coli* contain information related to final protein localization. A multivariate data analysis. *EMBO J.* **6**, 823–831 (1987).
  43. Waugh, D. S. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr. Purif.* **80**, 283–293 (2011).
  44. Shilling, P. J. *et al.* Improved designs for pET expression plasmids increase protein production yield in *Escherichia coli*. *Commun. Biol.* **3**, (2020).
  45. William Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).

- 
46. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **185**, (1986).
  47. Tabor, S. Expression Using the T7 RNA Polymerase/Promoter System. in *Current Protocols in Molecular Biology* **16**, 12–16 (2001).
  48. Faust, G., Stand, A. & Weuster-Botz, D. IPTG can replace lactose in auto-induction media to enhance protein expression in batch-cultured *Escherichia coli*. *Eng. Life Sci.* **15**, 824–29 (2015).
  49. Baneyx, F. & Mujacic, M. Recombinant protein folding and misfolding in *Escherichia coli*. *Nat. Biotechnol.* **10**, 411–21 (2004).
  50. Wang, Y. C., Peterson, S. E. & Loring, J. F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.* **24**, 143–160 (2014).
  51. Jenkins, N., Murphy, L. & Tyther, R. Post-translational modifications of recombinant proteins: Significance for biopharmaceuticals. *Mol. Biotechnol.* **39**, 113–8 (2008).
  52. Su, M. G. *et al.* Investigation and identification of functional post-translational modification sites associated with drug binding and protein-protein interactions. *BMC Syst. Biol.* **21**, 132–140 (2017).
  53. Darling, A. L. & Uversky, V. N. Intrinsic disorder and posttranslational modifications: The darker side of the biological dark matter. *Front. Genet.* **4**, 158–160 (2018).
  54. Humer, D. & Spadiut, O. Improving the Performance of Horseradish Peroxidase by Site-Directed Mutagenesis. *Int. J. Mol. Sci.* **20**, 916–920 (2019).
  55. Veitch, N. C. Horseradish peroxidase: A modern view of a classic enzyme. *Phytochemistry* **65**, 249–259 (2004).
  56. Transparency Market Research. Horseradish Peroxidase Market. *Mark. Res. Rep.* (2020).
  57. Jamadon, N. K., Busairi, N. & Syahir, A. Smart Phone Assisted Naked Eye Detection of Mercury (II) Ion using Horseradish Peroxidase Inhibitive Assays. *Protein Pept. Lett.* **25**, 90–95 (2017).
  58. Chung, Y., Tannia, D. C. & Kwon, Y. Glucose biofuel cells using bi-enzyme catalysts including glucose oxidase, horseradish peroxidase and terephthalaldehyde crosslinker. *Chem. Eng. J.* **39**, 238–43 (2018).
  59. Bansal, N. & Kanwar, S. S. Peroxidase(s) in environment protection. *Sci. World J.* **Dec 24**, 248–54 (2013).
  60. Bonifert, G., Folkes, L., Gmeiner, C., Dachs, G. & Spadiut, O. Recombinant horseradish peroxidase variants for targeted cancer treatment. *Cancer Med.* **5**, 1194–1203 (2016).
  61. Rajesh, S., Anvita, D., Vikas, P., Dilip, G. & Alok, P. J. Chromophore- An Utility in UV Spectrophotometer. *Inven. Journals* (2017).
  62. Pandey, V. P., Awasthi, M., Singh, S., Tiwari, S. & Dwivedi, U. N. A Comprehensive Review on Function and Application of Plant Peroxidases. *Biochem. Anal. Biochem.* **6**, 1000308–320 (2017).
  63. Asad, S., Khajeh, K. & Ghaemi, N. Investigating the structural and functional effects of mutating Asn glycosylation sites of Horseradish peroxidase to Asp. *Appl. Biochem. Biotechnol.* **164**, 454–63 (2011).
  64. Humer, D., Ebner, J. & Spadiut, O. Scalable high-performance production of recombinant horseradish peroxidase from *E. Coli* inclusion bodies. *Int. J. Mol. Sci.* **21**, 1–20 (2020).
  65. Huddy, S. M., Hitzeroth, I. I., Meyers, A. E., Weber, B. & Rybicki, E. P. Transient expression and purification of horseradish peroxidase C in *Nicotiana benthamiana*. *Int. J. Mol. Sci.* **19**, 115–120 (2018).
  66. Krainer, F. W. & Glieder, A. An updated view on horseradish peroxidases: recombinant production and biotechnological applications. *Applied Microbiology and Biotechnology* **99**, 1611–1625 (2015).
  67. Bartonek-Roxå, E. & Eriksson, H. Expression of a neutral horseradish peroxidase in *Escherichia coli*. *J. Biotechnol.* **37**, 133–142 (1994).

- 
68. Fujiyama, K., Takemura, H., Shimmyo, A., Okada, H. & Takano, M. Genomic DNA structure of two new horseradish-peroxidase-encoding genes. *Gene* **89**, 163–9 (1990).
  69. Strickland, E. H., Kay, E., Shannon, L. M. & Horwitz, J. Peroxidase isoenzymes from horseradish roots. 3. Circular dichroism of isoenzymes and apoisoenzymes. *J. Biol. Chem.* **243**, 3560–5 (1968).
  70. Capone, S. *et al.* Glyco-variant library of the versatile enzyme horseradish peroxidase. *Glycobiology* **24**, 852–863 (2014).
  71. Matsui, T. *et al.* High-efficiency secretory production of peroxidase C1a using vesicular transport engineering in transgenic tobacco. *J. Biosci. Bioeng.* **102**, 102–9 (2006).
  72. Utashima, Y., Matsumoto, H., Masaki, K. & Iefuji, H. Heterologous production of horseradish peroxidase C1a by the basidiomycete yeast *Cryptococcus* sp. S-2 using codon and signal optimizations. *Appl. Microbiol. Biotechnol.* **98**, 7893–900 (2014).
  73. Arnold, F. & Zhanglin, L. Expression of functional eukaryotic proteins. US Patent Application Publication. (2002).
  74. Ignatenko, O. V., Rubtsova, M. Y., Ivanova, N. L., Ouporov, I. V & Egorov, A. M. Analysis of the Antigenic Structure of Holo- and Apo-Forms of Horseradish Peroxidase. **41**, 102–105 (2000).
  75. Smith, A. T. *et al.* Expression of a synthetic gene for horseradish peroxidase C in *Escherichia coli* and folding and activation of the recombinant enzyme with Ca<sup>2+</sup> and heme. *J. Biol. Chem.* **265**, 13335–43 (1990).
  76. Kis, M. *et al.* An N-terminal peptide extension results in efficient expression, but not secretion, of a synthetic horseradish peroxidase gene in transgenic tobacco. *Ann. Bot.* **93**, 303–310 (2004).
  77. Morawski, B. *et al.* Functional expression of horseradish peroxidase in *Saccharomyces cerevisiae* and *Pichia pastoris*. *Protein Eng.* **13**, 377–387 (2000).
  78. Smith, A. T., Sanders, S. A., Thorneley, R. N. F., Burke, J. F. & Bray, R. R. C. Characterisation of a haem active-site mutant of horseradish peroxidase, Phe41 → Val, with altered reactivity towards hydrogen peroxide and reducing substrates. *Eur. J. Biochem.* **207**, 507–19 (1992).
  79. Gundinger, T. & Spadiut, O. A comparative approach to recombinantly produce the plant enzyme horseradish peroxidase in *Escherichia coli*. *J. Biotechnol.* **248**, 15–24 (2017).
  80. Spadiut, O., Rossetti, L., Dietzsch, C. & Herwig, C. Purification of a recombinant plant peroxidase produced in *Pichia pastoris* by a simple 2-step strategy. *Protein Expr. Purif.* **86**, 89–97 (2012).
  81. Gajhede, M., Schuller, D. J., Henriksen, A., Smith, A. T. & Poulos, T. L. Crystal structure of horseradish peroxidase C at 2.15 Å resolution. *Nat. Struct. Biol.* **4**, 1032–1038 (1997).
  82. Berglund, G. I. *et al.* The catalytic pathway of horseradish peroxidase at high resolution. *Nature* **417**, 463–468 (2002).
  83. Plieth, C. & Vollbehr, S. Calcium promotes activity and confers heat stability on plant peroxidases. *Plant Signal. Behav.* **7**, 650–660 (2012).
  84. Azevedo, A. M. *et al.* Horseradish peroxidase: A valuable tool in biotechnology. *Biotechnol. Annu. Rev.* **9**, 199–247 (2003).
  85. Hamilton, T. M., Dobie-Galuska, A. A. & Wietstock, S. M. The o-Phenylenediamine-Horseradish Peroxidase System: Enzyme Kinetics in the General Chemistry Laboratory. *J. Chem. Educ.* (1999). doi:10.1021/ed076p642
  86. Khan, P. *et al.* Luminol-based chemiluminescent signals: Clinical and non-clinical application and future uses. *Appl. Biochem. Biotechnol.* **173**, 333–55 (2014).
  87. Rodríguez-López, J. N. *et al.* Mechanism of reaction of hydrogen peroxide with horseradish peroxidase: Identification of intermediates in the catalytic cycle. *J. Am. Chem. Soc.* **123**, 11838–11847 (2001).

