



Understanding and Engineering Protein Oligomers

Rochelle Ahmed

A thesis submitted to Cardiff University for the degree

Doctor of Philosophy

December 2022

School of Biosciences

Cardiff University



Ysgoloriaethau Sgiliau Economi Gwybodaeth
Knowledge Economy Skills Scholarships

Acknowledgements

I'd firstly like to express my sincere gratitude to my supervisor Prof Dafydd Jones for all the guidance and support over the years, particularly during the writing of this thesis. I have learnt so much over this time, cementing my passion and desire to pursue a career in protein science. Thank you to Dr Ismail Taban who helped my transition from an undergraduate student to a PhD researcher. I also want to acknowledge other past and present members of the DDJ group: Dr Ben Bowen, Dr Rachel Johnson, Dr Harley Worthy, Ozan Aksakal, Lainey Williamson, and Athena Zitti. A special mention for my friend Rebecca Gwyther who has been on this PhD journey with me since day one. Thank you for all your kindness, support, and most of all your continued reassurance.

I also want to acknowledge several other members of the school who helped me throughout my PhD. A big thank you to Dr Pierre Rizkallah (School of Medicine, Cardiff University) who helped me solve several crystal structures and understand the principles of protein crystallography. I'm also grateful for the collaboration with Dr Ben Bax's lab and the help provided by Dr Magdalena Lipka-Lloyd.

I then want to thank all my friends and family who continued to motivate me throughout. Finally, a special acknowledgement to my partner Ciaran for whom I will be forever grateful. This thesis would not have been possible without your love, support and patience during this very long endeavour.

Abstract

Protein oligomerisation is a fascinating topic concerning the association of proteins to form a single multimeric complex, either naturally or artificially. The importance of protein association is highlighted by their prevalence, where protein oligomers are more prevalent in nature than monomers. They are critical for a wide range of biological processes such as providing diversity and specificity to physiological pathways, mediating gene expression, and the activity of enzymes and ion channels. They also play a key role in regulation of apoptosis and tumour formation. Thus, unveiling the structural and functional aspects of protein oligomerisation has acquired increasing importance.

The first portion of this thesis examines the oligomerisation state of Bcl3-p50, a I κ B - NF- κ B protein complex evidenced to play a key role in breast cancer metastasis. A range of *in silico*, structural and biophysical techniques were used to understand the binding and stability of PPIs governing the protein complex (Chapter 3). A novel small molecule inhibitor of Bcl3 was also tested for its ability to block interactions between Bcl3 and p50.

One key method that is widely used to study biomolecular interactions within cells is FRET (or Förster resonance energy transfer). However, the dependence of two fluorescent proteins limits this technique to monitoring heterodimerisation. Thus, the second aspect of this thesis sets out to engineer superfolder Green Fluorescent Protein (sfGFP) for use as a homodimer biosensor. This work was divided into separate sections. First, the design and incorporation of mutations into desired residue positions of sfGFP to alter the spectral properties (Chapter 4). Next, establishing PPIs between sfGFP subunits and detection of homodimerisation through resulting fluorescent changes (Chapter 5). Finally, fusing the engineered sfGFP to p50 to evidence its ability to detect homodimerisation (Chapter 6).

Associated papers

Papers associated with this thesis are listed below:

Chapter 3

Pope, J.R. et al. 2021. Association of Fluorescent Protein Pairs and Its Significant Impact on Fluorescence and Energy Transfer. *Advanced Science* 8(1). doi: 10.1002/advs.202003167.

Chapters 4 and 5:

Ahmed, R.D, Auhim, H.S, Worthy, H.L, Jones, D.D, 2022. Fluorescent Proteins: Crystallization, Structural Determination, and Nonnatural Amino Acid Incorporation. In: Mayank Sharma. eds. *Fluorescent Proteins: Methods and Protocols*. New York. Humana. pp 99-120

Contents

Acknowledgements.....	i
Abstract.....	ii
Associated papers.....	iii
Contents.....	iv
1 Introduction.....	1
1.1 NF- κ B signalling.....	1
1.1.1 NF- κ B dimers.....	1
1.1.2 Regulation by I κ B protein family	3
1.2 B-cell Lymphoma 3	5
1.2.1 Structure of Bcl3	6
1.2.2 Bcl3-NF- κ B interactions.....	7
1.2.3 Role of Bcl3 in Breast Cancer & Metastasis.....	8
1.3 Fluorescent proteins	9
1.3.1 Green fluorescent protein (GFP).....	9
1.3.2 Yellow fluorescent protein (Venus).....	14
1.4 Protein oligomerisation	16
1.4.1 Engineering protein oligomers	17
1.5 Aims and Objectives	19
2 Materials and methods	21
2.1 Materials.....	21
2.1.1 Chemicals.....	21
2.1.2 Bacterial Strains	21
2.1.3 Bacterial Growth Media.....	22

2.2	Molecular Biology	22
2.2.1	Vectors for protein expression	22
2.2.2	DNA Oligonucleotides.....	24
2.2.3	Amplification of DNA by PCR.....	26
2.2.4	Agarose gel electrophoresis	28
2.2.5	DNA purification and isolation	28
2.2.6	DNA Quantification	29
2.2.7	Restriction endonuclease digestion	29
2.2.8	Phosphorylation & ligation DNA.....	30
2.2.9	Bacterial Transformation.....	30
2.3	Protein production & purification.....	31
2.3.1	Protein production and cell lysis	31
2.3.2	Protein Purification	32
2.3.3	Metal Affinity Chromatography.....	33
2.3.4	Size Exclusion Chromatography.....	33
2.3.5	SfGFP dimers	34
2.4	Protein analysis	34
2.4.1	SDS-PAGE.....	34
2.4.2	Protein concentration determination	35
2.4.3	Calculation of extinction coefficient.....	36
2.4.4	Fluorescence Spectroscopy	36
2.4.5	Förster resonance electron transfer (FRET).....	36
2.4.6	Determination of molecular weight by gel filtration chromatography	38

2.4.7	Determination of quantum yield	40
2.4.8	Characterising dimerisation of sfGFP ^{C148}	41
2.4.9	Isothermal Calorimetry	41
2.5	Structural studies.....	42
2.5.1	Mass spectrometry	42
2.5.2	SEC – DLS	42
2.5.3	Thermal Melting Assay.....	43
2.5.4	X-ray Crystallography.....	44
2.6	<i>In Silico</i> Modelling.....	45
2.6.1	Input structures and Computational Requirements	45
2.6.2	Parametrisation.....	46
2.6.3	Molecular dynamics with GROMACS	47
2.6.4	Analysis of the Root Mean Square Fluctuation (RMSF) ..	48
2.6.5	Analysis of distance	48
3	The Bcl3-p50 complex	49
3.1	Introduction.....	49
3.2	Results.....	51
3.2.1	Structural Studies of Bcl3-p50 complex	51
3.2.2	Production of FP-tagged Bcl3-p50 complex.....	58
3.2.3	Testing small-molecule Bcl3 inhibitor, JS6	79
3.3	Discussion	83
3.4	Conclusion	86
4	Engineering residue 148 of sfGFP	87
4.1	Introduction.....	87

4.2	Results.....	90
4.2.1	Design and production of sfGFP ^{C148} , sfGFP ^{V148} , sfGFP ^{S148} and sfGFP ^{A148}	90
4.2.2	Dimerisation of sfGFP ^{X148}	102
4.2.3	X-ray crystallography of sfGFP ^{C148.F} monomer.....	106
4.3	Discussion.....	113
4.4	Conclusion.....	117
5	Modulation of sfGFP structure and function by homodimerisation.....	118
5.1	Introduction.....	118
5.2	Results.....	120
5.2.1	Design of mutations to promote dimerisation of sfGFP ^{C148} 120	
5.2.2	Production of sfGFP ^{C148.F} monomer.....	122
5.2.3	Dimerisation of sfGFP ^{C148.F}	126
5.2.4	X-ray crystallography of sfGFP ^{C148.F} dimer.....	139
5.2.5	Discussion.....	157
5.2.6	Conclusion.....	160
6	Design of a homodimer biosensor.....	162
6.1	Introduction.....	162
6.2	Results.....	164
6.2.1	Design and construction of p50GFP.....	164
6.2.2	Expression and purification of p50GFP ^{C148.F}	167
6.2.3	Dimerisation rates of p50GFP ^{C148.F}	168

6.2.4	Calculation of dissociation constant.....	170
6.2.5	Measuring homodimerisation in <i>E. coli</i>	171
6.3	Discussion	173
6.4	Conclusion	176
7	Discussion	177
7.1	General overview	177
7.2	Usefulness of <i>in silico</i> modelling.....	178
7.3	Significance of small structural perturbations on local conditions	180
7.4	Future work.....	180
	References.....	182

1 Introduction

1.1 NF- κ B signalling

The transcription factor NF- κ B was first discovered in 1986, by demonstration of its DNA-binding activity for the enhancer of the immunoglobulin κ light-chain gene in activated B cells (Sen and Baltimore 1986; Yu et al. 2009). NF- κ B functions to both induce and repress gene expression by binding to discrete DNA sequences, known as κ B elements, present in enhancers or promoters (Perkins 2007). The genes regulated by NF- κ B include those controlling cell death (apoptosis), proliferation, cell adhesion, the innate- and adaptive- immune responses, inflammation, the cellular-stress response and tissue remodelling (Pahl 1999; Bonizzi and Karin 2004; Hayden and Ghosh 2004). The expression of these genes is also tightly coordinated with the activity of several other signalling and transcription factor pathways (e.g. IKK). With such a vast range of biological functions, the deregulation or inappropriate activity of NF- κ B (and other associated transcription factor pathways) can contribute to the pathology of inflammatory diseases and cancer. Thus, improving our understanding of these pathways will help development treatments (e.g. NF- κ B or IKK inhibitors) for many diagnostic and prognostic applications.

1.1.1 NF- κ B dimers

Since the initial discovery, the NF- κ B protein family has expanded to 5 members including NF- κ B1 (p105/p50), NF- κ B2 (p100/p52), RelA (p65), RelB and c-Rel (Figure 1a)(Bonizzi and Karin 2004). They exist as either a homo- or heterodimer of proteins, which each share a conserved N-terminal feature known as the “Rel homology domain” (RHD). The RHD is comprised of 300 amino acids and is responsible for dimerisation (prerequisite to DNA binding), site-specific DNA binding, nuclear localisation and binding of I κ B proteins (Bonizzi and Karin 2004). The RHD folds into two domains where the C-terminal domain is predominantly responsible for dimer formation, and the N-terminal domain for DNA binding (Figure 1b). The three-dimensional (3D) structures of the RHD have been noted to

resemble a butterfly where the protein domains are wings connected to a cylindrical body of DNA (Ghosh et al. 1995; Moorthy et al. 2007).

Unlike the other subunits, p50 and p52 are produced by proteolytic cleavage from their precursors p105 and p100 respectively. These precursors also contain ankyrin (ANK) repeat motifs similar to those found in I κ B proteins (Figure 1a) (Perkins 2007). The p105/p50 and p100/p52 subunits are also unique in that they lack the transactivation domain (TAD) present in RelA, RelB and c-Rel.

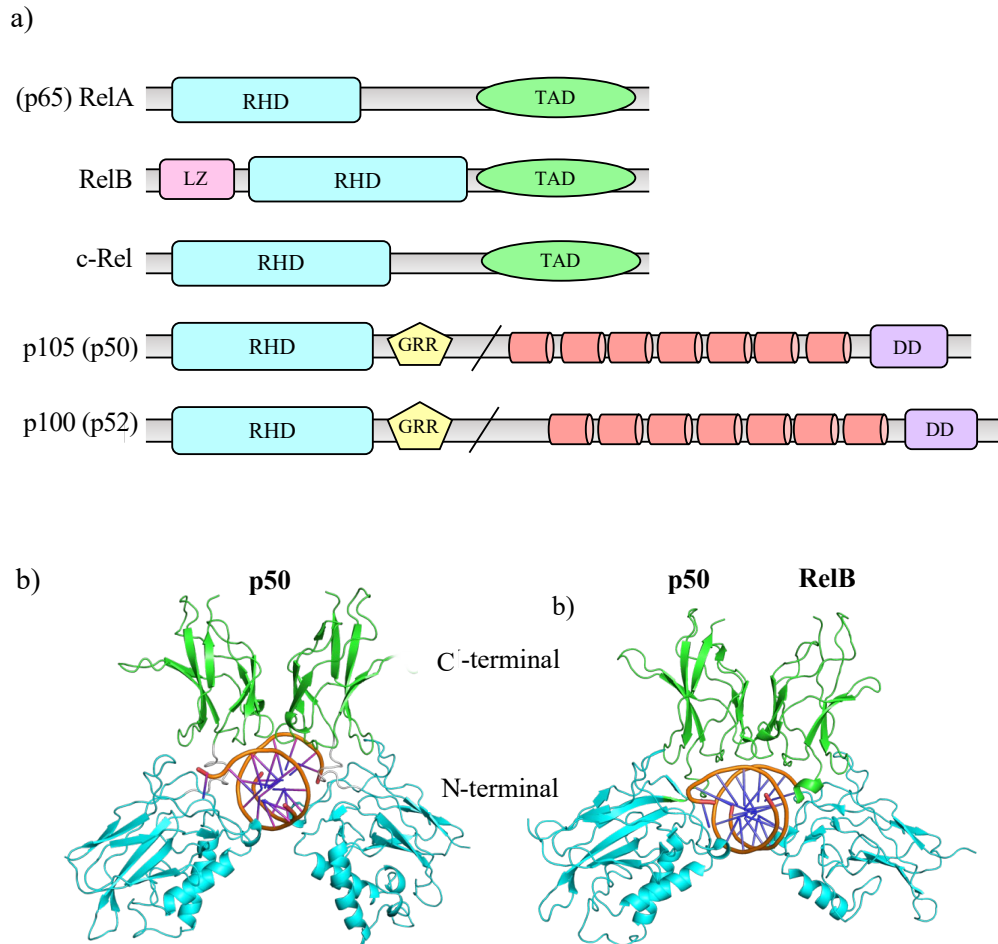


Figure 1. NF- κ B family members. a) Five members of NF- κ B containing RelA, RelB, c-Rel, p105 (p50), and p100 (p52). Rel homology domain (RHD) shown in blue. The leucine zipper (LZ) in pink, the transactivation domain (TAD) in green, glycine-rich region (GRR) in yellow, ANK in red, and death domain (DD) in purple. The processing of precursor molecules p100 and p105 to generate p52 and p50 proteins respectively is shown by the dash. Adapted from Perkins (2007). b) Structure of p50 homodimer (PDB ID: 1NFK) and p50/RelB heterodimer (PDB ID: 2V2T) complexed to DNA. N- and C-terminal domains shown in cyan and green respectively.

Through combinatorial associations, the NF- κ B protein family can form up to 15 different dimers, where each may exhibit distinct properties such as; differential preferences for certain κ B sites, transactivation potentials, kinetics of nuclear translocation, and levels of expression in tissue (Figure 2)(Siebenlist et al. 1994)(Hoesel and Schmid 2013). For example, p50 and p52 homodimers lack a TAD so function as transcriptional repressors unless activated by binding to a co-activator (Concetti and Wilson 2018). Whereas, p50/RelA heterodimer, the most abundant of NF- κ B complexes, is involved in the regulation of diverse cellular functions including immune response, cell growth and development (Chen et al. 1998). Thus, despite obvious structural similarities and their ability to bind related DNA sequences, NF- κ B subunits have distinct and non-overlapping functions (Perkins 2007).

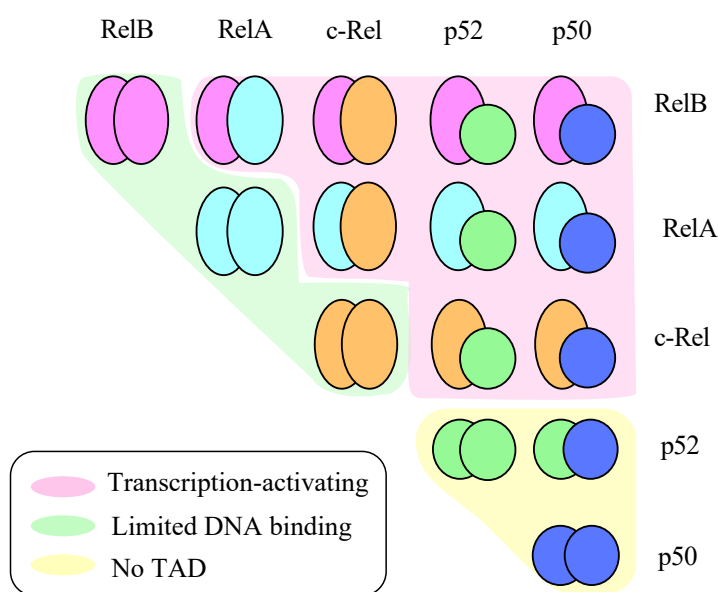


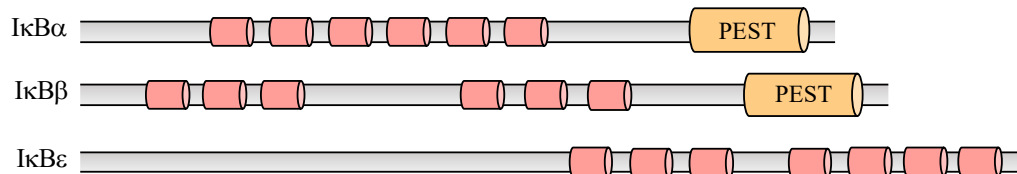
Figure 2. All 15 potential NF- κ B dimers. Hetero- or homodimers of NF- κ B including RelA (cyan), RelB (pink), c-Rel (orange), p52 (green) and p50 (blue). Adapted from Concetti and Wilson, (2018).

1.1.2 Regulation by I κ B protein family

Members of the NF- κ B protein family are tightly regulated by a group of proteins, known as the inhibitors of NF- κ B (I κ Bs). The I κ B family is comprised of I κ B α , I κ B β , I κ B ϵ , Bcl3, I κ B ζ , I κ BNS that all share a five or more 33-residue long central ankyrin (ANK) repeat domain (ARD) (Figure 3)(Oeckinghaus and Ghosh

2009). The ARD is a critical feature important for the interaction of the DNA-binding domains of NF- κ B proteins, and masking of their associated nuclear localisation signal (NLS). Although structurally similar, I κ B's are separated into two groups that show entirely different sub-cellular locations, activation kinetics and an unexpected functional diversity (Schuster et al. 2013).

a) **Classical I κ B**



b) **Nuclear I κ B**

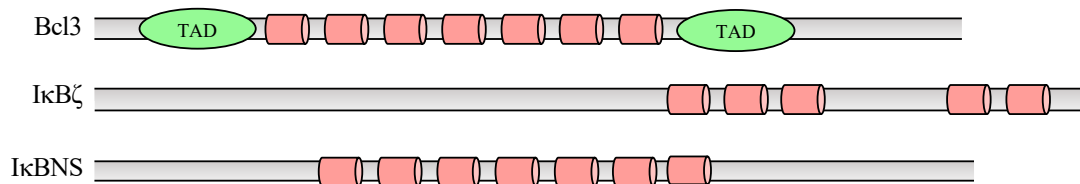


Figure 3. I κ B family members. a) Classical I κ Bs containing I κ B α , I κ B β , I κ B ϵ . b) Nuclear I κ Bs containing Bcl3, I κ B ζ , and I κ BNS. The transactivation domain (TAD) in green, ANK in red, and PEST (domain rich in proline, glutamate, serine and threonine) in orange. Adapted from Perkins (2007).

In most unstimulated, non-diseased mammalian cells, the NF- κ B protein family is regulated by the principal I κ B proteins; I κ B α , I κ B β and I κ B ϵ whose key role involves retaining the NF- κ B dimers in the cytosol, although this is largely dependent on the proteins in the I κ B:NF- κ B complex (Oeckinghaus and Ghosh 2009). The crystal structures of I κ B β -p65 homodimer revealed the NLS of p65 is masked, resulting in the complex sequestering in the cytoplasm and thus inhibiting binding of NF- κ B to DNA (Chen et al. 2003; Malek et al. 2003). Whereas, the crystal structure of I κ B α bound to p65/p50 heterodimers revealed that I κ B α only masks the nuclear localisation sequence of p65 (Bonizzi and Karin 2004; Hayden and Ghosh 2004). Thus, the accessible NLS of p50 coupled with the nuclear export sequence (NES) on both I κ B α and p65 results in constant shuffling of the complex

between the cytoplasm and nucleus in resting cells, although the steady-state localisation remains in the cytosol. Degradation of I κ B α unmasks the NLS of p65 and removes the contribution of I κ B NES resulting in a predominantly nuclear localisation of NF- κ B.

There are also atypical I κ B proteins namely Bcl3, I κ B ζ , I κ BNS whose interaction with NF- κ B proteins predominantly takes place in the nucleus of cells. Binding of atypical I κ B proteins with DNA-associated NF- κ B's can either enhance or diminish their transcriptional activity allowing them to modulate transcription either positively or negatively. This represents their most important and unique difference to that of the classical I κ B's.

1.2 B-cell Lymphoma 3

B-cell lymphoma 3 (Bcl3) is a proto-oncogene originally discovered through involvement in a repeated chromosomal translocation t(14;19)(q32;q13) present in a subset of human B-cell chronic lymphocytic leukaemia's (McKeithan and Ohno 1994). High expression of Bcl3 has since been reported with other cancers of the blood, lymph nodes, and bone marrow and is often accompanied with poor prognosis. It has also been detected in several solid tumours, including breast cancer, nasopharyngeal carcinoma, colorectal, cervical and ovarian cancers (Cogswell et al. 2000; Choi et al. 2010; Urban et al. 2016; Zhao et al. 2016; Zou et al. 2018; Ma et al. 2020). Thus, the oncogenic function of Bcl3 has been linked with cell survival, proliferation, and migration of cancer cells.

As previously mentioned, Bcl3 is an atypical member of the I κ B family of proteins that can be recruited to NF- κ B-responsive promoters, promote their localisation within the nucleus and modulate DNA binding activity. In cancer cells and transiently transfected cells, Bcl3 does mostly reside within the nucleus. Whereas, in many non-cancer cells (e.g. erythroblasts, keratinocytes, and hepatocytes) Bcl3 is primarily located within the cytoplasm and is translocated to the nucleus upon activation (Wang et al. 2017). Thus, depending on this and the subunit composition of NF- κ B complexes, the interaction with Bcl3 can result in either transcriptional activation or repression.

1.2.1 Structure of Bcl3

Human Bcl3 is a 447-residue protein containing a central ARD that is flanked by N- and C-terminal domains rich in proline and serine residues. The crystal structure of the Bcl3 ARD was solved by Fabrice Michel et al., (2001) which revealed an elongated structure (75 x 25 x 25 Å) of seven ANK repeats (Figure 4a). Each repeat consists of a β -hairpin loop (except for ANK1) followed by an ‘inner’ ($\alpha 1$) and ‘outer’ helix ($\alpha 2$). They stack together with a slight left-handed twist; where one complete turn of the superhelix would require a total of 32 repeats, with a 245 Å pitch and 35 Å radius.

A structure-based alignment shows that half of the residue positions within the ARD are highly conserved. For example, residues Ala (position 15) in helix $\alpha 1$ and Leu (position 27) in helix $\alpha 2$ are invariant across all seven repeats. There are also conserved hydrophobic residues (at positions 12, 14, 16, 24 and 28) that are necessary for packing interactions between each of the repeats, giving rise to the extended hydrophobic core.

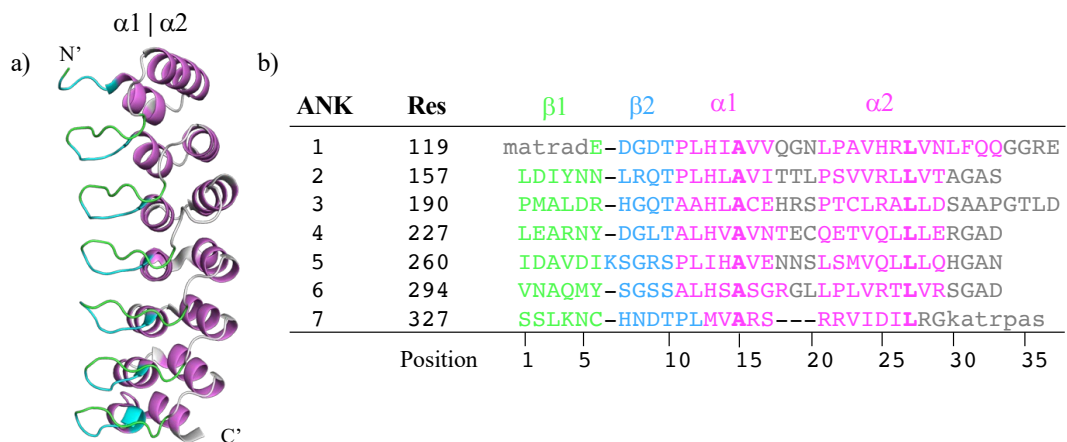


Figure 4. Structure of Bcl3. a) Cartoon representation of Bcl3 shown in ribbon representation (PDB ID: 1K1A). Helix $\alpha 1$ and $\alpha 2$ shown in pink, β -hairpin loop 1 and 2 shown in green and blue respectively. B) Sequence alignment of the 7 ANK repeats showing high sequence homology between each repeat.

1.2.2 Bcl3-NF- κ B interactions

Of the NF- κ B subunits, Bcl3 preferentially interacts with p50 and p52 homodimers to either activate or repress gene expression depending on the conditions (Wang et al. 2017). Binding of Bcl3 to p50 or p52 homodimer leads to the following possible results: (1) p50 and p52 lack TAD's and so are repressive, but Bcl3 containing a TAD possibly confers transcriptional activity, (2) Bcl3 could displace p50 or p52 homodimers from κ B sites allowing replacement with other transcriptionally active dimers (p50/p65), (3) Bcl3 inhibits NF- κ B-dependent transcription by stabilising p50 or p52 homodimers that are bound to κ B sites (Michel et al. 2001). The ability for Bcl3 to facilitate or inhibit NF- κ B dependent transcription is, in part dependent on posttranslational modification. For example, phosphorylated Bcl3 binds DNA that is bound to p50 and p52 homodimers and increases transcription (Wang et al. 2017). However, depending on the κ B sequence and the specific NF- κ B homodimer, phosphorylation has also been shown to inhibit p50/p52 binding to DNA (Bundy and McKeithan 1997). Bcl3 can protect p50 homodimers from ubiquitination and stabilise them when bound to κ B sites, repressing transcription (Carmody et al. 2007). As mentioned, in many non-cancer cells, Bcl3 resides in the cytoplasm and requires activation prior to nuclear transport (Zhang et al. 1998; Brasier et al. 2001). Given that, deubiquitination of Bcl3 can sequester p50/p52 heterodimers in the cytoplasm, preventing their nuclear transport and transcriptional regulation (Massoumi et al. 2006). As well as altering the biochemical properties of Bcl3, the posttranslational modifications are also suspected to influence the oligomeric state of Bcl3-NF- κ B complexes. Unlike I κ B α -p50/p65 (PDB: 1NFI) and I κ B β -p65/p65 (PDB: 1K3Z) complexes who exist solely in 1:2 ratio, Bcl3 and p50/p52 complexes are suggested to also form a higher order complex (2:2). This is true for studies carried out within both eukaryotic and bacterial cells (Wulczyn et al. 1992; Bundy and McKeithan 1997; Pang et al. 2004). Phosphorylation of Bcl3 is suggested to increase steric hindrance or electrostatic repulsion, and decrease stability of the 2:2 (Bundy and McKeithan 1997). Although the functional differences between each complex remains unclear, the mechanism of Bcl3 inhibition of NF- κ B protein binding to κ B sites is suggested to occur through a concentration and phosphorylation-dependent formation of a 2:2 Bcl3-p52 complex. With that, computational models have been determined and used to research both

stoichiometric forms (Manavalan et al. 2010; Soukupova et al. 2021). The structure of Bcl3-NF- κ B complexes will be explored further in chapter 3.

1.2.3 Role of Bcl3 in Breast Cancer & Metastasis

As mentioned previously, the deregulation of Bcl3 is associated with both haematopoietic and solid cancers (Urban et al. 2016; Zhao et al. 2016; Zou et al. 2018). Interestingly, the first solid cancer related with deregulation of Bcl3 was breast cancer, where an enhancement of NF- κ B dependent gene expression was observed (Cogswell *et al.* 2000). Since this discovery, overexpression of Bcl3 has evidenced its key role regulating cell death and proliferation.

A large proportion of breast cancers overexpress the HER2 receptor, leading to a clinically aggressive subtype of breast cancer. For these patients, the common cause of death associates with the development of metastatic disease. *In vivo* studies performed by Wakefield *et al.*, (2013) evidenced how the suppression of Bcl3 led to significant reductions in pulmonary metastasis in a HER2-positive mouse tumour models, without disruption to normal cell homeostasis (Wakefield *et al.* 2013). More recently, Turnham *et al.*, (2020) evidenced the mechanism of Bcl3 as an upstream regulator of the two principal forms of breast cancer cell motility, involving collective and single cell migration. Overexpression of mutant Bcl3, unable to bind p50, suppressed cell migration and metastatic tumour burden. It was therefore concluded that disruption of Bcl3-NF- κ B transcription complexes are sufficient to inhibit metastasis. Whilst there has been much interest in the use of NF- κ B inhibitors to suppress tumour progression, this has often had a profound effect on homeostasis of normal tissues, particularly those comprising the immune system. Thus, the significant role of Bcl3 in breast cancer and other cancers, offers a potential therapeutic target for metastatic disease.

1.3 Fluorescent proteins

1.3.1 Green Fluorescent Protein (GFP)

1.3.1.1 General overview

In the early 1960's, Osamu Shimomura discovered that the jellyfish *Aequorea victoria* fluoresced green when irradiated with ultraviolet light (Shimomura et al. 1962). The first protein responsible for its bioluminescence was isolated from *A. victoria* and later named aequorin. Interestingly, purified aequorin emitted bluish light, instead of the green fluorescence observed in the light-emitting tissues. This led to the identification of another protein accountable for the green fluorescence, which was named green fluorescent protein (GFP) (Hastings and Morin 1969). In 1971, the absorption of bluish light from aequorin was confirmed to cause the emission of green light from GFP (Morin and Hastings 1971). The exact mechanism was elucidated in 1974 suggesting Förster resonance energy transfer (FRET) mediates the energy transfer from aequorin to GFP (Morise et al. 1974). However, unlike aequorin which contains calcium-binding sites, GFP undergoes an intramolecular autocatalytic cyclisation that does not require an additional cofactor for its fluorescence (except the presence of oxygen). This permits fusion of GFP's DNA sequence with that of any protein whose expression can be readily monitored by sensitive fluorescence methods. This paved the way for experiments to assess its utility as an *in vivo* protein tag. GFP has since become one of the most commonly used tools in molecular biology, medicine, and cell biology. Its fluorescent properties have been exploited to study gene expression, live cell imaging, protein-protein interactions (PPIs), protein localisation, fusion proteins, and drug delivery.

1.3.1.2 Overall structure of GFP and the chromophore

GFP is a 27 kDa protein made up by a single polypeptide chain composed of 238 amino acids. It folds to a unique 11 β -sheet barrel-like structure with a diameter of ~ 24 Å and height of ~ 42 Å, with the barrel capped with helical and loop segments (Yang et al. 1996a; Remington 2011)(Figure 5a). A helix runs diagonally through the core of the structure. At its centre is the functional unit named the chromophore (p-hydroxybenzylideneimidazolinone) or CRO, formed from residues Ser65-Tyr66-Gly67 in the native protein, or now more commonly Thr65-Tyr66-

Gly67 in many engineered variants (Figure 5b). Since the discovery of GFP, fluorescent proteins from other predominantly marine organisms have been discovered that span into the red region of the UV-visible spectrum (Remington 2011; Chan et al. 2022). Notably, they all share a common β -barrel structure with the chromophore in the middle of the barrel. The Tyr-Gly dyad is strictly conserved in natural fluorescent proteins, while the residue equivalent to Ser65 can vary. In engineered variants, the residue equivalent to Tyr66 can be altered to other aromatic amino acids to generate variants with novel fluorescent properties; however Gly67 cannot be altered so is strictly necessary for chromophore formation.

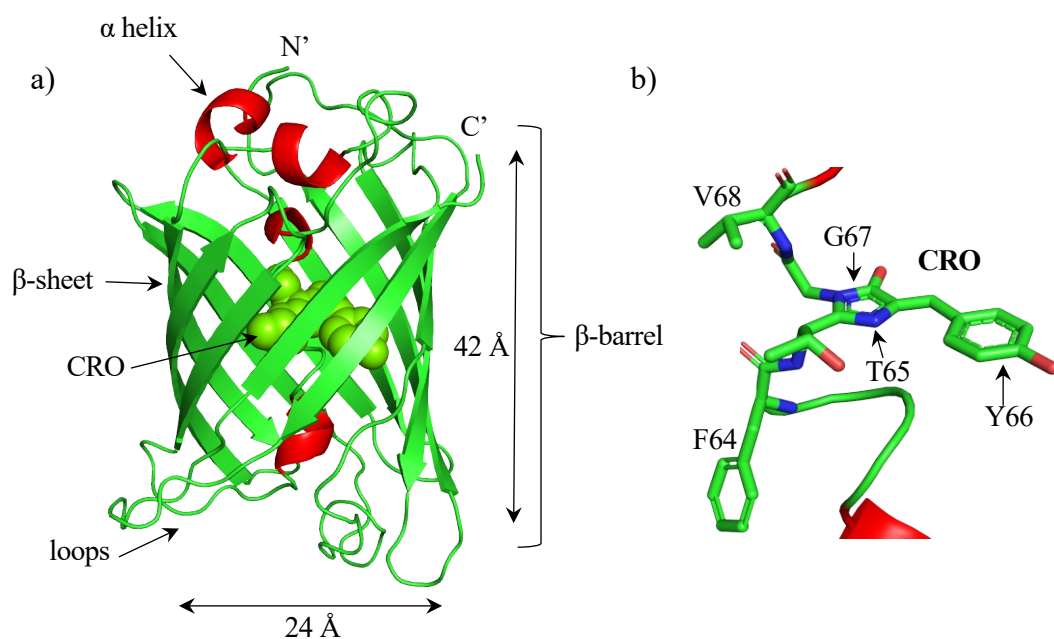


Figure 5. Structure of GFP. A) β -barrel of GFP comprised of 11 β -sheets, α -helices (red), loops and chromophore (spheres). PDB ID: 1EMA B) Stick representation of GFP chromophore comprised of T65, Y66 and G67.

The GFP chromophore is formed by the spontaneous covalent rearrangement of the tripeptide (Ser(or Thr)-Tyr-Gly). As the protein folds, the tight turn that is part of the central helix brings the amino group of Gly67 close to the carbonyl carbon of Ser65 (Branchini et al. 1998) (Figure 6a). The residues are then processed by a “cyclisation, dehydration and oxidation” pathway as proposed by Getzoff’s group (Barondeau et al. 2003; Barondeau et al. 2005; Wood et al. 2005; Auhim et al. 2021). First, during cyclisation, the backbone condenses allowing the nucleophilic attack of the amino group of Gly67 on the carbonyl group of Ser65 to form a

cyclopentyl tetrahedral intermediate (Figure 6b). The carbonyl carbon of Ser65 then undergoes dehydration followed by oxidation of the C_{α} - C_{β} bond in Tyr66 to form the mature chromophore (Figure 6c). Two potential protonation states are available for Tyr66: a neutral protonated state (termed CroA) and a charged phenolate (termed CroB) (Figure 6c). This is regulated through hydrogen bond contributions of the surrounding residues including residue 65 of the chromophore, H148, T203, S205, E222 and water molecules (Figure 7). Other residues that are vital for GFP fluorescence include V61, T62, Q69, Q94, R96, N146, R168 and Q183 (Tsien 1998).

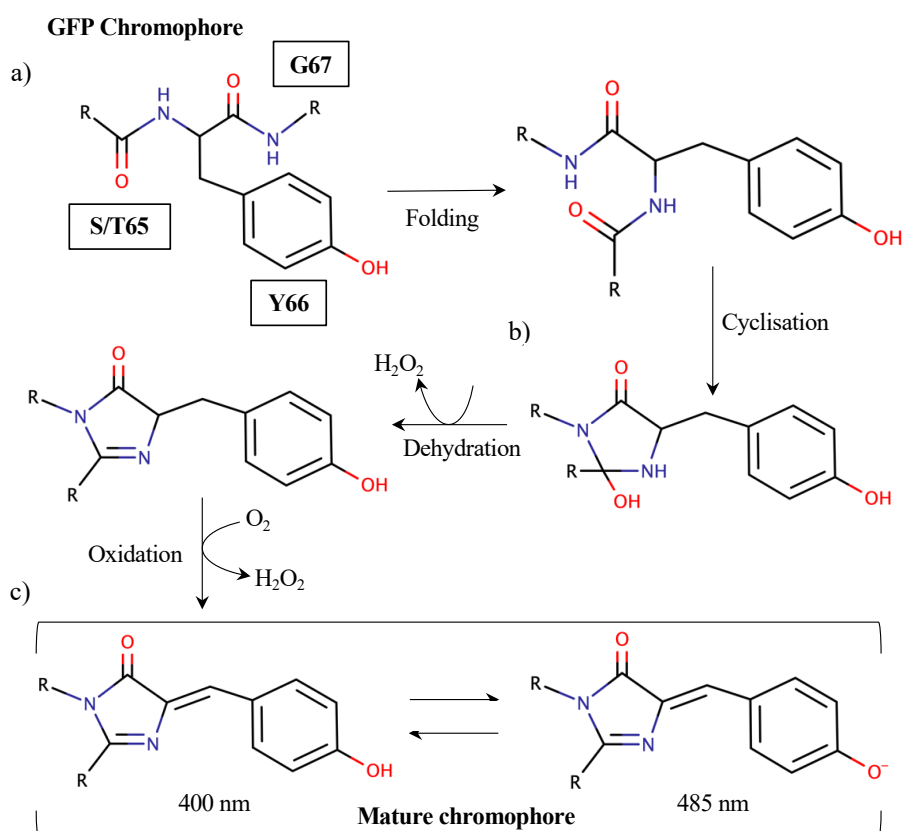


Figure 6. Proposed mechanism for chromophore maturation. Cyclisation, dehydration, and oxidation by Getzoff's group.