- 
88. Hashimoto, S. & Takeuchi, H. Protonation and hydrogen-bonding state of the distal histidine in the CO complex of horseradish peroxidase as studied by ultraviolet resonance Raman spectroscopy. *Biochemistry* **45**, 20352–20357 (2006).
  89. Nagano, S. *et al.* Catalytic Roles of the Distal Site Hydrogen Bond Network of Peroxidases. *Oxyg. Homeost. Its Dyn.* **8**, 68476–44 (1998).
  90. Fiege, K., Querebillo, C. J., Hildebrandt, P. & Frankenberg-Dinkel, N. Improved Method for the Incorporation of Heme Cofactors into Recombinant Proteins Using *Escherichia coli* Nissle 1917. *Biochemistry* **57**, 2747–2755 (2018).
  91. Krainer, F. W. *et al.* Optimizing cofactor availability for the production of recombinant heme peroxidase in *Pichia pastoris*. *Microb. Cell Fact.* **14**, 213–18 (2015).
  92. Lingg, N., Zhang, P., Song, Z. & Bardor, M. The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis. *Biotechnol. J.* **7**, 1462–72 (2012).
  93. Shental-Bechor, D. & Levy, Y. Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 8256–61 (2008).
  94. Brooks, S. A. Appropriate glycosylation of recombinant proteins for human use: Implications of choice of expression system. *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.* **28**, 241–55 (2004).
  95. Dwek, R. A. Biological importance of glycosylation. *Dev. Biol. Stand.* **23**, 234–54 (1998).
  96. Wuhrer, M., Balog, C. I. A., Koeleman, C. A. M., Deelder, A. M. & Hokke, C. H. New features of site-specific horseradish peroxidase (HRP) glycosylation uncovered by nano-LC-MS with repeated ion-isolation/fragmentation cycles. *Biochim. Biophys. Acta - Gen. Subj.* **25**, 229–39 (2005).
  97. Reily, C., Stewart, T. J., Renfrow, M. B. & Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **15**, (2019).
  98. Chui, D. *et al.* Genetic remodeling of protein glycosylation in vivo induces autoimmune disease. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1142–47 (2001).
  99. Lin, Z., Thorsen, T. & Arnold, F. H. Functional expression of horseradish peroxidase in *E. coli* by directed evolution. *Biotechnol. Prog.* **15**, 467–471 (1999).
  100. Gray, J. S. S., Yang, B. Y. & Montgomery, R. Heterogeneity of glycans at each N-glycosylation site of horseradish peroxidase. *Carbohydr. Res.* **311**, 61–9 (1998).
  101. Tams, J. W. & Welinder, K. G. Glycosylation and thermodynamic versus kinetic stability of horseradish peroxidase. *FEBS Lett.* **421**, 234–6 (1998).
  102. Bhanuramanand, K., Ahmad, S. & Rao, N. M. Engineering deamidation-susceptible asparagines leads to improved stability to thermal cycling in a lipase. *Protein Sci.* **23**, 1479–90 (2014).
  103. Saraswathy, N. & Ramalingam, P. Phosphoproteomics. *Concepts Tech. Genomics Proteomics* **98**, 23623–87 (2011).
  104. Flickinger, M. C., Aehle, W., Bott, R. & Graycar, T. Proteolytic Cleavage, Reaction Mechanisms. *Encycl. Ind. Biotechnol.* **298**, 893–98 (2009).
  105. Klein, T., Eckhard, U., Dufour, A., Solis, N. & Overall, C. M. Proteolytic Cleavage - Mechanisms, Function, and 'omic' Approaches for a Near-Ubiquitous Posttranslational Modification. *Chem. Rev.* **118**, 1137–68 (2018).
  106. Liu, M. *et al.* Biosynthesis, structure, and folding of the insulin precursor protein. *Diabetes, Obes. Metab.* **20**, 28–50 (2018).
  107. Chen, L. M., Yang, X. W. & Tang, J. G. Acidic residues on the N-terminus of proinsulin C-peptide are important for the folding of insulin precursor. *J. Biochem.* **131**, 855–9 (2002).

- 
108. Braakman, I. & Hebert, D. N. Protein folding in the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.* **10**, 443–454 (2013).
  109. Grigorenko, V. *et al.* New approaches for functional expression of recombinant horseradish peroxidase C in *Escherichia coli*. *Biocatal. Biotransformation* **12**, 359–379 (1999).
  110. Poole, L. B. The basics of thiols and cysteines in redox biology and chemistry. *Free Radic. Biol. Med.* **80**, 148–157 (2015).
  111. Nguyen, D. P. *et al.* Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry. *J. Am. Chem. Soc.* **131**, 8720–8721 (2009).
  112. de Marco, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.* **8**, 239–45 (2009).
  113. Kurokawa, Y., Yanagi, H. & Yura, T. Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 3960–5 (2000).
  114. Asad, S., Dabirmanesh, B., Ghaemi, N., Etezad, S. M. & Khajeh, K. Studies on the refolding process of recombinant horseradish peroxidase. *Mol. Biotechnol.* **54**, 484–492 (2013).
  115. Pál, C., Papp, B. & Lercher, M. J. An integrated view of protein evolution. *Nat. Rev. Genet.* **7**, 337–48 (2006).
  116. Rowlands, R. T. Industrial strain improvement: rational screens and genetic recombination techniques. *Enzyme Microb. Technol.* **6**, 290–300 (1984).
  117. Carrigan, P. E., Ballar, P. & Tuzmen, S. Site-directed mutagenesis. *Methods Mol. Biol.* **76**, 107–124 (2011).
  118. Hutchison, C. A. *et al.* Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–60 (1978).
  119. Arnold, F. H. Protein engineering for unusual environments. *Curr. Opin. Biotechnol.* **4**, 450–5 (1993).
  120. Jessen, T. H., Weber, R. E., Fermi, G., Tame, J. & Braunitzer, G. Adaptation of bird hemoglobins to high altitudes: Demonstration of molecular mechanism by protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 6519–6522 (1991).
  121. Barany, F. Single-stranded hexameric linkers: a system for in-phase insertion mutagenesis and protein engineering. *Gene* **37**, 111–123 (1985).
  122. Utsumi, S., Gidamis, A. B., Kanamori, J., Kang, I. J. & Kito, M. Effects of Deletion of Disulfide Bonds by Protein Engineering on the Conformation and Functional Properties of Soybean Proglycinin. *J. Agric. Food Chem.* **41**, 687–691 (1993).
  123. Higgins, S. A., Ouonkap, S. V. Y. & Savage, D. F. Rapid and Programmable Protein Mutagenesis Using Plasmid Recombineering. *ACS Synth. Biol.* **6**, 1825–33 (2017).
  124. Goh, Y. Y., Frecer, V., Ho, B. & Ding, J. L. Rational design of green fluorescent protein mutants as biosensor for bacterial endotoxin. *Protein Eng.* **15**, 493–502 (2002).
  125. Chica, R. A., Moore, M. M., Allen, B. D. & Mayo, S. L. Generation of longer emission wavelength red fluorescent proteins using computationally designed libraries. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20257–62 (2010).
  126. Eijssink, V. G. H., GÅseidnes, S., Borchert, T. V. & Van Den Burg, B. Directed evolution of enzyme stability. *Biomol. Eng.* **22**, 21–30 (2005).
  127. Kuchner, O. & Arnold, F. H. Directed evolution of enzyme catalysts. *Trends Biotechnol.* **15**, 523–30 (1997).

- 
128. Jäckel, C., Kast, P. & Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **37**, 153–173 (2008).
  129. Brannigan, J. A. & Wilkinson, A. J. Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.* **3**, 964–70 (2002).
  130. Woolfson, D. N. *et al.* De novo protein design: How do we expand into the universe of possible protein structures? *Curr. Opin. Struct. Biol.* **33**, 16–26 (2015).
  131. Walsh, G. Post-translational modifications of protein biopharmaceuticals. *Drug Discov. Today* **15**, 773–80 (2010).
  132. Fessner, W. D. How to meet the need for speed in protein evolution. *Nat. Catal.* **2**, 738–9 (2019).
  133. Kapoor, S., Rafiq, A. & Sharma, S. Protein engineering and its applications in food industry. *Crit. Rev. Food Sci. Nutr.* **57**, 2321–29 (2017).
  134. Liang, Q., Chen, L. & Fulco, A. J. An efficient and optimized PCR method with high fidelity for site-directed mutagenesis. *PCR Methods Appl.* **4**, 269–74 (1995).
  135. Ling, M. M. & Robinson, B. H. Approaches to DNA mutagenesis: An overview. *Anal. Biochem.* **254**, 157–78 (1997).
  136. McCullum, E. O., Williams, B. A. R., Zhang, J. & Chaput, J. C. Random mutagenesis by error-prone PCR. *Methods Mol. Biol.* **634**, (2010).
  137. Bachman, J. Site-directed mutagenesis. *Methods Enzymol.* **529**, 241–98 (2013).
  138. Luo, W. G., Liu, H. Z., Lin, W. H., Kabir, M. H. & Su, Y. Simultaneous splicing of multiple DNA fragments in one PCR reaction. *Biol. Proced. Online* **15**, 9222–29 (2013).
  139. Xia, Y., Chu, W., Qi, Q. & Xun, L. New insights into the QuikChange™ process guide the use of Phusion DNA polymerase for site-directed mutagenesis. *Nucleic Acids Res.* **34**, 3487–98 (2015).
  140. Dominy, C. N. & Andrews, D. W. Site-directed mutagenesis by inverse PCR. *Methods Mol. Biol.* (2003). doi:10.1385/1-59259-409-3:209
  141. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, *Aequorea*. *J. Cell. Comp. Physiol.* **59**, 223–39 (1962).
  142. Morise, H., Shimomura, O., Johnson, F. H. & Winant, J. Intermolecular energy transfer in the bioluminescent system of *aequorea*. *Biochemistry* 237–45 (1974). doi:10.1021/bi00709a028
  143. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science (80-. )*. **263**, 802–805 (1994).
  144. Kain, S. R. *et al.* Green fluorescent protein as a reporter of gene expression and protein localization. *Biotechniques* 3454–60 (1995).
  145. Cormack, B. Green fluorescent protein as a reporter of transcription and protein localization in fungi. *Curr. Opin. Microbiol.* **1**, 406–10 (1998).
  146. Hoffman, R. M. Application of GFP imaging in cancer. *Lab. Investig.* **1194**, 3247–3458 (2015).
  147. Kordyum, E. L., Shevchenko, G. V., Yemets, A. I., Nyporko, A. I. & Blume, Y. B. Application of GFP technique for cytoskeleton visualization onboard the International Space Station. *Acta Astronaut.* **56**, 613–21 (2005).
  148. Zimmer, M. GFP: From jellyfish to the Nobel prize and beyond. *Chem. Soc. Rev.* **38**, 2823–32 (2009).
  149. Stepanenko, O. V. *et al.* Comparative studies on the structure and stability of fluorescent proteins EGFP, zFP506, mRFP1, ‘dimer2’, and DsRed1. *Biochemistry* **43**, 14913–14923 (2004).

- 
150. Chattopadhyay, A. & Haldar, S. Dynamic insight into protein structure utilizing red edge excitation shift. *Acc. Chem. Res.* **47**, 12–19 (2014).
  151. Jung, K., Park, J., Maeng, P. J. & Kim, H. Fluorescence quenching of green fluorescent protein during denaturation by guanidine. *Bull. Korean Chem. Soc.* **26**, 413–17 (2005).
  152. Mitra, M. *et al.* Organization and dynamics of tryptophan residues in brain spectrin: Novel insight into conformational flexibility. *J. Fluoresc.* **25**, 707–17 (2015).
  153. Ormö, M. *et al.* Crystal Structure of the Aequorea victoria Green Fluorescent Protein The structure of GFP was determined by. *Science (80-. )*. **7**, 7132–8 (1996).
  154. Tsien, R. Y. THE GREEN FLUORESCENT PROTEIN. *Annu. Rev. Biochem.* **67**, 509–44 (1998).
  155. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12501–4 (1994).
  156. Rosenow, M. A., Huffman, H. A., Phail, M. E. & Wachter, R. M. The Crystal Structure of the Y66L Variant of Green Fluorescent Protein Supports a Cyclization-Oxidation-Dehydration Mechanism for Chromophore Maturation. *Biochemistry* **43**, 4464–72 (2004).
  157. Brejc, K. *et al.* Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 206–2311 (1997).
  158. Stepanenko, O. V. *et al.* Understanding the role of Arg96 in structure and stability of green fluorescent protein. *Proteins Struct. Funct. Genet.* **73**, 539–551 (2008).
  159. Shaner, N. C., Patterson, G. H. & Davidson, M. W. Advances in fluorescent protein technology. *J. Cell Sci.* **120**, 4247–60 (2007).
  160. Scheyhing, C. H., Meersman, F., Ehrmann, M. A., Heremans, K. & Vogel, R. F. Temperature-pressure stability of green fluorescent protein: A Fourier transform infrared spectroscopy study. *Biopolymers* **65**, 244–53 (2002).
  161. Siemerling, K. R., Golbik, R., Sever, R. & Haseloff, J. Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr. Biol.* **6**, 1653–63 (1996).
  162. Von Stetten, D., Noirclerc-Savoye, M., Goedhart, J., Gadella, T. W. J. & Royant, A. Structure of a fluorescent protein from Aequorea victoria bearing the obligate-monomer mutation A206K. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **68**, 878–82 (2012).
  163. Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat. Methods* **15**, 47–51 (2018).
  164. Kilgard, R., Heim, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **23373**, 663–3 (1995).
  165. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–8 (1996).
  166. Staudacher, V. *et al.* Redox-sensitive GFP fusions for monitoring the catalytic mechanism and inactivation of peroxiredoxins in living cells. *Redox Biol.* **14**, 549–556 (2018).
  167. Stepanenko, O. V., Stepanenko, O. V., Kuznetsova, I. M., Uverskyo, V. N. & Turoverov, K. K. Peculiarities of the super-folder GFP folding in a crowded milieu. *Int. J. Mol. Sci.* **17**, 1805–12 (2016).
  168. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
  169. Pedelacq, J. D. & Cabantous, S. Development and applications of superfolder and split fluorescent protein detection systems in biology. *Int. J. Mol. Sci.* **20**, 3479 (2019).
  170. Hanson, G. T. *et al.* Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural characterization and preliminary application. *Biochemistry* **41**, 15477–88 (2002).

- 
171. Cubitt, A. B. *et al.* Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20**, 448–55 (1995).
  172. Stepanenko, O., Verkhusha, V., Kuznetsova, I., Uversky, V. & Turoverov, K. Fluorescent Proteins as Biomarkers and Biosensors: Throwing Color Lights on Molecular and Cellular Processes. *Curr. Protein Pept. Sci.* **9**, 338–69 (2008).
  173. Jensen, E. C. Use of Fluorescent Probes: Their Effect on Cell Biology and Limitations. *Anat. Rec.* **295**, 2031–2036 (2012).
  174. Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H. & Phillips, G. J. Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *J. Bacteriol.* **182**, 4068–76 (2000).
  175. Leibly, D. J. *et al.* A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **23**, 1754–68 (2015).
  176. Arpino, J. A. J. *et al.* Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. *J. Am. Chem. Soc.* **134**, 13632–40 (2012).
  177. Park, S. Y., Moon, H. C. & Park, H. Y. Live-cell imaging of single mRNA dynamics using split superfolder green fluorescent proteins with minimal background. *RNA* **14**, 1894–95 (2020).
  178. Albertazzi, L., Arosio, D., Marchetti, L., Ricci, F. & Beltram, F. Quantitative FRET analysis with the EOGFP-mCherry fluorescent protein pair. *Photochem. Photobiol.* **85**, 287–97 (2009).
  179. Periasamy, A. & Day, R. N. Visualizing protein interactions in living cells using digitized GFP imaging and FRET microscopy. *Methods Cell Biol.* **58**, 293–314 (1999).
  180. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma sp.* red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
  181. Segal, M. F., Chu, H., Christian, J. A. & Low, P. S. Fluorescence assay of the interaction between hemoglobin and the cytoplasmic domain of erythrocyte membrane band 3. *Blood Cells, Mol. Dis.* **55**, 266–71 (2015).
  182. Akerboom, J. *et al.* Optimization of a GCaMP calcium indicator for neural activity imaging. *J. Neurosci.* **32**, 13819–13830 (2012).
  183. Mutoh, H., Perron, A., Akemann, W., Iwamoto, Y. & Knöpfel, T. Optogenetic monitoring of membrane potentials. *Exp. Physiol.* **96**, 13–18 (2011).
  184. Watts, S. D., Suchland, K. L., Amara, S. G. & Ingram, S. L. A sensitive membrane-targeted biosensor for monitoring changes in intracellular chloride in neuronal processes. *PLoS One* **7**, 35373 (2012).
  185. Cannon, M. B. Re-engineering redox-sensitive green fluorescent protein for improved response rate. *Protein Sci.* **15**, 45–57 (2006).
  186. Llopis, J., McCaffery, J. M., Miyawaki, A., Farquhar, M. G. & Tsien, R. Y. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6803–8 (1998).
  187. Østergaard, H., Henriksen, A., Hansen, F. G. & Winther, J. R. Shedding light on disulfide bond formation: Engineering a redox switch in green fluorescent protein. *EMBO J.* **20**, 5853–5862 (2001).
  188. Leibly, D. J. *et al.* A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **23**, 1754–68 (2015).
  189. Gwyther, R. E. A., Dafydd Jones, D. & Worthy, H. L. Better together: Building protein oligomers naturally and by design. *Biochem. Soc. Trans.* **47**, 1773–1780 (2019).
  190. Janin, J., Miller, S. & Chothia, C. Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155–64 (1988).

- 
191. Hu, Z., Ma, B., Wolfson, H. & Nussinov, R. Conservation of polar residues as hot spots at protein interfaces. *Proteins Struct. Funct. Genet.* **39**, 331–342 (2000).
  192. Brodin, J. D. *et al.* Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–382 (2012).
  193. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9 (1998).
  194. Jones, D. D. & Barker, P. D. Controlling self-assembly by linking protein folding, DNA binding, and the redox chemistry of heme. *Angew. Chemie - Int. Ed.* (2005). doi:10.1002/anie.200463035
  195. Song, W. J. & Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science (80-. )*. (2014). doi:10.1126/science.1259680
  196. Thomson, A. R. *et al.* Computational design of water-soluble  $\alpha$ -helical barrels. *Science (80-. )*. **346**, 485–88 (2014).
  197. Mou, Y., Huang, P. S., Hsu, F. C., Huang, S. J. & Mayo, S. L. Computational design and experimental verification of a symmetric protein homodimer. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10714–9 (2015).
  198. Kobayashi, N. & Arai, R. Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Curr. Opin. Biotechnol.* **46**, 57–65 (2017).
  199. Huang, P. S. *et al.* High thermodynamic stability of parametrically designed helical bundles. *Science (80-. )*. (2014). doi:10.1126/science.1257481
  200. Thamri, A. *et al.* Peptide modification results in the formation of a dimer with a 60-fold enhanced antimicrobial activity. *PLoS One* **12**, 017383–7 (2017).
  201. Ali, M. H. & Imperiali, B. Protein oligomerization: How and why. *Bioorganic Med. Chem.* **13**, 5013–5020 (2005).
  202. Worthy, H. L. *et al.* Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry. *Commun. Chem.* **2**, 1–12 (2019).
  203. Ljubetič, A., Gradišar, H. & Jerala, R. Advances in design of protein folds and assemblies. *Curr. Opin. Chem. Biol.* **40**, (2017).
  204. Alberts, B., A, J. & Lewis J. Analyzing Protein Structure and Function. in *Molecular Biology of the Cell* 432–65 (2002).
  205. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–53 (2000).
  206. Marianayagam, N. J., Sunde, M. & Matthews, J. M. The power of two: Protein dimerization in biology. *Trends Biochem. Sci.* **29**, 618–25 (2004).
  207. Mei, G., Di Venere, A., Rosato, N. & Finazzi-Agrò, A. The importance of being dimeric. *FEBS J.* **272**, 16–27 (2005).
  208. Larsen, T. A., Olson, A. J. & Goodsell, D. S. Morphology of protein-protein interfaces. *Structure* **6**, 421–27 (1998).
  209. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7877–82 (2002).
  210. Schilter, D. Thiol oxidation: A slippery slope. *Nat. Rev. Chem.* **1**, 1–13 (2017).
  211. Bischoff, R. & Schlüter, H. Amino acids: Chemistry, functionality and selected non-enzymatic post-translational modifications. *J. Proteomics* **75**, 2275–96 (2012).
  212. Smolskaya, S. & Andreev, Y. A. Site-specific incorporation of unnatural amino acids into escherichia coli recombinant protein: Methodology development and recent achievement. *Biomolecules* **8**, 2837–9 (2019).