The protonation state of the tyrosyl (phenol) group of the GFP chromophore will dictate the absorbance properties and thus energy of electron excitation as part of the fluorescence process. The neutral chromophore (CroA) absorbs with a λ_{max} of 395 nm whereas the deprotonated anionic phenolate form (CroB) absorbs at 475 nm (Tsien 1998). Both ionisation forms exist in wild-type GFP giving rise to a bimodal absorbance peak, where a 6:1 ratio (CroA:CroB) is present. Comparison of wild-type

GFP and a few of its relatives revealed that regardless of the presence of either chromophore and the two widely separated excitation maxima, the chromophore emits green fluorescence (Remington 2006; Remington 2011). Although it was argued CroA excitation should instead emit in the blue region. This led to the discovery of excited anionic intermediate (CroI) involving a process commonly referred to as ‘excited state proton transfer’, or ESPT (Figure 7). When GFP undergoes CroA excitation, the proton in the hydroxyl group of Tyr66 is lost by shuttling through a ‘proton wire’, carried by a water molecule and Ser205, and finally accepted by Glu222. This results in the formation of CroI, before the charged phenolate chromophore ground state is restored. In GFP’s with S65T mutation, the hydrogen bonding network of Glu222 is changed, suppressing the negative charge on Glu222 and allowing the phenolate CroB form to dominate (Remington 2011). Residues H148 and T203 are then important for stabilising the charge on phenolate.

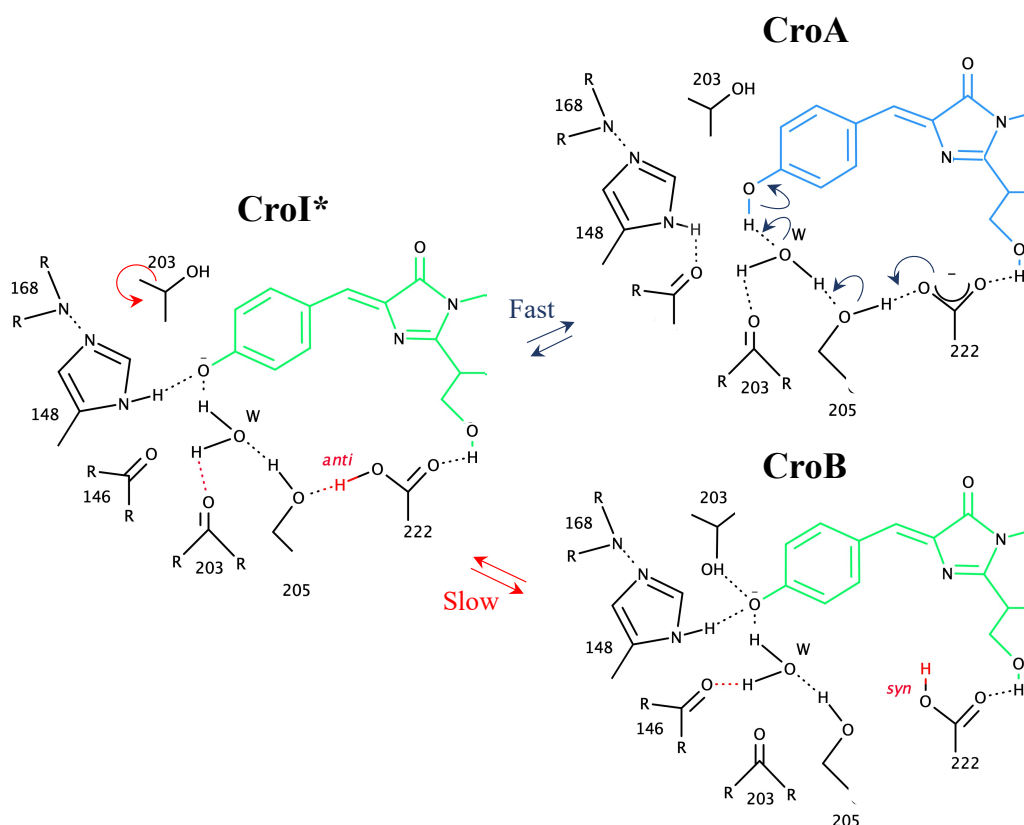


Figure 7. Suggested mechanism for the ESPT of GFP based on structural data and the spectroscopic work.

While the chromophore is essential for the core fluorescent properties of GFP, the protein as a whole plays an important role. As a preceding step in chromophore maturation, protein folding ensures the formation of the β -barrel and

the necessary protection for the chromophore from the solvent. Unfolded GFP is largely non-fluorescent due to quenching by the external environment. Also, as will be discussed in the next section, the residues surrounding the chromophore fine tune the spectral properties, allowing new and useful variants to be generated with more appropriate properties for downstream applications.

1.3.1.3 Engineering of GFP

Despite many early successes, wild type GFP was sub-optimal for most applications due to low fluorescence intensity (upon excitation at 485 nm), a significant lag in fluorescence following protein synthesis, complex photoisomerization of the chromophore and poor expression in various higher eukaryotes (Yang et al. 1996b). Thus, over the years scientists focused on re-engineering wild type GFP to shift the excitation and emission wavelengths, creating different colours and new applications. The first advantageous mutations F64L and S65T gave rise to EGFP, an enhanced GFP that displayed increased fluorescence and more efficient folding (Cubitt et al. 1995; Cormack et al. 1996; Yang et al. 1996b). Replacement of Phe64 with Leu results in the subtle rearrangement of hydrophobic core packing close to the chromophore, including the reduction in surface exposed hydrophobic residues (Arpino et al. 2012). The replacement of Ser65 with Thr impacts the local hydrogen bond network in the vicinity of the chromophore. Directed evolution approaches led to the generation of super-folder GFP (sfGFP) with six additional mutations (S30R, Y39N, N105T, Y145F, I171V and A206V) (Pédélecq et al. 2006). SfGFP features high resistance to chemical denaturation and displays rapid and robust folding. Fusion with sfGFP can improve protein solubility, expanding the range of proteins amenable to cell trafficking experiments through fluorescent protein tagging. This makes it an exemplar model for protein engineering efforts. For this reason, sfGFP was used throughout this study in order to identify key structural changes that can influence protein function.

The sfGFP chromophore exists predominately in the CroB form giving rise to a quantum yield of 0.75 and a molar extinction coefficient of $49,000 \text{ mM}^{-1}\text{cm}^{-1}$ (Figure 8)(as calculated by the Jones group (Reddington et al. 2012; Reddington et al. 2013; Reddington et al. 2015; Hartley et al. 2016). The deprotonated ionisation

state of the hydroxyl group of Tyr66 is mainly due to the hydrogen bond contributions from His148, Thr203, Ser205 and Glu222. Together these key residues (His148, Thr203, Ser205, and Glu222) and ordered water molecules control the complex hydrogen bond network surrounding the sfGFP chromophore and are critical for maintenance of the deprotonated ground state.

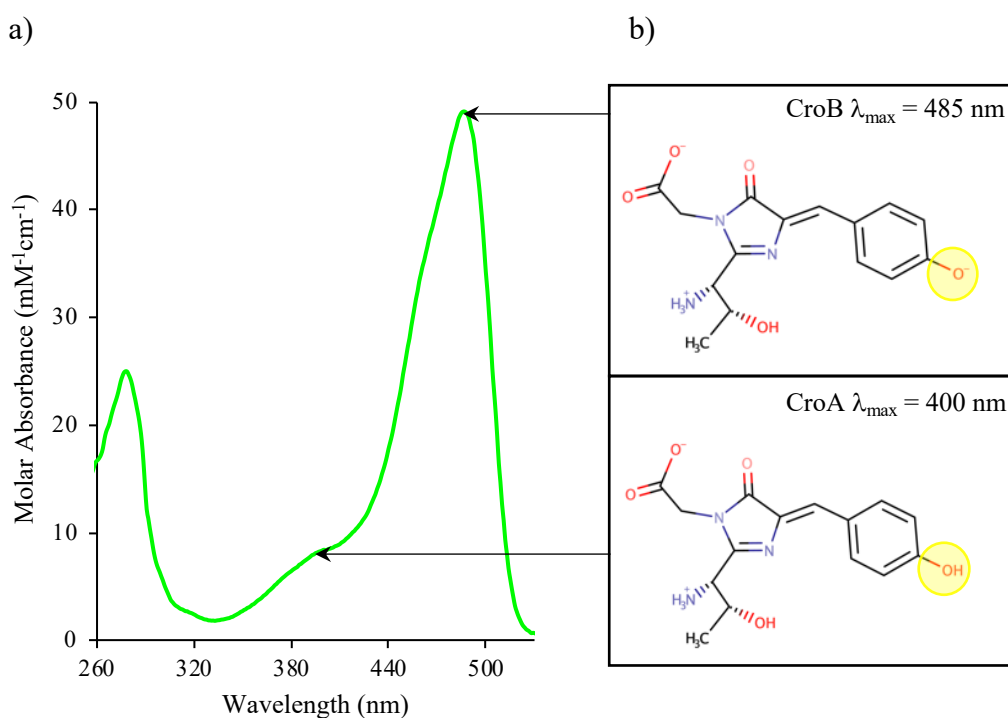


Figure 8. Characteristics of sfGFP. A) Molar absorbance of sfGFP. B) Two protonation states of sfGFP chromophore.

1.3.2 Yellow fluorescent protein (Venus)

Mutagenesis of GFP has produced a wide variety of mutants with shifted wavelengths of excitation and emission. Such mutants have become a valuable tool throughout areas of biological research including reducing the effect of cellular autofluorescence at short wavelengths, simultaneous multicolour reporting of the activity of two or more cellular processes, and for exploitation of FRET to monitor protein-protein interactions (Ormo 1996). Such engineering efforts can take two main routes: altering the composition of the chromophore itself or residues surrounding the chromophore (Chan et al. 2022). For example, the generation of cyan and blue fluorescent proteins starts with changing the tyrosine in the

chromophore to tryptophan or histidine, respectively (Mena et al. 2006; Goedhart et al. 2012).

One particularly useful colour variant are the yellow fluorescent proteins (YFPs) with red shifted excitation and emissions into the yellow region. The key mutation for generating YFPs is the mutation of residue 203 to an aromatic residue, most commonly tyrosine. The T203Y mutation results in pi-stacking between residue 203 and the tyrosyl group of the chromophore. As a result, the energy required to excite an electron decreases causing the red shift in the spectral properties. Additional mutations have been introduced into the basic YFP to optimise its spectral properties, folding and stability. One such variant is termed enhanced YFP (EYFP) that displays excitation and emission peaks at 513 and 527 nm, sufficiently different from those of previous GFP descendants. In addition to the T203Y mutation, EYFP also contains S65G (promotes CroB), V68L, S72A (Cormack et al. 1996; Miyawaki and Tsien 2000). However, use of EYFP was largely restricted due to its high sensitivity to acidic environments and chloride ions. Thus, further mutations were incorporated to generate super-EYFP, or SEYFP (EYFP – F64L, M153T, V163A, S175G) and eventually SEYFP – F64L, with the improved chromophore mutation (F64L). The final version (SEYFP-F64L) had reduced environmental sensitivity, and improved protein folding and chromophore maturation at 37 °C (Nagai et al. 2002). This protein is now referred to as ‘Venus’ for the comparability of its bright yellow appearance against the brightest object in the night sky (except for the moon) (Nagai et al. 2002).

Structural alignment of Venus with sfGFP and GFP^{WT} highlights their structural similarity (Figure 9a)(Rekas *et al.*, 2002). The β -barrels overlay almost perfectly, except for termini ends and loops regions as would be expected given their inherent flexibility. The residues surrounding the chromophore responsible for forming critical interactions are near identical to those in GFP^{WT} and sfGFP (V61, T62, Q69, Q94, R96, N146, H148, R168, Q183, S205 and E222). However, in Venus there is a T65G and T203Y mutation, where a 3.8 Å centroid-centroid distance is observed with the phenol ring of Y66, within the optimum range for π - π stacking interactions (Figure 9b).

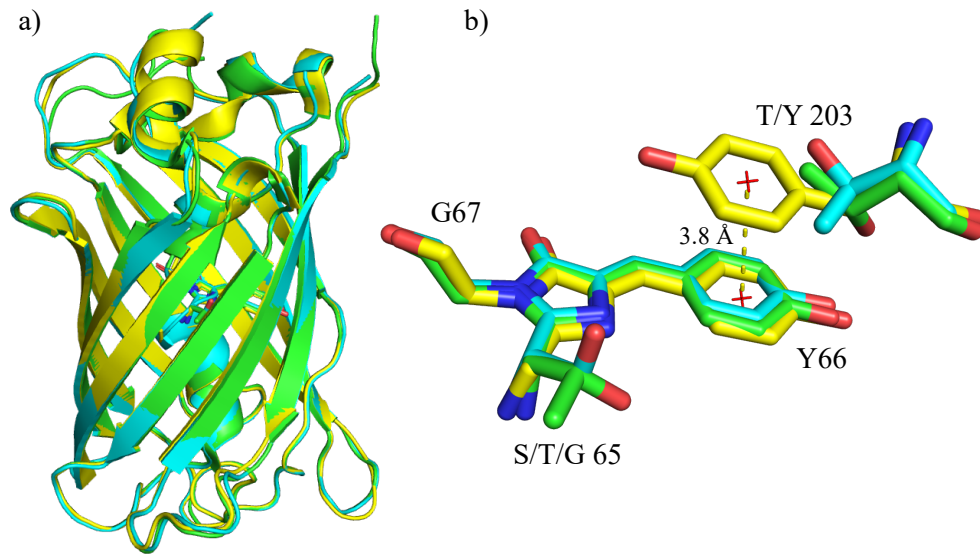


Figure 9. Structural comparison of GFP^{WT} (cyan), sfGFP (green) and Venus (Yellow). A) Cartoon representation. B) Stick representation of chromophores and residue 203. Residue 65 is a serine, threonine or glycine in GFP^{WT}, sfGFP and Venus respectively. Residue 203 is threonine in GFP^{WT} and sfGFP, and tyrosine in Venus. PDB ID: 1GFL (GFP^{WT}), 2B3P (sfGFP), 1MYW (Venus).

1.4 Protein oligomerisation

Protein oligomerisation refers to the formation of a single multimeric complex through defined intermolecular interactions. As the most abundant structural state, protein dimers and higher-order oligomers are more prevalent in nature than monomers, as evidenced in the Protein Data Bank (Gwyther et al. 2019). The ability of proteins to self-associate allows the development of new structural and functional properties that cannot be achieved by those in a simple monomeric system. Recent structural and biophysical studies have revealed the necessity of protein oligomers and their key role in the regulation of proteins such as ion channels, enzymes, receptors and transcription factors (Marianayagam et al. 2004). Natural protein oligomers are formed by a variety of mechanisms which include the formation of a mutually compatible non-covalent interface comprised of hydrophobic interactions, hydrogen bonds and electrostatic interactions, with covalent links such as intermolecular disulphide bonds making the oligomerisation event more permanent.

Given the abundance of oligomeric proteins, the key question arises as to why protein oligomerisation is so common? Through combined efforts, Goodsell and Olsen concluded a few of the key biological features responsible for the evolutionary success of protein oligomers (Goodsell and Olson 2000) For example, multi-subunits show a greater resistance to denaturation and degradation through reduced solvent to surface area exposure. There is also the advantage of functional cooperativity where RNA polymerase has an increased turnover rate compared to the subunits acting independently. Multi-subunit association can also reduce the risk of misfolding by avoiding the complex folding energy landscapes of longer polypeptide chains. Lastly, multiple shorter sequences conserve the cells energy through more efficient replication, transcription, and translation.

1.4.1 Engineering protein oligomers

Inspired by nature's selection pressure, the design and production of artificial self-assembling protein complexes is currently of great interest to protein engineering. A variety of techniques are proven to achieve this; including split proteins/domains (Shekhawat and Ghosh 2011), fusion proteins/domains (Kobayashi and Arai 2017), metal ion bridging, co-factor bridging, helix-helix interactions, and disulphide bridging (Gwyther et al. 2019). Through these methodologies, new structure and function can be designed beyond those achieved in nature to try and solve technological, medical and scientific problems. For example, directed evolution approaches employed a split GFP system to improve the thermodynamic stability of a variant of the B1 domain of protein G (Lindman et al. 2010), and Wehr *et al.*, (2006) showed how split TEV can be used to monitor and regulate protein-protein interactions at the membrane, and in the cytosol of living mammalian cells. Meanwhile Scott *et al.*, (2021) designed peptides that can self-assemble into water soluble α -helical barrels and function as well-defined ion channels.

Regardless of technique, all engineering methods are based upon structural information of naturally forming oligomers in the Protein Data Bank. Thus, generation of a protein interface for self-assembling higher-order oligomers can require an intensive design process and significant engineering. One approach to mediate this complexity is disulphide crosslinking via engineered cysteine residues.

Inspired by their fundamental role in nature (protein folding and stability), disulphide bonds can also be used to form novel inter-protein cross-links that provide more stability and selectivity than non-covalent interactions. A modified cell-free protein synthesis system was developed in *E.coli* that stimulated disulphide bond formation and was used to produce self-assembling virus-like particle for vaccination/drug delivery (Bundy and Swartz 2011). Recently Yao *et al.*, (2022) combined computational *de novo* designs and directed disulphide pairing to synthesise a set of peptide heterodimers with mutual orthogonality. Compared to non-covalent heterodimers, interchain disulphide bonded dimers possessed increased structural stability, displayed resistance to chemical and thermal denaturation. The potential applications for such dimers include general chemical tools or building blocks for protein labelling (e.g. fluorescent probes), and construction of proteins assemblies. Thus, this unexplored dimeric structure space could provide solutions for protein functionalisation, drug delivery, vaccine design and the development of dimeric peptide therapeutics. However, there are multiple limitations to using disulphide bridges for protein engineering and its applications in biotechnology and therapeutics including its reversibility and redox sensitivity. The cysteine side chain is also relatively short so depending on the structure of the protein and position of the mutated residue, a mutually compatible interface is often still required. Another concern includes the increased risk of misfolded proteins where mutated cysteines form incorrect disulphide bonds with native cysteine residues. Although various computational approaches can now be used to predict the likelihood of forming engineered disulphide bonds to ~ 70% accuracy (Gao *et al.* 2020). Thus, improving the selectivity of cysteine mutations and subsequent disulphide bonds will increase their use for both intra-subunit and inter-subunit interactions.

Worthy *et al.*, (2019) used sfGFP to demonstrate the use of non-canonical amino acids (ncAA) for the generation of protein dimers by structure-guided biorthogonal chemistry, using genetically encoded strain promoted azide-alkyne cycloaddition chemistry (SPAAC) (Figure 10). The benefit to such an approach is the simple design process, defined linkage position, and that it can be easily combined with other protein interfacing approaches. NcAA's like those used by Worthy *et al.*, (2019) (azF and SCO) are ideal given their relatively long side chains as this reduces steric clashes, while maintaining structural intimacy to promote and

stabilise favourable non-covalent interactions between the protein interface. The engineered dimers demonstrated functional synergy by switching “on” once assembled and displayed enhanced spectral properties (Figure 10). This occurred as a result of forming interunit communication and networking beyond the direct interface region. This will be discussed in more detail throughout chapters 4 and 5, as this work forms the basis for this research project.

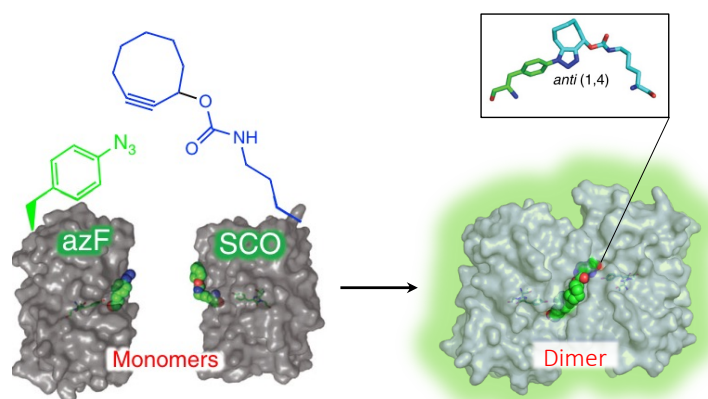


Figure 10. Bioorthogonal chemistry driven protein dimerisation. SPAAC reaction between two genetically encoded ncAA’s (azF, green and SCO, blue). The resulting dimer containing a triazole link had enhanced spectral properties in comparison to the monomers.

1.5 Aims and Objectives

The thesis can be broken down into two main sections where the first focus will be on understanding the PPIs between Bcl3 and p50, including how developed drugs affect these interactions. The second is to address one of the major challenges in monitoring PPIs; that of monitoring homo-oligomerisation events. Processes such as FRET help researchers understand and monitor PPIs but is currently largely restricted to hetero-oligomerisation due to the need for two different fluorescent proteins. As outlined in section 1.1.1, the composition of the core NF- κ B complex can vary between homodimers and heterodimers with the nature of final species dictating cellular events. Thus, there is a need to be able to monitor homodimerisation processes as well as heterodimerisation. The second aspect of the thesis will therefore focus on developing a fluorescent protein-based approach for monitoring homodimerisation that can also potentially complement existing heterodimerisation approaches.

The first aim of this project is to study the biochemical characteristics of Bcl3-NF- κ B complexes. To achieve this, genetic constructs for both Bcl3 and NF- κ B needed to be designed and generated for recombinant protein expression. Proteins will then be expressed and purified using a variety of techniques. Various downstream analytical processes will be used to detect protein-protein interactions (PPI) and protein-drug interactions including: isothermal calorimetry (ITC), size exclusion chromatography (SEC), dynamic light scattering (DLS), and fluorescence. To further understand the molecular basis of the interaction and to facilitate advanced drug design, atomic level information of Bcl3 complexes is required. Whilst structures of Bcl3 and p50/p52 have been determined separately (Ghosh et al. 1995; Cramer et al. 1997; Michel et al. 2001), we currently lack information regarding their structures when complexed. To do this, x-ray crystallography will be used where protein crystals will be taken to the national synchrotron source (Diamond Light Source) to generate diffraction data for structure determination.

The second aim of this project involves the protein engineering of sfGFP, particularly focusing on controlling the ionic state of the chromophore, giving rise to the two separate absorbance peaks (CroA = 400 nm and CroB = 485 nm). Computational simulations will be used to test alternative residues at specific positions of sfGFP. Predictions will then be made and used to select particular residues for replacement. After expression and purification, new sfGFP variants will be characterised and compared to those of wild type. Residues will also be selected based on their potential to facilitate protein oligomerisation. The ability to regulate sfGFP fluorescence through association of two identical monomeric units opens new avenues for its use to monitor homodimerisation events. Worthy *et al.*, (2019) demonstrated the use of ncAA's to generate synergistic dimers that switched protonation state on and thus signal output ratio on dimerisation. This will be the basis of the design and work carried throughout the remaining chapters (discussed further in Chapter 4 and 5). As part of this process, p50 will used as a model for application of sfGFP-based homodimerisation.

2 Materials and methods

The materials and methods carried out throughout this work are described below. Any deviations from the methods will be mentioned in the main text.

2.1 Materials

2.1.1 Chemicals

Antibiotics were prepared as 1000x stock solutions using ultra-pure water and filter sterilised through a 0.22 μ M syringe filter unit (Sartorius). Final working concentrations of each antibiotic were as follows; 50 μ g/mL Carbenicillin (Melford, Suffolk, UK), 50 μ g/mL Chloramphenicol (Melford) and 35 μ g/mL Ampicillin (Melford).

Novel small molecule inhibitor compound JS6 (MW – 371.46) was designed by Dr Richard Clarkson's group (School of Biosciences, Cardiff University) to disrupt interactions of Bcl3 with p50 and p52 homodimers. It was synthesised and supplied by The School of Pharmacy, Cardiff University in powder form for storage at room temperature. The drug was dissolved in 100 % DMSO before use and diluted with 20 mM HEPES, 300 mM NaCl buffer.

For the redox dependent dimerisation studies, copper sulphate (Fisher Scientific, Leicestershire, UK), copper chloride (Sigma-Aldrich) and Dithiothreitol (DTT) (Melford) were prepared as 100 mM stock solutions by dissolving in ultra-pure water. Unless otherwise mentioned, the final working concentration of CuSO₄ was 1 mM dissolved in 50 mM Tris-HCl, the same buffer as the protein.

2.1.2 Bacterial Strains

Propagation of DNA was performed by transforming plasmids into *E. coli* cells BL21 (DE3) (New England Biolabs (NEB), Hertfordshire, UK) and Top 10TM (Invitrogen, Paisley, UK). The same cells were also used for protein expression. The genotypes of the competent cells are shown in Table 1.

Table 1. List of cell strains and Genotypes

<i>E.coli</i> cell strain	Genotype
BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] Δ hsdS λ DE3 = λ sBamHIo Δ EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δ in5
Top 10™	F- mcrA Δ (mrr-hsdRMS-mcrBC) ϕ 80lacZ Δ M15 Δ lacX74 recA1 araD139 Δ (araleu)7697 galU galK rpsL (StrR) endA1 nupG

2.1.3 Bacterial Growth Media

Solid media lysogeny broth (LB) agar (Melford), liquid media 2xYT (Melford) and super optimal broth (SOB) (Melford) were prepared according to manufacturer's instruction using ultra-pure water. SOB was supplemented with 4% (w/v) glucose (Melford) to produce super optimal condition (SOC) broth; used for cell recovery following DNA transformation. All medium were sterilised by autoclaving at 121°C before use. Any components added to medium were prepared using ultra-pure water and filter sterilised using a 0.22 μ M syringe filter unit.

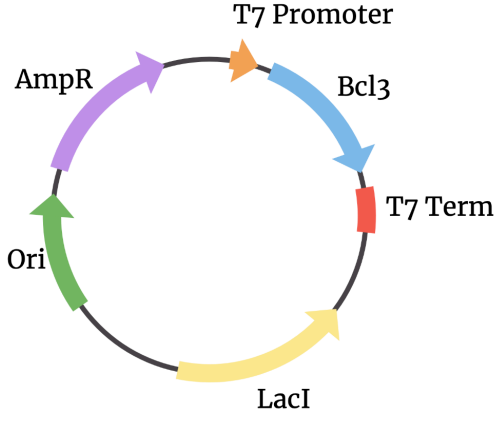
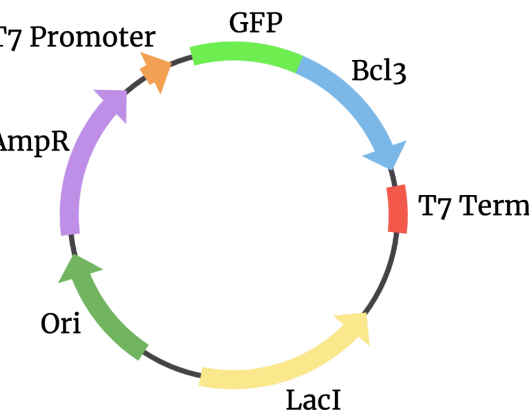
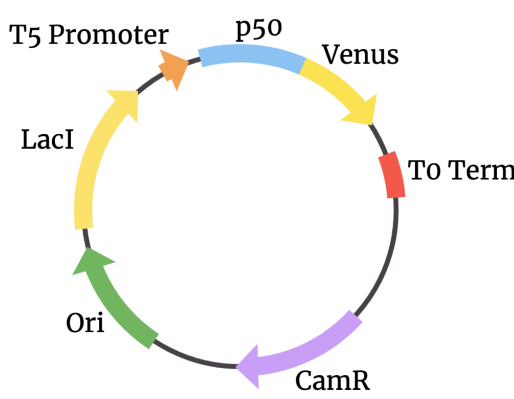
The following lists the auto-induction media used to produce proteins encoded in the pBAD plasmid: 1% (w/v) tryptone; 0.5 % (w/v) yeast extract; 0.5% (v/v) glycerol; 0.05 % (w/v) glucose; 0.2% lactose; 25 mM Na₂HPO₄; 25 mM KH₂PO₄; 50 mM NH₄Cl; 5 mM NaSO₄; 2 mM MgSO₄; and 0.05 % (w/v) L-arabinose. All chemicals were supplied by Melford.

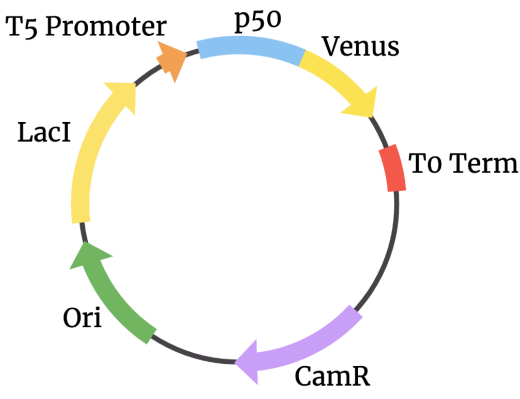
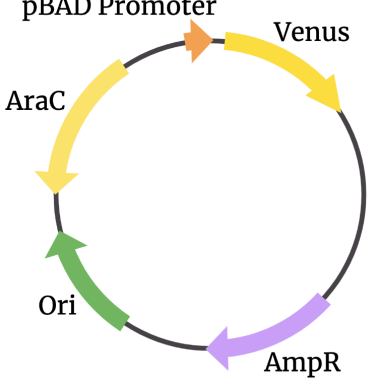
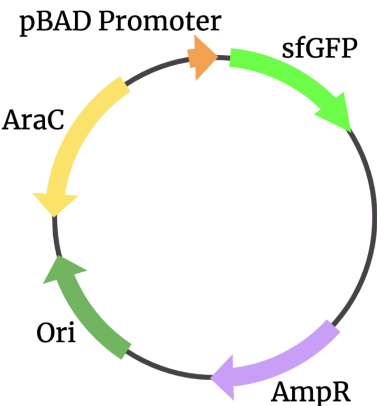
2.2 Molecular Biology

2.2.1 Vectors for protein expression

Throughout this work numerous vectors were used for recombinant protein expression as shown in Table 2.

Table 2. Vectors used for recombinant protein expression in *E.coli*.

Vectors	Plasmid map	Source
pET11a - Bcl3		<p>Donated from Dr Harley Worthy, Cardiff School of Biosciences, Cardiff University, Cardiff, UK</p>
pET11a-sfGFPBcl3		<p>Donated from Dr Harley Worthy, Cardiff School of Biosciences, Cardiff University, Cardiff, UK</p>
pCA24N-p50		<p>Donated from Dr Ismail Taban, Cardiff School of Biosciences, Cardiff University, Cardiff, UK</p>

<p>pCA24N- p50Venus</p>		<p>Donated from Dr Ismail Taban, Cardiff School of Biosciences, Cardiff University, Cardiff, UK</p>
<p>pBAD-Venus</p>		<p>Donated from DDJ group, Cardiff School of</p>
<p>pBAD-sfGFP</p>		<p>Biosciences, Cardiff University, Cardiff, UK</p>

2.2.2 DNA Oligonucleotides

Various sequence manipulations were performed including mutagenesis, restriction digests, ligations, and sequence alignments. Serial Cloner (http://serialbasics.free.fr/Serial_Cloner.html) and Amplifx were used for *in silico* design of oligonucleotides. DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). The list of oligonucleotides used throughout this study

are mentioned in Table 3. Annealing temperatures (T_m's) were determined using New England Biolabs (NEB) online T_m calculator (<https://tmcalculator.neb.com/#!/main>).

Table 3. Primer sequences and T_m used in this study. Mutations highlighted in bold.

Primer name	Sequence (5' → 3')	T _m (°C)
GFP_H148C mutation		
F_sfGFP_148C	GCGGTAATATACACATT GC AGCTGTTG	68
R_sfGFP_148C	CGATAAACAGAAAAATGGCATCAAAGCG	
GFP_H148A mutation		
F_sfGFP_148A	GGAATATAATTTCAACAG CG CTAATG	63
R_sfGFP_148A	AGTTTATGACCCAGAATGTTGCCATCT	
GFP_H148V mutation		
F_sfGFP_148V	TGTATATTACCGCCGATAAACAGA	64
R_sfGFP_148V	CATT GAC GCTGTTGAAATTATATTCCAG	
GFP_H148S mutation		
F_sfGFP_148S	CTGGAATATAATTTCAACAG CAG TAATG	63
R_sfGFP_148S	TTTATGACCCAGAATGTTGCCATC	
GFP_V206F mutation		
F_sfGFP_206F	GAGCAAAGATCCGAATGAAAA AC GTG	68
R_sfGFP_206F	AG AA AGCTCTGGGTGCTCAGATA	
Insert GFP into pCA24N vector		
F_sfGFP_SacI	ccgtGAGCTCgcatATGATGGTTAGCAAAGGTG	72
R_sfGFP_HindIII	ttatcAAGCTTttaAATGCCCGCGGCGGTAAC	
sfGFP_seq		
F_sfGFP_seq	GTTGTGCCGATTCTGGTGGAACT	69

R_sfGFP_seq	ACAAATTCCAGCAGCACCATATGATCA	
T5_seq		
F_T5_seq	CGTCTTCACCTCGAGAAATCATAAA	65
R_T5_seq	AACCGAGCGTTCTGAACAAAT	
Bcl3_seq		
F_T7_seq	CCCATTTGCTGTCCACCAGTCATGCTA	65
R_T7_seq	TTGTGAGCGGATAACAATTCCC	

2.2.3 Amplification of DNA by PCR

Polymerase chain reaction (PCR) was used to amplify genes and whole plasmid for site directed mutagenesis (SDM). Reaction mixtures and PCR conditions for whole plasmid PCR using Q5 DNA polymerase (NEB) are described in Table 4. All PCR reaction mixtures were set up in sterile PCR tubes (Starlab, UK) and performed in Techne thermocycler.

Table 4. Composition of PCR reaction mixture and PCR conditions. *T_m for each reaction is listed in Table 3.

Materials	Volume (μL) (Total – 50 μL)
Q5 reaction buffer	10
dNTP's (10 mM)	1
Forward primer (10 μM/μL)	2.5
Reverse primer (10 μM/μL)	2.5
Template plasmid (0.5 ng/μL)	1
Q5 high GC enhancer	2
Q5 DNA polymerase	0.5
Ultra-pure water	30.5

PCR conditions		
Step	Temperature (°C)	Time (s)
Initial denaturation	98	30
30 cycles	98	10
	*52-72	30
	72	30 /kb
Final extension	72	120
Hold	4	

GoTaq® Green Master Mix (Promega, USA) was used for colony screening to identify clones containing the gene of interest. Colonies were selected and transferred into a 50 µL aliquot of ultra-pure water before boiling at 98 °C for 10 mins. The mixture was subjected to centrifugation at 13,000 RPM for 3 mins. The supernatant was used to set up the PCR reaction (described in Table 5).

Table 5. Composition of colony PCR reaction mixture and PCR conditions. *T_m for each reaction is listed in Table 3.

Materials	Volume (µL) (total – 25 µL)	
GoTaq® Green Master Mix	12.5	
Forward primer (10 µM/µL)	1	
Reverse primer (10 µM/µL)	1	
Bacterial Supernatant	2	
Ultra-pure water	8.5	
PCR conditions		
Step	Temperature (°C)	Time (s)
Initial denaturation	98	120

30 cycles	98	10
	*52-72	30
	72	60 /kb
Final extension	72	300
Hold	4	

2.2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse and separate DNA samples. A 1 % (w/v) agarose gel was prepared by dissolving 1 g of agarose powder (Melford) in 100 mL of 1x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.8) then supplemented with 4 µL ethidium bromide (0.5 µg/mL) (Sigma-Aldrich). Sample loading dye (NEB) was mixed with 10 µL of DNA and ran against 1 µL marker ladder (1 kB DNA ladder, NEB). Electrophoresis was performed at 140 V for 50 min. DNA bands were visualised with UV light in a GelDoc-It UV-Transilluminator (Ultra-Violet Products Ltd, Cambridge, UK).

2.2.5 DNA purification and isolation

Principle

DNA purification uses a bind-wash-elute procedure where DNA is absorbed onto a silica membrane, impurities are efficiently washed away, and purified DNA is eluted with ultra-pure water.

2.2.5.1 Extraction and purification of plasmid DNA from bacterial cells

Plasmid DNA was isolated and purified from bacterial cells using QIAquick Miniprep Kit (Qiagen, West Sussex, UK). A 10 mL overnight culture (incubated at 37 °C, 200 RPM) was subjected to centrifugation at 5000 xg for 20 min. Cell pellets were resuspended and lysed using solutions supplied with the kit according to manufacturer's instructions.

2.2.5.2 Purification of PCR product

Purification of PCR product was performed according to the QIAGEN protocol using the QIAquick PCR purification kit (Qiagen).

2.2.5.3 Extraction and purification of DNA from agarose gel

An agarose gel can be used to separate DNA strands following digestion with restriction enzymes. Specific bands are isolated from the agarose gel using the QIAquick gel extraction kit (Qiagen). A sterile scalpel is used to cut the desired band whilst visualising with a UV-transilluminator. The fragment of gel was then solubilised at 50 °C in a specialised buffer to release the DNA for purification.

2.2.5.4 DNA purification from enzymatic reactions

Following enzymatic reactions, DNA was purified and isolated using the QIAquick MinElute Kit (Qiagen).

2.2.6 DNA Quantification

DNA concentration was determined using the NanoDrop® ND1000 spectrophotometer (Thermo Fisher Scientific). An appropriate solution was used as a blank before recording the absorbance at 260 nm. The DNA concentration can then be calculated using the Beer-Lambert law.

2.2.7 Restriction endonuclease digestion

DNA fragments generated from PCR will have blunt ends. Restriction enzymes were used to produce cohesive ends. The vector also requires digestion with the same restriction enzymes to generate linear DNA with complementary sticky ends. SacI-HF (NEB) and HindIII-HF (NEB) restriction enzymes were used to generate the p50-GFP construct. Reaction mixtures and conditions are depicted in Table 6. The reactions were stopped by running the DNA on an agarose gel for extraction and purification as mentioned in section 2.2.5.3.

Table 6. Reaction mixture and conditions for restriction endonuclease digestion.

Purified PCR product digestion		Vector digestion	
Materials	Volume (μL)	Materials	Volume (μL)
PCR product	25	Vector	25
PCR H ₂ O	20	PCR H ₂ O	20
Cutsmart®	5	CutSmart®	5
SacI-HF enzyme	1	SacI-HF enzyme	1
Incubate for 60 min at 37 °C		Incubate for 60 min at 37 °C	
Add 1 μL HindIII-HF and incubate for 60 min at 37 °C		Add 1 μL HindIII-HF and incubate for 60 min at 37 °C	

2.2.8 Phosphorylation & ligation DNA

For SDM, whole plasmid PCR is performed according to section 2.2.3. DNA fragments then need to be phosphorylated at the 5' end before ligation and recircularization. The reaction was set up by mixing 12 μL purified PCR product, 12 μL quick ligase buffer (NEB), 1 μL T4 polynucleotide kinase (PNK) (10 units/μL) (NEB) and incubated at 37 °C for 30 mins. The phosphorylated DNA was then ligated by addition of 1 μL quick ligase (NEB), incubating for 8 min and stopping the reaction by DNA purification as mentioned in section 2.2.5.2. For the insertion of sfGFP into p50 vector, a 3:1 molar ratio of insert:vector was used according to Equation 1 below.

Equation 1.

$$\text{Insert mass (ng)} = \frac{\text{Vector mass (ng)} \times \text{Size of insert (kb)}}{\text{Size of vector (kb)}} \times \text{Desirable molar ratio}$$

2.2.9 Bacterial Transformation

2.2.9.1 Preparation of electrocompetent cells

All materials and solutions used were sterilised by autoclaving and prechilled to 4 °C before use. The original cells mentioned in Table 1 were used to prepare

competent cells. Cells were transferred onto an agar plate and grown at 37 °C overnight. A colony was selected and grown in 2xYT medium incubating at 37 °C with shaking overnight. The overnight culture was used to inoculate a 500 mL 2xYT solution with incubation at 37 °C with shaking until an OD₆₀₀ between 0.4 – 0.6 was reached. Cells were harvested by centrifugation at 5000 xg at 4 °C (Beckman Coulter rotor JLA-16.25) for 20 min. Cells were washed twice with 50% glycerol with centrifugation at 5000 xg at 4 °C (Beckman Coulter rotor JLA-16.25) for 20 min. Cells were finally resuspended in 50% glycerol and divided into 50 µL aliquots in sterile 0.5 mL microfuge tubes. Tubes were flash frozen with liquid nitrogen and stored at -80 °C before use.

2.2.9.2 Electroporation

Electroporation was used to transform pET-11a and pCA24N plasmids into *E. coli* BL21 (DE3) cells. The pBAD plasmid was transformed into Top 10™ cells for protein expression. The competent cells (50 µL) were mixed with 1-5 µL of plasmid DNA (20-100 ng) and transferred into an electroporation cuvette to apply a short pulse of 2500 V (CelljetcT, Flowgen). Cells were mixed with 450 µL of SOC and placed at 37 °C in a shaking incubator (20 rpm) for 60 minutes. Cells were spread onto LB plates containing the appropriate antibiotic conditions for selection and incubated overnight at 37 °C.

2.2.9.3 Sequencing

All constructed vectors were sequenced to verify mutations and gene inserts using the Eurofins Genomics DNA sequencing service (<https://www.eurofins.com/>).

2.3 Protein production & purification

2.3.1 Protein production and cell lysis

Plasmids were transformed into *E. coli* BL21 (DE3) (NEB) and Top 10™ (Invitrogen, Paisley, UK) cells for protein expression as described in section 2.2.9.2. A single colony was taken from the transformation and used to inoculate a 10 mL 2xYT starter culture supplemented with ampicillin (50 µg/mL), carbenicillin (50 µg/mL) or chloramphenicol (35 µg/mL) for pBAD, pET-11a and pCA24N vectors

respectively. The starter cultures were placed in a shaking incubator overnight at 37 °C, 200 RPM.

For Bcl3 and p50 variants, 1L of 2xYT medium was prepared and an initial 1 mL aliquot was taken as a blank, recording the optical density (OD) using the Cary spectrophotometer in a 1 cm pathlength cuvette (Hellma, Müllheim, Germany). The medium supplemented with appropriate antibiotic selection was inoculated with the starter culture (100 µL per 100 mL of expression culture) and incubated at 37 °C with shaking (200 RPM). At an OD₆₀₀ of approximately 0.8, IPTG is added (final concentration of 0.4 mM) and incubated overnight with shaking (200 RPM) at 20 or 28 °C for Bcl3 or p50 variants respectively. For Bcl3-p50 complex expression, medium was incubated at 27 °C overnight.

For protein expression of the pBAD vector, an overnight culture was used to inoculate a 1L culture supplemented with antibiotic and autoinduction medium as described in section 2.1.3. The culture was inoculated overnight at 37 °C with shaking at 200 RPM.

At an OD₆₀₀ of ~ 2.0, cells were pelleted by centrifugation in a Beckman JLA 16.250 rotor at 5000 xg for 15 mins at 4 °C. Cells were resuspended in the following buffers: 20 mM HEPES, 300 mM NaCl, pH 8.0 for Bcl3 and p50 variants (except p50-GFP in 50 mM Tris-HCl, pH 8.0) and 50 mM Tris-HCl, pH 8.0 for sfGFP variants. Cells were lysed using the French Pressure Cell. A maximum of 35 mL was lysed each time at 1250 psi of pressure and collected into 50 mL falcon tubes. Soluble cell lysate was then separated from insoluble fractions via centrifugation using a Beckman JLA 25.50 rotor at 25,000 xg for 40 mins.

2.3.2 Protein Purification

Protein purification was carried out with either an ÄKTA Prime Plus, ÄKTA Purifier FPLC or ÄKTA Pure systems and used with various columns purchased from GE Healthcare. A UV wavelength of 280 nm was used to detect total protein. Where appropriate, wavelengths of ~ 400, 485 and 515 nm were used to detect presence of sfGFP and Venus respectively. Protein samples were concentrated using Vivaspin™ 10 kDa molecular weight cut-off spin filters (VWR) by centrifugation at 4000 x g until the desired volume is reached.

2.3.3 Metal Affinity Chromatography

Principle

A column is used containing sepharose beads coordinated to Ni^{2+} . Crude cell lysate is passed through the column, where proteins containing a N- or C-terminal hexa-histidine tag will bind to the column via coordination with the cation. Other proteins will elute from the column. Bound target protein is released by the addition of the elution buffer containing imidazole, which is the functional group of histidine. Increasing the concentration of imidazole out competes with the bound protein for the cation and causes the protein to elute.

2.3.3.1 Nickel affinity chromatography

Clarified cell lysate was passed through a 5 mL His Trap™ HP column (binding capacity ~200 mg protein) equilibrated in buffer containing minimal (10 / 20 mM) or no imidazole. Bound target protein was then eluted by the addition of the elution buffer containing imidazole at a gradient from 10 to 500 mM and 1.4 mL fractions were collected. Samples of each fraction were taken and used to check for purity via SDS-PAGE analysis.

2.3.4 Size Exclusion Chromatography

Various columns were used to perform size exclusion chromatography including: HiLoad™ 16/600Superdex™ S75 pg (preparative grade, 120 mL bed volume, separation range 3-75 kDa), HiLoad™ 26/600Superdex™ S200 pg (preparative grade, 320 mL bed volume, separation range 10-600 kDa), and HiLoad™ 16/600Superdex™ S200 pg (preparative grade, 120 mL bed volume, separation range 10-600 kDa). For separation of the Bcl3 and p50 complex, the HiLoad™ 26/600Superdex™ S200 pg (preparative grade, 320 mL bed volume, separation range 10-600 kDa) is used. All columns are equilibrated in either 20 mM HEPES, 300 mM NaCl, pH 8.0, or 50 mM Tris-HCl, pH 8.0. Samples (0.5 – 2.0 mL) were loaded onto the column with a continuous flow rate of 1 mL/min. Fractions (1.4 mL) were collected according to the UV trace and checked for purity via SDS-PAGE analysis.

2.3.5 SfGFP dimers

Copper sulphate (CuSO₄) (Sigma-Aldrich), copper chloride (CuCl₂) (Sigma-Aldrich), or manganese sulphate (Sigma-Aldrich) was added to sfGFP protein samples to encourage the dimerisation by disulphide bonding. Unless stated in the text, CuSO₄ was prepared as mentioned in section 2.1.1 and used at a final concentration of 1 mM by diluting in 50 mM Tris-HCl, pH 8.0.

Prior to dimerisation studies, all protein was exchanged into 50 mM Tris-HCl, pH 8.0. Residual monomer was removed from sfGFP dimers prior to x-ray crystallography studies and to determine accurate spectral properties. For this, CuSO₄ was added to the sample and incubated for 1 hr with mixing by a rotator (Stuart). The sample was then passed through a SEC column as described in section 2.3.4 to remove remaining CuSO₄.

2.4 Protein analysis

2.4.1 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein samples with a Mini-PROTEAN Tetra cell tank and plates (Bio-Rad). The gel was prepared and setup according to Bio-Rad (2011) protocol (Table 7).

The protein samples were prepared using 4x SDS-PAGE loading buffer with a final composition of 2% w/v SDS, 200mM Tris-HCl pH 6.8, 0.04% w/v bromophenol blue, 8% v/v glycerol and 10% v/v β-mercaptoethanol and denatured at 95 °C for 10 minutes. For non-reducing SDS-PAGE gels, the β-mercaptoethanol was absent from the loading buffer. A broad range protein ladder (New England Biolabs) was used as a molecular weight reference. Gels were run at 200V for 40 minutes.

Gels were stained in a solution consisting of 0.1% w/v coomassie blue, 40% v/v methanol and 10% v/v acetic acid. Followed by de-staining in 40% v/v methanol and 10% v/v acetic acid. Gels were visualised using a GelDoc-It system (Bio-Imaging Systems).

Table 7. Components of SDS-PAGE separating and stacking gel. *40% - acrylamide/bis acrylamide 37.5:1.

Materials	12.5 % Separating gel	Stacking gel
	Volume	Volume
Water	4.275 mL	3.69 mL
1.5 M Tris-HCl, pH 8.8	2.5 mL	-
0.5 M Tris-HCl, pH 6.8	-	650 μ L
Acrylamide ^[a]	3.125 mL	500 μ L
10 % SDS (w/v)	100 μ L	100 μ L
10% APS (w/v)	50 μ L	50 μ L
TEMED	20 μ L	20 μ L

2.4.2 Protein concentration determination

Protein concentrations were calculated using the Bio-Rad *DC* Protein Assay. Purified protein (5 μ L) was pipetted in triplicate into a 96-well plate. SfGFP^{WT} or BSA were used as a standard at known concentrations (0 – 2 mg/mL) and used to calibrate the assay each time. DC reagents were added following manufacturer's instructions. The absorbance at 750 nm was recorded for each well on a CLARIOstar Plus.

Where mentioned, the nanodrop (Thermo Scientific) was used as a method to approximate protein concentrations. Droplets between 1 – 2 μ L are added onto a pedestal. Initially water is added to ensure the device is clean and then blanked by addition of the relevant buffer solution. The nanodrop directly records the absorbance at 280 nm and calculates the protein concentration by using the Beer Lambert equation (Equation 2) as seen below.

Equation 2.
$$\varepsilon = A / cl$$

where ε is the extinction coefficient ($M^{-1}cm^{-1}$), A is the absorbance value at λ_{max} , c is the protein concentration (M) and l is the pathlength (cm).

2.4.3 Calculation of extinction coefficient

UV-visible (UV-vis) absorption spectra were recorded on a Cary spectrophotometer in a 1 cm pathlength cuvette (Hellma, Müllheim, Germany). Spectra was recorded from 200 – 600 nm at a rate of 300 nm/min. Absorption data was recorded for the known concentration of protein (using Bio-Rad DC protein assay, section 2.4.2). These values were then substituted into the Beer Lambert equation (Equation 2) to determine the extinction coefficient of the protein.

2.4.4 Fluorescence Spectroscopy

Emission and excitation spectra were measured via two methods. For the Varian Cary Eclipse Fluorimeter, samples (400 μ L) were added to a 5 mm x 5 mm QS quartz cuvette (Hellma) and data was collected with 5 nm slit width at a rate of 300 nm/min. Emission spectra were recorded at a fixed excitation wavelength according to the excitation maximum of the variant. Excitation spectra were recorded at the fixed wavelength according to the maximum emission. Unless otherwise stated, spectral scans were recorded for purified sfGFP and Venus at concentrations of 0.5 μ M in 50 mM Tris-HCl, pH 8.0, except for dimers at 0.25 μ M.

For sfGFP dimerisation experiments, the CLARIOstar Plus was also used. The bandwidth of filters used was 8 nm which dictates the amount of light which can pass through. Considering the minimum distance allowed between filters is 30 nm, the excitation was recorded at 516-8 upon emission at 388-8 nm and 483-8 nm. Both wavelengths were consistent when testing both sfGFP and p50-GFP variants. For each sample, 200 μ L was prepared in triplicate in a Nunc™ F96 Microwell™ black polystyrene plate (Thermo fisher). To calculate CroB/CroA ratios, emission values upon excitation at 483 nm were divided by emission values when recorded at 388 nm.

2.4.5 Förster resonance energy transfer (FRET)

The Förster resonance energy transfer (FRET) was measured between sfGFP (the donor) and Venus (the acceptor) using the Varian Cary Eclipse Fluorimeter or CLARIOstar Plus. Testing was performed on two different sample types: (1) mixing of the individual components GFP-Bcl3 and p50-Venus, or (2) co-expression of

GFP-Bcl3 and p50-Venus. For the mixing of samples, 1 μM of GFP-Bcl3 was prepared with 1, 2 and 4 μM of p50-V to give final ratios of 1:1, 1:2 and 1:4 (GFP-Bcl3:p50-Venus). Samples were incubated at 1 and 24 hour time periods at both 20 and 37 $^{\circ}\text{C}$. The emission spectrum was then recorded for the samples (containing both GFP-Bcl3 and p50-Venus) as described in section 2.4.4, at a fixed excitation of 420 nm. The spectra was also recorded for the individual components (GFP-Bcl3 [1 μM] and p50-Venus [1, 2 and 4 μM]) as a control. For calculation of the FRET efficiency, the emission spectra of test samples were deconvoluted using FluorTools (<http://www.fluortools.com/home>), to determine the fluorescence intensity of acceptor (p50-Venus) and fluorescence intensity of donor (GFP-Bcl3). To account for background p50-Venus signal, the emission spectra of p50-Venus (control sample) was subtracted from the emission of the p50-Venus (acceptor) in the test sample. For the co-expressed proteins, the concentration of p50-Venus in the sample was approximated using ImageJ by comparing the band intensity against known concentrations. The integrated fluorescence of the donor and acceptor were then taken and supplemented into Equation 3 to determine the relative FRET efficiency (E_{rel}).

Equation 3.
$$E_{rel} = \frac{I_A}{(I_D + I_A)}$$

where I_A is the integrated fluorescence of acceptor and I_D is the integrated fluorescence of donor. The distance at which energy transfer is 50% efficient is referred to as Förster radius (R_0), calculated using Equation 4.

Equation 4.
$$R_0 = 0.211 \sqrt[6]{k^2 n^{-4} Q_D J(\lambda)}$$

Where k^2 is the dipole orientation factor, n is the solvent refractive index, Q_D is the quantum yield of the donor, and $J(\lambda)$ is the overlap integral between the donor emission and the acceptor molar absorbance. K^2 is arbitrarily set as 0.667 to reflect two randomly orientated chromophores (Pope et al. 2021). The refractive index (n) was determined as 1.4 when accounting for a combined protein-water environment. With a known distance between FRET chromophores (r), the energy transfer efficiency (E) can be then determined with Equation 5.

Equation 5.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

2.4.6 Determination of molecular weight by gel filtration chromatography

The oligomeric state of proteins and the estimated molecular weight was determined by gel filtration chromatography. The Hiload™ 26/600 Superdex™ S200 was calibrated using a 15 – 600 kDa (670, 150, 44.3 and 13.7 kDa) protein standard mix (Sigma-Aldrich) at a flow rate of 1.0 mL/min in 20 mM HEPES, 300 mM NaCl, pH 8.0. Dr Husam Sabah Auhim calibrated the Hiload™ 16/60 Superdex™ S200 using gel filtration standard proteins (Bio-Rad) of molecular weight (670, 158, 44, 17 and 1.35 kDa) at a flow rate of 0.5 ml/min with 50 mM Tris-HCl buffer, pH 8. The partition coefficient (K_{av}) for each protein (standard and sample) was calculated using Equation 6.

Equation 6.

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

Where K_{av} is the partition coefficient, V_e is elution volume, V_0 is the void volume, V_t is the total column volume. The V_0 of the Hiload™ 26/600 Superdex™ S200 column was determined using blue dextran (2000 kDa), V_t was 320 ml (column volume) and the elution volume V_e for each sample was monitored by absorption at 280 nm (λ_{max} of elution peak) as shown in Figure 11a. The estimated molecular mass of the test protein sample was calculated using molecular weight standard curve as shown in Figure 11b.

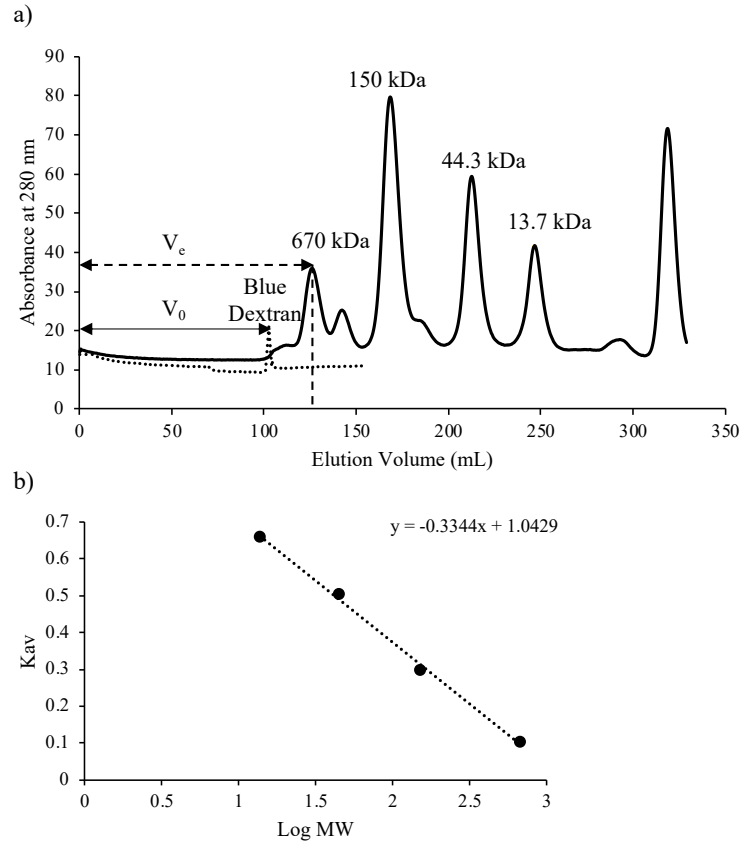


Figure 11. Determination of molecular weight by gel filtration chromatography. a) Elution profile of standard proteins of molecular weight 670, 150, 44.3 and 13.7 kDa in a Hiload™ 26/600 Superdex™ S200 pg column at a flow rate of 1.0 ml/min with 20 mM HEPES, 300 mM NaCl, pH 8. B). The plot of the log of molecular weight (log MW) against the partition coefficient (K_{av}).

The void volume (V_0) of the Hiload™ 16/60 Superdex™ S200 column was determined using blue dextran (2000 kDa), V_t was 120 ml (column volume) and the elution volume V_e for each sample was monitored by absorption at 280 nm (λ_{max} of elution peak) as shown in Figure 12. The estimated molecular mass of the test protein sample was calculated using molecular weight standard curve as shown in Figure 12b.

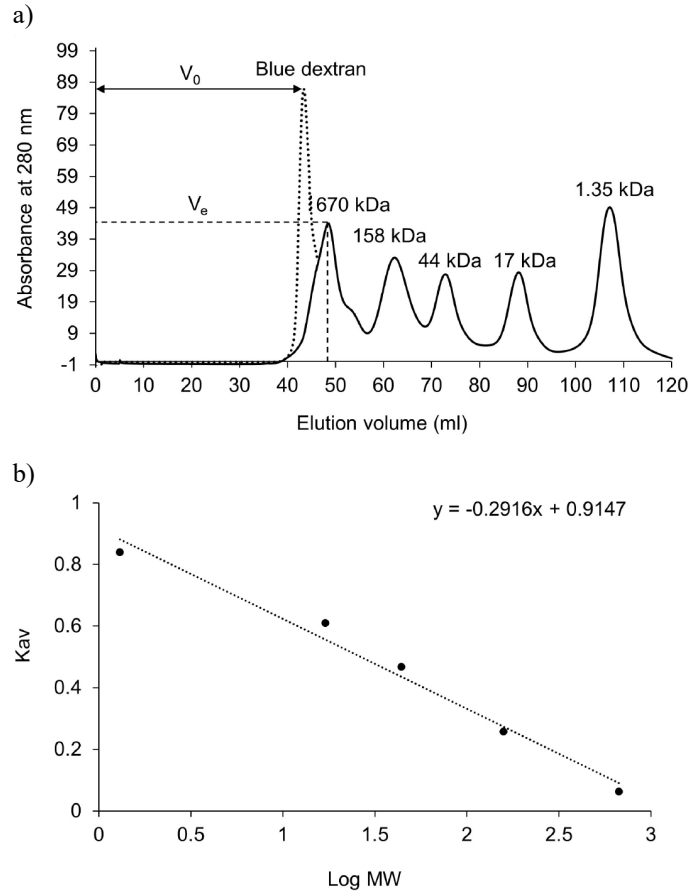


Figure 12. Determination of molecular weight by gel filtration chromatography. a) Elution profile of standard proteins of molecular weight 670, 158, 44, 17 and 1.35 kDa in a Hiload™ 16/60 Superdex™ S200 pg column at a flow rate of 0.5 ml/min with 50 mM Tris-HCl buffer, pH 8. B). The plot of the log of molecular weight (log MW) against the partition coefficient (K_{av}).

2.4.7 Determination of quantum yield

Quantum yield (QY) with a theoretical maximum of 1.0 is defined as the ratio of photons emitted to photons absorbed. A standard molecule is used as a comparative to determine the QY of sfGFP variants. The standard fluorescein with a QY of 0.75 (dissolved in 0.1 M NaOH) was used as it has similar excitation and emission spectra's as sfGFP^{WT}. Variants and fluorescein were diluted to a maximum absorbance of 0.1 and used to record emission spectra according to section 2.4.4, when excited at λ_{max} . Areas under fluorescence emission spectra traces were integrated from 503 nm ($\lambda_{ex} + 5$ nm from the most redshifted variant – sfGFP^{C148}) to 650 nm and input into Equation 7.

Equation 7.
$$\frac{QY(x)}{QY(s)} = \frac{\int Em(x)}{\int Em(s)} \cdot \frac{\eta(x)}{\eta(s)}$$

Where x is the unknown sample, s is the standard and η is the refractive index of the solvent (for most aqueous buffers η is ~ 1.33) (Hartley et al. 2016). The refractive index of both samples is almost negligible and is such ignored, since $\frac{\eta(x)}{\eta(s)} = 1$.

2.4.8 Characterising dimerisation of sfGFP^{C148}

The dimerisation of sfGFP variants used in this study was of the second order rate reaction, as homodimerisation is recorded as $A + A \rightarrow A_2$ (Garratt et al. 2013). To calculate the percentage of sfGFP that has dimerised, the emission ratios of CroB/CroA were recorded (using the CLARIOstar Plus, section 2.4.4). Values for pure monomer and pure dimer were used to calibrate 0 and 100 % dimerised respectively. The second order rate constant was calculated using Equation 8. As a rapid reaction, the lower concentrations (0.5, 0.25 and 0.1 μM) for both sfGFP^{C148} and p50-sfGFP^{C148.F} were used for this.

Equation 8.
$$Rate = k[A]^2$$

Where k is the rate constant and A is the reactant.

2.4.9 Isothermal Calorimetry

Purified protein was placed in a dialysis membrane (3.5 kDa) (Spectrum Chemical, US) and added to a beaker filled with suitable buffer (20 mM HEPES, 300 mM NaCl). Dialysis was performed overnight at 4 °C with stirring. The protein samples were concentrated and supplemented with DMSO for final concentrations of 5 or 50 μM (1% DMSO). The drug JS6 was prepared according to section 2.1.1 and diluted to final 1% DMSO concentration with the same previously dialysed buffer (20 mM HEPES, 300 mM NaCl). A final drug concentration of 100 μM was used.

Isothermal Calorimetry (ITC) was performed using MicroCal PEAQ-ITC (Malvern Panalytical) to investigate potential binding of JS6 to Bcl3, p50 and Bcl3-p50 complex. The experiment was run at 25 °C with a reference power set to 10 $\mu\text{cal/sec}$ and an injection volume of 2 μL . Initial controls were performed titrating

buffer into buffer, drug into buffer and buffer into complex, before titrating the drug into protein. For drug-protein titrations, the injection syringe (300 μ L) was filled a total of three times to ensure the drug concentration was great enough to achieve saturation. Heat changes were recorded over 49 minutes and analysed using MicroCal PEAQ-ITC Analysis software.

2.5 Structural studies

2.5.1 Mass spectrometry

The dimerised protein sample was prepared at 10 μ M in 20 mM Tris-HCl, pH 8.0 and sent for analysis at Cardiff University School of Chemistry's Mass Spectrometry Suite. Theoretical masses of the sfGFP variant were calculated using ExPASy ProtParam tool (<https://web.expasy.org/protparam/>) substituting cysteine and phenylalanine residues at positions 148 and 206 respectively (Gasteiger *et al.* 2005). Predicted masses were altered to allow for chromophore maturation (-20 kDa), doubled to account for the protein dimer and minus a further 2 kDa for disulphide bond formation.

2.5.2 SEC – DLS

Size exclusion chromatography–dynamic light scattering (SEC-DLS) was performed using equipment provided by the Protein Technology Hub at Cardiff School of Biosciences. Particle size and approximate molecular weight were determined using an ÄKTA Pure FPLC (GE Healthcare) system by passing samples through a Superdex™ 200 10/300 GL column (GE healthcare) to a Zetasizer MicroV dynamic light scattering (DLS) system (Malvern Panalytical). The sample continued through a Viscotek Refractive Index detector to determine an accurate protein concentration based on differences between the sample itself and a reference cell (filled with buffer). Bovine serum albumin (BSA) was used to calibrate the column.

The Bcl3-p50 protein complex (100 μ L at 3 mg/ml) was injected onto the column equilibrated in 20 mM HEPES, 300 NaCl, pH 8.0 and flow controlled at 0.5 mL/min. Analysis of data was performed using OmniSEC software (Malvern Panalytical) calculating the Mw of the protein complex from the hydrodynamic radius, which is directly derived from the observed light scattering.

2.5.3 Thermal Melting Assay

The thermal melting assay was used to investigate stability effects caused by different protein concentrations, various pH's and the addition of DNA.

The Bcl3-p50 protein complex was expressed and purified via nickel affinity chromatography. The buffer suggested by Rupesh et al. (2014) was amended for use for DNA binding studies (20 mM HEPES, 150 mM NaCl, 3 mM MgCl₂, 1 mM EDTA) and used for the second purification step of the complex by SEC, using the HiLoad™ 26/600Superdex™ S200 pg.

A dsDNA sequence that binds p50 with high affinity was chosen and used to design the single stranded oligonucleotides (5' – GGGGATTCCC - 3') (Udalova et al. 2002; Nijnik et al. 2003). The ssDNA was synthesised by IDT and prepared to final concentration of 2 mM by the addition of nuclease free water. The DNA was annealed to produce a 1 mM duplex by mixing the ssDNA (1:1 molar ratio) followed by incubation at 20 °C for 1 hour with light shaking at 100 RPM.

A hard-shell 384-well PCR microplate (Bio-Rad) was set up with final protein concentrations of 2.0, 2.5 and 3.0 mg/ml at the various pH's, with and without DNA (1:1 molar ratio with protein), and a variety of drug concentrations (25, 50, 75 and 100 µM) and pH's (7.0, 7.5, 8.0, 8.5 and 9.0). SYPRO orange (5x) (Sigma-Alrich) was added last to make up the final sample volume of 25 µL. An adhesive microseal® (Bio-Rad) was applied followed by centrifugation at 500 RPM for 2 mins to collect solutions in the bottom of the wells and remove any bubbles.

The microplate was then placed in a real-time PCR (RT-PCR) machine (Bio-Rad) and set up with the fluorescent properties of SYPRO orange (λ_{ex} 470 nm / λ_{em} 570 nm). The program had an initial hold at 25 °C for 10 mins before taking readings for increasing temperature increments of 0.5 °C followed by a 15 second hold. The steps were repeated until reaching a temperature of 95 °C. Data was saved as an excel file and analysed as described in Huynh and Partch (2015) which removes post-peak fluorescence quenching and performs non-linear fitting of the truncated dataset to a Boltzmann Sigmoidal curve with Equation 9.

Equation 9.
$$Y = Bottom + (Top - Bottom) / (1 + \exp\left(T_m - \frac{x}{Slope}\right))$$

Where y = fluorescence emission in arbitrary units; X = temperature; Bottom = baseline fluorescence at low temperature; Top = maximal fluorescence at top of the truncated dataset; Slope = describes the steepness of the curve, with larger values denoting shallower curves; and T_m = melting temperature of the protein.

2.5.4 X-ray Crystallography

2.5.4.1 Bcl3-p50 Crystal formation

Purified protein complex in 20 mM HEPES, 150 mM NaCl, 3 mM MgCl₂, 1 mM EDTA, pH 7.0 was concentrated by centrifugation at 4000 RPM in 30 kDa spin columns until it reached ~ 3 mg/ml, approximated using the nanodrop. Crystal formation was screened using sitting drop vapour diffusion across a wide variety of conditions as described by the PACT *premier*TM, JCSG-*plus*TM, SG1TM, Morpheus®, BSC and MIDASplusTM HT-96 broad crystallisation screens (Molecular Dimensions, Suffolk, UK). A multichannel pipette was used to aliquot (40 μ L) of the crystallisation conditions into the 96-well plate. Drops were set up with various protein-crystallisation buffer ratio's including 1:1, 1:2 and 1:3, to give a final drop volume of 200 nL using mosquito® crystal, sptlabtech. The plates were sealed with a ClearVueTM sheet (Molecular Dimensions, Suffolk, UK) and stored at 4 and 20 °C. Crystal growth was monitored once a week.

2.5.4.2 SfGFP Crystal formation

Purified proteins samples in 50 mM Tris-HCl were concentrated by centrifugation at 4000 RPM in 30 kDa spin columns (Merck) until it reached ~ 10 mg/mL. Protein concentration was estimated using the nanodrop. Crystal formation was screened using sitting drop vapour diffusion across a wide variety of conditions as described by the PACT *premier*TM and JCSG-*plus*TM HT-96 broad crystallisation screens (Molecular Dimensions). A multichannel pipette was used to aliquot (50 μ L) of the crystallisation conditions into the 96-well plate. Drops were then manually set up with equal volumes of protein and crystallisation buffer (1 μ L each). Two rows were set up each time and temporarily sealed with an adhesive film (Molecular dimensions) to avoid evaporation of the droplet. When all 96 wells are complete, the temporary seals are removed, and the entire plate is covered with a ClearVueTM

sheet. Crystallisation plates were stored at 23 °C. Crystal growth was monitored once a week.

2.5.4.3 Data acquisition and refinement

Crystals were harvested by Dr Pierre Rizkallah (Cardiff Medical School, Cardiff University) by addition of 1 mM ethylene glycol and flash freezing in liquid nitrogen. Crystals were stored in liquid nitrogen during transit to Diamond Light Source, Harwell, UK for x-ray diffraction.

2.6 *In Silico* Modelling

In silico modelling was carried out throughout this thesis, using a variety of computational software as listed in Table 8. It was used for two purposes: (1) model the sfGFP-Bcl3 – p50-Venus complex, and for energy minimisation of sfGFP^{H148X} variants.

Table 8. List of software packages used for *in silico* modelling.

Software package	Use
ACEPYPE	File preparation for adding to GROMACS database
Marvin Sketch	Molecular design
Robetta	Protein structure prediction service
PyMOL	Protein viewer – inserting mutations into residues
GROMACS	Energy minimisation and Molecular dynamics simulations

2.6.1 Input structures and Computational Requirements

For generation of sfGFP-Bcl3 and p50-Venus, the following PDB ID's were downloaded and used in PyMOL: 2B3P, 1K1A, 1NFK, and 1MYW respectively. The 3-D structure of a linker between p50 and Venus was also predicted using the online software Robetta (<https://robetta.bakerlab.org/>). Construction of proteins (sfGFP-Bcl3 and p50-Venus) was performed using the 'fuse' tool in PyMOL where the nitrogen atom from one chain is selected followed by the carbon from the second

chain. For the MD simulation of the Bcl3-p50 protein complex; Bcl3 and p50 (fused to FPs) were then superimposed onto the Bcl3₂-p50₂ model determined by Prof Andrea Brancale (School of Pharmacy, Cardiff University). Only one sfGFP-Bcl3 was used, as our data suggested one Bcl3 molecule bound to p50 homodimer.

For generation of sfGFP^{H148X} mutants, the mutagenesis tool was used in PyMOL, selecting the rotamer with the least amount of steric clashes.

2.6.2 Parametrisation

A forcefield refers to a collection of equations that are used to describe the energy of a protein as a function of its atomic coordinates. They are required in order to run a MD simulation as it will determine the way in which all atoms, bonds, and angles of a system are treated. In this work, the AMBER (Assisted Model Building with Energy Refinement) forcefield was used to perform MD simulations in GROMACS (specifically the Amber99sb forcefield) (Bayly et al. 1995; Hornak et al. 2006). The potential energy function used by AMBER forcefield is described is shown in Equation 10.

Equation 10. *AMBER potential energy function as described in the Amber 2022 Reference Manual.*

$$V_{\text{AMBER}} = \sum_{\text{bonds}} k(r - r_{eq})^2 + \sum_{\text{angles}} k(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] / \\ + \sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \sum_{i < j} \left[\frac{q_i q_j}{\epsilon R_{ij}} \right]$$

Where Potential energy (v) is characterised by the individual sums of bonds, angles, dihedrals, and non-bonded interactions. Additional parameters for non-canonical biological structures such as the chromophores of sfGFP and Venus, were calculated for use with the Amber99sb forcefield constraints (Bayly et al. 1995; Hornak et al. 2006). Parametrisation requires spatial coordinates of each atom within a structure for geometrical calculations. Spatial coordinates for the chromophore of sfGFP and Venus were extracted using their PDB IDs (2B3P and 1MYW respectively) in PyMOL. The chromophores were then constructed using Marvin

Sketch molecular editing tool (freely downloaded from <https://chemaxon.com/marvin>). The atomic charge calculator was used to assign partial charges (<https://bio.tools/atomic-charge-calculator-ii>). Atom names were assigned via standard naming conventions using a text editor. The antechamber python parser interface (ACEPYPE) was then used to determine bond lengths, angle sizes and dihedrals for input into Amber99sb forcefield for use in GROMACS.

2.6.3 Molecular dynamics with GROMACS

GROMACS (GRONingen Machine for Chemical Simulations) was the software package used for the MD simulations (freely downloaded from www.gromacs.org). All MD used throughout this project was performed according to GROMACS Documentation, with version 2021.2. The proteins were submitted in a PDB file format for conversion to the .gro structural file, with a Gromos87 format. An appropriately sized octahedron box was determined for the addition of solvent molecules around the protein or protein complex. The Simple Point Charge water (SPC) with coordinates from spc16.gro was used. Monoatomic ions (Na^+ or Cl^-) were then added (by random replacement with solvent molecules) to neutralise the system. The neutralised system requires an energy minimisation (EM); the first most important stage of any simulation to relieve any clashes in the file that would cause problems in later stages. It also minimises the structural file to its lowest energy so identifying the most stable confirmation for input into the MD simulation. The steepest algorithm was presented with a step size of 0.01 nm and a maximum force of 1000 kJ/mol/nm. The neighborlist determined using the grid framework was updated every step of the minimisation, with a Verlet cut-off scheme. Electrostatics were considered using Particle Mesh Ewald (PME) with a 1 nm Coulomb cut-off. Van der Waal (vdW) interactions were measured to a 1 nm vdW cut-off. Periodic Boundary Conditions (PBC) were considered in all three dimensions (xyz).

The system requires parameters that define the relationship between temperature, pressure and density of the simulation. This was processed through a two-step equilibrium involving leap-frog integrator to allow for Newtons laws of motion. Each equilibrium is programmed for 100 ps with a 2 fs time step integration. Holonomic constraints are accounted with the LINCS algorithm, allowing for matrix inversion correction (corrects bond lengthening when rotation occurs). The

lengthening correction was set to a single step process, with highest order (4) constraints. All bonds involving hydrogen are also converted to a constraint. Electrostatic parameters use the PME coulomb-type with a cubic interpolation grid dimension expressing 0.16 nm Fourier-space grid points. For both equilibrium steps, temperature cooling was set using velocity rescaling. Coupling groups involved both protein and non-protein groups for improved accuracy. Time constants for temperature coupling was limited to 0.1 ps with a standard reference temperature of 300 K (27 °C). The second equilibrium step also included isotropic coupling, using Parrinello-Rahman extended ensemble. Simulation box vectors were therefore subject to Newton's laws of motion and uniform scaling. The time constant for pressure coupling was set to 2 ps with a 1 bar reference pressure. Isothermal compressibility of solutes was referenced at $4.5^{-5} \text{ bar}^{-1}$. All parameters were used for the final MD simulation.

2.6.4 Analysis of the Root Mean Square Fluctuation (RMSF)

The root mean square fluctuation (RMSF) is a measure of the displacement of a particular atom, or group of atoms, relative to the reference structure, averaged over the number of atoms. It was calculated for the C_{α} atom of each residue with the function *gmx rmsf* using Equation 11. The RMSF was analysed for separate domains of p50, Venus, Bcl3 and GFP by creating an index file and selecting the range of atoms for each protein.

$$RMSF_i = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle}$$

Equation 11.

Where atom i has an atomic position of x and averages are over all input frames.

2.6.5 Analysis of distance

The function *gmx distance* calculates distances between pairs of positions as a function of time. For calculation of the average chromophore distances between GFP and Venus, an index file was created to select the centre of geometry from all atoms within the chromophore.