- 
213. Spassov, V. Z., Yan, L. & Flook, P. K. The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: A side-chain prediction algorithm based on side-chain backbone interactions. *Protein Sci.* **16**, 494–506 (2007).
214. Luo, Q., Tao, Y., Sheng, W., Lu, J. & Wang, H. Dinitroimidazoles as bifunctional bioconjugation reagents for protein functionalization and peptide macrocyclization. *Nat. Commun.* **10**, 1–10 (2019).
215. Das, T., Shin, S. C., Song, E. J. & Kim, E. E. Regulation of deubiquitinating enzymes by post-translational modifications. *Int. J. Mol. Sci.* **21**, 4028–34 (2020).
216. Groban, E. S., Narayanan, A. & Jacobson, M. P. Conformational changes in protein loops and helices induced by post-translational phosphorylation. *PLoS Comput. Biol.* **2**, 32 (2006).
217. Wuttke, D. S., Gray, H. B., Fisher, S. L. & Imperiali, B. Semisynthesis of Bipyridyl-Alanine Cytochrome c Mutants: Novel Proteins with Enhanced Electron-Transfer Properties. *J. Am. Chem. Soc.* **23**, 908–23 (1993).
218. Toseland, C. P. Fluorescent labeling and modification of proteins. *J. Chem. Biol.* **6**, 85–95 (2013).
219. Krall, N., Da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113 (2016).
220. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **115**, 2174–95 (2015).
221. Macmillan, D., Bill, R. M., Sage, K. A., Fern, D. & Flitsch, S. L. Selective in vitro glycosylation of recombinant proteins: Semi-synthesis of novel homogeneous glycoforms of human erythropoietin. *Chem. Biol.* **8**, 133–45 (2001).
222. Lim, C. Y. *et al.* Succinimidyl ester surface chemistry: Implications of the competition between aminolysis and hydrolysis on covalent protein immobilization. *Langmuir* **30**, 12868–78 (2014).
223. Verma, V., Kaur, C., Grover, P., Gupta, A. & Chaudhary, V. K. Biotin-tagged proteins: Reagents for efficient ELISA-based serodiagnosis and phage display-based affinity selection. *PLoS One* **13**, 0191315 (2018).
224. Mishra, P., Nayak, B. & Dey, R. K. PEGylation in anti-cancer therapy: An overview. *Asian J. Pharm. Sci.* **11**, 337–48 (2016).
225. Köhrer, C. & Rajbhandary, U. L. Proteins with One or More Unnatural Amino Acids. in *The Aminoacyl-tRNA Synthetases* 435–67 (2005).
226. Lang, K. & Chin, J. W. Bioorthogonal reactions for labeling proteins. *ACS Chem. Biol.* **9**, 16–20 (2014).
227. Hartley, A. M., Worthy, H. L., Reddington, S. C., Rizkallah, P. J. & Jones, D. D. Molecular basis for functional switching of GFP by two disparate non-native post-translational modifications of a phenyl azide reaction handle. *Chem. Sci.* **7**, 6484–91 (2016).
228. Reddington, S. C. *et al.* Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chem. Sci.* **6**, 1159–1166 (2015).
229. Villa, J. K. *et al.* Fluorescence Modulation of Green Fluorescent Protein Using Fluorinated Unnatural Amino Acids. *Molecules* **22**, 1194–98 (2017).
230. Guerrero, S. A., Hecht, H. J., Hofmann, B., Biebl, H. & Singh, M. Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl. Microbiol. Biotechnol.* **56**, 718–23 (2001).
231. Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three dimensional structure. *EMBO J.* **9**, 1665–72 (1990).
232. Budisa, N. *et al.* High-level Biosynthetic Substitution of Methionine in Proteins by its Analogs 2-

- 
- Aminohexanoic Acid, Selenomethionine, Telluromethionine and Ethionine in *Escherichia coli*. *Eur. J. Biochem.* **230**, 788–96 (1995).
233. Hendrickson, T. L., Crécy-Lagard, V. de & Schimmel, P. Incorporation of Nonnatural Amino Acids Into Proteins. *Annu. Rev. Biochem.* **6**, 809–15 (2004).
234. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science (80-. )*. **244**, 182–8 (1989).
235. Hecht, S. M., Alford, B. L., Kuroda, Y. & Kitano, S. ‘Chemical aminoacylation’ of tRNA’s. *J. Biol. Chem.* **34**, 1864–8 (1978).
236. Smolskaya, S., Zhang, Z. J. & Alfonta, L. Enhanced Yield of Recombinant Proteins with Site-Specifically Incorporated Unnatural Amino Acids Using a Cell-Free Expression System. *PLoS One* **8**, 68363 (2013).
237. Gao, W., Cho, E., Liu, Y. & Lu, Y. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Front. Pharmacol.* **10**, 23988–98 (2019).
238. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
239. Capone, J. P., Sedivy, J. M., Sharp, P. A. & RajBhandary, U. L. Introduction of UAG, UAA, and UGA nonsense mutations at a specific site in the *Escherichia coli* chloramphenicol acetyltransferase gene: use in measurement of amber, ochre, and opal suppression in mammalian cells. *Mol. Cell. Biol.* **6**, 3059–67 (1986).
240. Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H. & Sisido, M. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an *Escherichia coli* in vitro translation system. *Biochemistry* **40**, 11060–64 (2001).
241. Hohsaka, T., Ashizuka, Y., Murakami, H. & Sisido, M. Five-base codons for incorporation of nonnatural amino acids into proteins. *Nucleic Acids Res.* **29**, 3646–51 (2001).
242. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883–890 (2001).
243. Jakubowski, H. & Goldman, E. Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* **56**, 412–29 (1992).
244. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science (80-. )*. **292**, 498–500 (2001).
245. Zhang, Y., Wang, L., Schultz, P. G. & Wilson, I. A. Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O -methyl-L-tyrosine. *Protein Sci.* **14**, 1340–9 (2005).
246. Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science (80-. )*. **301**, 964–7 (2003).
247. Sakamoto, K. *et al.* Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30**, 4692–4699 (2002).
248. Greiss, S. & Chin, J. W. Expanding the genetic code of an animal. *J. Am. Chem. Soc.* **133**, 14196–99 (2011).
249. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **23**, 235–76 (2010).
250. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nat. Biotechnol.* **22**, 1297–301 (2004).
251. Tucker, M. J., Oyola, R. & Gai, F. A novel fluorescent probe for protein binding and folding studies: p-cyano-phenylalanine. *Biopolymers* **83**, 571–6 (2006).
252. Reddington, S. C. *et al.* Genetically encoded phenyl azide photochemistry drives positive and negative

- 
- functional modulation of a red fluorescent protein. *RSC Adv.* **5**, 77734–76 (2015).
253. Wang, Y. S. *et al.* Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol.* **23**, 9778–89 (2013).
254. Preston, G. W. & Wilson, A. J. Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem. Soc. Rev.* **42**, 3289–94 (2013).
255. Reddington, S., Watson, P., Rizkallah, P., Tippmann, E. & Dafydd Jones, D. Genetically encoding phenyl azide chemistry: New uses and ideas for classical biochemistry. *Biochem. Soc. Trans.* **41**, 1177–82 (2013).
256. Thomas, S. K. *et al.* Site-Specific Protein Photochemical Covalent Attachment to Carbon Nanotube Side Walls and Its Electronic Impact on Single Molecule Function. *Bioconjug. Chem.* **31**, 584–594 (2020).
257. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **40**, 2004–21 (2001).
258. Kim, E. & Koo, H. Biomedical applications of copper-free click chemistry:: In vitro, in vivo, and ex vivo. *Chem. Sci.* **10**, 7835–51 (2019).
259. Le Droumaguet, B. & Velonia, K. Click chemistry: A powerful tool to create polymer-based macromolecular chimeras. *Macromol. Rapid Commun.* **29**, 1073–89 (2008).
260. Mbuja, N. E., Guo, J., Wolfert, M. A., Steet, R. & Boons, G. J. Strain-Promoted Alkyne-Azide Cycloadditions (SPAAC) Reveal New Features of Glycoconjugate Biosynthesis. *ChemBioChem* **12**, 1912–21 (2011).
261. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E. A. Genetically encoded copper-free click chemistry. *Angew. Chemie - Int. Ed.* **50**, 3878–3881 (2011).
262. Pang, Y. L. J., Poruri, K. & Martinis, S. A. tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA* **5**, 461–80 (2014).
263. Dommerholt, J., Rutjes, F. P. J. T. & van Delft, F. L. Strain-Promoted 1,3-Dipolar Cycloaddition of Cycloalkynes and Organic Azides. *Top. Curr. Chem.* **374**, 1244–56 (2016).
264. Heber, D., Rösner, P. & Tochtermann, W. Cyclooctyne and 4-cyclooctyn-1-ol - Versatile building blocks in organic synthesis. *European J. Org. Chem.* **2005**, 4231–47 (2005).
265. Reddington, S. C., Tippmann, E. M. & Dafydd Jones, D. Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry. *Chem. Commun.* **48**, 8419–21 (2012).
266. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–5 (1970).
267. Rasband, W. S. ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA. <http://imagej.nih.gov/ij/> (2014).
268. Hindson, S. A., Kwok, A. & Pudney, C. R. Creating a dynamic hydration map in denatured states using the red edge excitation shift. *Dep. Biol. Biochem. Cent. Ther. Innov.* **34**, 1983–98 (2019).
269. Aumiller, W. M., Davis, B. W., Hatzakis, E. & Keating, C. D. Interactions of macromolecular crowding agents and cosolutes with small-molecule substrates: Effect on horseradish peroxidase activity with two different substrates. *J. Phys. Chem. B* **118**, 10624–32 (2014).
270. Tarcha, P. J., Chu, V. P. & Whittern, D. 2,3-diaminophenazine is the product from the horseradish peroxidase-catalyzed oxidation of o-phenylenediamine. *Anal. Biochem.* **165**, 230–3 (1987).
271. Gasteiger, E. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. in *The Proteomics Protocols Handbook* 571–607 (2005). doi:10.1385/1-59259-890-0:571
272. Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server\’98 Escherichia coli

- 
- SWISS-2DPAGE database update\rLarge-scale identification of proteins of Haemophilus influenzae by amino acid composition analysis. *Methods Mol Biol* (1999).
273. Cooper, C. A., Gasteiger, E. & Packer, N. H. GlycoMod - A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**, 340–9 (2001).
274. Cooper, C. A., Gasteiger, E. & Packer, N. H. Predicting Glycan Composition from Experimental Mass Using GlycoMod. in *Handbook of Proteomic Methods* 225–231 (2003).
275. AL-Maslookhi, H. S. A. Structure, function and engineering of biotechnologically important proteins. (Cardiff University, 2019).
276. Ortho Clinical Diagnostics. Tropl INSTRUCTIONS FOR USE. *Immunodiagn. Prod. Vitr.* (2019).
277. Song, Y. *et al.* High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–42 (2013).
278. Raman, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins Struct. Funct. Bioinforma.* **9**, 89–99 (2009).
279. Abraham, M.J., van der Spoel, D., Lindahl, E., Hess, B., and the G. development team. Gromacs 2019.2. *GROMACS User Man. version 2019* (2019).
280. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, (2017).
281. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically useful: What the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987–2998 (2010).
282. Liu, Y. & Kuhlman, B. RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, 235–8 (2006).
283. Lyskov, S. & Gray, J. J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **38**, 233–8 (2008).
284. DeLano, W. L. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* (2002).
285. DeLano, W. L. The PyMOL Molecular Graphics System, Version 1.8. *Schrödinger LLC* **1**, 14–5 (2014).
286. Arroyuelo, A., Vila, J. A. & Martin, O. A. Azahar: a PyMOL plugin for construction, visualization and analysis of glycan molecules. *J. Comput. Aided. Mol. Des.* **30**, 619–624 (2016).
287. Gfeller, D., Michielin, O. & Zoete, V. SwissSidechain: A molecular and structural database of non-natural sidechains. *Nucleic Acids Res.* **41**, 327–32 (2013).
288. UniProt. P00433. *UniProt* Available at: <https://www.uniprot.org/uniprot/P00433>.
289. Matsui, T., Nakayama, H., Yoshida, K. & Shinmyo, A. Vesicular transport route of horseradish C1a peroxidase is regulated by N- and C-terminal propeptides in tobacco cells. *Appl. Microbiol. Biotechnol.* **52**, 517–22 (2003).
290. Asad, S., Dastgheib, S. M. M. & Khajeh, K. Construction of a horseradish peroxidase resistant toward hydrogen peroxide by saturation mutagenesis. *Biotechnol. Appl. Biochem.* **63**, 789–794 (2016).
291. Kery, V., Elleder, D. & Kraus, J. P.  $\delta$ -Aminolevulinic acid increases heme saturation and yield of human cystathionine  $\beta$ -synthase expressed in Escherichia coli. *Arch. Biochem. Biophys.* **316**, 24–9 (1995).
292. Kis, M. *et al.* An N-terminal peptide extension results in efficient expression, but not secretion, of a synthetic horseradish peroxidase gene in transgenic tobacco. *Ann. Bot.* **93**, 303–10 (2004).
293. Ortlepp, S., Polard-Knight, D. & Chiswell, D. Expression and characterisation of a protein specified by a synthetic horseradish peroxidase gene in Escherichia coli. *J. Biotechnol.* **11**, (1989).
294. Khan, A. R. & James, M. N. G. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7**, 815–36 (2008).
295. Veitch, N. C., Gao, Y., Smith, A. T. & White, C. G. Identification of a critical phenylalanine residue in

- 
- horseradish peroxidase, Phe179, by site-directed mutagenesis and <sup>1</sup>H-NMR: Implications for complex formation with aromatic donor molecules. *Biochemistry* **236**, 14751–61 (1997).
296. Heering, H. A., Smulevich, G. & Smith, A. T. Role of the distal phenylalanine 41 on the properties of horseradish peroxidase C. *Spectrosc. Biol. Mol. New Dir.* **2**, 89–93 (1999).
297. Sockolosky, J. T. & Szoka, F. C. Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. *Protein Expr. Purif.* **87**, 129–135 (2013).
298. Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 18301–29 (2012).
299. Reddington, S. C. Introducing novel protein functionality using unnatural amino acids. (Cardiff University, 2013).
300. Xu, M Q., Chong, S. Protein Splicing. in *Encyclopedia of Genetics* 1565–1567 (2001). doi:<https://doi.org/10.1006/rwgn.2001.1039>
301. Shah, N. H. & Muir, T. W. Inteins: Nature’s gift to protein chemists. *Chem. Sci.* **5**, 446–61 (2014).
302. Groves, R. & Groves, R. Ship To : Ship To : **98607**, 1–2 (2015).
303. Ryan, B. J. & Ó’Fágáin, C. Arginine-to-lysine substitutions influence recombinant horseradish peroxidase stability and immobilisation effectiveness. *BMC Biotechnol.* **7**, 7–86 (2007).
304. Ryan, B. J., O’Connell, M. J. & Ó’Fágáin, C. Consensus mutagenesis reveals that non-helical regions influence thermal stability of horseradish peroxidase. *Biochimie* **90**, 1389–96 (2008).
305. Zheng, K., Bantog, C. & Bayer, R. The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* **3**, 568–76 (2011).
306. Yi, C. W. *et al.* Glycosylation Significantly Inhibits the Aggregation of Human Prion Protein and Decreases Its Cytotoxicity. *Sci. Rep.* **8**, 12603–4 (2018).
307. Peters, T. W. & Huang, M. Protein aggregation and polyasparagine-mediated cellular toxicity in *Saccharomyces cerevisiae*. *Prion* **1**, 144–53 (2007).
308. Rogers, J. M. Peptide Folding and Binding Probed by Systematic Non-canonical Mutagenesis. *Front. Mol. Biosci.* **24**, 100 (2020).
309. Scott, K. A., Alonso, D. O. V., Sato, S., Fersht, A. R. & Daggett, V. Conformational entropy of alanine versus glycine in protein denatured states. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2661–2666 (2007).
310. Lefèvre, F., Rémy, M. H. & Masson, J. M. Alanine-stretch scanning mutagenesis: A simple and efficient method to probe protein structure and function. *Nucleic Acids Res.* **25**, 447–8 (1997).
311. Tohjo, M. & Shibata, K. Absorption spectra of hematin complexes with polyhistidine and copoly-(histidine, glutamic acid). *Arch. Biochem. Biophys.* **103**, 401–408 (1963).
312. Van Rosmalen, M., Krom, M. & Merckx, M. Tuning the Flexibility of Glycine-Serine Linkers to Allow Rational Design of Multidomain Proteins. *Biochemistry* **56**, 6565–74 (2017).
313. Wilkins, M. R. *et al.* Detailed peptide characterization using PEPTIDEMASS - A World-Wide-Web-accessible tool. *Electrophoresis* **18**, 403–8 (1997).
314. Weinfurtner, D. CHAPTER 1.4. Analysis of Disulfide Bond Formation in Therapeutic Proteins. in 235–54 (2018). doi:[10.1039/9781788013253-00081](https://doi.org/10.1039/9781788013253-00081)
315. Jain, R. G., Rusch, S. L. & Kendall, D. A. Signal peptide cleavage regions. Functional limits on length and topological implications. *J. Biol. Chem.* (1994).
316. Paetzel, M. Structure and mechanism of Escherichia coli type I signal peptidase. *Biochim. Biophys. Acta - Mol. Cell Res.* **87**, 27–49 (2014).

- 
317. Han, S., Green, L. & Schnell, D. J. The signal peptide peptidase is required for pollen function in arabidopsis. *Plant Physiol.* **149**, 1289–1301 (2009).
318. Auclair, S. M., Bhanu, M. K. & Kendall, D. A. Signal peptidase I: Cleaving the way to mature proteins. *Protein Sci.* **21**, 13–25 (2012).
319. Wu, A. H. B. A selected history and future of immunoassay development and applications in clinical chemistry. *Clin. Chim. Acta* **23369**, 119–24 (2006).
320. Yalow, R. S. & Berson, S. A. Assay of plasma insulin in human subjects by immunological methods. *Nature* **184**, 1648–9 (1959).
321. Aydin, S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides* **72**, 4–15 (2015).
322. Hosseini, S., Vázquez-Villegas, P., Rito-Palomares, M. & Martínez-Chapa, S. O. Fundamentals and history of ELISA: The evolution of the immunoassays until invention of ELISA. in *SpringerBriefs in Applied Sciences and Technology* 1–18 (2018).
323. Marks, V. False-positive immunoassay results: A multicenter survey of erroneous immunoassay results from assays of 74 analytes in 10 donors from 66 laboratories in seven countries. *Clin. Chem.* (2002).
324. Tate, J. & Ward, G. 5-3 Interferences in immunoassay - Google Scholar. *Clin. Biochem. Rev.* **25**, 105–20 (2004).
325. Selby, C. Interference in immunoassay. *Ann. Clin. Biol.* **36**, 704–721 (1999).
326. Cole, L. A., Rinne, K. M., Shahabi, S. & Omrani, A. False-positive hCG assay results leading to unnecessary surgery and chemotherapy and needless occurrences of diabetes and coma [7]. *Clinical Chemistry* (1999).
327. Rotmensch, S. & Cole, L. A. False diagnosis and needless therapy of presumed malignant disease in women with false-positive human chorionic gonadotropin concentrations. *Lancet* **9205**, 712–5 (2000).
328. Martin, J. WO1986007462A1 Enzyme Assay of Body Fluids. (1985).
329. Lum, G., Solarz, D. E. & Farney, L. False Positive Cardiac Troponin Results in Patients Without Acute Myocardial Infarction. *Lab. Med.* **37**, 546–550 (2006).
330. Bolstad, N. *et al.* Heterophilic antibody interference in commercial immunoassays; A screening study using paired native and pre-blocked sera. *Clin. Chem. Lab. Med.* **49**, 2001–6 (2011).
331. Levinson, S. S. & Miller, J. J. Towards a better understanding of heterophile (and the like) antibody interference with modern immunoassays. *Clin. Chim. Acta* **325**, 1–15 (2002).
332. Ismail, A. A. A., Walker, P. L., Cawood, M. L. & Barth, J. H. Interference in immunoassay is an underestimated problem. in *Annals of Clinical Biochemistry* 45–60 (2002). doi:10.1258/000456302760042128
333. Park, K. C., Gaze, D. C., Collinson, P. O. & Marber, M. S. Cardiac troponins: From myocardial infarction to chronic disease. *Cardiovasc. Res.* **113**, 1708–1718 (2017).
334. Snoke, R. E. United States Patent US005776714A. (1998).
335. Van Avondt, K., Nur, E. & Zeerleder, S. Mechanisms of haemolysis-induced kidney injury. *Nature Reviews Nephrology* 671–692 (2019). doi:10.1038/s41581-019-0181-0
336. Kremer, M. L. The effect of dimerization on the catalytic activity of haemin. *BBA - Gen. Subj.* **297**, 268–75 (1973).
337. Brown, S. B. & Jones, P. Reactions between haemin and hydrogen peroxide. Part 3. - Catalytic decomposition of hydrogen peroxide. *Trans. Faraday Soc.* **64**, 999–1005 (1968).
338. Brown, S. B., Dean, T. C., Jones, P. & Kremer, M. L. Catalytic activity of haemin. Influence of dissolution