3 The Bcl3-p50 complex

3.1 Introduction

When complexed with p50 homodimer, Bcl3 is associated with tumorigenesis in several malignancies, likely via the upregulation of NF- κ B target genes (detailed in section 1.2). Whilst this protein-protein interaction has been evidenced both *in vitro* and *in vivo*, it is critical that we improve our understanding of the biomolecular role of Bcl3 and its mechanism of action when complexed. However, currently the crystal structure of Bcl3 in a complex with its binding partners is not available. Various *in silico* approaches have been utilised that model both stoichiometric forms; the Bcl3:p50₂ (Michel et al. 2001; Manavalan et al. 2010) and Bcl3₂:p50₂ (Pang et al. 2004; Soukupova et al. 2021) structures. Pang et al. (2003) constructed a model of Bcl3₂:p50₂ using the I κ B α - NF- κ B structure as a template where the Bcl3 ARD (PDB ID: 1K1A) and p50 homodimer (PDB ID: 1NFK) were aligned and superimposed onto the I κ B α - NF- κ B complex (PDB ID: 1NFI). The second Bcl3 molecule was added to the model using the pseudodyad symmetry of the p50 homodimer. This model was then further refined by Prof Brancale (School of Pharmacy, Cardiff University) and Prof Clarkson's group (School of Biosciences, Cardiff University) using a short (5ns) MD simulation to relax the contacts between Bcl3 and p50 (later published as Soukupov et al. (2021)) (Figure 13a). The dimer interface between Bcl3 and p50 revealed an extensive network of hydrophobic interactions and hydrogen bonds. Of particular interest, was a hydrophobic binding pocket within ANK repeats 6 and 7. Compared to I κ B α which only contains 6 ANK repeats, it posed a unique target on Bcl3 for the development of small molecule inhibitors. The binding pockets comprised p50 residues K275, D297, S299, P300, T301, D302, V303, H304 and R305 interacting with Bcl3 ANK repeat 6 residues M298, Y299, S300, G301, S302, S303, H306 and S307, and residues from the unique ANK repeat 7 (N331, C332, H333, N334, D335, V240 and R342)(Figure 13b). Virtual screening was then performed to identify specific compounds that fit the binding pocket in a similar manner as the corresponding p50 residues. A selection of compounds were tested for their efficacy using cell cultures, which evidenced their role blocking Bcl3-p50 binding interactions. Of those trialled, the

compound denoted JS6 demonstrated its ability to inhibit the interaction between Bcl3 and p50.

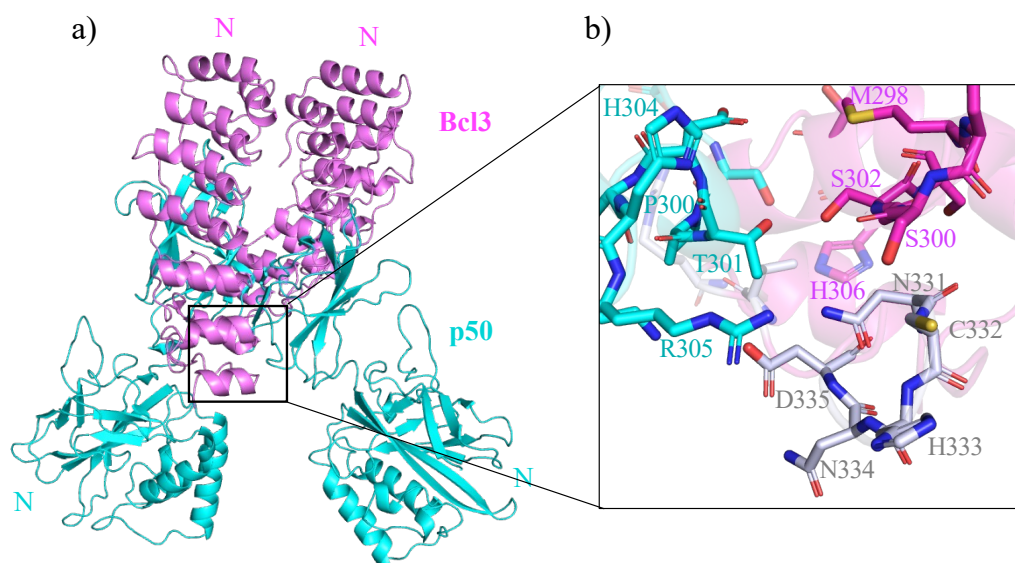


Figure 13. Model of Bcl3₂-p50₂ by Prof Andrea Branciale. A) Cartoon representation of two Bcl3 ARD (magenta) bound by a p50 homodimer (cyan). B) Binding pocket identified between p50 residues (cyan), Bcl3 ARD 6 residues (magenta), Bcl3 ARD 7 residues (grey).

In silico analysis was performed by Manavalan et al. (2010) to model the Bcl3-p50₂ complex. However, rather than using the structure of IκBα- NF-κB as a template for structural alignment (method used by Michel et al. 2001, Pang et al. 2003; Soukupov et al. 2021), they performed molecular docking for Bcl3 and p50 homodimer using the PPI docking software GRAMM-X and ZDOCK. Two models of the Bcl3-p50₂ complex were determined, where two different orientations of Bcl3 were bound with the p50 homodimer. It was hypothesised that the two structures illustrate the dual roles in NF-κB signalling: mediating transcription via Bcl3 TAD and facilitating transcription by displacing p50 homodimer. Comparison of structures determined by docking approaches and structural alignment revealed the models are largely similar, where the main difference is the orientation of Bcl3. Thus, the consistencies throughout all models highlight the key features expected for the binding of Bcl3 to p50 including; the C-terminal dimerisation region of p50 plays a crucial role in the binding of Bcl3, and the majority of PPIs are made between Bcl3 ANK repeats 4-7.

Whilst the models of Bcl3-p50 complex provide insight into the overall protein-protein interactions, to fully understand the molecular basis of the interaction, atomic level information of Bcl3 complexes is required. The aim of this chapter will therefore first focus on the generation and characterisation of Bcl3 and p50 proteins. PPIs will then be investigated between Bcl3 and p50 via a variety of techniques including; thermal melting assay, SEC-DLS, and FRET. The drug compound JS6, identified by Prof Clarkson's group via experiments in cell culture screens, will then be tested with purified proteins for its ability to bind Bcl3 and its subsequent effect on complex formation. As a purified complex, Bcl3-p50 will then be used to set up x-ray crystallography trials to determine its structure and compare against previously determined models. With the structural data of Bcl3-p50 complex, this can then be used as the basis for further computational drug design.

3.2 Results

3.2.1 Structural Studies of Bcl3-p50 complex

3.2.1.1 Expression and purification of Bcl3-p50

Plasmids containing either Bcl3 or p50 encoding genes (Table 2) were transformed into *E.coli* as described in section 2.2.9. Protein expression was performed at 25 and 28 °C for Bcl3 and p50 respectively. The proteins contained a hexahistidine tag at the N-terminal (GSHHHHHH) which was used for purification by nickel affinity chromatography using the ÄKTA FPLC system (section 2.3.3). Fractions were collected according to a rise in absorbance (monitoring at 280 nm) and analysed by SDS-PAGE. A second SEC purification step was then performed to remove free imidazole and any remaining contaminants. Bands were observed corresponding to the approximate size of Bcl3 (28 kDa) and p50 (39 kDa)(Figure 14), as determined by ExPASy (ExPASy ProtParam tool. 2008). Protein expression and purification were confirmed successful.

To determine the stability of Bcl3 and p50, the purified protein samples were stored at both 4°C and 20 °C and checked daily by SDS-PAGE. Precipitation was also detected from samples stored at both temperatures after approximately two days. Thus, as individual proteins, both p50 and particularly Bcl3 were unstable. This factor was considered during the planning of future experiments.

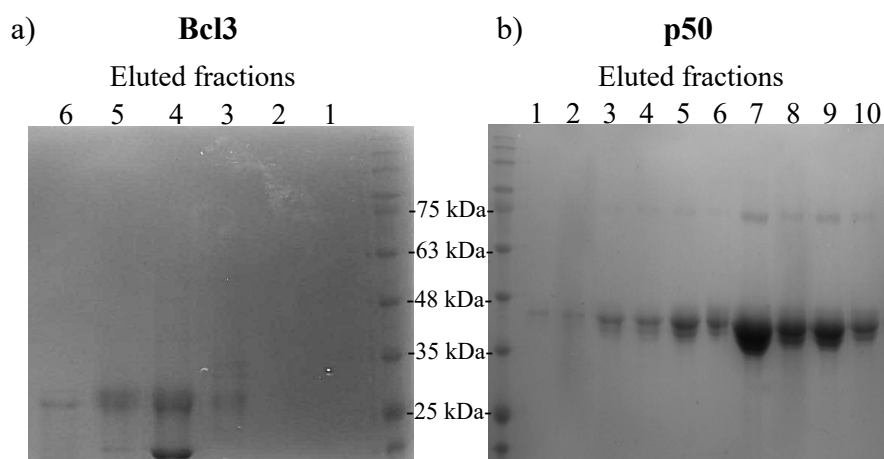


Figure 14. SDS-PAGE analysis of protein fractions that eluted following SEC. Fractions labelled in order of their collection from the SE column. a) Bcl3. B) p50

To overcome the apparent instabilities of the individual components, the plasmids containing genes encoding Bcl3 and p50 were then used together, to co-transform into *E.coli* and thus co-express them. Naturally, binding of Bcl3 and p50 occurs inside the cell, possibly stabilising the proteins. Co-expression of proteins was carried out in a 1 L culture overnight at 27 °C. Purification was first performed by nickel affinity chromatography where two elution peaks were observed; a major peak (peak 1) and a minor peak (peak 2) (Figure 15a). Analysis by SDS-PAGE showed that both elution peaks contained bands the expected size of Bcl3 (28 kDa) and p50 (39 kDa) (Figure 15b). Assuming Bcl3 is bound to p50, then at least 3 (Bcl3:p50₂), or 4 His-tags (Bcl3₂:p50₂) could be bound to the column, per complex. Complexed proteins would likely require a higher concentration of imidazole for elution in comparison to individual non-complexed proteins which have just 1 (Bcl3) or 2 (p50 homodimer) his-tags. Therefore, although Bcl3 and p50 both eluted during the major peak (peak 1), there is greater confidence that proteins eluted from the minor peak (peak 2) with the higher imidazole concentration (500 mM) are complexed. Thus, samples were pooled from the minor peak and then subjected to SEC. Two elution peaks were observed: a minor peak at ~155 mL (peak 1) and a major peak at 180 mL (peak 2) (Figure 15c). SDS-PAGE analysis showed both Bcl3 and p50 had eluted in the same fractions (Figure 15d). Based on the calibrated SEC column and molecular weight estimation described in section 2.4.6, the major elution peak (peak 2) has a molecular weight of approximately 118 kDa, suggesting a 1:2 protein complex of Bcl3:p50. Whereas, the minor elution peak (peak 1) has a

molecular weight of 259 kDa, suggesting a higher order complex. ImageJ was used to analyse the band intensities present on the SDS-PAGE. Results confirmed the minor peak (peak 1) contained a higher order complex with Bcl3:p50 in a 1:3 or 1:4 ratio respectively. Whereas, the main elution peak (peak 2) suggests Bcl3 is bound to a p50 homodimer.

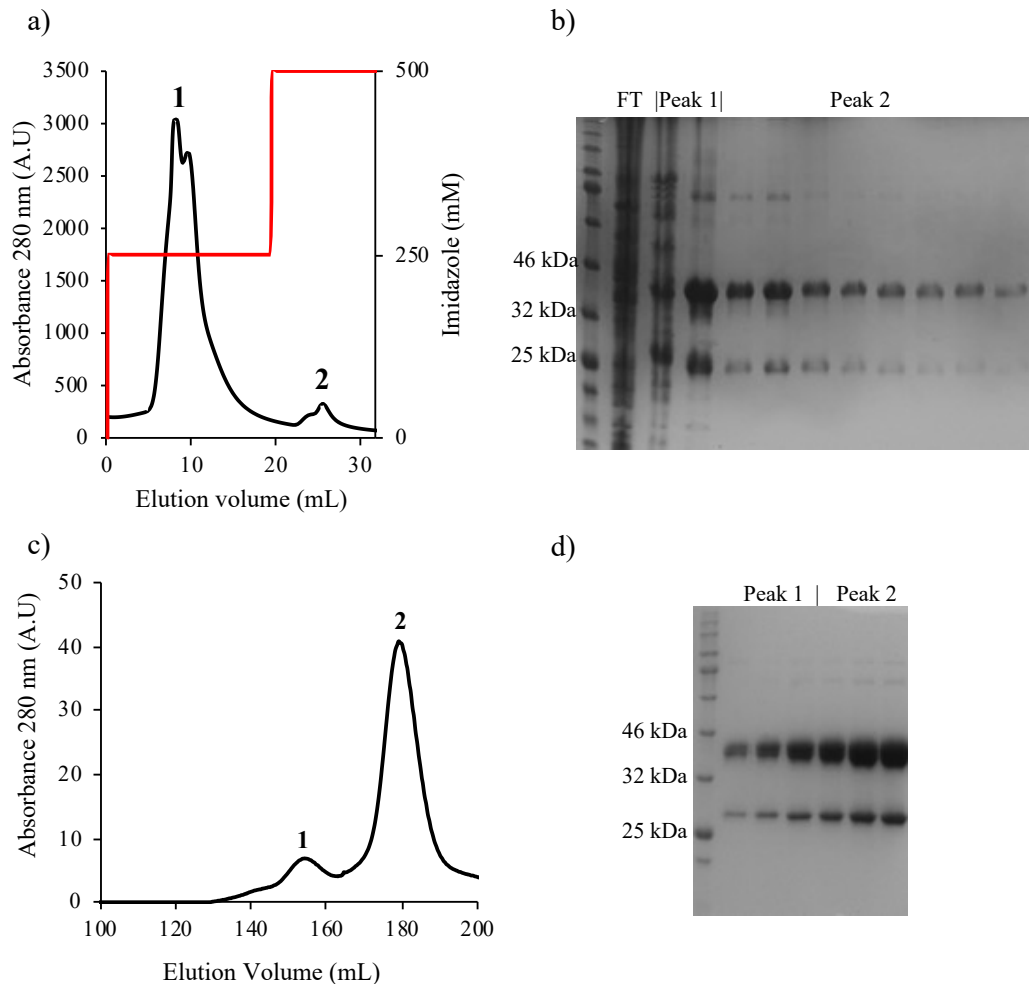


Figure 15. Purification of Bcl3-p50 complex by IMAC and SEC. a) Elution profile from Nickel affinity chromatography. Absorbance monitoring at 280 nm shown in black, imidazole concentration shown in red. b) SDS-PAGE gel following nickel affinity chromatography. c) Elution profile from size exclusion chromatography. d) SDS-PAGE gel following size exclusion.

3.2.1.2 Determining stoichiometry of Bcl3-p50 complex by SEC-DLS

It was important to confirm the stoichiometry of the protein complex and determine whether our results matched previous evidence of suggested 1:2 or 2:2 ratios (Bcl3 to p50)(Wulczyn et al. 1992). To do this, SEC-DLS was performed

according to section 2.5.2. The protein complex, concentrated to ~ 2 mg/mL was applied to the SEC column, recording scattering curves of the eluent. The absorbance and refractive index show a shoulder and minor peak eluting at ~ 12 mL and a major peak eluting at ~ 14 mL (Figure 16). Strong right-angled scattering is present for both peaks, indicating they are protein. The major peak (peak 1) has a hydrodynamic radius of 4.95 nm which equated to a MW of 114 kDa, suggesting 1:2 stoichiometry of Bcl3 to p50 (Figure 16). Peak 2 had a hydrodynamic radius of 6.05 nm with a MW of ~ 200 kDa, likely to be a 1:4 ratio. The total protein recovered from the column was 1.6 mg/ml suggesting an 80 % recovery. Of this, majority (92%) was in a 1:2 stoichiometric ratio with the remainder (8%) in a higher order complex (possibly 1:4).

These results match those suggested by SDS-PAGE analysis and confirm the main elution peak likely contains Bcl3 bound to a p50 homodimer. The higher order complex where two Bcl3 subunits are bound to a p50 homodimer, may therefore require additional factors; such as DNA association, or phosphorylated residues.

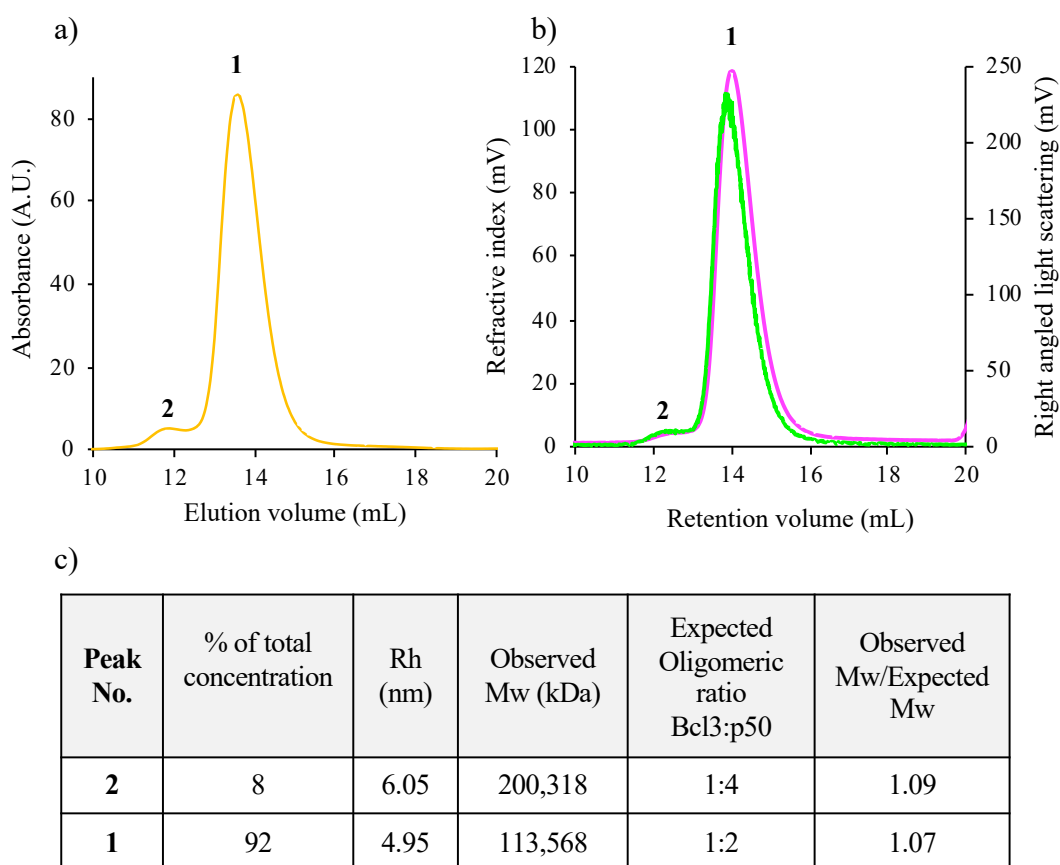


Figure 16. Analysis by SEC-DLS. a) Absorbance trace at 280 nm shows protein elution peaks at 12 and 14 mL. b) Plot of refractive index (RI)(pink line) and right angled light scattering (RALS)(green line) against elution volume. C) Peak table highlighting observed molecular weight and the oligomeric ratio. DLS was calibrated to Bovine Serum Albumin.

3.2.1.3 Thermal melting assay

Multiple techniques have been used to determine protein stability, most often by chemical or thermal denaturation (Nicholson and Scholtz 1996; Mei et al. 2005; Sinha and Surolia 2005). While a lot of these studies have focused on smaller monomeric proteins, a large number of proteins exist as multimers and so it has become necessary to monitor the role of subunit interactions for increasing the stability of oligomeric proteins (Sinha et al. 2005). Previous reports have suggested p53 homodimers change their conformation upon binding to either non-specific or specific DNA sequences (Hay and Nicholson 1993)(Matthews et al. 1995). A thermal melting assay was therefore performed according to section 2.5.3 to determine protein stability and the possible effects of DNA binding on the Bcl3-p53 complex. For this, SYPRO orange dye is added to bind to hydrophobic residues as the protein undergoes thermal denaturation (Figure 17). The fluorescence emission of the dye is monitored to determine the melting temperature, or T_m , of the protein.

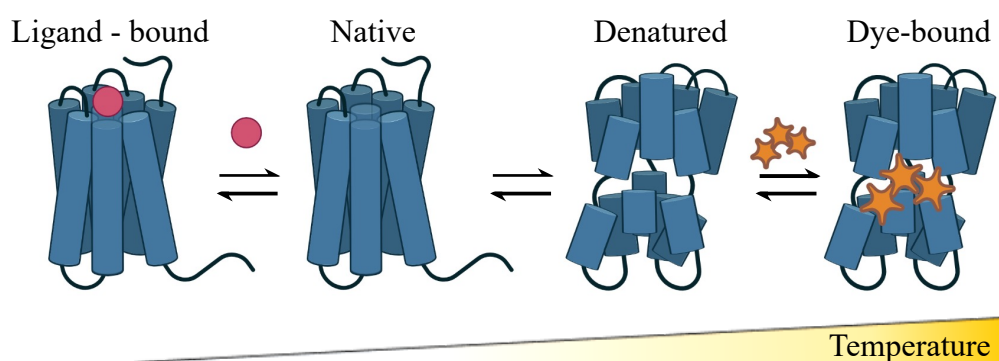


Figure 17. Process flow diagram of thermal melting assay. Purified protein bound to ligand slowly undergoes thermal denaturation. SYPRO orange interacts with the exposed hydrophobic residues increasing the fluorescence emission. Figure adapted from Huynh and Partch (2015).

A large dataset was collected which required automated processing prior to analysis. Post peak quenching was removed and results were fit to a non-linear curve using the Boltzmann Equation, as suggested by Huynh and Partch (2015). Initial analysis of the melting curves shows a typical thermal denaturation profile of recombinant proteins (Figure 18). The low fluorescence signal at room temperature indicates well-folded proteins. The SYPRO orange fluorescence gradually increases with temperature starting at ~ 40 °C to produce a sigmoidal curve that represents protein unfolding.

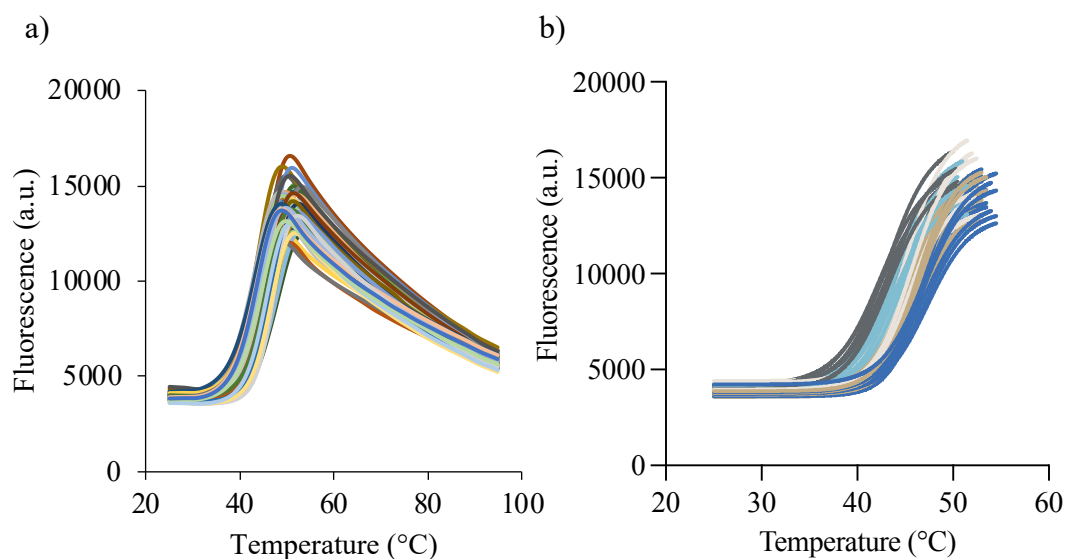


Figure 18. Thermal melting assay data for Bcl3-p50 complex. a) Raw data showing thermal denaturation profile of protein complex. Fluorescence emission increases with increasing temperature, giving rise to the sigmoidal curve representing unfolding of the protein. b) Automated processing of thermal denaturation curves truncates the dataset to remove post-peak quenching. The resulting sigmoidal curve is fit to a non-linear Boltzmann Equation.

To determine the optimum conditions for protein stability, the datasets were then separated into appropriate groups. The melting curves representing protein concentrations of 2.0, 2.5 and 3.0 mg/mL were very similar. It was likely that the concentration range was not great enough to demonstrate the effects of protein aggregation. However, through previous experience handling the complex, concentrations greater than 3.0 mg/mL had often precipitated. This was therefore chosen as the maximum concentration to sample for this experiment. These results however, have confirmed that this concentration is appropriate for use. Subsequent data analysis was therefore performed at a protein concentration of 3 mg/mL.

Where values were closely matched, data was plotted as a bar graph to ease visualization of calculated T_m 's. Results showed that as pH decreased, the T_m of the proteins increased from 43 to 47 °C (Figure 19 a,b). The addition of DNA also slightly increased the T_m of the protein complex, as observed across all results (Figure 19 b).

Overall, these results have confirmed that the Bcl3-p50 complex can be used at a protein concentration of 3.0 mg/mL in a suitable buffer, at pH 7.0. DNA can be added, although the binding and subsequent structural effects on the protein complex are not clear.

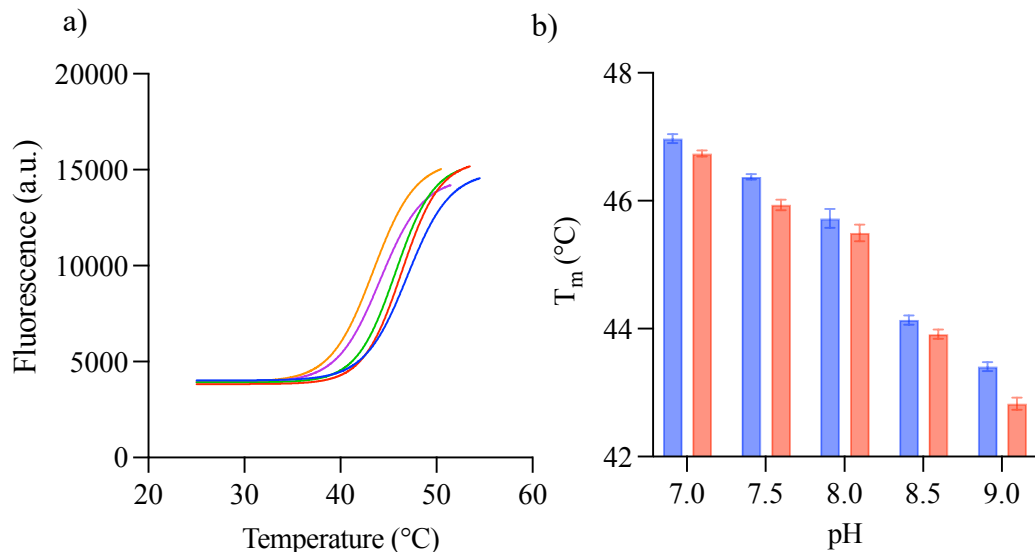


Figure 19. Thermal denaturation of Bcl3-p50 complex. Protein complex at 3 mg/ml. a) Effects of varying pH on the thermal denaturation of the protein complex. Data shown as sigmoidal curve where pH 7.0, 7.5, 8.0, 8.5 and 9.0 are shown in blue, red, green, purple and orange respectively. b) Effects of pH and addition of DNA on the melting temperature of the protein complex. Protein complex incubated in 1:1 molar ratio with DNA shown in blue. Protein complex without DNA is shown in red. Data represented as bar graph with SEM error bars.

3.2.1.4 Crystallography

Many protein structures are available on the Protein Data Bank (PDB) for the IκB and NF-κB family of proteins, both individually and complexed (<https://www.rcsb.org/>). The crystal structures of Bcl3 ANK domain (PDB: 1K1A) and p50 homodimer bound to DNA (PDB: 1NFK), have also been solved (Ghosh et al. 1995; Michel et al. 2001). The models of Bcl3-p50 complex provide a good indication of the overall protein-protein interactions, however important details regarding complex assembly and key residues may be missed. The crystal structure of the Bcl3-p50 complex therefore needs to be determined.

Multiple crystal trials were set up as described in section 2.5.4.1. A wide variety of conditions were used to try and discover the best possible crystallisation screen. Other factors were also sampled including; protein concentration, addition of DNA, ratio of protein to crystallisation buffer, and the incubation temperature. After 8 weeks, a crystal was visualised in well A1 of JCSG which contained Bcl3-p50 complex (-DNA) in a 1:1 protein-buffer ratio, grown at 20 °C (Figure 20 a). The crystal was harvested and sent to Diamond Light Source for x-ray diffraction. The

results did show diffraction with diffraction spots clearly visible within the inner resolution rings (Figure 20b). It was therefore likely that the sample harvested was a protein crystal. However, the diffraction data recorded was of poor quality, and not sufficient for structure solution so unfortunately was discarded.

After multiple attempts, considering and testing all possible varying factors, protein crystals of Bcl3-p50 complex did not grow. It was therefore concluded, that for reasons unknown, the protein complex will not crystallise and thus a crystal structure could unfortunately not be solved.

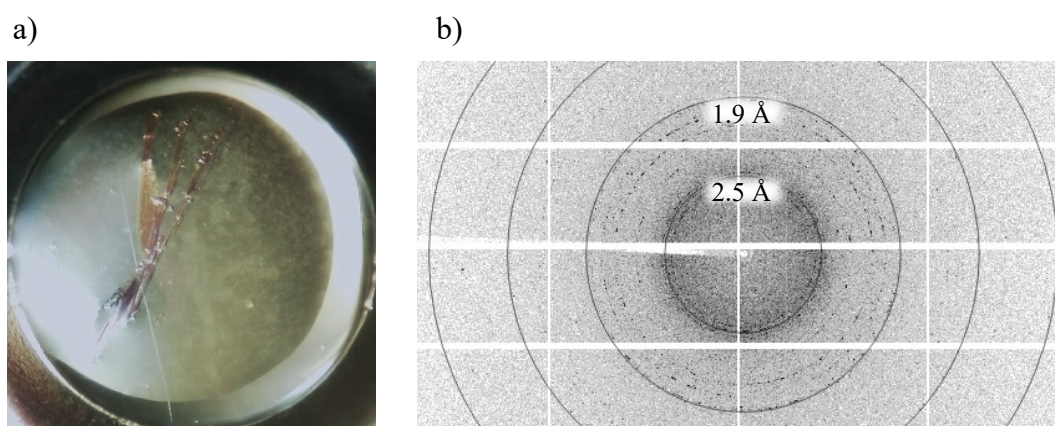


Figure 20. Bcl3-p50 crystal and diffraction pattern. a) Crystal in A1 JCSG well prior to harvesting. b) X-ray diffraction pattern.

3.2.2 Production of FP-tagged Bcl3-p50 complex

PPIs have been studied via multiple different methods including those *in vitro*, *in vivo* and *in silico* (Rao et al. 2014). One popular approach used to investigate PPIs is FRET. FRET is a distance-dependent photophysical process where energy is transferred from one excited molecular fluorophore (the donor) to a second fluorophore (the acceptor) by means of intermolecular long-range dipole-dipole coupling mechanism (Sekar and Periasamy 2003). Due to its sensitivity to distance, FRET can be used to investigate biomolecular interactions by monitoring changes in fluorescence. The transfer of energy leads to a reduction in the donors fluorescence intensity and the subsequent increase in the acceptors emission intensity.

While there are many factors that influence FRET, there are a few important conditions to consider. The absorption or excitation spectrum of the acceptor must overlap (>30%) the fluorescence emission spectrum of the donor (Sekar and Periasamy 2003). The degree of overlap is commonly referred to as the spectral overlap integral (J). There should also still be reasonable separation in emission spectra between donor and acceptor, to allow independent measurement of the fluorescence of each fluorophore. However, one of the most critical elements necessary for FRET is the close proximity of pairs; where the donor and acceptor molecules must be within 10-100 Å of one another (Shrestha et al. 2015). The distance at which energy transfer is 50% efficient is referred to as Förster radius (R_0). Note that R_0 itself is dependent on a number of factors, as determined in Equation 4.

Equation 4.
$$R_0 = 0.211 \sqrt[6]{k^2 n^{-4} Q_D J(\lambda)}$$

where R_0 is the Förster radius, k^2 is the dipole orientation factor, n is the solvent refractive index, Q_D is the quantum yield of the donor, and $J(\lambda)$ is the overlap integral between the donor emission and the acceptor molar absorbance. Often, k^2 is arbitrarily set as 0.667 to reflect two randomly orientated chromophores (Pope et al. 2021). The refractive index (n) was determined as 1.4 when accounting for a combined protein-water environment. With a known distance between FRET chromophores (r), the energy transfer efficiency (E) can be then determined with Equation 5.

Equation 5.
$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

3.2.2.1 Design of FP-tagged Bcl3 and p50 for FRET experiments

For this study, FRET was used to detect binding between Bcl3 and p50. However, before FRET studies could begin, FP-tagged versions of Bcl3 and p50 had to be designed. All factors mentioned above were considered when selecting fluorescent proteins (FPs) including; the spectral overlap, the QY of the donor, the extinction coefficient of the acceptor (Bajar et al. 2016). Based on this, sfGFP and

Venus were chosen as they possess the desired characteristics of suitable FRET partners. There was sufficient overlap between the excitation spectrum of the acceptor and the fluorescence emission spectrum of the donor; where the overlap integral (J) was calculated as $3.03 \times 10^{14} \text{ M}^{-1} \text{ cm}^{-1} \text{ nm}^4$, according to FPbase (Table 9). The quantum yield of sfGFP is also relatively high, of 0.75. Thus, by supplementing these values into Equation 4, the Förster radius (R_0), distance at which energy transfer is 50% efficient, was calculated as 57.14 Å (using the arbitrarily set k^2 of 0.667 and n of 1.4). The donor-to-acceptor stoichiometry of suggested 1:2 or 2:2 ratio for Bcl3-p50 was also considered. It was therefore decided that sfGFP (the donor) and Venus (the acceptor) would be fused to Bcl3 and p50 respectively.

Table 9. Properties of sfGFP and Venus as FRET partners as determined by FPbase (<https://www.fpbase.org>). R_0 calculated using the arbitrarily set k^2 of 0.667. [a] Quantum yield determined experimentally.

Donor	Acceptor	J ($\text{M}^{-1} \text{ cm}^{-1} \text{ nm}^4$)	QY ^[a] (Donor)	R_0 (Å) $k^2 = 0.667$
sfGFP	Venus	3.03×10^{14}	0.75	57.14

As a distance sensitive approach, it was important to determine which terminus of Bcl3 and p50 would be suitable for FP attachments. The experimental data has suggested a stoichiometric ratio of 1:2 where Bcl3 is bound to a p50 homodimer. Thus, the 2:2 (Bcl3:p50) model determined by Prof Richard Clarksons group (later published by Soukupov et al. (2021)) was taken and amended accordingly to remove one Bcl3 subunit. According to this model, the termini ends of Bcl3 and p50 face away from one another. Placement of FP's on the N-terminus of Bcl3 and p50 would therefore position them at opposite ends of the protein complex, at least 90 Å apart (Figure 21a). This distance is nearing the maximum limit of ~100 Å for FRET to occur (Sekar and Periasamy 2003). Complications would similarly arise if FPs were fused to the C-terminus, as sfGFP would still be at a far distance apart from Venus (Figure 21b). Given this, the N- terminus of Bcl3 and the C- terminus of p50 were deemed most suitable for FP fusion (Figure 21c). Going forward, these proteins will be referred to as G-Bcl3 and p50-V in accordance with FP site attachment.

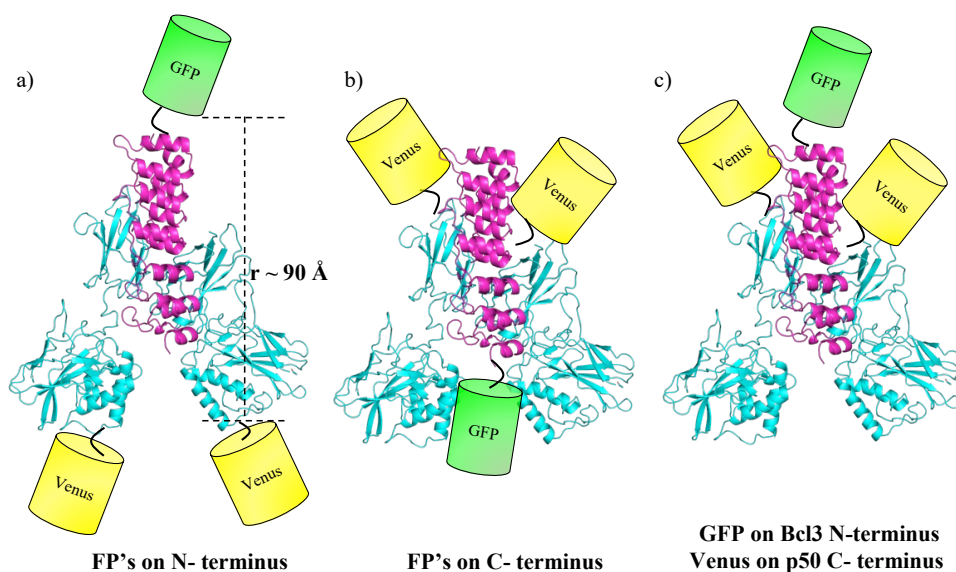


Figure 21. Model to show attachment of sfGFP and Venus to terminus ends of Bcl3 and p50 respectively. Cartoon representation of Bcl3 shown in pink and p50 shown in cyan. GFP and Venus represented by cylindrical barrels. a) FP's placed on the N-terminus of Bcl3 and p50. b) FP's fused to the C-terminus of Bcl3 and p50. c) GFP on Bcl3 N-terminus, Venus on p50 C-terminus.

To confirm appropriate placement and proximity between FPs, computational approaches were used to create a model structure of G-Bcl3 - p50-V. The 1:2 Bcl3-p50 model (amended from Prof Richard Clarksons group 2:2 model) was used as a template for FP attachment in PyMOL (detailed in section 2.6). The addition of a 'linker' between p50-V was deemed necessary, to ensure sufficient space and appropriate positioning of Venus. It was noted however, that 20 amino acids were absent from p50's C-terminus, due to the inherent poor resolution of terminus ends during structure solution. As already part of the p50 gene, these residues could be sufficient for use as the extendable 'link' to increase distance between p50-V.

To model p50's missing residues (353 – 372), the amino acid sequence was submitted to the online server, Robetta (<https://robetta.bakerlab.org/>). The Rosetta *ab initio* (RosettaAB) application was used to perform a two-step relaxation; a course-grained fragment-based search and an all-atom refinement. The results provided the top 5 models with a 0.76 *ab initio* confidence score, which suggests a reasonable topological similarity between protein structures. Alignment of models revealed residues 356 ~ 364 of p50 would likely form a helix shape (Figure 22). This seemed a suitable structure for use as a 'linker' as it is sufficient to lengthen the distance

between p50-V for the two to act independently, whilst not too long as to ensure FPs remain within a suitable distance to one another. It was also decided to include a ‘GSSH’ sequence to aid flexibility and allow angular rotation at the C-terminus end of the linker. This connecting sequence was therefore chosen and used to attach Venus to p50 (Figure 22).

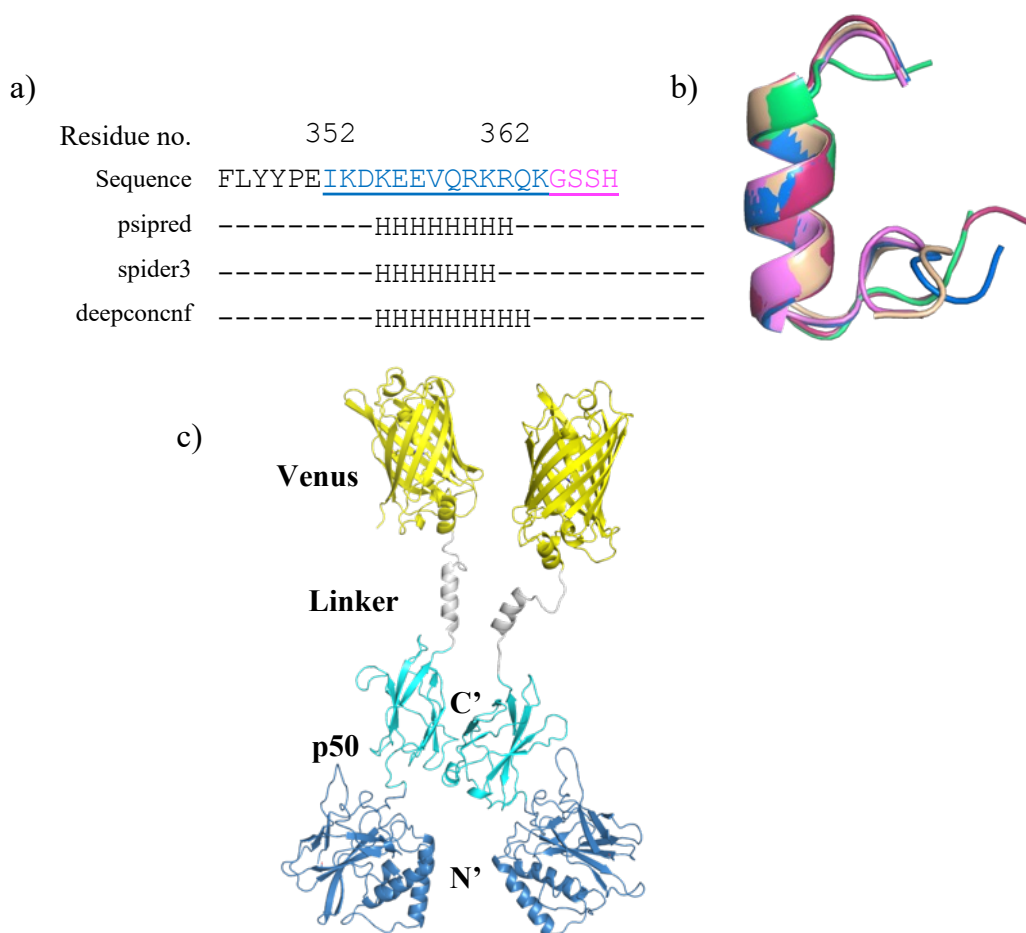


Figure 22. Structure prediction results from Robetta for p50-V linker. Residues highlighted in blue represent predicted structure of additional residues 353-365 of p50. Additional ‘GSSH’ sequence shown in magenta. A) Amino acid alignment showing expected helical structure as denoted as ‘H’. B) Cartoon representation showing alignment of 5 models. C) Cartoon representation of p50-V.

To refine the structure of G-Bcl3-p50-V, the model was subjected to a 100 ns molecular dynamic (MD) simulation. As well as the G-Bcl3-p50-V protein complex, both G-Bcl3 and p50-V homodimer non-complexed were also separately subjected to a 100 ns molecular dynamic (MD) simulation (Figure 23). This allowed key differences in protein dynamics to be identified between proteins when complexed and non-complexed. The MD simulations were performed in duplicate according to section 2.6.3 with the MD package GROMACS. Further parametrisation was

required due to the presence of non-standard amino acids (the chromophores of Venus and GFP) in the structure. For this, the atomic charges, bond lengths and angles were computed using ACPYPE and input into the Amber forcefield (amber99sb.ff) as described in section 2.6.2. The structure was then subjected to an initial energy minimisation step before submission to the 100 ns MD simulation.

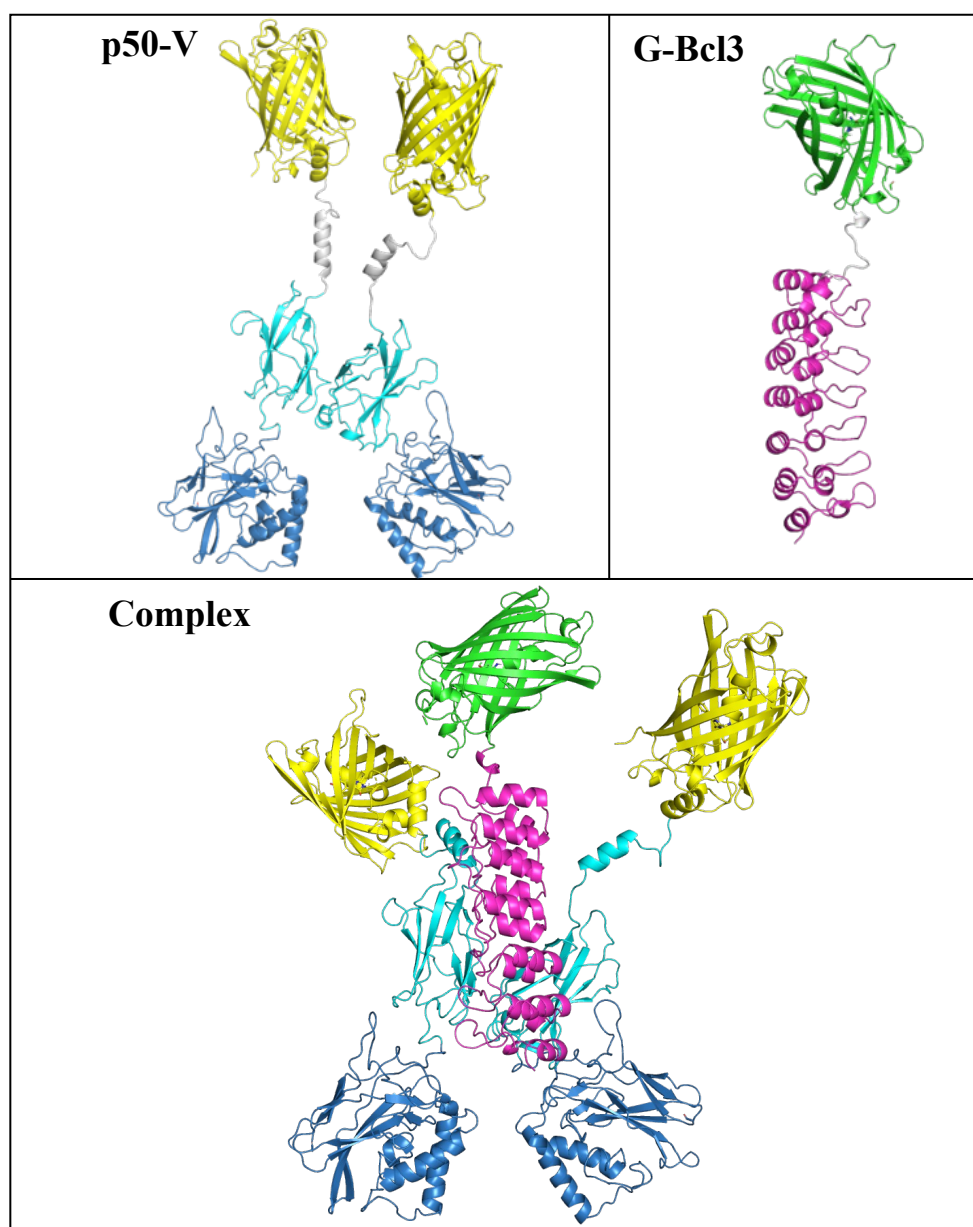


Figure 23. Starting models of 100 ns MD simulation. Non-complexed; G-Bcl3 and p50-homodimer. Complexed; G-Bcl3 - p50-V

The root mean square fluctuation (RMSF) is a measure of the displacement of a particular atom, or group of atoms, relative to a reference structure, averaged over the number of atoms. An area of the structure with high RMSF values indicates a region with high mobility. Firstly, the RMSF was calculated for the C_{α} of G-Bcl3, both when complexed and non-complexed. This will determine the effect of complexation on the flexibility of both proteins. As separate domains, the RMSF was calculated separately for sfGFP and Bcl3. The results showed that across all residues of sfGFP, the C_{α} RMSF was approximately 0.05 nm when complexed and non-complexed (Figure 24a). This suggests that the mobility of sfGFP was not affected by the binding of Bcl3 to p50. On the contrary, complexation of Bcl3 and p50 resulted in the C_{α} RMSF of Bcl3 to decrease from approximately 0.3 nm to 0.1 nm (Figure 24). This suggests that when bound to p50, the mobility of Bcl3 is reduced.

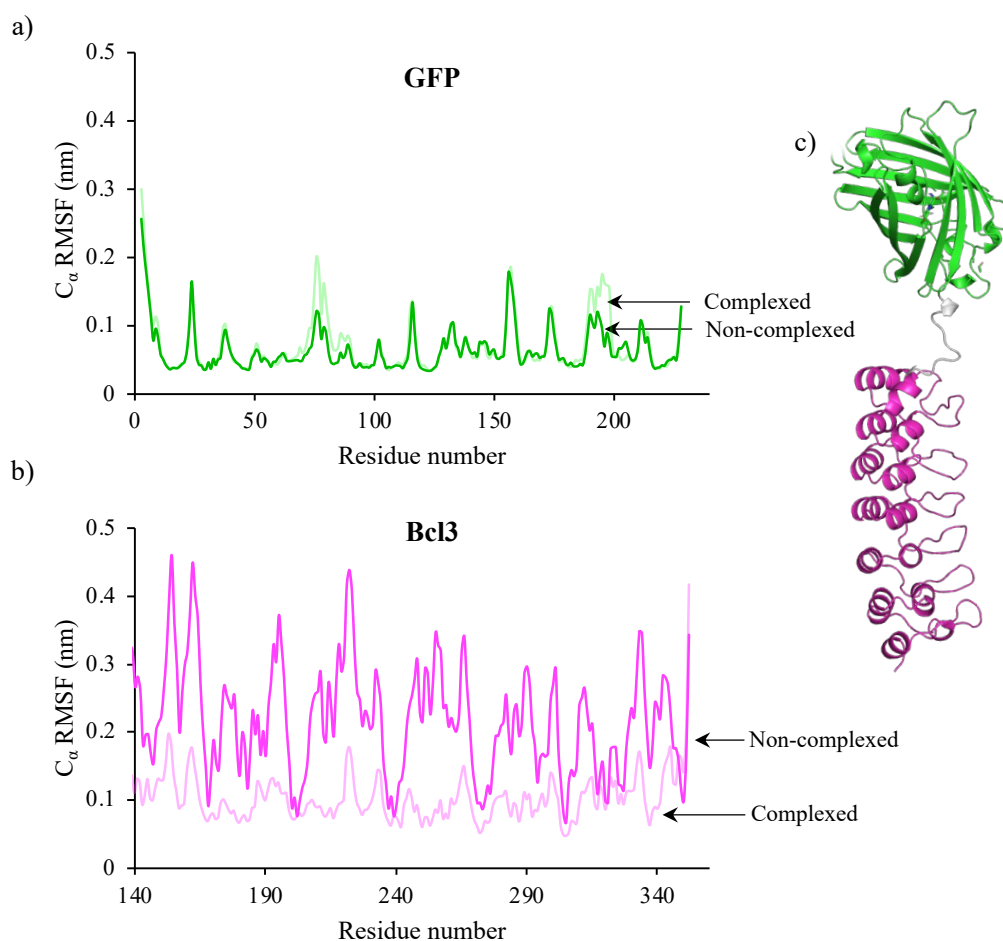


Figure 24. The C_{α} RMSF values of G-Bcl3 when complexed and non-complexed. A) RMSF calculated for residues 1-228 of sfGFP. Dark green, non-complexed; light green, complexed. B) RMSF calculated for residues 131-352 of Bcl3. Dark magenta, Non-complexed; light magenta, complexed. C) Cartoon representation of G-Bcl3.

Next, the C_{α} RMSF was determined for p50-V homodimer when complexed and non-complexed to Bcl3. Analysis was performed for the separate domains; N- and C-terminus of p50, and Venus. The results showed that for both terminus of p50, the C_{α} RMSF remained at approximately 0.1 nm (Figure 13). The results were also similar for both the N' or C' terminus of p50 when complexed, and non-complexed. This suggests that mobility of p50 homodimer was largely unaffected by binding to Bcl3. This is probable given that the dimerisation domain of C' terminus is most likely already stabilised by the p50 homodimer interface and the N' terminus of p50 has few contacts to Bcl3, so would unlikely be affected by complexation. The C_{α} RMSF of Venus was recorded at 0.05 nm, suggestive of a stable protein.

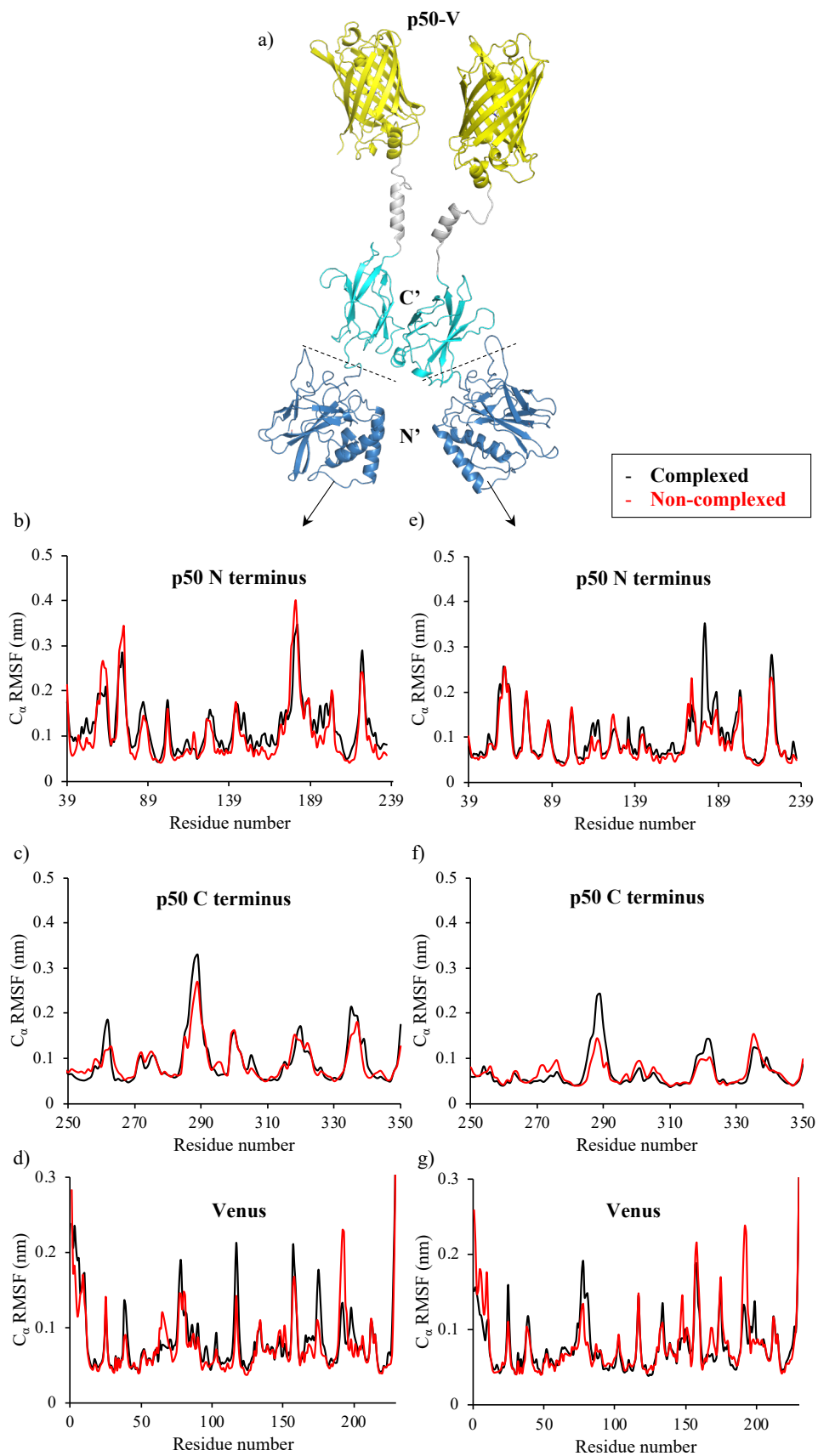


Figure 25. C_α RMSF of p50-V. A) Cartoon representation of p50-V showing p50 N-terminus (dark blue), C-terminus (light blue), and Venus (yellow). b-d) C_α RMSF of p50 molecule 1 of homodimer. Black line represents the protein when complexed to Bcl3. Red line is non-complexed. E-g) C_α RMSF of p50 molecule 2 of homodimer.

The distance between FP chromophores pairs within G-Bcl3 – p50-V complex was then calculated. The results vary across the two MD runs and for distances between GFP and each of the Venus molecules. For the first Venus molecule; MD run 1 shows a chromophore distance to sfGFP of ~ 60 Å for 30 ns of the simulation and then a subsequent decrease to approximately 35 Å for the remainder of the simulation (Figure 26 a,b). For MD run 2, the chromophore distances remain between 60 – 80 Å over the 100 ns. For the second Venus molecule, the chromophore distances to sfGFP were between 40 – 60 Å for the duration of the simulation (Figure 26 a,c). Thus, these results suggest that the distance between sfGFP and Venus, when complexed to Bcl3 and p50, can vary considerably. We can therefore predict that the chromophores will for a portion of time exceed the calculated R_0 of 57.14 Å. This will subsequently affect the FRET efficiency; where a interchromophore distance above 57.14 Å will result in a FRET efficiency below 50 %. Thus, by supplementing the hypothesised interchromophore distance (r) into Equation 5, and using the calculated R_0 of 57.14 Å, the expected FRET efficiency can be predicted. For an r of 35 – 40 Å, the FRET efficiency is expected to be 89 - 95 %. Whereas an r of 60 or 80 Å is expected to give a FRET efficiency of 12 – 43 %. These results highlight the impact of the interchromophore distance on FRET. We may therefore expect the FRET efficiency to vary between these values when Bcl3 is bound to p50.

It should also be mentioned here that MD simulations can be significantly influenced by the initial submitted model. The exact starting position/angle of FPs in relation to Bcl3-p50 could not be known. Whilst the Bcl3-p50 FP-tagged structure was used to model the overall complex, the exact spatial configurations of FPs will be viewed with caution.

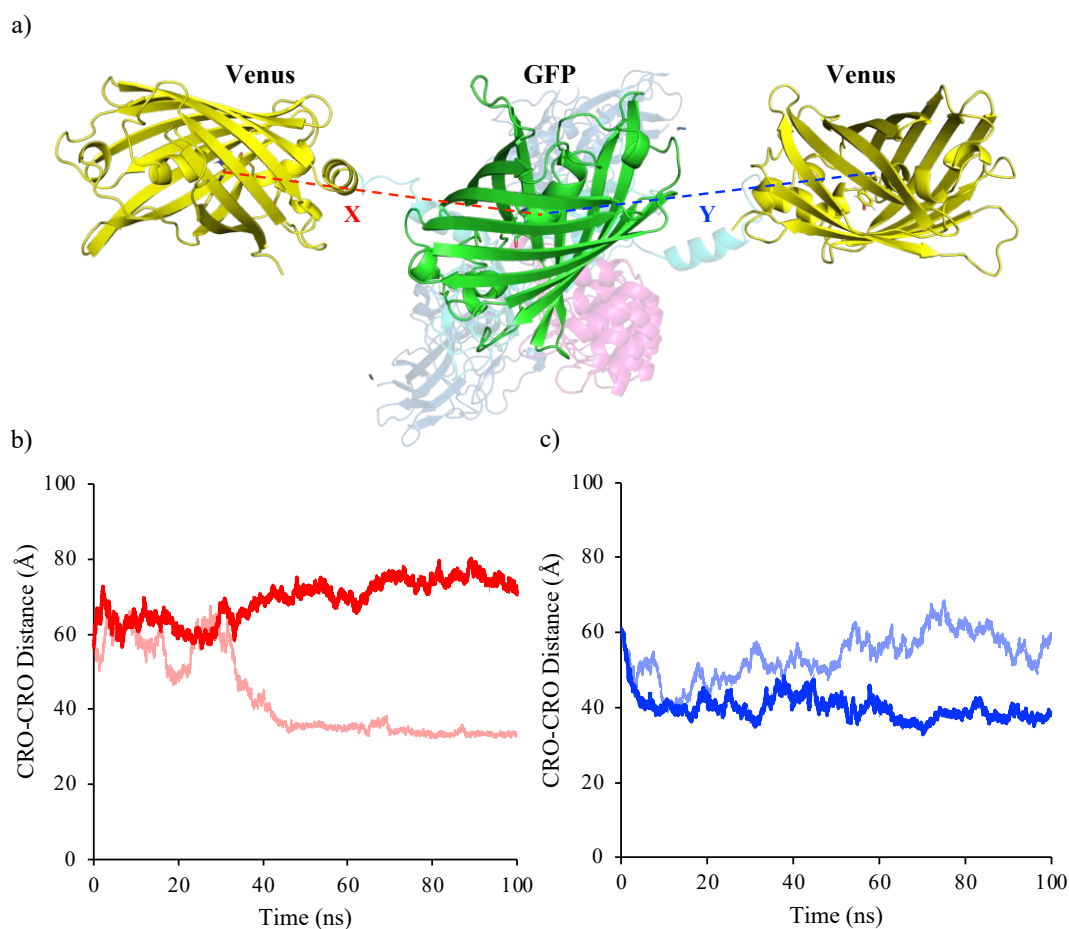


Figure 26. Top view of G-Bcl3 – p50-V starting model prior to 100 ns MD simulation. a) GFP and Venus attached to the N- and C-terminus of Bcl3 and p50 respectively. Bcl3 is coloured purple, p50 is cyan, GFP is green and Venus is yellow. Distance between FP chromophores is highlighted. b) Average distance between GFP and Venus molecule 1 labelled in red (X). Lighter colour is MD run 1 and darker colour is MD run 2. c) Average distance between GFP and Venus molecule 2 labelled in blue (Y). MD run 1 and darker colour is MD run 2.

3.2.2.2 Expression and purification of G-Bcl3 and p50-V

Prior to FRET studies, it is important to characterise and determine the key parameters (e.g. molar absorbance, brightness) of G-Bcl3 and p50-V proteins separately, to confirm that fusion of the FP to the partner protein did not have an adverse effect on their fluorescence properties. Plasmids encoding G-Bcl3 and p50-V were constructed by Dr Ismail Taban and used to recombinantly generate protein in *E. coli*. Proteins were expressed overnight at 25 and 28 °C for Bcl3 and p50 respectively. Overexpression was confirmed by the presence of coloured cells in the growth media; green for G-Bcl3 and yellow for p50-V. Cells were pelleted and lysed

using French press. The proteins contained a hexahistidine tag at the N-terminal (GSHHHHHH) which was used for purification by nickel affinity chromatography as described in 2.3.3. Both proteins eluted using ~ 250 mM imidazole and fractions were analysed by SDS-PAGE. Bands of 66 and 52 kDa confirmed presence of p50-V and G-Bcl3 respectively (Figure 27).

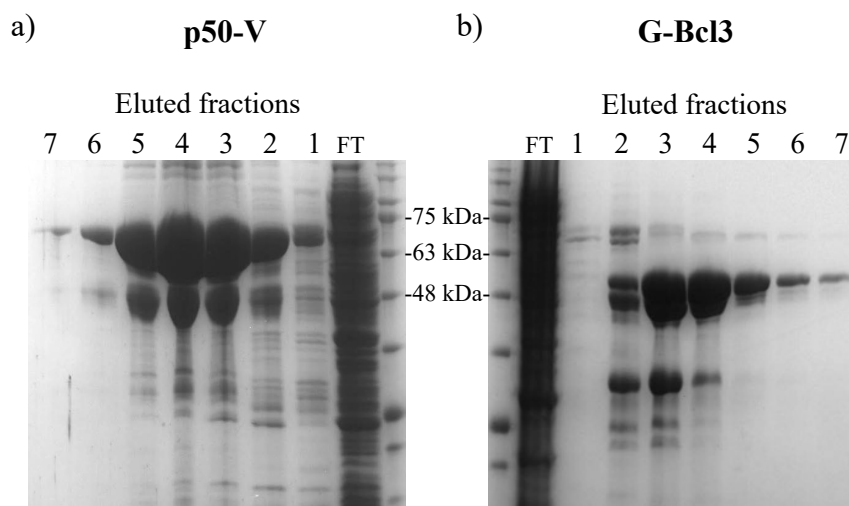


Figure 27. SDS-PAGE analysis of protein fractions that eluted following nickel affinity chromatography. Fractions labelled in order of their collection. FT – flow through. a) p50-V b) G-Bcl3.

Size exclusion chromatography was performed to remove free imidazole and any remaining contaminants (section 2.3.4). Major elution peaks were observed at ~ 60 and 70 mL for p50-V and G-Bcl3 respectively (Figure 28, a,d). Based on the calibrated SEC column and molecular weight estimation (described in section 2.4.6), the main peak (peak 2) for p50-V eluting at ~ 65 mL has a molecular weight of ~134 kDa. This suggests elution of a p50-V homodimer. For G-Bcl3, the minor elution peak at ~ 64 mL (peak 1) has a molecular weight of ~150 kDa, potentially a G-Bcl3 trimer. However, the main elution peak (peak 2) at ~ 74 mL has a molecular weight of 56 kDa, indicative of G-Bcl3 in a monomeric state. The coloured fractions were analysed by SDS-PAGE analysis, confirming elution of purified p50-V (Figure 28 b,c) and G-Bcl3 (Figure 28 e,f).

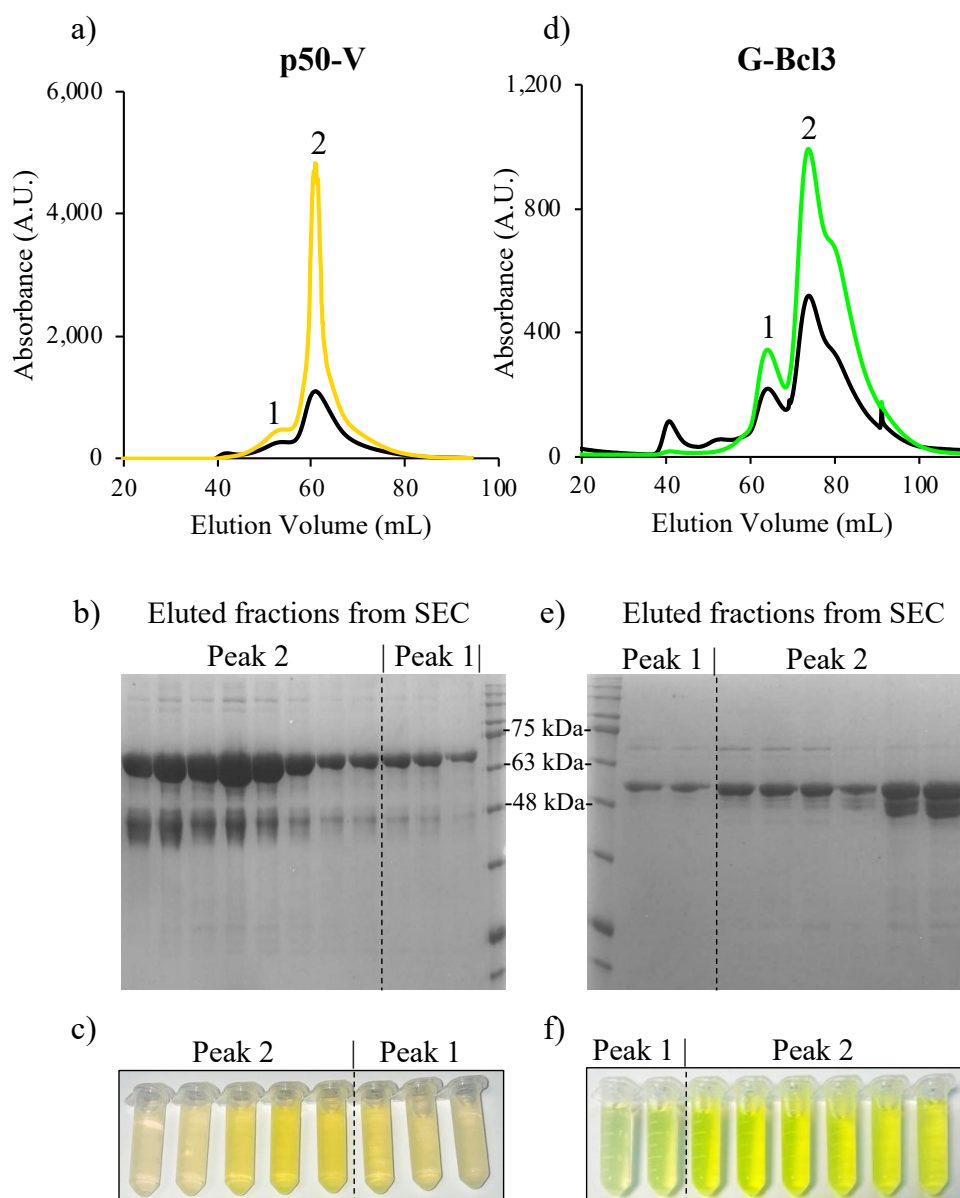


Figure 28. Purification of p50-V and G-Bcl3. Absorbance monitored at 280, 485 and 510 nm shown in black, green and yellow respectively. a) Elution profile from size exclusion of p50-V. b) SDS-PAGE analysis of p50-V. c) Fractions eluted from SEC of p50-V coloured yellow. d) Elution profile from size exclusion of G-Bcl3. e) SDS-PAGE analysis of G-Bcl3. f) Fractions eluted from SEC of G-Bcl3 coloured green.

The absorbance and fluorescence spectra for G-Bcl3 and p50-V were recorded to determine the expected spectral characteristics of the fused FP's. G-Bcl3 had a molar extinction coefficient of $49,000 \text{ M}^{-1}\text{cm}^{-1}$ and maximum excitation and emission wavelengths of 485 and 511 nm respectively (Figure 29). Whereas p50-V had a higher extinction coefficient of $92,000 \text{ M}^{-1}\text{cm}^{-1}$ and λ_{max} and λ_{em} of 515 and 528 nm respectively. These results matched the spectral characteristics determined for free sfGFP and Venus proteins proving that the fusion had no effect on the FPs intrinsic properties (Figure 29).

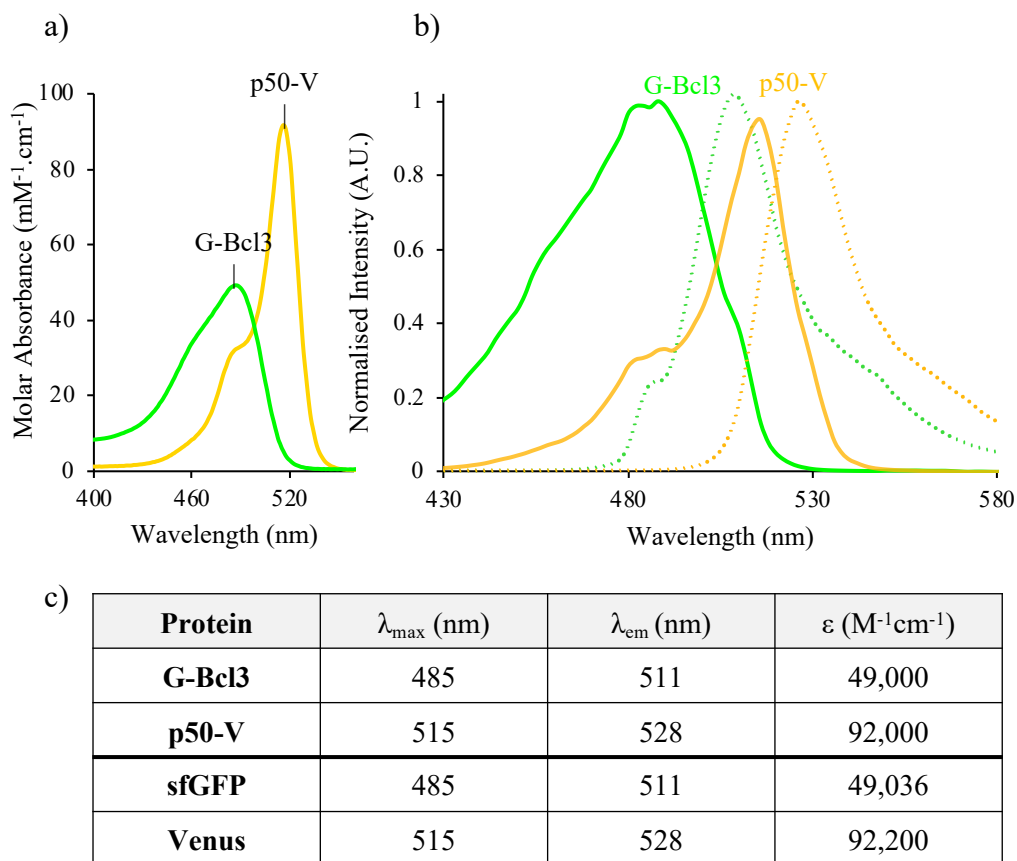


Figure 29. Spectral characterisation of G-Bcl3 and p50-V. a) Absorbance spectra for G-Bcl3 (Green) and p50-V (yellow). b) Fluorescence spectra for G-Bcl3 (Green) and p50-V (yellow). Excitation spectra (solid lines) recorded at a fixed wavelength according to λ_{\max} emission. Emission spectra (dotted lines) recorded upon excitation at λ_{\max} . c) Table listing spectral characteristics of G-Bcl3, p50-V and unbound sfGFP and Venus.

To confirm the suitability of sfGFP and Venus as FRET partners, the fluorescence spectra of the FPs were then analysed collectively. The results highlight a significant overlap between the excitation spectra of the donor (sfGFP) and acceptor (Venus) (Figure 29 b). This could cause a potential problem as there would potentially be cross excitation issues (i.e. excitation of GFP [donor] could also result in excitation of the acceptor so generating a false positive in terms of interactions). It was important that excitation selectively targeted the donor fluorophore so that emission would be solely based on the acceptor's emission, as a direct measure of FRET and thus the PPI. Excitation at fixed wavelengths shorter than λ_{\max} of G-Bcl3 were sampled (400, 410, 420 and 430 nm) to try and reduce the contamination signal from p50-V. The spectra highlighted that shorter excitation wavelengths led to a

lower emission signal for both proteins (Figure 30). However, the signal output observed for G-Bcl3 was much greater than that for p50-V; where a ~12-fold difference was recorded when excited at 420 nm. This is expected since λ_{max} of GFP (485 nm) is blue shifted compared to that for Venus (511 nm). A fixed excitation wavelength of 420 nm was therefore selected as it allowed selective excitation of G-Bcl3 (donor) whilst maintaining a reduced signal from p50-V (acceptor). Thus, emission recorded for p50-V will be a direct measure of FRET (and PPI).

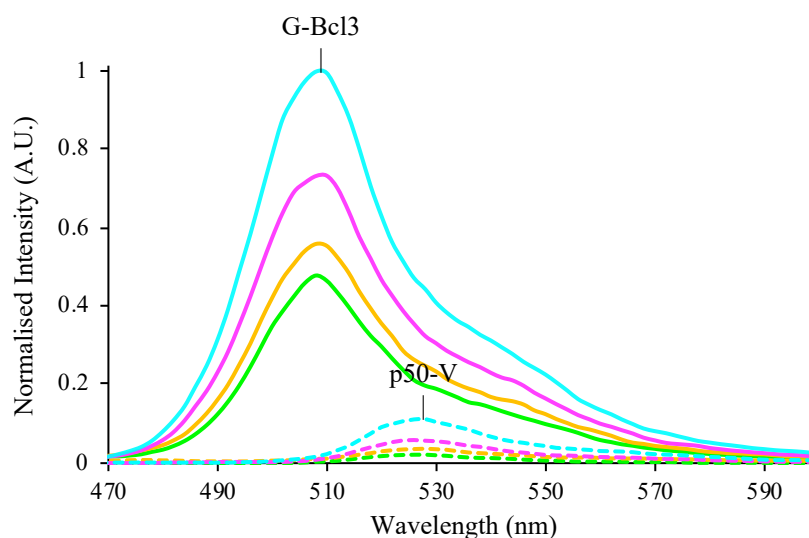


Figure 30. Fluorescence spectra of G-Bcl3 and p50-V proteins. Spectra for G-Bcl3 shown in solid lines and for p50-V in dotted lines. Proteins excited at a wavelength of 400, 410, 420 and 440 nm shown in green, orange, pink and blue respectively. Data are normalised to the maximum emission for G-Bcl3 when excited at 440 nm.

3.2.2.3 Stability of G-Bcl3 and p50-V

The stability of G-Bcl3 and p50-V proteins was determined by measuring fluorescence over time at two different temperatures: 20 and 37 °C. These temperatures were selected to test their stability for experiments performed under *in vitro* (20 °C) and *in vivo* (37 °C) conditions. Fluorescence spectra were recorded at different time intervals over 48 hours, upon incubation at either 20 or 37 °C. For G-Bcl3, the main emission peak remained at 510 nm at both experimental temperatures and over 48 hours (Figure 31 a,b). However, the peak intensity decreases after 48 hours by 17 and 20 % when incubated at 20 and 37 °C respectively. The higher temperature might have caused protein aggregation, although the difference in results is deemed minimal. The same is true for p50-V; the main emission peak

remains at 530 nm across results. Peak intensity decreased by 16 and 17 % after 48 hours incubation at 20 and 37 °C.