- 
- conditions on activity. *Trans. Faraday Soc.* **66**, 1485–90 (1970).
339. Kremer, M. L. Catalytic activity of haemin on the decomposition of hydrogen peroxide. *Nature* **205**, 384–85 (1965).
340. Sodi, R., Darn, S. M., Davison, A. S., Stott, A. & Shenkin, A. Mechanism of interference by haemolysis in the cardiac troponin T immunoassay. *Ann. Clin. Biochem.* **43**, 49–56 (2006).
341. Wenk, R. E. Mechanism of interference by hemolysis in immunoassays and requirements for sample quality [1]. *Clinical Chemistry* (1998).
342. Sturgeon, C. M. & Viljoen, A. Analytical error and interference in immunoassay: Minimizing risk. *Ann. Clin. Biochem.* **48**, 418–32 (2011).
343. Dimeski, G. Interference testing. *Clin. Biochem. Rev.* (2008).
344. Luong, J. H. T., Male, K. B. & Glennon, J. D. Biotin interference in immunoassays based on biotin-streptavidin chemistry: An emerging threat. *Biotechnol. Adv.* **37**, 634–41 (2019).
345. Näätsaari, L., Krainer, F. W., Schubert, M., Glieder, A. & Thallinger, G. G. Peroxidase gene discovery from the horseradish transcriptome. *BMC Genomics* **15**, 238–43 (2014).
346. Jones, D. D. & Barker, P. D. Design and characterisation of an artificial DNA-binding cytochrome. *ChemBioChem* **5**, 964–71 (2004).
347. Lee, H. S., Qi, Y. & Im, W. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Rep.* **5**, 1 (2015).
348. Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays Biochem.* **15**, 1–41 (2015).
349. Barker, P. D., Nerou, E. P., Freund, S. M. V. & Feamley, I. M. Conversion of Cytochrome b562 to c-Type Cytochromes. *Biochemistry* **34**, 15191–203 (1995).
350. Ortho Clinical Diagnostics. ZPR023 SIPOC Analysis ZPR023 SIPOC Analysis. **1**, 1–5
351. Dunford, H. B. Horseradish Peroxidase: Structure and Kinetic Properties. *Peroxidases Chem. Biol.* (1991).
352. Nienhaus, K. & Nienhaus, G. U. Probing heme protein-ligand interactions by UV/visible absorption spectroscopy. *Methods Mol. Biol.* **305**, 215–42 (2005).
353. Rimington, C. Spectral-absorption coefficients of some porphyrins in the Soret-band region. *Biochem. J.* **75**, 620–3 (1960).
354. Yang, F., Moss, L. G. & Phillips, G. N. The Molecular Structure of Green Fluorescent Protein. *Nat. Biotechnol.* **14**, 1246–51 (1996).
355. Mesters, J. R. & Hilgenfeld, R. Protein glycosylation, sweet to crystal growth? *Cryst. Growth Des.* **7**, 2251–3 (2007).
356. Itagaki, E. & Hager, L. P. Studies on cytochrome b-562 of Escherichia coli. I. Purification and crystallization of cytochrome b-562. *J. Biol. Chem.* (1966).
357. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Comput. Biol.* **11**, e1004049 (2015).
358. Tupper, J., Tozer, G. M. & Dachs, G. U. Use of horseradish peroxidase for gene-directed enzyme prodrug therapy with paracetamol. *Br. J. Cancer* **90**, 1858–62 (2004).
359. Han, M. *et al.* The role of N-glycosylation sites in the activity, stability, and expression of the recombinant elastase expressed by *Pichia pastoris*. *Enzyme Microb. Technol.* **10**, 32–7 (2014).
360. Wujek, P., Kida, E., Walus, M., Wisniewski, K. E. & Golabek, A. A. N-Glycosylation Is Crucial for Folding, Trafficking, and Stability of Human Tripeptidyl-peptidase I. *J. Biol. Chem.* **279**, 12827–39 (2004).

- 
361. Sola, R. J. & Griebenow, K. Effects of glycosylation on the stability of protein pharmaceuticals. *J. Pharm. Sci.* **98**, 1223–45 (2009).
  362. Sadat, M. A. *et al.* Glycosylation, hypogammaglobulinemia, and resistance to viral infections. *N. Engl. J. Med.* **370**, 1615–1625 (2014).
  363. Al Teneiji, A. *et al.* Phenotypic and genotypic spectrum of congenital disorders of glycosylation type I and type II. *Mol. Genet. Metab.* **120**, 235–242 (2017).
  364. Parekh, R. B. *et al.* Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature* **316**, 452–7 (1985).
  365. Mestecky, J. *et al.* Defective galactosylation and clearance of IgA1 molecules as a possible etiopathogenic factor in IgA nephropathy. *Contrib. Nephrol.* **104**, 172–82 (1993).
  366. Kameyama, A., Dissanayake, S. K. & Thet Tin, W. W. Rapid chemical de-N-glycosylation and derivatization for liquid chromatography of immunoglobulin N-linked glycans. *PLoS One* **56**, 124643–5 (2018).
  367. Landfried, D. A., Vuletich, D. A., Pond, M. P. & Lecomte, J. T. J. Structural and thermodynamic consequences of heme binding for monomeric apoglobins and other apoproteins. *Gene* **398**, 12–28 (2007).
  368. Strasser, R. Plant protein glycosylation. *Glycobiology* **9**, (2016).
  369. Spadiut, O. & Herwig, C. Production and purification of the multifunctional enzyme horseradish peroxidase. *Pharm. Bioprocess.* **1**, 283–295 (2013).
  370. Lige, B., Ma, S. & Van Huystee, R. B. The effects of the site-directed removal of N-glycosylation from cationic peanut peroxidase on its function. *Arch. Biochem. Biophys.* **386**, 17–24 (2001).
  371. Tams, J. W. & Welinder, K. G. Mild chemical deglycosylation of horseradish peroxidase yields a fully active, homogeneous enzyme. *Anal. Biochem.* **228**, 48–55 (1995).
  372. Hanna, D. A. *et al.* Heme dynamics and trafficking factors revealed by genetically encoded fluorescent heme sensors. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7539–44 (2016).
  373. Feng, Y., Sligar, S. G. & Wand, A. J. Solution structure of apocytochrome B562. *Nat. Struct. Biol.* **1**, 30–35 (1994).
  374. Arnesano, F. *et al.* The solution structure of oxidized Escherichia coli cytochrome b562. *Biochemistry* **38**, 8657–70 (1999).
  375. Allen, J. W. A., Leach, N. & Ferguson, S. J. The histidine of the c-type cytochrome CXXCH haem-binding motif is essential for haem attachment by the Escherichia coli cytochrome c maturation (Ccm) apparatus. *Biochem. J.* **389**, 587–92 (2005).
  376. Yang, B. Y., Gray, J. S. S. & Montgomery, R. The glycans of horseradish peroxidase. *Carbohydr. Res.* **32**, 124–30 (1996).
  377. Lakowicz, J. R. *Principles of fluorescence spectroscopy. Principles of Fluorescence Spectroscopy* (2006).
  378. Sanderson, M. J., Smith, I., Parker, I. & Bootman, M. D. Fluorescence microscopy. *Cold Spring Harb. Protoc.* **10**, 2138–40 (2014).
  379. Chattopadhyay, A. & Mukherjee, S. Fluorophore Environments in Membrane-Bound Probes: A Red Edge Excitation Shift Study. *Biochemistry* **32**, 3804–11 (1993).
  380. Demchenko, A. P. The red-edge effects: 30 years of exploration. *Luminescence* (2002). doi:10.1002/bio.671
  381. Lakowicz, J. R. & Keating-Nakamoto, S. Red-Edge Excitation of Fluorescence and Dynamic Properties of Proteins and Membranes. *Biochemistry* **324**, 238–43 (1984).