Although the peak intensities did slightly decrease during the 48-hour experiment, G-Bcl3 and p50-V proteins were still considered stable to perform future experiments at both 20 and 37 °C. It was noted however, that for both constructs, spontaneous removal of FPs from its fusion partner began to occur after 24 hours. To prevent this impacting future experiments, tests were performed within 24 hours following purification.

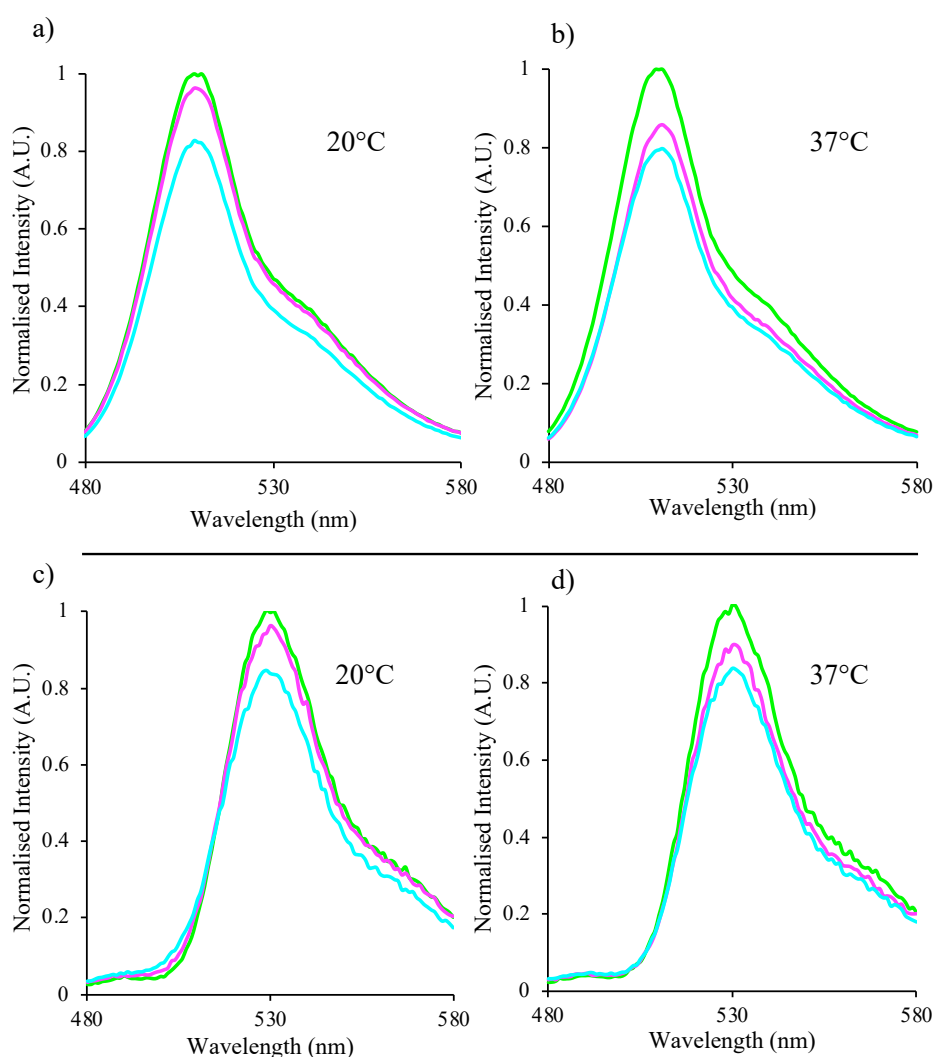


Figure 31. Temperature effect on G-Bcl3 and p50-V proteins. Proteins incubated at 20 and 37 °C. Spectra was recorded after 1 hour (green), 24 hour (magenta) and 48 hour (cyan). a) G-Bcl3 incubated at 20 °C. b) G-Bcl3 incubated at 37 °C. c) p50-V incubated at 20 °C. d) p50-V incubated at 37 °C.

3.2.2.4 FRET analysis between G-Bcl3 and p50-V

To determine whether there is binding between Bcl3 and p50, the purified fusion proteins were used for FRET analysis. Testing was performed using 1 μM of G-Bcl3 with 1, 2 and 4 μM of p50-V to give final mixing ratios of 1:1, 1:2 and 1:4 (G-Bcl3:p50-V). Experiments were performed according to section 2.4.5, recording the fluorescence at 1 and 24 hr intervals at both 20 and 37°C.

Across all datasets, results showed that increasing the concentration of p50-V caused a decrease in signal intensity at 510 nm and an increased intensity at 530 nm (Figure 32). The increase in emission at 530 nm was greater when incubated at 37°C than at the lower incubation temperature (20°C). The emergence of fluorescence at 530 nm is due to FRET, however emission at 510 nm is still prominent across all spectra. This suggests that there is an increase in the amount of G-Bcl3:p50-V complex, more so the case at a 1:4 ratio.

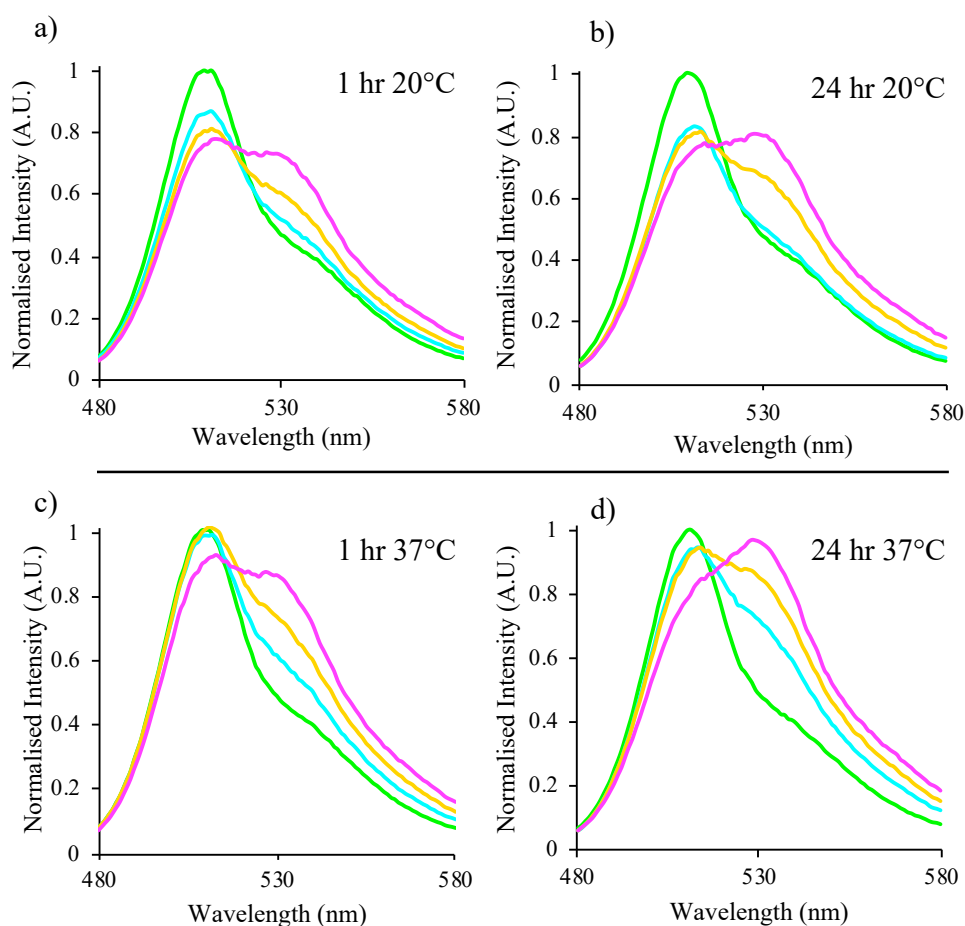


Figure 32. FRET analysis of G-Bcl3 and p50-V. Emission spectra recorded at a fixed excitation wavelength of 420 nm. G-Bcl3 mixed with p50-V at ratio's of 1:1 (cyan), 1:2 (orange) and 1:4 (magenta). Data normalised to G-Bcl3 (green). a) Results recorded after 1 hr incubation period at 20 °C. b) 24hr incubation at 20 °C. c) 1 hr incubation period at 37 °C. d) 24 hr incubation period at 37 °C.

A commonly used approach based on Equation 3 was then utilised to calculate the apparent relative FRET efficiency.

Equation 3
$$E_{\text{rel}} = \frac{I_A}{(I_D + I_A)}$$

Where E_{rel} is relative FRET efficiency, I_A is integrated fluorescence of acceptor, I_D is integrated fluorescence of donor. The spectra were decomposed using the online [a|e-UV-vis-IR spectral software](http://www.fluortools.com/software/ae) (<http://www.fluortools.com/software/ae>), as described in section 2.4.5. Residual fluorescence from background p50-V signal was removed before integration of fluorescence values for the donor and acceptor. The results showed an increase in relative FRET efficiency with higher concentrations of p50-V (Table 10). At both temperatures (20 and 37 °C), the relative FRET efficiency was greater after 24 hours in comparison to just a 1-hour incubation period. The higher temperature (37 °C) increased the relative FRET efficiency; after 24 hours FRET was recorded at 27 % compared to 20 % at 20 °C, for a 1:2 (Bcl3:p50) ratio. This suggests an increase in Bcl3-p50 complexation at 37 °C. At both temperatures, the greatest FRET efficiency was recorded at a 1:4 Bcl3:p50 ratio. This is likely due to the increase in p50-V concentration, shifting the equilibrium to form the protein complex. Also, the FRET efficiency can also be affected by many other factors including R_0 and r , which will be discussed in section 3.3.

Table 10. Relative FRET efficiency when mixing Bcl3 and p50 in 1:1, 1:2 and 1:4 ratio respectively. Proteins incubated for 1 hour or 24 hours and left at 20 or 37°C prior to spectral analysis.

Incubation time (hours)	Temperature (°C)	Relative FRET efficiency (%)		
		Bcl3:p50 ratio		
		1:1	1:2	1:4
1	20	5.53	11.89	15.54
24		8.04	19.57	25.78
1	37	7.37	12.55	19.33
24		19.75	26.91	34.69

3.2.2.5 Co-expression of proteins

To promote binding of Bcl3 and p50, G-Bcl3 and p50-V plasmids were then used for co-transformation and co-expression in *E.coli*. Nickel affinity chromatography was performed based on the same principle used to purify Bcl3-p50 complex; where complexed G-Bcl3-p50-V will have 3 Histags in comparison to individual components (1 histag – Bcl3, 2 histags – p50). Two different elution peaks were recorded: a minor peak at 250 mM and a major peak at 500 mM imidazole (Figure 33a). Fractions (1.4 mL) were collected from both peaks and analysed by SDS-PAGE. Bands corresponding to the size of G-Bcl3 (52 kDa) and p50-V (66 kDa) were present, confirming presence of both target proteins in both elution peaks (not shown here).

Fractions that eluted with 500 mM imidazole were pooled and subjected to size exclusion chromatography using the HiLoad™ 26/600Superdex™ S200 pg column, as the molecular weight of the complex would be ~184 kDa and most likely to resolve better on this column. Two elution peaks were collected and checked by SDS-PAGE (Figure 33b, c). The presence of G-Bcl3 and p50-V was confirmed within both peaks. Based on the calibrated SEC column and molecular weight estimation (described in section 2.4.6), the minor peak (peak 1) eluting at ~130 mL has a molecular weight of ~ 564 kDa, so is likely due to protein aggregation. Analysis of the SDS-PAGE gel suggests that the major elution peak (peak 2) contains two different stoichiometric ratio's of G-Bcl3-p50-V. ImageJ estimated the fractions eluting between ~140 – 155 mL contain p50-V at a 4-fold higher concentration than G-Bcl3. The molecular weight was estimated at ~ 320 kDa which would account for one G-Bcl3 and 4 p50-V molecules. Within this elution volume, the presence of high molecular proteins (> 100 kDa) within the SDS-PAGE gel was also noted. It is possible that these are involved in further PPI with Bcl3 and p50, and thus cause elution of additional proteins with the protein complex.

According to ImageJ, the tail portion of the major peak (elution at ~155 -190 mL) contained G-Bcl3 and p50-V at a 1:2 ratio respectively (Figure 33 b,c). The calibrated SEC column also confirms this; where elution of protein at ~165 mL is estimated to have a molecular weight of ~185 kDa, Thus, the fractions collected here were taken forward for analysis.

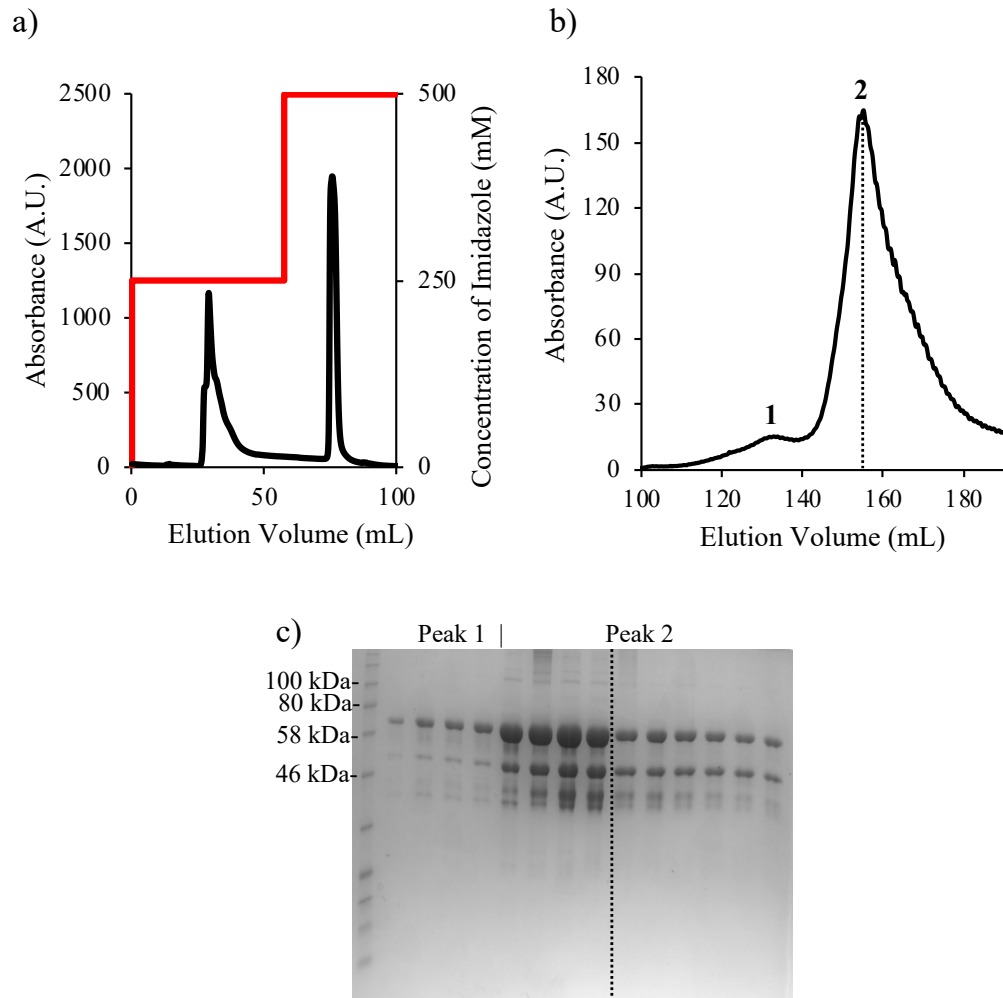


Figure 33. Purification and analysis of G-Bcl3 and p50-V. a) Elution profile from nickel affinity chromatography. Absorbance monitored at 280 shown in black. Imidazole gradient shown in red. b) Elution profile from size exclusion chromatography. Dotted line through peak 2 to highlight elution of two oligomeric states of Bcl3:p50. c) SDS-PAGE analysis from SEC. Black dashed line separating fractions eluted from peak 2 that have a 1:4 or 1:2 ratio of Bcl3:p50 respectively.

To confirm whether proteins were complexed, the emission spectra was recorded to detect possible FRET. The spectra showed the main emission wavelength was at 530 nm when excited at a fixed excitation of 420 nm (Figure 34a). This suggests FRET has occurred causing emission for p50-V. Deconvolution of spectra using the online [a|e – UV-vis-IR spectral software](#) revealed the spectra of the individual components (G-Bcl3 and p50-V) (Figure 34b). The concentration of p50-V in the sample was estimated using ImageJ, by comparing the intensity of the band against known concentrations in an SDS-PAGE. With an approximate concentration, the background fluorescence of p50-V was determined and subtracted from p50-V emission peak. The integrated fluorescence of G-Bcl3 and p50-V was

then calculated and supplemented into Equation 3. The relative FRET efficiency was calculated at 83 %.

Overall, the co-expression of G-Bcl3 and p50-V evidenced a successful method for complex formation as observed by FRET.

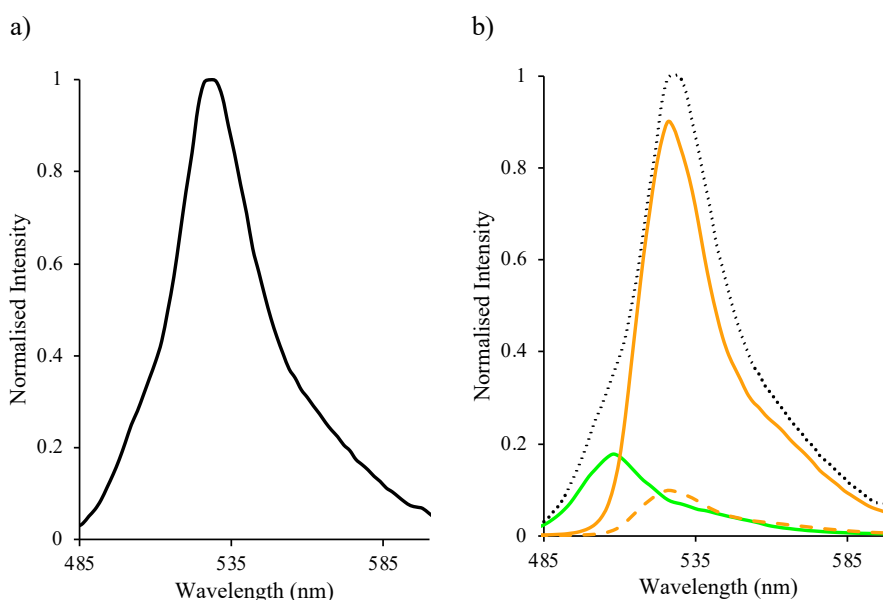


Figure 34. Emission spectra of G-Bcl3-p50-V. a) Emission spectra upon excitation at a fixed wavelength of 420 nm. b) Deconvolution of emission spectra (black dotted) showing G-Bcl3 (green) and p50-V components (orange). Residual p50-V shown as orange dashed line.

The effect of temperature on the stability of the Bcl3-p50 complex was determined. The samples were incubated at 4 and 37 °C over a total period of 2 days. The spectra was monitored each day to detect changes in the emission peaks. The results showed the spectra remained constant throughout the incubation at 4 °C, suggesting the complex is stable at this temperature (Figure 35). Whereas at 37 °C, the spectra changes over the time course with a rise in peak at 510 nm and subsequent decrease at 530 nm. This suggests that there is a loss in PPI between Bcl3 and p50 and that the complex is less stable at this higher temperature. However, section 3.2.2.4 evidenced that the higher incubation temperature of 37 °C encouraged formation of the Bcl3-p50 protein complex. It is therefore more likely that during the incubation period, FPs are being cleaved from their fusion partner, as previously mentioned. The close proximity of FPs will therefore be lost, having a direct impact on FRET.

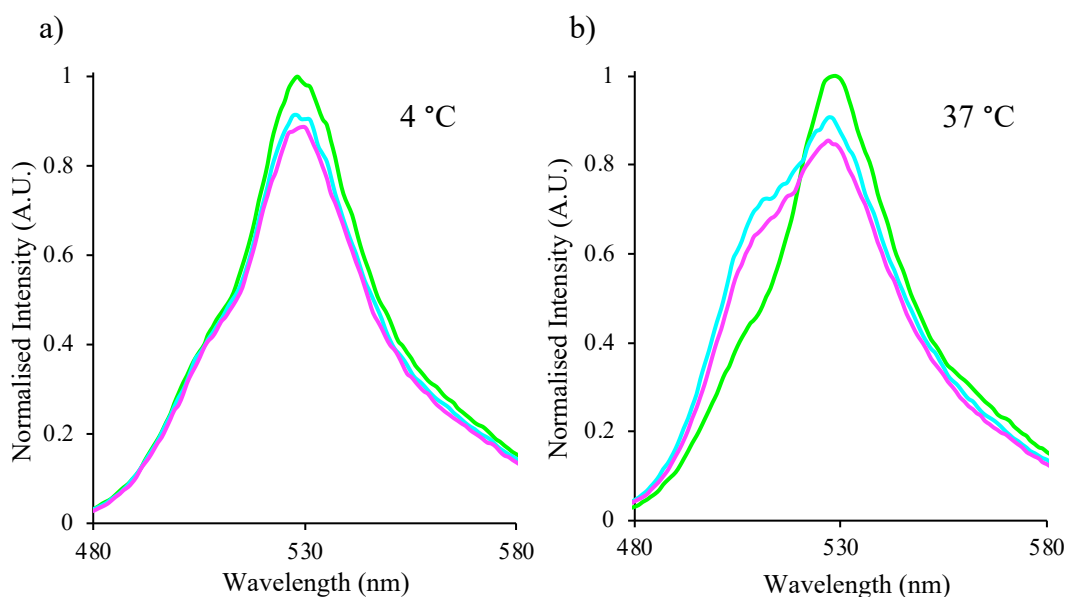


Figure 35. Emission spectra of complex whilst incubated at 4 and 20 °C. a) Emission spectra recorded upon excitation at 420 nm after 1 (green), 24 (cyan) and 48 hours (magenta) at 4 °C. c) Proteins incubated at 37 °C.

3.2.3 Testing small-molecule Bcl3 inhibitor, JS6

In the last few decades, the use of computational modelling for drug development has significantly grown, advanced by the availability of high throughput datasets and new data analysis strategies allowing reliable predications to be generated as well as new knowledge on the mode of action of drugs (Piñero et al. 2018). The structure of the Bcl3₂-p50₂ complex modelled by Dr Andrea Brancale (School of Pharmacy, Cardiff University) was critical for drug development studies and developing a treatment for breast cancer metastasis (as published in Soukupov et al. 2021). The hydrophobic binding pocket within ANK repeat 6 and 7 was determined as a unique target for Bcl3. Collaboration with Cardiff School of Pharmacy spurred the development of small-molecule inhibitors to target the Bcl3-p50 (and Bcl3-p52) protein-protein interactions (Yang 2018). One drug that appeared to work both *in situ* and *in vivo* was JS6. However, the exact mechanism of action and the specific target protein was unknown. This drug was therefore used in defined conditions *in vitro* as part of this study.

3.2.3.1 Testing drug binding with Bcl3 by FRET

To determine whether JS6 blocks binding of Bcl3 to p50, FRET experiments were performed. It was important G-Bcl3 and p50-V were used as individual proteins (expressed and purified as individual proteins), as opposed to the already generated complex (via the co-expression method). To encourage binding of Bcl3 with the drug, there was an initial 30 min incubation of G-Bcl3 with the drug, prior to addition of p50-V. As the highest relative FRET efficiency was previously recorded at a 1:4 ratio of Bcl3 bound to p50 respectively (section 3.2.2.4), 1 μ M of Bcl3 was mixed with 4 μ M of p50. A range of drug concentrations were sampled (1, 5, 10, 25, 50 μ M), recording data over 1, 24 and 48 hour intervals at both 20 and 37 °C.

Following a 1-hour incubation with JS6, similar spectra were observed across all concentrations sampled, with near equal emission peaks at 510 and 530 nm (Figure 36). Both the peak intensity and peak shift closely matched that recorded for the control, G-Bcl3 and p50-V without drug. The same was true for data recorded after 24 and 48 hours at the two experimental temperatures (20 and 37 °C). Deconvolution of spectra was carried out to determine the integrated fluorescence of G-Bcl3 (donor) and p50-V (acceptor) for use in equation 3. The relative FRET efficiency was estimated at 25%. As explored above in section 3.2.2.4, the relative FRET efficiency between G-Bcl3 and p50-V may vary. This is again likely due multiple factors including the distance and R0 value for the system. However, this ultimately impacted the ability to observe the spectral changes caused by the drug. Thus, given the limitations of this experiment, this result is not clear to report on the inhibitory action of JS6.

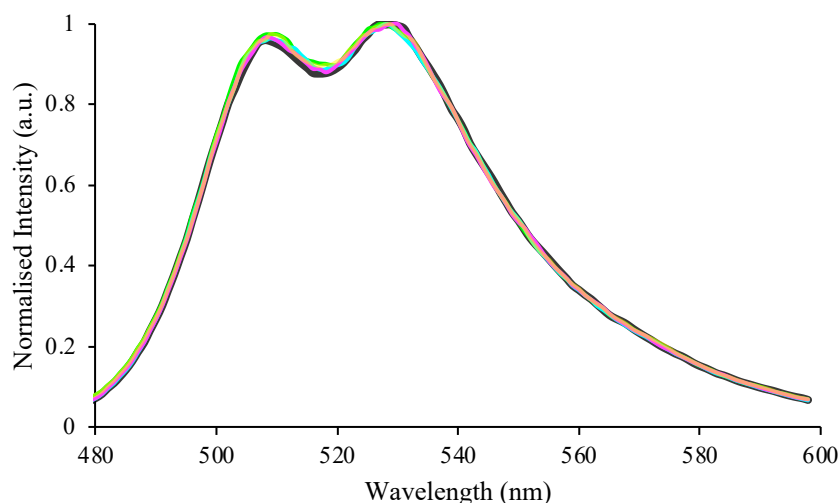


Figure 36. Incubation of G-Bcl3-p50-V with JS6 at 20 °C. Spectral data recorded after 1 hour for G-Bcl3-p50-V incubated with JS6 at concentrations of 1 (green), 5 (cyan), 25 (magenta) and 50 μ M (yellow). G-Bcl3 at 1 μ M, p50 at 4 μ M. G-Bcl3-p50-V without JS6 is shown in black.

3.2.3.2 Testing drug binding with Bcl3-p50 complex by FRET

Given the limited FRET efficiency when mixing G-Bcl3 and p50-V, drug testing continued with the preformed complex (by co-expression method, section 3.2.2.5). The complex was incubated with the drug at a variety of concentrations (1, 5, 25, and 50 μ M), recording data over 1, 24 and 48 hour intervals at both 20 and 37 °C. After a 1-hour incubation with JS6, results showed no change in spectra across different JS6 concentrations (Figure 37). The maximum emission peak was recorded at 530 nm indicating full complexation. The subtle decrease in peak intensity compared to the control (G-Bcl3-p50-V without drug) was deemed negligible as it did not transform into a peak shift towards 510 nm. Similar results were recorded after a 48-hour incubation time for both temperatures.

Although, no spectral change was observed across all samples, this experiment involved the incubation of drugs with preformed complex. As JS6 is designed to block the interaction of Bcl3 to p50, this could have impacted the binding of JS6 to Bcl3 when already bound to p50. Due to this limitation, this dataset will also not be used to confirm/deny the mode of action of these small molecule inhibitors.

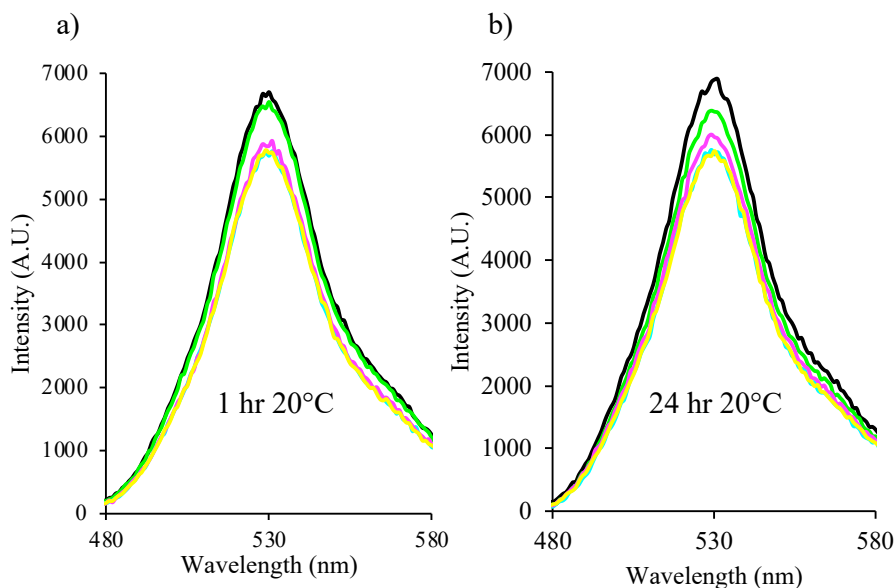


Figure 37. Incubation of G-Bcl3 and p50-V complex with JS6. The complex without JS6 (black) and with JS6 at concentrations of 1, 5, 25 and 50 μM shown in green, cyan, magenta, and yellow.

3.2.3.3 Isothermal Titration Calorimetry

Isothermal calorimetry is a key technique used to determine specific interactions between molecules in aqueous solution (Paketurytė et al. 2019). ITC works by directly measuring the heat that is either released or absorbed during a biomolecular binding event. It is an invaluable tool used to characterise the thermodynamic and kinetic association of ligands with proteins throughout drug development studies (Su and Xu 2018). Whilst the inhibitory action of JS6 could not be confirmed in prior studies, ITC was performed according to section 2.4.9, to determine whether drug binding occurs with Bcl3.

To prevent artefactual signals in the ITC measurements caused by small molecule impurities, dialysis was used to prepare protein and drug samples (Duff et al. 2011). Across all titrations, the differential power stabilised near the set reference power of 10 $\mu\text{cal}/\text{sec}$ ($\pm 1 \mu\text{cal}/\text{sec}$), indicating correct setup of equipment. Control titration experiments were performed (buffer into buffer, buffer into protein and drug into buffer) to confirm buffer compositions and determine the heat of dilution of the drug. Injection of the JS6 into Bcl3 resulted in small heat peaks that were equal in magnitude to one another (not shown here). A maximum signal output of 0.2 $\mu\text{cal}/\text{s}$ was observed which closely matched the control peaks. No heat from association

was detected so it could therefore be assumed that no biomolecular binding event had occurred.

It was then necessary to confirm whether drug binding inadvertently targeted Bcl3's partner proteins, or Bcl3 when complexed to partner proteins. ITC experiments were performed as above with p50 and the Bcl3-p50 complex. Again, small heats were detected mimicking those of controls (Figure 38). It was therefore concluded that binding interactions between the drug and protein were not detected.

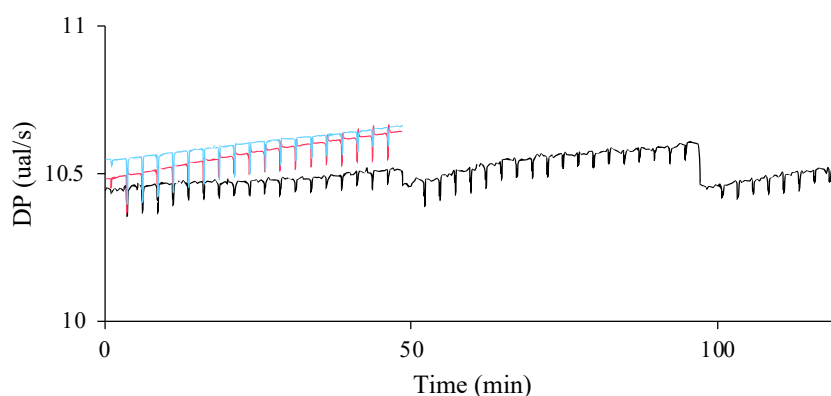


Figure 38. Raw ITC data testing JS6 drug binding to Bcl3-p50 complex. Black line showing titration of 100 μ M JS6 into 50 μ M Bcl3. The following controls are included for reference: buffer into buffer shown in blue, JS6 into buffer shown in red. All solutions were made up of 20 mM HEPES, 300 mM NaCl and contained 1% DMSO.

3.3 Discussion

Multi-protein complexes are involved in essentially all cellular processes. A protein's function or activity will be defined by a combination of its inherent properties, cellular localisation, interacting partners, and stoichiometry of each (Kerrigan et al. 2011). In some instances, targeting PPIs offers a viable approach for therapeutic intervention, where PPI inhibitors can be designed to achieve highly specific drug binding to a target protein complex. Bcl3 is a member of the I κ B family of proteins that preferentially binds with p50 and p52 homodimers and can function either as an activator or a repressor of gene expression (Michel et al. 2001). High expression of Bcl3 has been linked with a variety of haematopoietic and solid cancers. However, to fully elucidate the biomolecular role of Bcl3, its interaction with binding partners should be fully understood.

The interactions within a protein complex are believed to benefit the proteins in terms of stability and its function. Many experimental techniques have been developed to identify PPIs; by either high-throughput screening (yeast two-hybrid (Y2H) system, tandem affinity purification (TAP), protein microarray and phage display) or low throughput methods (confocal microscopy, co-immunoprecipitation (co-IP), x-ray crystallography). Over-expression of proteins performed in a host system (such as bacteria, yeast, insect or mammals) can be useful for biochemical, biophysical, structural studies and drug discovery. Within this chapter, Bcl3 and p50 proteins were first expressed and purified separately, followed by reconstitution *in vitro*. However, the expression yield of both individual proteins was low, they were unstable and degraded rapidly. Co-expression of Bcl3 and p50 was then performed within *E.coli* cells. Analysis by gel electrophoresis and SEC-DLS confirmed successful production and purification of Bcl3-p50 complex. As a result, the proteins were not only more stable but the yield was greatly increased.

The co-expression method was then used for FP-tagged version of Bcl3-p50. The addition of fluorescent proteins at the terminus of Bcl3 and p50 also improved stability of individual proteins. Thus, this enabled separate expression and purification of G-Bcl3 and p50-V, so that different concentrations of each could be mixed and tested for binding. The maximum FRET efficiency was recorded at 35 % after a 24 hour incubation period with proteins at a 1:4 ratio (Bcl3:p50 respectively) (Table 10). This is almost a 50 % reduction in comparison to the FRET efficiency of 83 % recorded for co-expressed G-Bcl3 - p50-V (section 3.2.2.5). There are multiple reasons that could explain this including that formation of protein complexes from purified proteins can often require re-folding of at least one of the subunits. Kerrigan *et al.*, (2010) detailed a number of case studies where co-expression is a requirement in forming functional multi-protein complexes, such as the heterodimeric T-cell receptors (TCR). Separate expression of the TCR chains in *E.coli* typically leads to the formation of inclusion bodies and requires re-folding in order to form the heterodimer. Another reason could have been that FPs interfered with Bcl3 and p50 binding. However, the linker between the fusion proteins is relatively short in length so is not likely the case here. It is also possible that other proteins have bound to Bcl3 and/or p50, but analysis by SDS-PAGE suggests our proteins are of high purity. Aggregation of one of the proteins may also have occurred, where a significant GFP signal may still be recorded but prevents complexation from

occurring, and thus a low FRET efficiency. However, the most likely reason is due to the spontaneous cleavage of FPs from their fusion partner. Thus, FPs would not be brought into close proximity, causing the decreased FRET efficiency.

FRET is one of the most important applications of FP pairs as it can be used to investigate PPIs in live or fixed specimens. Whilst there are many advantages to this technique including the instant reporting of PPI, there are also several factors that can affect FRET and lead to inaccurate reporting of the PPI. For example, k^2 is arbitrarily set as 0.667 to reflect two randomly orientated chromophores. In reality, two chromophores are unlikely to be truly freely rotating with respect to one another when fused to a protein of interest (Hellenkamp et al. 2018). It is therefore difficult to accurately equate FRET efficiency to distance and, more importantly, relative amounts of complex and non-complexed components. Also, the physical interaction of FP pairs can result in inherent functional changes (Alford et al. 2012b; Worthy et al. 2019). Thus, when investigating FRET between FPs there may not just be simple spatial proximity at work but molecular interactions leading to more defined distance and dipole alignment, which may in turn influence inherent fluorescence (Pope et al. 2021). To predict the distance between FPs in the G-Bcl3 – p50-V complex, MD simulations were performed. The results showed that FPs are capable of either closely associating to one another (r of 35-40 Å) or further apart (r of 60 – 80 Å). This highlights a key concern over the assumptions of FRET where it is possible to have two different systems relating to FP association, which would almost certainly impact k^2 and R_0 . Using the apparent relative FRET efficiency of 83 % (section 3.2.2.5), and the standard R_0 of 57.14, (where k^2 is 0.667) the two chromophore should be 44 Å apart. Whilst this is true for a portion of the MD simulations, at an r of 60 Å, the relative FRET efficiency is expected to be 43%. Pope *et al.*, (2021) utilized the structure of an engineered sfGFP dimer covalently linked at residue 204, to investigate how discrepancies between observed and theoretical FRET efficiency can occur. Suggestions include; residual monomer FPs, water molecules present at the domain interface playing a role in fluorescence quenching, and mixed population of dimers where the FRET efficiency is more highly efficient in one population than another. This latter suggestion may well be the case for our G-Bcl3 – p50-V complex, as FPs are predicted to be both closely associated or freely rotating. Thus, in conjunction with structural analysis, we can conclude that FRET was useful for confirming the formation of Bcl3-p50 complex by co-expression. However, the

different r values between FP pairs highlighted by our modelling could have impacted the reporting of Bcl3 interactions with p50 when mixing purified proteins. Overall, fluorescence changes to measure FRET may not necessarily fully reflect the molecular events in protein complexes, without understanding of positional relationships of FP pairs.

3.4 Conclusion

Overall, this chapter generated and characterised the Bcl3-p50 protein complex following recombinant expression in *E.coli*. The stoichiometry of 1:2 was confirmed by a variety of methods including SDS-PAGE and SEC-DLS. However, the drug testing of JS6 was particularly difficult given that individual proteins of Bcl3 and p50 (without FPs) were unstable, and formation of the complex was performed by co-expression, meaning that Bcl3 binding pocket was already preoccupied. However, since the initial discovery of JS6, Prof Richard Clarkson's group continued to evidence its role inhibiting the interaction between Bcl3 and p50 (now published in Soukupova et al. (2021)). Another important aspect of this work was the use of *in silico* modelling. It highlighted some of the key limitations to FRET where assumptions are made based on the association of FP pairs and the orientations of chromophores. Thus, it poses the idea that MD simulations are a suitable method to use in conjunction with FRET studies in order for accurate FRET efficiencies to be calculated.

4 Engineering residue 148 of sfGFP

4.1 Introduction

Since its discovery, GFP has been at the centre of one of the most extensive engineering efforts ever made on a single protein (Campanini *et al.*, 2013). A huge armoury of GFP variants now exist, offering a wide range of desirable characteristics including improved protein solubility, rapid and robust folding and chromophore maturation, high resistivity to chemical denaturation, and a variety of spectral and photodynamic properties. A key target area concerns the excited state proton transfer (ESPT) and controlling proton shuttling to and from the chromophore (Remington 2011). Original *Aequorea victoria* GFP (avGFP) excited at ~ 400 nm due to the dominant presence of the protonated phenol group of the chromophore (known as CroA), but underwent EPST to emit at ~ 510 nm via the anionic phenolate chromophore. One of the first efforts in engineering avGFP was to promote the anionic chromophore ground state (known as CroB) so shifting excitation more into the visible range. Enhanced GFP (EGFP) is an early example of promoting the ground state CroB in which Phe64 and Ser65 are mutated to leucine and threonine, respectively. This resulted not only in a more stable and faster maturing protein, but a dominant excitation wavelength at ~ 490 nm. The S65T mutation was considered the critical mutation for promoting the CroB ground state by changing the nature of the charged state of E222 and thus the network linked through to chromophore phenol group (Arpino *et al.* 2012). The role of neighbouring residues including H148, S205 and E222, and water molecules have also been shown to regulate the proton transfer system and maintain the anionic ground state of the chromophore, and thus some of the key properties of GFP (Figure 39)(Zhang *et al.* 2005). This concept will be used to drive the work in this chapter, with a focus on residue 148 in the context of the superfolder GFP variant.

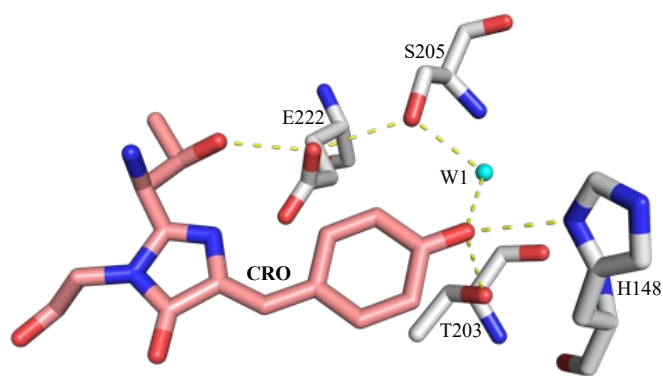


Figure 39. Hydrogen bond network surrounding the chromophore (pink) of sfGFP including neighbouring residues H148, T203, S205 and E222. The critical water molecule (W1) shown as a sphere in cyan.

H148 plays a critical role in defining the fluorescent properties of GFP. It forms a hydrogen bond (H-bond) with the tyrosyl hydroxyl group of the chromophore and is thought to contribute to the charge transfer network that modulates chromophore protonation state equilibrium (Figure 39). Replacement of H148 with a non-polar glycine (H148G) causes the loss of the H-bond to the chromophore and results in a dual emission GFP, or deGFP, with a predominant blue emission ($\lambda_{\text{max}} \sim 400$ nm, CroA) (Hanson et al. 2002). H148 is also thought to be dynamic, sampling a minor “open” conformation that shifts the residue away from the chromophore as well as the predominant “closed” form. The open form leaves a “hole” in the protein that is suggested to favour structural flexibility (Auhim et al. 2021)(Seifert et al. 2003) and provides a channel to the chromophore that allows diffusion of solvent molecules, as well as large metal ions (Barondeau et al. 2002). The H148K and H148R mutations lowers expression yields and stability of the mature protein (Campanini et al. 2013). Two absorbance bands were recorded that were attributed to the neutral and anionic chromophore. The incorporation of ncAA has also been demonstrated with residue 148 of sfGFP; where replacement of H148 with azidophenylalanine (azF) resulted in the loss of two H-bonds from the CRO (from H148 and T203)(Figure 40 a). This caused a blue shift in the major absorbance peak to a predominantly neutral chromophore ground state (Figure 40 b)(Hartley et al. 2016). However, non-native modifications of 148azF (e.g. photochemical conversion and bioorthogonal modification) can reverse the effects of the mutation to promote the CroB form.

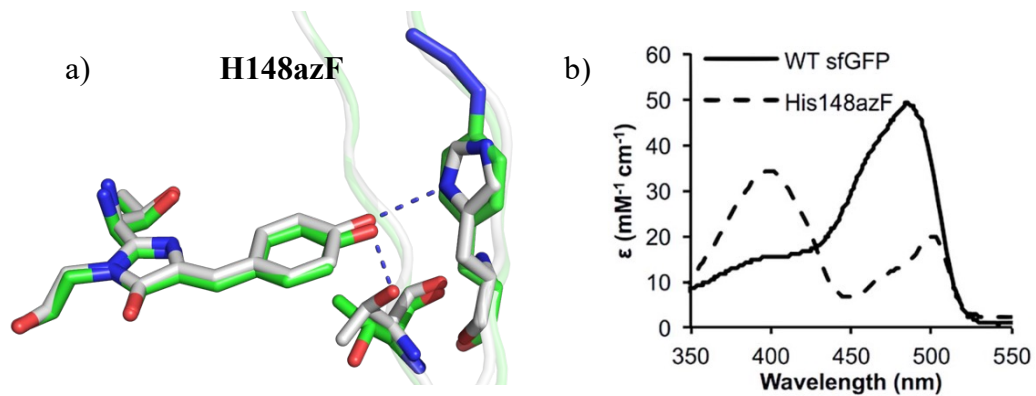


Figure 40. Replacement of H148 with azF in sfGFP. a) The chromophore, residue 148 and T203 shown as sticks. SfGFP^{WT} shown in grey and sfGFP^{148azF} shown in green. PDB ID: 5BT0. b) Absorbance spectra showing sfGFP^{WT} (solid line) and sfGFP^{148azF} (dashed line).

Alongside its widespread use as a reporter of gene expression, GFP and selected mutants have demonstrated their potential to sense biologically important intracellular functions. For example, redox-sensitive variants of GFP were developed that allowed real-time monitoring of the redox status of cellular compartments by fluorescence excitation ratiometry. Two cysteine residues were incorporated into adjacent β -strands (residue positions 147 and 204) that formed a disulphide bond under oxidative conditions. The resulting disulphide caused small structural rearrangements which perturbed the chromophore environment and influenced the population ratio of the neutral to anionic chromophore (Hanson et al. 2004; Cannon 2006). GFP has also been engineered to function as a non-invasive intracellular pH indicator, by the incorporation of S65T mutation. Both crystallographic and phenolic data suggest the pH sensitivity is reliant on the protonation state of the chromophore's phenolic group (Kneen et al. 1998)(Hanson et al. 2002). Barondeau *et al.*, (2002) designed and characterised a GFP mutant that has potential to monitor *in vivo* spatial and temporal distributions of metal ions. The binding of Zn(II) was shown to enhance fluorescence intensity whereas Cu(II) caused fluorescence quenching. One particular biomolecular event that is of significance is the monitoring of protein dimerisation or oligomerisation. Although two separate FPs can be used to monitor hetero-oligomerisation by FRET, this highlighted a current limitation for monitoring

the equally important process of homo-oligomerisation; as no single FP can detect dimerisation. This aspect will be addressed in Chapter 5.

In this chapter, residue position 148 will be targeted in an attempt to modify the spectral properties of sfGFP through both direct and indirect changes to the local CRO hydrogen bond network, including modulating the chromophore ground state charge equilibrium. Mutations will replace H148 with alternative side chains of different length, polarity, and functional groups. Spectral changes will be monitored to determine the effects of residue replacement on protein function. The hypothesis stands that with a shorter, non-polar residue at position 148, the chromophore will exist in a predominantly neutral form (CroA, $\lambda_{\max} \sim 400$ nm). Based on these findings, assuming a CroA variant is identified, the potential use of the novel GFP mutant will be explored for the monitoring of dimerisation, through alteration of its spectral properties. Previously, docking methods were used by (Hartley et al. 2016) who highlighted the compatible interface of sfGFP comprising residues 145-148, 202-207 and 221-224. They also showed that by linking the two monomers together by bioorthogonal, ncAA-based covalent click chemistry, spectral shifting and enhanced brightness could be achieved on dimerisation. Thus, the question arises as to if dimerisation of sfGFP occurs at this interface promoted by naturally occurring amino acids, could the dimerisation restore the hydrogen network surrounding the chromophore (either through local conformational changes, or via the entrapment of water molecules), allowing switching between the CroA (protonated chromophore state, $\lambda_{\max} \sim 400$ nm) and CroB (deprotonated chromophore state, $\lambda_{\max} \sim 485$ nm). This would add an additional biologically important function of GFP; to monitor dimerisation through the use of a single protein variant.

4.2 Results

4.2.1 Design and production of sfGFP^{C148}, sfGFP^{V148}, sfGFP^{S148} and sfGFP^{A148}

To investigate the functional influence of residue 148 of sfGFP, H148 was replaced with four amino acids: alanine, valine, serine, and cysteine. These amino acids were selected based on two considerations: (1) To ensure the neutral form of chromophore ground state predominates ($\lambda_{\max} \sim 400$ nm, CroA), and; (2) For compatibility with protein interfaces to encourage dimerisation and a potential

switch in chromophore ground state back to the anionic form ($\lambda_{\text{max}} \sim 495 \text{ nm}$, CroB). Thus, a variety of properties were sampled including short side chains to prevent local steric clashes and different side chain polarities. This should impact the charge transfer system surrounding the chromophore in sfGFP^{WT} and alter the equilibrium state of CroA:CroB. These residues should also offer suitable characteristics for dimerisation based on hydrophobic interactions, H-bonding, disulphide bonding (H148C), and van der Waals. For simplicity, these proteins will be referred to as sfGFP^{A148}, sfGFP^{V148}, sfGFP^{S148} and sfGFP^{C148} according to the mutation at residue 148.

Larger bulkier side chains or those carrying a charge were not investigated here as they are unlikely to fit within the protein, close to the chromophore. This could then result in steric clashes or electrostatic repulsion at the dimer interface. For example, the suggested conformations within PyMOL of H148R (probability score of 22 %) and H148F (probability score of 66 %) are shown as in Figure 41, where steric clashes are highlighted in red.

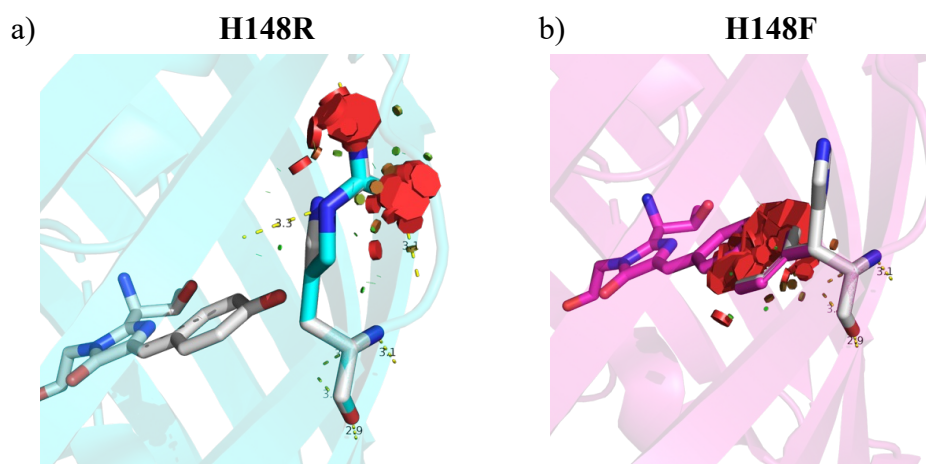


Figure 41. Mutation of H148 to arginine and phenylalanine in sfGFP. SfGFP^{WT} is shown in grey with H148 highlighted as sticks. Steric clashes shown in red. A) H148R in cyan. b) H148F in magenta.

To assess the impact of these mutations on sfGFP, the mutagenesis tool in PyMOL was used. For the purposes of the current simple modelling, the major closed conformation of H148 was used as the starting point. For three residue substitutions (valine, cysteine, and serine), multiple side chain rotamer conformations were available (Figure 42). It is possible that all suggested rotamers can exist, although each may occupy one conformation more often than another, based on steric clashes and surrounding interactions. For sfGFP^{V148}, rotamer 1 was

given a 97 % probability score over 2 other possible forms, and is shown in Figure 42c. The effect was a loss in hydrogen bond from replacement of H148, as similarly observed with H148A. The side chain of sfGFP^{C148} was more tolerant to different rotamer forms, with a 61 to 39 % ranked score for rotamer 1 vs. rotamer 2 respectively. Rotamer 1 faced away from the chromophore, whereas rotamer 2 was within appropriate distance to participate in hydrogen bonding with Tyr66 (2.4 Å). For sfGFP^{S148}, a hydrogen bond could form between S148 and the hydroxyl group of the chromophore, for the most probable rotamer form (76% likelihood). Similarly to that for H148 in sfGFP^{WT}, this interaction would likely stabilise this residue, making them the most preferred configuration.

The proteins were then submitted for energy minimisation (EM) using the molecular dynamics software GROMACS, to allow the proteins to relax and local structural changes to be sampled. Each of the models were successfully energy minimised where the change in potential energy approached zero (Figure 43). The energy minimisation curves were similar for each sfGFP^{X148} variant, where each of the different rotamers of residue 148 were submitted. The results are also similar across proteins variants (containing either valine, alanine, cysteine and serine at residue position 148).

The EM models were then compared against the initial models in PyMOL (Figure 42). These EM results suggest that just one of the mutations, S148, had the possibility to form H-bonds with the sfGFP chromophore. This would most likely play a role in stabilising the anionic ground state of the chromophore. The absorbance spectra was expected to evidence this, with a major CroB form ($\lambda_{\max} \sim 485$ nm), as seen with sfGFP^{WT}. Although suggested within PyMOL's static structure, rotamer 2 of H148C cannot form the H-bond with the chromophore following EM. The H-bond is also absent for H148V and H148A variants. Without a H-bond to stabilise the anionic chromophore, evidence of neutral chromophore, or CroA ($\lambda_{\max} \sim 400$ nm) was expected.

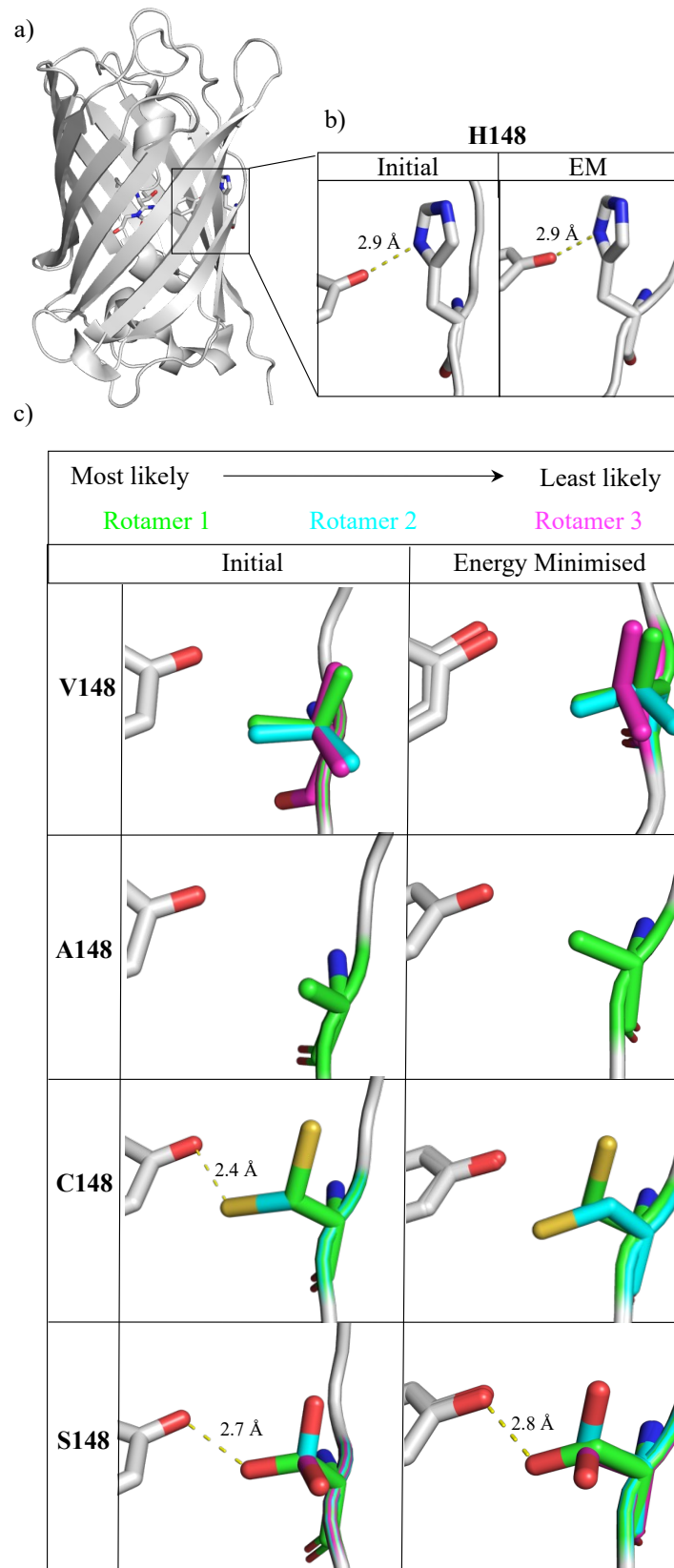


Figure 42. Mutation of residue 148 of sfGFP. A) Cartoon representation of sfGFP^{WT}. b) H148 residue hydrogen bonding to chromophore. c) Stick representation of the incorporation of valine, alanine, cysteine and serine at position 148 of sfGFP. Rotamer 1-3 shown in green, cyan and magenta. Starting pdb and following energy minimisation performed in GROMACS.

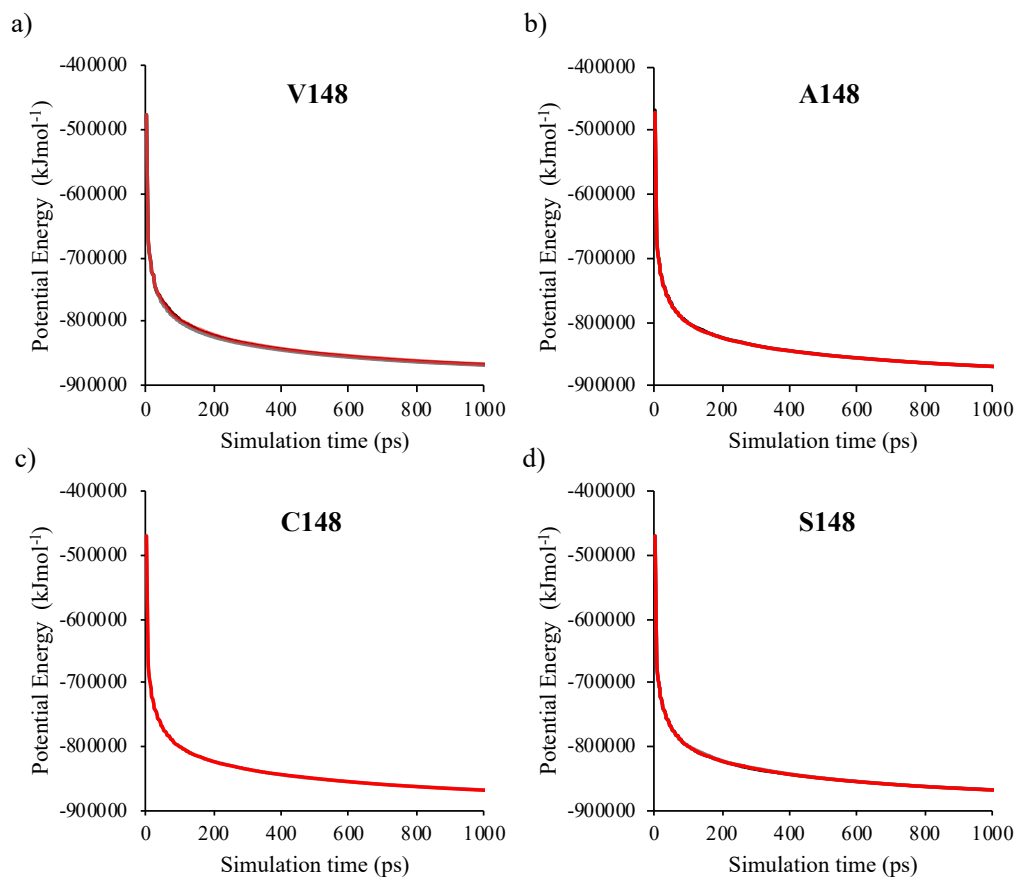


Figure 43. Energy minimisation curves for sfGFP^{X148} variants. EM for rotamer 1 (black), rotamer 2 (grey) and rotamer 3 (red). A) SfGFP^{V148}. b) SfGFP^{A148}. C) SfGFP^{C148}. D) SfGFP^{S148}.

4.2.1.1 Site directed mutagenesis of sfGFP^{WT}

To incorporate the desired amino acids at residue position 148 of sfGFP, site directed mutagenesis (SDM) was performed on sfGFP^{WT}, present within the pBAD plasmid. The sfGFP^{C148} variant was previously constructed by Dr Harley Worthy (Worthy et al. 2019). For production of sfGFP^{S148}, sfGFP^{A148} and sfGFP^{V148} mutants, primer pairs were designed containing the mutations for residue replacement, as listed in Table 3. The template plasmid DNA was amplified by whole vector PCR at the optimum binding temperature for each of the primer pairs (Figure 44 a). Agarose gel electrophoresis confirmed presence of DNA bands ~ 4.7 Kb, the approximate size of the pBAD plasmid containing the sfGFP gene (Figure 44 b). The DNA was purified, phosphorylated, and ligated to re-circularise the vector. The plasmid was then transformed into *E. coli* Top10TM electrocompetent cells and subsequently placed onto LB agar plates, supplemented with antibiotics and 0.1% arabinose

(Figure 44 c). The agar plates were illuminated under UV light to allow selection of green coloured colonies. The final construct with desired DNA mutation was confirmed by sequencing.

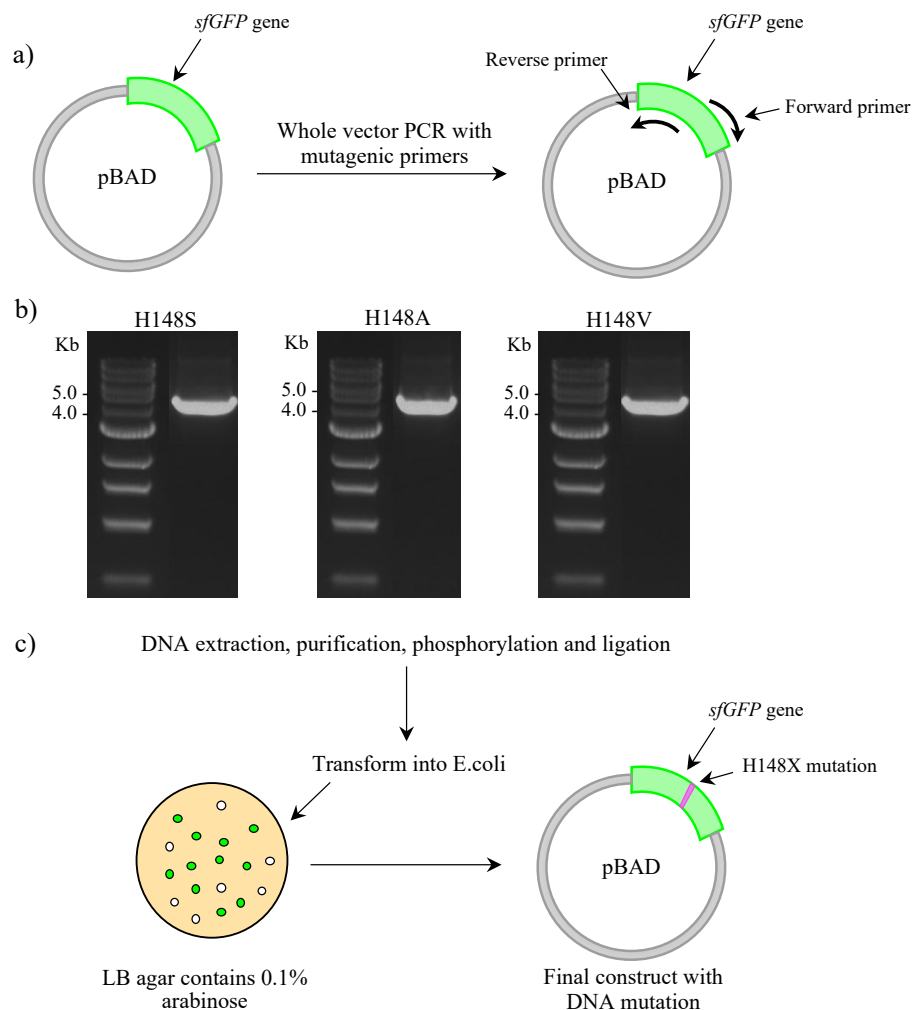


Figure 44. Site directed mutagenesis of sfGFP^{WT} to incorporate a serine, alanine and valine mutation at residue position 148. a) Whole vector PCR with mutagenic primers. b) Gel electrophoresis. c) DNA ligation and transformation into *E. coli*.

4.2.1.2 Expression and purification of sfGFP^{X148}

Before detailed analysis could be performed, sfGFP^{X148} protein variants had to be expressed and purified. Here, the purification of sfGFP^{C148} was described, although the same method was applied to produce all four sfGFP^{X148} variants.

Overexpression of sfGFP^{C148} was confirmed by the presence of green coloured cultures. This provides the first indication that proteins are functional with maturation of the chromophore occurring. Cells were lysed by French press and the

soluble lysate was applied to a nickel column for purification by immobilised metal affinity chromatography (IMAC). SfGFP contains a C-terminal hexahistidine-tag allowing the protein to bind onto nickel column. Absorbance was monitored at 400 and 485 nm to examine both the neutral ($\lambda_{\text{max}} \sim 400$ nm, CroA) and anionic ($\lambda_{\text{max}} \sim 485$ nm, CroB) ground state of the sfGFP chromophore. After contaminating proteins had eluted from the column, an imidazole gradient was applied to elute sfGFP^{C148}. Fractions were collected according to a rise in absorbance at the above wavelengths and by the presence of green colour (Figure 45 a). Considering the potentially redox-sensitive nature of the cysteine variant, fractions were checked by both reducing and non-reducing SDS-PAGE gels. Bands of the expected monomer size of sfGFP (~27 kDa) were observed in both gels (Figure 45 b). However, higher bands were also present in the non-reducing SDS-PAGE gel. Thus, this is likely to represent covalently linked sfGFP^{C148} dimers as the band is not observed in the reducing SDS-PAGE gel. There were no obvious bands the size of sfGFP^{C148} in the flow through, confirming that majority of target protein bound to the column.

Purified protein fractions were pooled and concentrated to a maximum volume of 2 mL. Size exclusion chromatography using a HiLoadTM 16/600SuperdexTM S75 pg (preparative grade, 120 mL bed volume, separation range 3-75 kDa) column was performed according to section 2.3.4, to remove residual impurities and desalt the protein sample. To ensure purification of monomeric sfGFP^{C148}, DTT (5mM) was added to the SEC buffer as a reducing agent. A major elution peak with a dominant 400 nm absorbance was detected at ~ 60 mL (Figure 45 c). Given the absorbance ratio at 400:485 nm, it was likely the chromophore exists predominantly in the CroA form ($\lambda_{\text{max}} \sim 400$ nm) for sfGFP^{C148}. Fractions were checked via non-reducing gel electrophoresis, confirming successful purification of monomeric sfGFP^{C148} (Figure 45 d).

For clarity, SDS-PAGE gels of sfGFP^{S148}, sfGFP^{A148} and sfGFP^{V148} are included in Figure 45e to show purified protein fractions after SEC.

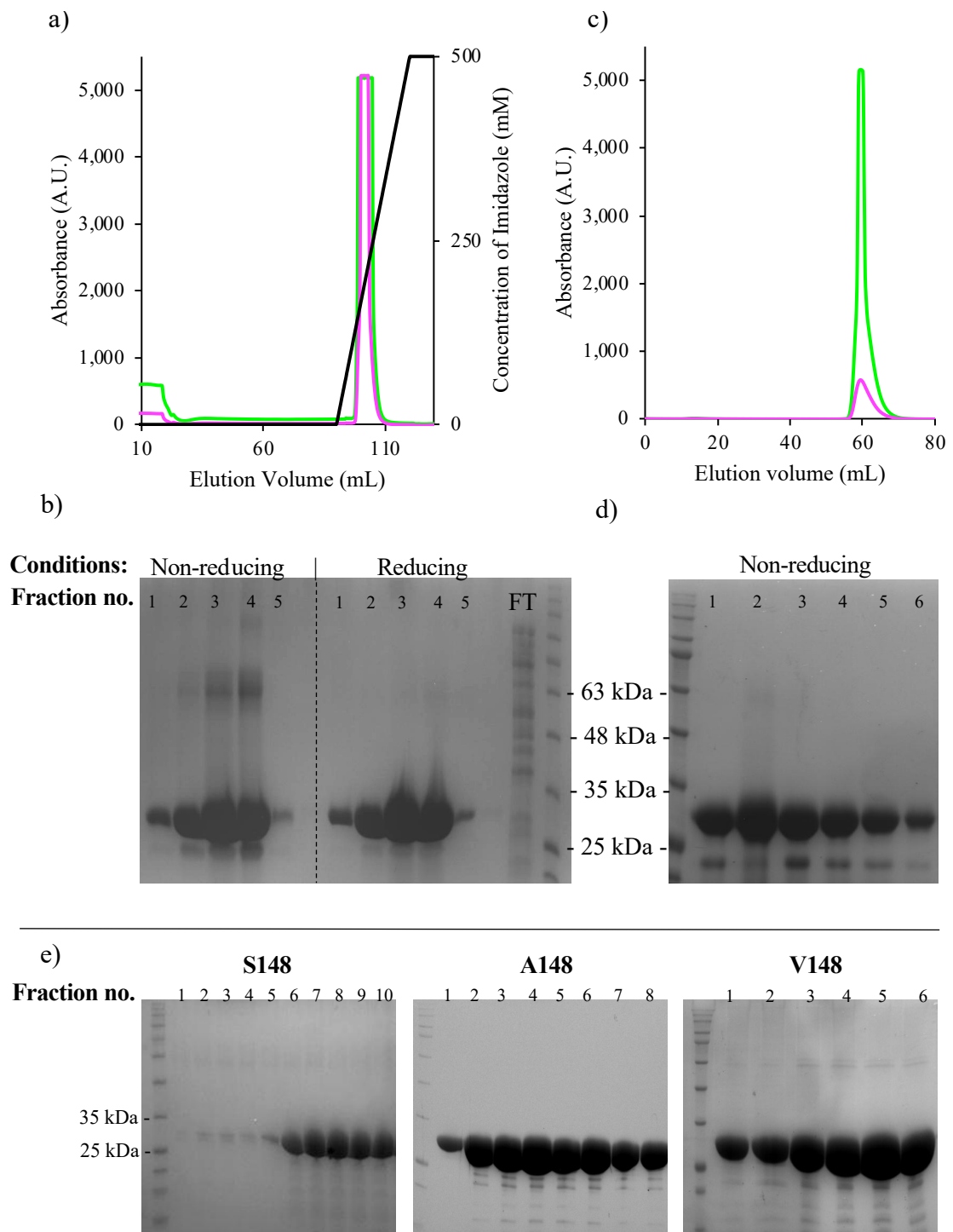


Figure 45. Purification of sfGFP^{C148}. Absorbance monitoring at 400 and 490 nm shown in green and magenta respectively. a) Elution profile from nickel affinity chromatography with imidazole gradient shown with a black line. b) Non reducing and reducing SDS-PAGE of eluted fractions, including the flow through (FT). c) Elution profile from size exclusion chromatography performed under reducing conditions. d) Non-reducing SDS-PAGE gels showing purified sfGFP^{C148} after SEC. e) Purified sfGFP variants after SEC with serine, alanine and valine at residue position 148.

4.2.1.3 Characterisation of sfGFP^{X148} variants

To understand the effect of the selected mutations to residue 148, the spectral properties of the sfGFP^{X148} variants were characterised. The fluorescence and absorbance spectra were recorded for purified protein samples at 0.5 and 10 μM , respectively (Figure 46). For monomeric sfGFP^{C148}, characterisation was performed on the protein sample prepared as described in section 4.2.1.2, in DTT.

Overall, the X148 variants fall into two major groups: bimodal absorbance with peaks at ~ 400 nm and 495 nm and a dominant absorbance at ~ 490 nm. The variants sfGFP^{A148}, sfGFP^{V148} and sfGFP^{C148} display a bimodal absorbance spectrum (peaks at ~ 400 and 495 nm) (Figure 46a). The molar extinction coefficient was *circa* 35,000 $\text{M}^{-1}\text{cm}^{-1}$ at 400 nm for all three protein variants (see Table 11 for actual values). However, there are differences in the molar extinction coefficients calculated at ~ 495 nm, which are due to different CroA:CroB ratios; 3:1 for sfGFP^{C148}, and 2:1 for sfGFP^{A148} and sfGFP^{V148}. Replacing H148 has caused the expected switch from the CroB dominated chromophore of sfGFP^{WT} (CroA:CroB = 1:6) over to CroA. For sfGFP^{A148} and sfGFP^{V148} variants, this is likely to be due the disruption of the charged transfer network by the loss of the H-bond to the chromophore. For the polar cysteine residue of sfGFP^{C148}, capable of participating in hydrogen bonding, the side chain may sample multiple orientations, or an orientation and configuration unfavourable for hydrogen bonding to the chromophore. The modelling described in section 4.2.1 above suggests that this could be the case (Figure 42). The larger atomic radius of S is likely to play a critical role with steric clashes preventing a suitable side chain orientation capable of H-bonding with the chromophore.

The shift in the major absorbance peak was also mirrored in the fluorescence emission spectra (Table 11 c). Each variant had a similar 511 nm emission maximum wavelength (λ_{EM}), irrespective of the excitation wavelength. This is due to the ESPT system, where a proton is lost upon excitation of the neutral chromophore resulting in emission due the phenolate charged form. For sfGFP^{A148} and sfGFP^{C148} variants, the highest fluorescence intensity was observed on excitation at ~ 400 nm rather than on excitation at ~ 490 nm, as is the case for sfGFP^{WT}. The level of emission on excitation at 400 nm was lower but this is largely down to the decrease in quantum yield. Calculated quantum yields at 400 nm excitation gave ~ 32 -35 %, which is

below the QY for sfGFP^{WT} (75 %) on excitation at its optimal wavelength (Table 11). For sfGFP^{V148}, excitation was slightly higher at 495 than 400 nm. This is due to the 60 % QY at ~ 495 nm for sfGFP^{V148} in comparison to sfGFP^{A148} (54%) and sfGFP^{C148} (50%) variants. Thus, not only are the sfGFP^{A148}, sfGFP^{V148} and sfGFP^{C148} variants shifted in terms of their optimal excitation wavelength, they are generally less bright when excited at ~ 400 nm.

SfGFP^{C148} was subjected to a second round of SEC, to remove the reducing agent and compare characteristics in 50 mM Tris, the buffer used further throughout this study. Analysis of results revealed the presence of DTT had no significant impact on protein characterisation, as summarised in Table 11.

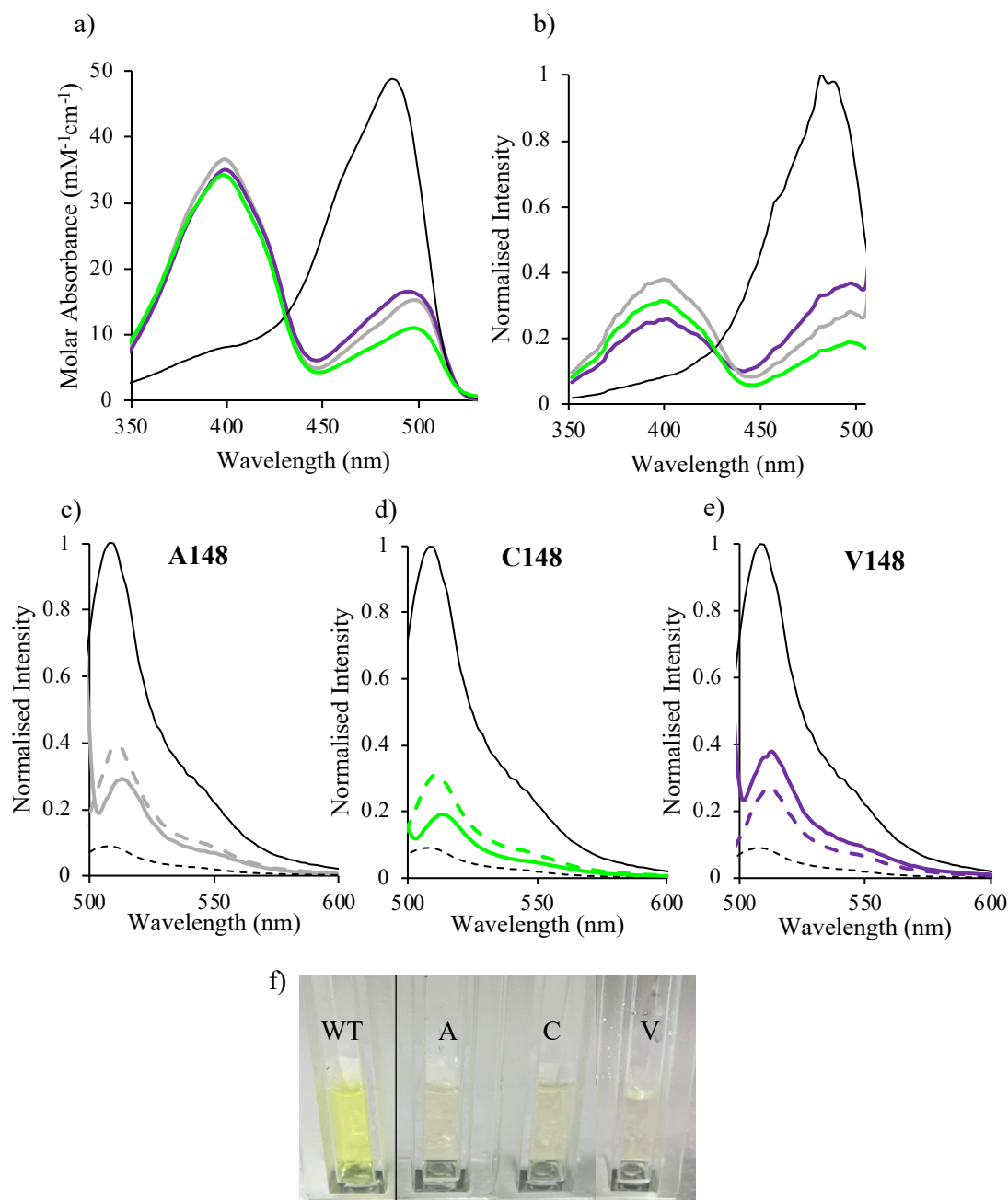


Figure 46. Spectral properties of sfGFP^{X148} variants. a) Absorbance spectra of sfGFP^{A148} (grey line), sfGFP^{C148} (green line) and sfGFP^{V148} (purple line), compared with sfGFP^{WT} (black line). b) Excitation spectra on measuring emission at 511 nm. c-e) Fluorescence on excitation at ~ 400 nm (dashed lines) and ~ 490 nm (solid lines) (497, 498 and 495 for sfGFP^{A148}, sfGFP^{C148}, sfGFP^{V148} respectively). Normalised to excitation of sfGFP^{WT} at 485 nm. f) Image of sfGFP^{X148} protein at 10 μ M. Mutation at residue 148 is annotated. Left to right – sfGFP^{WT}, sfGFP^{A148}, sfGFP^{C148}, and sfGFP^{V148}.