- 
382. Demchenko, A. P. Chapter 4 Site-Selective Red-Edge Effects. *Methods in Enzymology* 245–9 (2008). doi:10.1016/S0076-6879(08)03404-6
383. Rawat, S. S., Kelkar, D. A. & Chattopadhyay, A. Monitoring gramicidin conformations in membranes: A fluorescence approach. *Biophys. J.* **87**, 831–43 (2004).
384. Demchenko, A. P. Dependence of human serum albumin fluorescence spectrum on the excitation wavelength. *Ukr. Biokhim. Zh.* **53**, 22–27 (1981).
385. Catci, D. A. M., Amos, H. E., Yang, Y., van den Elsen, J. M. H. & Pudney, C. R. The red edge excitation shift phenomenon can be used to unmask protein structural ensembles: implications for NEMO–ubiquitin interactions. *FEBS J.* **283**, 2272–84 (2016).
386. Kelkar, D. A., Chaudhuri, A., Haldar, S. & Chattopadhyay, A. Exploring tryptophan dynamics in acid-induced molten globule state of bovine  $\alpha$ -lactalbumin: A wavelength-selective fluorescence approach. *Eur. Biophys. J.* **39**, 1453–63 (2010).
387. Cheng, Z., Kuru, E., Sachdeva, A. & Vendrell, M. Fluorescent amino acids as versatile building blocks for chemical biology. *Nat. Rev. Chem.* **4**, 275–290 (2020).
388. Demchenko, A. P. The red-edge effects: 30 years of exploration. *Luminescence* **17**, 19–42 (2002).
389. Burshtein, E. Intrinsic luminescence of protein as a tool for studying fast structural Dynamics. *Biology (Basel)*. (1983).
390. Reshetnyak, Y. K., Koshevnik, Y. & Burstein, E. A. Decomposition of protein tryptophan fluorescence spectra into log-normal components. III. Correlation between fluorescence and microenvironment parameters of individual tryptophan residues. *Biophys. J.* **81**, 1735–58 (2001).
391. Knight, M. *et al.* Monoclonal antibody stability can be usefully monitored using the excitation-energy-dependent fluorescence edge-shift. *Biochem. J.* **477**, 3599–3612 (2020).
392. Jones, H. B. L. *et al.* A complete thermodynamic analysis of enzyme turnover links the free energy landscape to enzyme catalysis. *FEBS J.* **65**, 198–9 (2017).
393. Pham, T. A. *et al.* Heat-induced formation of one-dimensional coordination polymers on Au(111): An STM study. *Chem. Commun.* **51**, 14473–6 (2015).
394. Alpert, B. & Lopez-delgado, R. Fluorescence lifetimes of haem proteins excited into the tryptophan absorption band with synchrotron radiation. *Nature* **263**, 445–6 (1976).
395. Wei, Y., Thyparambil, A. A. & Latour, R. A. Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230 nm. *Biochim. Biophys. Acta - Proteins Proteomics* **1844**, 2331–7 (2014).
396. Chemes, L. B., Alonso, L. G., Noval, M. G. & De Prat-Gay, G. Circular dichroism techniques for the analysis of intrinsically disordered proteins and domains. *Methods Mol. Biol.* **895**, 387–404 (2012).
397. Mohamed, S. A., Darwish, A. A. & El-Shishtawy, R. M. Immobilization of horseradish peroxidase on activated wool. *Process Biochem.* **48**, (2013).
398. Wang, T., Tian, J., Li, T. & Li, J. Enzyme properties and thermal stability of horseradish peroxidase (HRP-DL). **23**, 234–9 (2016).
399. Henriksen, A. Structure of soybean seed coat peroxidase: A plant peroxidase with unusual stability and haem-apoprotein interactions. *Protein Sci.* **10**, 108–15 (2001).
400. Henriksen, A., Welinder, K. G. & Gajhede, M. Structure of barley grain peroxidase refined at 1.9-Å Resolution: A plant peroxidase reversibly inactivated at neutral pH. *J. Biol. Chem.* **68**, 19834–7 (1998).
401. Moyal, T. W. *et al.* Characterization of class III peroxidases from switchgrass. *Plant Physiol.* **173**, 417–433 (2017).

- 
402. Nielsen, K. L. *et al.* Differential activity and structure of highly similar peroxidases. Spectroscopic, crystallographic, and enzymatic analyses of lignifying *Arabidopsis thaliana* peroxidase A2 and horseradish peroxidase A2. *Biochemistry* **40**, 11013–21 (2001).
403. Watanabe, L. *et al.* Crystal structure and statistical coupling analysis of highly glycosylated peroxidase from royal palm tree (*Roystonea regia*). *J. Struct. Biol.* **169**, 226–42 (2010).
404. Li, T., Bonkovsky, H. L. & Guo, J. T. Structural analysis of heme proteins: Implications for design and prediction. *BMC Struct. Biol.* **11**, 138–9 (2011).
405. Hernández-Cancel, G. *et al.* Chemically glycosylation improves the stability of an amperometric horseradish peroxidase biosensor. *Anal. Chim. Acta* **854**, 129–39 (2015).
406. Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **23**, 105–53 (2000).
407. Wiens, M. D. *et al.* A Tandem Green–Red Heterodimeric Fluorescent Protein with High FRET Efficiency. *ChemBioChem* **17**, 2361–7 (2016).
408. Hashimoto, K. & Panchenko, A. R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20352–20357 (2010).
409. Fallas, J. A. *et al.* Computational design of self-assembling cyclic protein homo-oligomers. *Nat. Chem.* **9**, 353–60 (2017).
410. Raina, S. & Missiakas, D. MAKING AND BREAKING DISULFIDE BONDS. *Annu. Rev. Microbiol.* **51**, 179–202 (1997).
411. Aroul-Selvam, R., Hubbard, T. & Sasidharan, R. Domain insertions in protein structures. *J. Mol. Biol.* **338**, 633–41 (2004).
412. Yarbrough, D., Wachter, R. M., Kallio, K., Matz, M. V. & Remington, S. J. Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-Å resolution. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 462–67 (2001).
413. Shagin, D. A. *et al.* GFP-like Proteins as Ubiquitous Metazoan Superfamily: Evolution of Functional Features and Structural Complexity. *Mol. Biol. Evol.* **21**, 841–50 (2004).
414. Gross, L. A., Baird, G. S., Hoffman, R. C., Baldridge, K. K. & Tsien, R. Y. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 11990–11995 (2000).
415. Terskikh, A. V., Fradkov, A. F., Zaraksky, A. G., Kajava, A. V. & Angres, B. Analysis of DsRed mutants: Space around the fluorophore accelerates fluorescence development. *J. Biol. Chem.* **277**, 7633–6 (2002).
416. Mizukami, S., Hori, Y. & Kikuchi, K. Small-molecule-based protein-labeling technology in live cell studies: Probe-design concepts and applications. *Acc. Chem. Res.* **47**, 247–56 (2014).
417. Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7877–7882 (2002).
418. Lin, M. Z. *et al.* Autofluorescent Proteins with Excitation in the Optical Window for Intravital Imaging in Mammals. *Chem. Biol.* **16**, 1169–74 (2009).
419. Shen, Y., Chen, Y., Wu, J., Shaner, N. C. & Campbell, R. E. Engineering of mCherry variants with long Stokes shift, red-shifted fluorescence, and low cytotoxicity. *PLoS One* **12**, 171257–65 (2017).
420. Subach, F. V. & Verkhusha, V. V. Chromophore transformations in red fluorescent proteins. *Chemical Reviews* 123–7 (2012).
421. Subach, F. V. *et al.* Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat. Methods* **6**, 153–9 (2009).

- 
422. Shcherbakova, D. M. & Verkhusha, V. V. Chromophore chemistry of fluorescent proteins controlled by light. *Curr. Opin. Chem. Biol.* **1**, 60–68 (2014).
423. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2**, (2005).
424. Day, R. N. & Davidson, M. W. Fluorescent proteins for FRET microscopy: Monitoring protein interactions in living cells. *BioEssays* **34**, 341–50 (2012).
425. Clegg, R. M. Chapter 1 Förster resonance energy transfer-FRET what is it, why do it, and how it's done. *Lab. Tech. Biochem. Mol. Biol.* **33**, 1–57 (2009).
426. Sekar, R. B. & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160**, 629–33 (2003).
427. Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* **32**, 96–99 (2004).
428. Ritchie, D. W. & Kemp, G. J. L. Protein docking using spherical polar Fourier correlations. *Proteins Struct. Funct. Genet.* **34**, 278–9 (2000).
429. Shu, X., Shaner, N. C., Yarbrough, C. A., Tsien, R. Y. & Remington, S. J. Novel chromophores and buried charges control color in mFruits. *Biochemistry* **45**, 9639–47 (2006).
430. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein Structure Prediction Using Rosetta. *Methods Enzymol.* **3**, 297–305 (2004).
431. Worthy, H. L. Novel routes to defined post translational modifications using non-canonical amino acids. (Cardiff University, 2018).
432. Rohl, C. A. Protein structure estimation from minimal restraints using Rosetta. *Methods Enzymol.* **288**, 2466–8 (2005).
433. Conchúir, S. *et al.* A Web resource for standardized benchmark datasets, metrics, and rosetta protocols for macromolecular modeling and design. *PLoS One* **3**, 0130433–5 (2015).
434. Gobbo, P., Novoa, S., Biesinger, M. C. & Workentin, M. S. Interfacial strain-promoted alkyne-azide cycloaddition (I-SPAAC) for the synthesis of nanomaterial hybrids. *Chem. Commun.* **49**, 3982–3984 (2013).
435. Kuzichkina, E. O., Shilova, O. N. & Deyev, S. M. The mechanism of fluorescence quenching of protein photosensitizers based on miniSOG during internalization of the HER2 receptor. *Acta Naturae* **10**, 87–94 (2018).
436. Sanchez, K. M., Schlamadinger, D. E., Gable, J. E. & Kim, J. E. Förster resonance energy transfer and conformational stability of proteins: An advanced biophysical module for physical chemistry students. *J. Chem. Educ.* **85**, 1253–6 (2008).
437. Wu, B., Chen, Y. & Müller, J. D. Fluorescence fluctuation spectroscopy of mCherry in living cells. *Biophys. J.* **96**, 2391–404 (2009).
438. Strack, R. Sensors and probes: Super-resolution imaging with mCherry. *Nat. Methods* **114**, 7013–18 (2017).
439. van Zyl, W. F., Deane, S. M. & Dicks, L. M. T. Use of the mcherry fluorescent protein to study intestinal colonization by *Enterococcus mundtii* ST4SA and *Lactobacillus plantarum* 423 in mice. *Appl. Environ. Microbiol.* **81**, 5993–6002 (2015).
440. Prangma, J. C. *et al.* Quantitative Determination of Dark Chromophore Population Explains the Apparent Low Quantum Yield of Red Fluorescent Proteins. *J. Phys. Chem. B* **24**, 346–8 (2020).
441. Cloin, B. M. C. *et al.* Efficient switching of mCherry fluorescence using chemical caging. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7013–18 (2017).

- 
442. Sattarzadeh, A., Saberianfar, R., Zipfel, W. R., Menassa, R. & Hanson, M. R. Green to red photoconversion of GFP for protein tracking in vivo. *Sci. Rep.* **5**, 11771–5 (2015).
443. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. & Leibler, S. Photoactivation turns green fluorescent protein red. *Curr. Biol.* **7**, 809–12 (1997).
444. Li, H. & Vaughan, J. C. Switchable Fluorophores for Single-Molecule Localization Microscopy. *Chem. Rev.* **341**, 9412–54 (2018).
445. Dean, K. M. *et al.* Analysis of red-fluorescent proteins provides insight into dark-state conversion and photodegradation. *Biophys. J.* **101**, 961–969 (2011).
446. Brown, J. H. Breaking symmetry in protein dimers: Designs and functions. *Protein Sci.* **15**, 1–13 (2006).
447. O’Sullivan, M. J., Bridges, J. W. & Marks, V. Enzyme immunoassay: A review. *Ann. Clin. Biochem.* **16**, 221–40 (1979).
448. Hosseinzadeh, P. & Lu, Y. Design and fine-tuning redox potentials of metalloproteins involved in electron transfer in bioenergetics. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 557–81 (2016).
449. Battersby, A. Tetrapyrroles: the pigments of life. *RSC Adv.* (2000).
450. Johansson, M. P., Blomberg, M. R. A., Sundholm, D. & Wikström, M. Change in electron and spin density upon electron transfer to haem. *Biochim. Biophys. Acta - Bioenerg.* **4**, 479–84 (2002).
451. Diana, R. & Panunzi, B. The role of zinc(II) ion in fluorescence tuning of tridentate pincers: A review. *Molecules* **25**, 4984–8 (2020).
452. Pamei, M. & Puzari, A. Luminescent transition metal–organic frameworks: An emerging sensor for detecting biologically essential metal ions. *Nano-Structures and Nano-Objects* **19**, (2019).
453. Jones, G. A. & Bradshaw, D. S. Resonance energy transfer: From fundamental theory to recent applications. *Front. Phys.* **434**, 341–98 (2019).
454. Mathews, F. S., Bethge, P. H. & Czerwinski, E. W. The structure of cytochrome b562 from *Escherichia coli* at 2.5 Å resolution. *J. Biol. Chem.* **254**, 1699–706 (1979).
455. Barker, P. D., Butler, J. L., De Oliveira, P., Hill, H. A. O. & Hunt, N. I. Direct electrochemical studies of cytochromes b562. *Inorganica Chim. Acta* **252**, 71–77 (1996).
456. NIKKILA, H., GENNIS, R. B. & SLIGAR, S. G. Cloning and expression of the gene encoding the soluble cytochrome b562 of *Escherichia coli*. *Eur. J. Biochem.* **2**, (202AD).
457. Takeda, S., Kamiya, N. & Nagamune, T. Rational design of a protein-based molecular device consisting of blue fluorescent protein and zinc protoporphyrin IX incorporated into a cytochrome b562 scaffold. *Biotechnol. Lett.* **26**, 121–5 (2004).
458. Takeda, S., Kamiya, N., Arai, R. & Nagamune, T. Design of an artificial light-harvesting unit by protein engineering: Cytochrome b562-green fluorescent protein chimera. *Biochem. Biophys. Res. Commun.* **289**, 299–304 (2001).
459. Lee, B. *et al.* Rectified photocurrent in a protein based molecular photo-diode consisting of a cytochrome b562-green fluorescent protein chimera self-assembled monolayer. *Biosens. Bioelectron.* **19**, 1169–74 (2004).
460. Takeda, S., Kamiya, N. & Nagamune, T. A novel protein-based heme sensor consisting of green fluorescent protein and apocytochrome b562. *Anal. Biochem.* **317**, 116–9 (2003).
461. Arpino, J. A. J. Single amino acid deletions to whole domain insertions; Engineering GFP through polypeptide backbone mutations. (Cardiff University, 2011).
462. Förster, T. Förster - Intermolecular Energy Transference and Fluorescence.pdf. *Ann Phys.* (1948).
463. Pickens, C. J., Johnson, S. N., Pressnall, M. M., Leon, M. A. & Berkland, C. J. Practical Considerations,

- 
- Challenges, and Limitations of Bioconjugation via Azide-Alkyne Cycloaddition. *Bioconjug. Chem.* **29**, 686–701 (2018).
464. Robinson, C. R., Liu, Y., Thomson, J. A., Sturtevant, J. M. & Sligar, S. G. Energetics of heme binding to native and denatured states of cytochrome b562. *Biochemistry* **36**, 16141–6 (1997).
465. Scott, D. J. *et al.* A Novel Ultra-Stable, Monomeric Green Fluorescent Protein for Direct Volumetric Imaging of Whole Organs Using CLARITY. *Sci. Rep.* **8**, 872–8 (2018).
466. Aronson, D. E., Costantini, L. M. & Snapp, E. L. Superfolder GFP Is Fluorescent in Oxidizing Environments When Targeted via the Sec Translocon. *Traffic* **12**, 543–8 (2011).
467. Şener, M. *et al.* Förster energy transfer theory as reflected in the structures of photosynthetic light-harvesting systems. *ChemPhysChem* **34**, 984–9 (2011).
468. Hanson, G. T. *et al.* Investigating Mitochondrial Redox Potential with Redox-sensitive Green Fluorescent Protein Indicators. *J. Biol. Chem.* **279**, 13044–53 (2004).
469. Della Pia, E. A. *et al.* Fast electron transfer through a single molecule natively structured redox protein. *Nanoscale* **4**, (2012).
470. Pia, E. A. Della *et al.* Redox tuning of cytochrome b562 through facile metal porphyrin substitution. *Chem. Commun.* **48**, 10624–6 (2012).
471. Pecci, L., Montefoschi, G., Musci, G. & Cavallini, D. Novel findings on the copper catalysed oxidation of cysteine. *Amino Acids* **13**, (1997).
472. Matthews, B. W. & Bernhard, S. A. Structure and symmetry of oligomeric enzymes. *Annual review of biophysics and bioengineering* (1973). doi:10.1146/annurev.bb.02.060173.001353
473. Shah, S. I. A., Kostyuk, L. W. & Kresta, S. M. The effects of mixing, reaction rates, and stoichiometry on yield for mixing sensitive reactions - Part I: Model development. *Int. J. Chem. Eng.* **4**, 863–8 (2012).
474. Meyer, A. J. & Dick, T. P. Fluorescent protein-based redox probes. *Antioxidants Redox Signal.* **56**, 348–9 (2010).
475. Fan, Y. & Ai, H. W. Development of redox-sensitive red fluorescent proteins for imaging redox dynamics in cellular compartments. *Anal. Bioanal. Chem.* **408**, 2901–11 (2016).
476. Roma, L. P., Deponte, M., Riemer, J. & Morgan, B. Mechanisms and applications of redox-sensitive green fluorescent protein-based hydrogen peroxide probes. *Antioxidants Redox Signal.* **29**, 552–68 (2018).
477. Rosenwasser, S. *et al.* A fluorometer-based method for monitoring oxidation of redox-sensitive GFP (roGFP) during development and extended dark stress. *Physiol. Plant.* **138**, 493–502 (2010).
478. Schillberg, S., Raven, N., Spiegel, H., Rasche, S. & Buntru, M. Critical analysis of the commercial potential of plants for the production of recombinant proteins. *Front. Plant Sci.* **10**, 1–10 (2019).
479. Huang, C. J., Lin, H. & Yang, X. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J. Ind. Microbiol. Biotechnol.* **39**, 383–99 (2012).
480. Jin, S. & Kennedy, R. T. New developments in Western blot technology. *Chinese Chem. Lett.* **26**, 416–8 (2015).
481. McCann, N. *et al.* Molecular interactions between amine and carbonate species in aqueous solution - kinetics and thermodynamics. *Energy Procedia* **2**, 27–9 (2009).
482. Hand, E. S. & Jencks, W. P. Mechanism of the Reaction of Imido Esters with Amines. *J. Am. Chem. Soc.* **84**, 3505–14 (1962).
483. Oliveira, B. L., Guo, Z. & Bernardes, G. J. L. Inverse electron demand Diels-Alder reactions in chemical biology. *Chem. Soc. Rev.* **46**, (2017).
484. Flocco, C. G. & Giulietti, A. M. Effect of chitosan on peroxidase activity and isoenzyme profile in hairy

- 
- root cultures of *Armoracia lapathifolia*. *Appl. Biochem. Biotechnol. - Part A Enzym. Eng. Biotechnol.* **110**, 175–83 (2003).
485. Hartmann, C. & Ortiz de Montellano, P. R. Baculovirus expression and characterization of catalytically active horseradish peroxidase. *Arch. Biochem. Biophys.* **291**, 61–72 (1992).
486. Greco, O., Folkes, L. K., Wardman, P., Tozer, G. M. & Dachs, G. U. Development of a novel enzyme/prodrug combination for gene therapy of cancer: Horseradish peroxidase/indole-3-acetic acid. *Cancer Gene Ther.* **324**, 1873–7 (2000).
487. Barondeau, D. P., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* **45**, 34676–8 (2005).
488. Remington, S. J. Green fluorescent protein: A perspective. *Protein Sci.* **34**, 8745–9 (2011).
489. Tomosugi, W. *et al.* An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nat. Methods* **6**, 351–3 (2009).
490. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6**, 178–182 (1996).
491. Cubitt, A. B., Woollenweber, L. A. & Heim, R. Understanding structure - Function relationships in the *Aequorea victoria* green fluorescent protein. *Methods Cell Biol.* **58**, (1999).
492. Lam, A. J. *et al.* Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat. Methods* **9**, 1005–1012 (2012).
493. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.* **20**, 87–90 (2002).