Table 11. Spectral properties of sfGFP^{X148}. [a] Excitation at 400 nm was negligible and so QY was not determined. [b] Value differs from described in Pédélecq *et al.*, (2006).

Variant	λ_{\max} (nm)	λ_{em} (nm)	ϵ (M ⁻¹ cm ⁻¹)	Φ (%)	Brightness	
					(M ⁻¹ cm ⁻¹)	% sfGFP
sfGFP ^{A148}	398	511	36,500	35	12,775	35
	497		15,200	54	8,208	22
sfGFP ^{V148}	399	511	35,000	32	11,200	30
	495		16,500	60	9,900	27
sfGFP ^{S148} [a]	492	511	61,400	90	55,125	150
sfGFP ^{C148} (+DTT)	400	511	34,000	32	10,880	30
	498		10,700	50	5,350	15
sfGFP ^{C148} (-DTT)	400	511	34,000	32	10,880	30
	498		11,100	50	5,550	15
sfGFP ^{WT}	400	511	8,000	N/D	N/D	N/D
	485		49,000 ^[b]	75	36,750	100

For sfGFP^{S148}, the major absorbance peak was at 492 nm was used to calculate a molar extinction coefficient of 61,400 M⁻¹cm⁻¹ (Table 11 a). The fluorescence spectra showed a 50 % increase in emission intensity upon excitation at the λ_{\max} , in comparison to sfGFP^{WT} (Table 11 b). The modelling above suggests that the side chain of serine and its polar hydroxyl group could form a hydrogen bond with the chromophore, not present in sfGFP^{C148}, sfGFP^{A148} and sfGFP^{V148}. It is therefore likely that this hydrogen stabilises the anionic ground state of chromophore, as is the case for sfGFP^{WT}. The difference in QY between sfGFP^{S148} and other variants is likely due to multiple factors including chromophore rigidity, where the hydrogen bond would stabilise its position within the β -barrel, so it samples less thermal motion and allows more energy to be used for light emission. This protein was not used further in this study, as variants with a predominant CroA form were required. However, investigation of this residue replacement is being continued by other members of Prof. Dafydd Jones group.

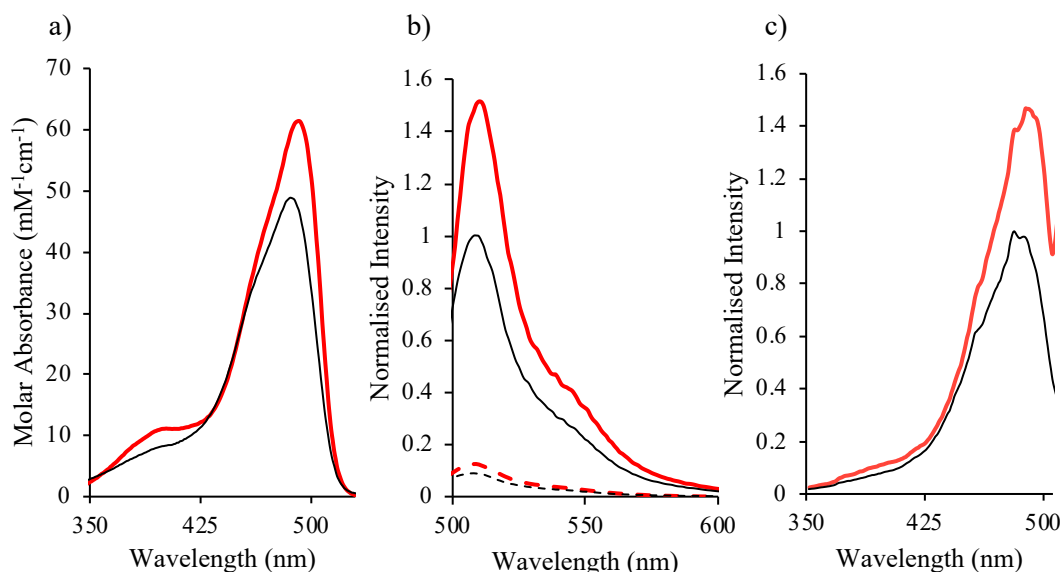


Figure 47. Spectral characteristics of sfGFP^{S148} (red). Normalised against sfGFP^{WT} (black). a) Molar Absorbance. b) Emission spectra upon excitation at 400 nm (dashed line) and 492 nm (solid line). c) Excitation spectra on measuring emission at 511 nm.

4.2.2 Dimerisation of sfGFP^{X148}

The X148 amino acids (valine, alanine, cysteine and serine) were selected for investigation based on their side chain properties (see section 4.2.1 above). Three of the variants (sfGFP^{A148}, sfGFP^{V148} and sfGFP^{C148}) have successfully been identified to have a major absorbance peak at ~400 nm indicative of a predominantly neutral chromophore ground state. Now these proteins will be tested for their ability to form sfGFP dimers at the 148 interface, through formation of hydrophobic interactions, van der Waals, or disulphide bonding. The spectral effects will then be monitored to investigate the potential impact of dimerisation and whether the charged state of the chromophore can be restored as a result of the PPI (i.e. entrapment of water molecules.)

4.2.2.1 Concentration dependent test of sfGFP^{A148}, sfGFP^{V148} and sfGFP^{C148}

To determine whether functional effects occur as a result of dimerising sfGFP^{X148} variants, the concentration of protein was increased up to 100 μM . As dimerisation is a biomolecular process, higher concentrations should promote dimerisation. For the variant to be a useful sensing system, the spectral properties

should change on dimerisation, which should be reflected in changes to observed spectra. The emission intensity upon excitation of ~ 400 and 495 nm was recorded for each variant at 1 and 100 μM . Values were used to calculate the ratio of ~ 495 nm emission over ~ 400 nm emission. That is, if the population of CroB increases, the relative fluorescence intensity increases on excitation ~ 495 nm compared to ~ 400 nm, so the CroB/CroA emission ratio will also increase. This of course is based on the assumption that CroB is promoted on dimerisation. Due to limitations of the Clariostar plate reader (slit width restrictions), the emission was recorded at 516 nm upon excitations at 388 and 483 nm, wavelengths closest to λ_{max} for CroA and CroB. This was kept consistent throughout this work.

The results revealed a ~ 1.6 -fold increase in CroB/CroA ratio for sfGFP^{C148} from 1 μM to 100 μM ($1.3 \rightarrow 2.1$) (Table 12). This suggests that dimerisation of sfGFP^{C148} had played a role in the partial restoration of the deprotonated chromophore state. However, the reverse effect was recorded for sfGFP^{V148} and sfGFP^{A148} with a slight decrease (~ 1.1 – 1.7 -fold respectively) in CroB/CroA ratio upon increasing protein concentration. These results suggest that only sfGFP^{H148C} has the potential to promote dimerisation. In comparison to sfGFP^{V148} and sfGFP^{A148A}, sfGFP^{C148} is capable of forming a covalent interaction by the formation of a disulphide bridge. This is therefore likely to encourage a persistent PPI rather than relying on non-covalent hydrophobic interactions and van der waals with V148 and A148. Assuming a disulphide bond has formed, C148 would provide the reaction handle so increasing selectivity of the protein dimerisation interface around residue 148. Thus, sfGFP^{C148} was taken forward for further investigation.

Table 12. Endpoint fluorescence of sfGFP^{X148} variants at 100 and 1 μ M. Emission was recorded at 516 nm upon excitation at 388 and 483 nm. [a] Gain adjustments for Clariostar plate reader were different for each protein variant.

SfGFP variant	Protein concentration (μ M)	Emission at 516 nm ^[a]		483/388 nm Ratio
		Exc at 388 nm	Exc at 483 nm	
sfGFP ^{V148}	1	266	580	2.2
	100	25402	33059	1.3
sfGFP ^{A148}	1	319	355	1.1
	100	27841	26755	1.0
sfGFP ^{C148}	1	1062	1341	1.3
	100	23722	50275	2.1

4.2.2.2 Addition of CuSO₄ to promote dimerisation of sfGFP by C148

The complex interplay between cysteine and redox-active metals has been used to regulate oxidation of thiol groups (Park and Bauerle 1999; Giles et al. 2003). Worthy *et al.*, (2019) had briefly explored the dimerisation capability of sfGFP^{C148} through Cu²⁺-induced dimerisation. Thus, the use of Cu (II) will be further tested throughout this study, for its ability to promote dimerisation of sfGFP^{C148} by the formation of the disulphide bridge.

To confirm the dimerisation potential and functional properties of sfGFP^{C148}, spectral data was recorded in the presence and absence of CuSO₄ according to section 2.4.4. It should be mentioned, the concentration of CuSO₄ used here depended on the associated protein concentration. For absorbance recordings, protein was used at 10 μ M with 1 mM CuSO₄. However, when recording the fluorescence spectra with lower protein concentrations (0.5 – 2 μ M), 5 mM of CuSO₄ was used. The results showed that during CuSO₄ incubation, absorbance at ~ 400 nm decreased concomitantly with an increase at ~ 490 nm. This results in a shift from 1:2.5 to 1:1 (488:400 nm) after 3 hours (Figure 48 a). Although the absorbance at ~ 400 nm still predominates, suggesting a considerable amount of protein still contains the neutral

chromophore ground state (CroA) and thus presumably remains monomeric. These absorption changes are also reflected in the fluorescence (Figure 48 b). Taking into account both the increase in emission at 488 nm and the decrease at 400 nm, we see a 2-fold change in the 488:400 nm emission ratio at 510 nm after just 60 mins from 0.60 to 1.28. After 22 hours, this change is 5.0-fold (488:400 nm).

To confirm this spectral “switch” reflects sfGFP^{C148} dimerisation, gel electrophoresis was performed using a non-reducing SDS-PAGE gel (Figure 48 c). For the pre-treatment sample (without CuSO₄), bands can be observed the size of monomeric sfGFP (~ 27 kDa). For the sample in the presence of CuSO₄, bands corresponding to dimeric sfGFP (~ 54 kDa) can be observed as time precedes. According to ImageJ, ~ 5 % of sfGFP^{C148} had dimerised after 60 minutes. This suggests that monitoring of wavelengths that reflect on relative populations of the chromophore’s ionic state can be used as a direct reflection of sfGFP^{C148} dimerisation.

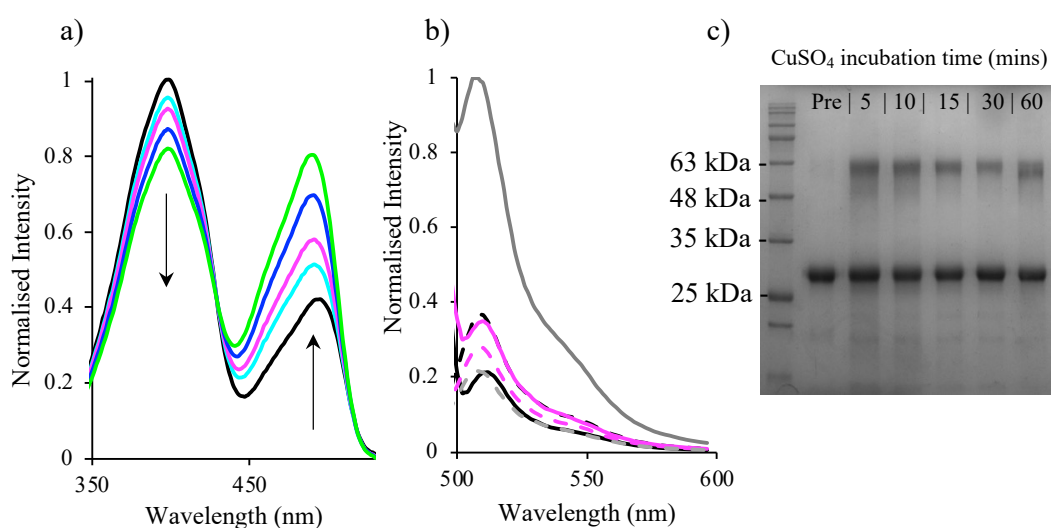


Figure 48. Trial dimerisation of sfGFP^{C148}. Data normalised to sfGFP^{C148} without CuSO₄ (black). a) Absorbance spectra following incubation in 1 mM CuSO₄ for 30, 60, 120, and 180 mins shown in cyan, magenta, blue and green. b) Fluorescence spectra upon excitation at 400 nm (dashed lines) and 495 nm (solid lines) following incubation at 5 mM CuSO₄. Protein concentration at 2 μM following incubation in 1 mM CuSO₄ for 60 mins (pink) and 22 hours (grey). C) Non-reducing SDS-PAGE.

As a control, the other bimodal sfGFP variants (sfGFP^{A148} and sfGFP^{V148}) were also exposed to CuSO₄. However, no significant difference was recorded for the absorbance or fluorescence spectra as a result of CuSO₄ incubation (Figure 49).

This suggests Cu(II) does not alter the spectral properties of sfGFP unexpectedly and that spectral changes are detected as a result of Cu(II) promoting the formation of the disulphide bridge between sfGFP^{C148} dimers.

Overall, the dimerisation of sfGFP^{C148} has evidenced its ability to ‘switch’ protein function. However, structural evidence is required to determine the exact mechanism for restoring the deprotonated ground state of the sfGFP^{C148} chromophore. Also, the residual 400 nm absorbance, and lack of complete dimerisation could pose problems as a homo-oligomerisation sensing system. These factors will be addressed in Chapter 5.

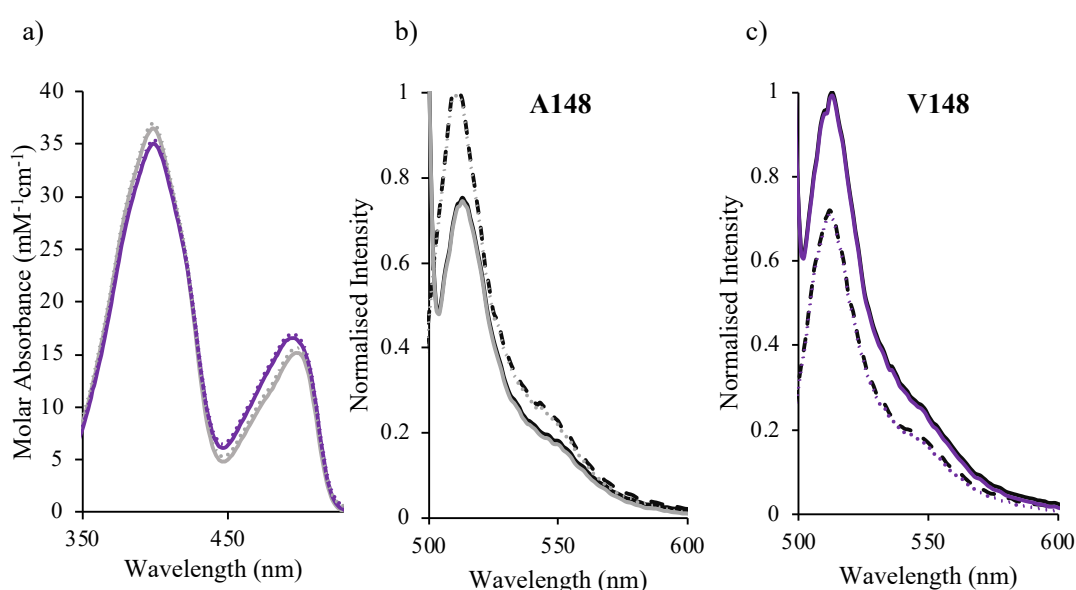


Figure 49. Spectral characteristics of sfGFP^{A148} and sfGFP^{V148} in the presence and absence of CuSO₄. a) Absorbance spectra without (solid line) and with (dashed line) CuSO₄. SfGFP^{A148} shown in grey and sfGFP^{V148} shown in purple. b) Emission spectra of sfGFP^{A148}. Without (black line) and with (grey line) CuSO₄. Excitation at 400 and 490 nm are shown in dotted and solid lines respectively. c) Emission spectra of sfGFP^{V148}. Without (black line) and with (purple line) CuSO₄. Excitation at 400 and 490 nm are shown in dotted and solid lines respectively.

4.2.3 X-ray crystallography of sfGFP^{C148.F} monomer

To determine the structural modifications of incorporating cysteine into position 148 of sfGFP, the structure was solved by x-ray crystallography. As the monomeric structure was required, no copper sulphate was added to the sample. Crystals were harvested and sent to Diamond light source to employ x-ray scattering

and diffraction. For structure solution, molecular replacement was performed using 2B3P (sfGFP^{WT}) as a model, incorporating H148C mutation as required. As well as H148C, valine was also substituted for phenylalanine at position 206; the rationale for this will be addressed in Chapter 5. This protein will now be referred to as sfGFP^{C148.F}. However, it seems logical to address the structure of the monomeric state here in this chapter given that the focus of the chapter is on the monomers.

4.2.3.1 Crystal refinement

Crystals of sfGFP^{C148.F} in the P 1 2₁ 2 space group grew in condition G12 on the PACT Premier screen (0.2 M Sodium malonate dibasic monohydrate, 0.1 M Bis-Tris propane, pH 7.5, 20 % w/v PEG 3350) after 14 days of incubation at 20 °C. The crystals diffracted to the highest resolution of 1.89 Å. The structure went through multiple rounds of refinement to reach R-factor and R_{free} values of 0.206 and 0.247 % respectively. Full diffraction data and refinement statistics can be found in Table 13.

Table 13. Crystallographic statistics from x-ray diffraction data and final bond parameters for crystal structure of monoclinic sfGFP^{C148.F}. PDB ID: 8C1X

SfGFP^{C148.F} Monoclinic	
Data collection/reduction statistics	
Wavelength (Å)	0.81532
Beamline	I03
Space group	P 1 2 ₁ 2
a (Å)	67.49
b (Å)	73.76
c (Å)	127.28
Resolution range (Å)	39.25 – 1.89
Total reflections measured	707,104
Unique reflections	100,181
Completeness (%) (last shell)	100 (100)
I/σ (last shell)	10.4 (0.3)
R(merge)(%) (last shell)	9.3 (329.2)
B (iso) from Wilson (Å ²)	35.3
Refinement statistics	
R-factor ^b (%)	0.206
R-factor ^c (%)	0.247
RMSD bond lengths (Å)	0.0121
RMSD bond angles (°)	1.806
Ramachandran Plot statistics	
Favored region (%)	93.0
Allowed region (%)	6.4
Disallowed region (%)	0.6

Analysis of the contents of the crystal cell showed there were 4 molecules in the unit cell (chains A-D) (Figure 50). The structure was submitted to PDBePISA; an online interactive tool for the exploration of macromolecular interfaces (<https://www.ebi.ac.uk/pdbe/pisa/>). The results identified a protein interface area of $\sim 117 - 315 \text{ \AA}^2$ between proteins in the asymmetric unit, i.e. A-B, A-C, B-D and C-D. However, each of the interfaces scored 0.0 on the complex significance score (CSS) so determined these associations to be insignificant for assembly formation.

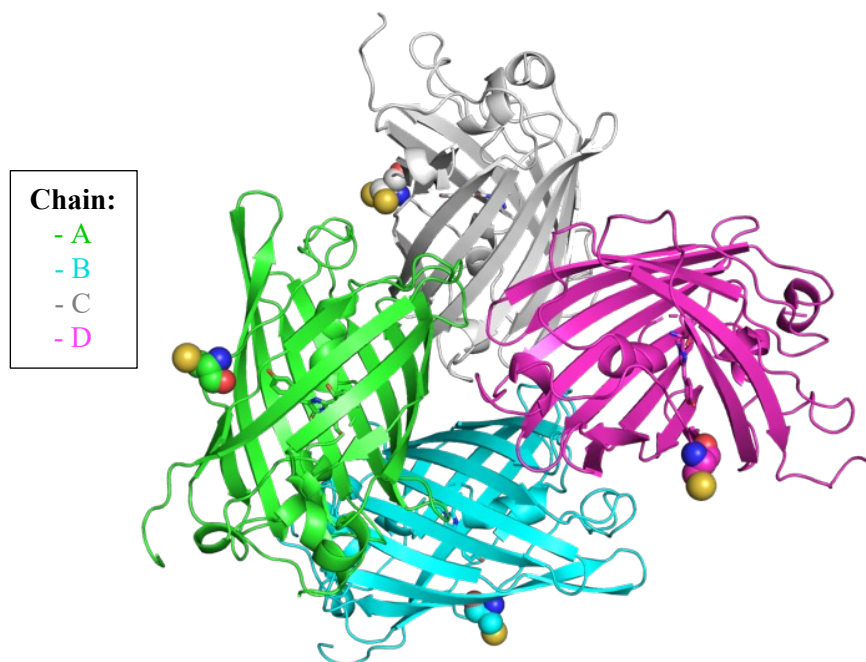


Figure 50. Cartoon representation of molecules making up the unit cell. The unit cell comprises chain A (green), B (cyan), C (grey), and D (magenta). Cysteine residues at position 148 are highlighted as spheres.

4.2.3.2 Structure of monomeric sfGFP^{C148.F}

The structure of chains A-D present within the asymmetric unit were analysed. Structural alignment of chains A, B, C, and D within PyMOL yielded an RMSD score $< 0.21 \text{ \AA}$. This suggests that there are minor structural differences between chains. The structural alignment revealed that all chains maintain the central β -barrel structure with no major structural differences to the overall protein scaffold (Figure 51a). The chromophores align perfectly showing no perturbations within the core of the protein (Figure 51b). Residue F206 occupies one conformation across all four chains (Figure 51c). There is one conformation of residue C148 for chains A, B and D where the side chain faces away from the chromophore (Figure 51d). For

chain C, two possible rotamers for the C148 are fit to the observed electron density, where there is a $\sim 120^\circ$ rotation about $C\beta$.

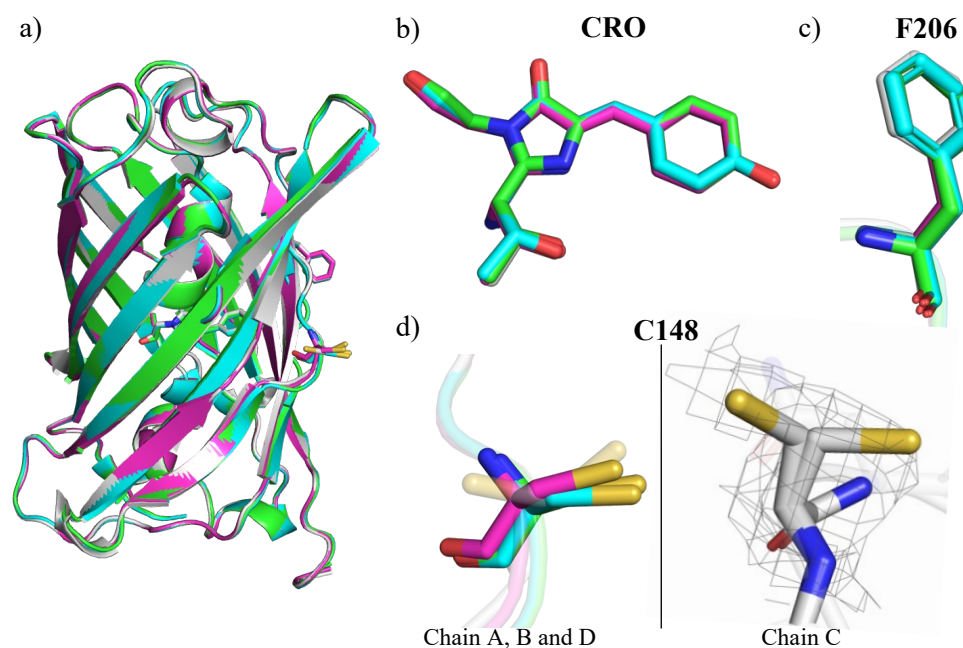


Figure 51. Alignment of sfGFP^{C148.F} monomers. A) Cartoon representation of chain A (green), chain B (cyan), chain C (grey), and chain D (magenta). B) Stick representation of chromophore. C) Residue F206. d) Residue C148 showing 2Fo-2Fc electron density map at 1.0 sigma.

4.2.3.3 Comparison of sfGFP^{C148.F} to sfGFP^{WT}

To determine the key structural differences upon incorporation of cysteine and phenylalanine into residue positions 148 and 206 of sfGFP respectively, comparisons will be drawn against sfGFP^{WT}. As all chains of sfGFP^{C148.F} are similar, chain A will be used for comparison. Where there are differences between chains of the variant, these will be highlighted.

Structural alignment of sfGFP^{WT} onto chain A yielded a RMSD score of 0.424 Å (0.415, 0.294, 0.296 Å for chains B, C and D respectively). This suggests that overall, the structures are similar. Upon examination, the backbone of the loop housing residue C148 extends further away from the chromophore than for sfGFP^{WT} (Figure 52 a). This is likely to be caused by the conformation of C148, where side chain faces away from the central chromophore. This causes the subsequent loss of the hydrogen bond originally present with H148 and would likely increase flexibility

of the β -strand (Figure 52 b). There is a $\sim 120^\circ$ rotation of S147 where the hydroxyl group in sfGFP^{C148.F} is of appropriate distance (2.7 Å) to participate in hydrogen bonding with the chromophore (Figure 52 c). There is a slight offset in position of chromophores, with a 0.7 Å difference between the OH group of Tyr66 (Figure 52 d). Residue T203 undergoes a $\sim 107^\circ$ rotation and 1.0 Å shift in C β position (Figure 52 e). There is a 0.8 Å offset in positioning of residue E222. For chains B, C and D, there is also a rotation of E222 side chain (Figure 52 f). The phenylalanine residue at position 206 is similarly orientated to valine in sfGFP^{WT}, where the aromatic ring protrudes away from the central core of the protein (Figure 52 g).

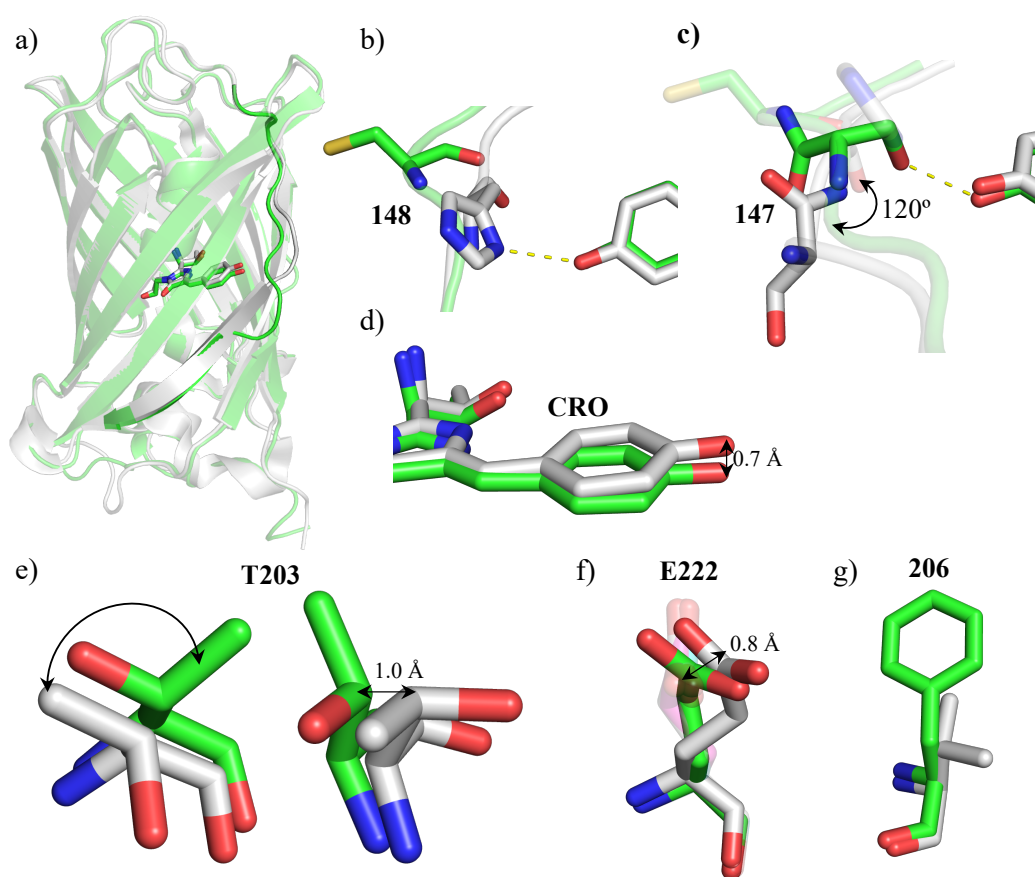


Figure 52. Structural differences between sfGFP^{C148.F} and sfGFP^{WT}. Chain A of sfGFP^{C148.F} shown in green and sfGFP^{WT} shown in grey. Cartoon representation highlighting β -strand 7 housing residue C148. b) Residues 148. C) Residue 147. D) SfGFP chromophores. E) Residue T203. F) Residue E222. Chains B, C and D shown in background. G) Residue 206.

The hydrogen bond network surrounding the chromophores was then compared. For sfGFP^{WT}, Tyr66 is involved in hydrogen bonding to H148, a structured water molecule (W1) and T203 (Figure 53a). Whereas for sfGFP^{C148.F}, hydrogen bonding of Tyr66 occurs with S147 and W1 (Figure 53b, c). However, given the spectral properties of the monomeric sfGFP^{C148}, this appears insufficient to promote the formation of the phenolate CroB. The $\sim 107^\circ$ rotation of T203 results in an increased distance (4.2 Å) from the chromophore and thus a loss of hydrogen bonding capability. T203 plays a critical role stabilising the charge of the phenolate.

Another residue critical in defining the protonated state of the chromophore is E222 – it acts as the acceptor for the proton that is lost by the hydroxyl group of Tyr66. For chain A of sfGFP^{C148.F}, the alternative conformation of E222 increases the distance between S205 and E222 to 4.0 Å (Figure 53 b). This could disrupt the proton shuttling system, or ESPT, responsible for maintaining the chromophores deprotonated state. However, the remaining key residues and structured water molecules (W1, S205 and E222) of the ESPT are within appropriate distance for hydrogen bonding within chains B, C and D of sfGFP^{C148.F} (Figure 53c).

The combination of structural changes within sfGFP^{C148.F} (particularly the loss of H-bonds to H148 and T203) are likely to destabilise the deprotonated state of the chromophore. Thus, the resulting CroA predominates in the ground state as the hydroxyl group of Tyr66 is protonated, giving rise to the major absorbance peak at 400 nm.

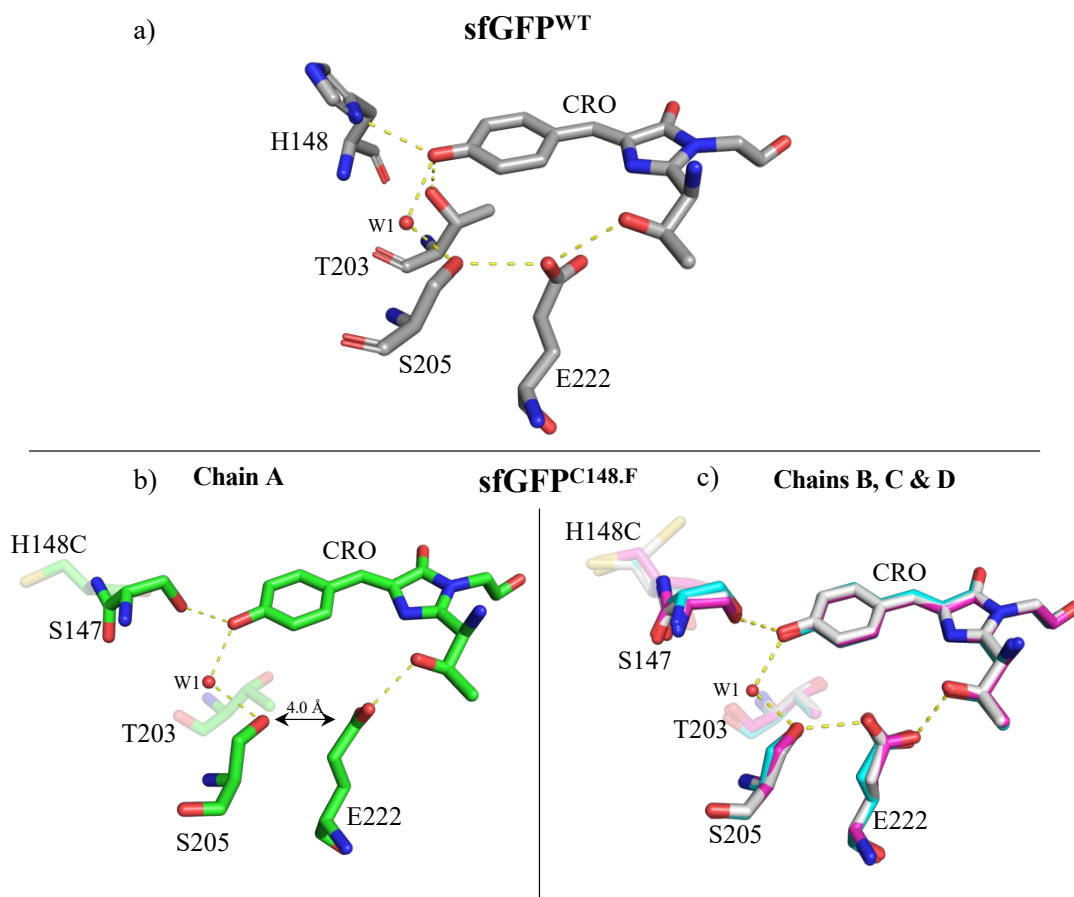


Figure 53. Comparison of hydrogen bond network surrounding the chromophore for $\text{sfGFP}^{\text{C148.F}}$ and sfGFP^{WT} . Key residues shown as sticks with hydrogen bonds shown in yellow. a) sfGFP^{WT} . b) Chain A of $\text{sfGFP}^{\text{C148.F}}$. c) Chains B, C, D of $\text{sfGFP}^{\text{C148.F}}$.

4.3 Discussion

In this chapter, residue 148 of sfGFP was selected as a key target position, due to its proximity and contribution towards the hydrogen bonding network surrounding the chromophore. Together with T203, S205, E222, and structured water molecules, H148 is an integral component of the ESPT, and stabilises the deprotonated charge of the chromophore, giving rise to the 490 nm absorbance peak. This chapter evidenced how replacement of H148 with either a cysteine, valine, alanine, or serine significantly effects the spectral properties of sfGFP. By investigating each of the residue's properties, we can hypothesise the effect on hydrogen bonding, and local structural effects based on each of the functional implications.

Despite extensive engineering efforts, H148 remains well conserved within the *Aequorea Victoria* GFP-like proteins. Thus, one of the most surprising findings within this chapter was the discovery of a 50% brighter sfGFP variant by the incorporation of serine into position 148. The question therefore arises as to why replacement of H148 with a serine has improved both absorbance and fluorescence properties of sfGFP? In terms of how S148 is imparting its beneficial effects, we can speculate as to this structural role. Firstly, the hydrogen bond with residue 148 is deemed critical for stabilising the deprotonated state of the chromophore, giving rise to the maximum absorbance peak at ~ 490 nm. For a hydrogen bond to form, the distance between the donor and acceptor atoms is generally considered to be from 2.7 to 3.1 Å (Lehrman 2017). Following energy minimisation of S148 (Figure 42), the OH side chain is within a 2.8 Å distance from the chromophore, whereas the nitrogen (ND1) atom of H148 is at a distance of 2.9 Å apart. Considering oxygen is more highly electronegative than nitrogen and at a closer distance, a stronger hydrogen bond is therefore likely to form with the chromophore. This could result in a more persistent conformation of S148; which is another factor to consider given the evidence that H148 can occupy two alternative configurations, where the hydrogen bond is lost as a result of the “open” formation (Seifert et al. 2003). Other considerations include the geometry of the hydrogen bond where the most frequently observed geometry is between 90° and 180° for the donor-hydrogen-acceptor (Hubbard and Kamran Haider 2010). Although, the exact configurations of S148 are currently unknown without a crystal structure.

The quantum yield of a protein relates to the efficiency at which a fluorescent molecule converts absorbed photons into emitted photons (Wall et al. 2014). It largely depends on the local environment of the chromophore and its associated dynamics and flexibility. For the high quantum yield of sfGFP^{S148}, the local environment of the chromophore should reduce its dynamics and associated flexibility. A hydrogen bond provided by S148 is likely to stabilise the position of sfGFP chromophore. This may also impact local structural changes, bringing surrounding residues (e.g. F145, T203) closer to the chromophore for better packing.

By comparison to other FP related structures in the protein databank, it appears that within other species, serine is fairly well conserved at this corresponding position. For example, S142 (equivalent to residue position 148 in

GFP^{WT}) is a critical residue in Dronpa, a photochromatic fluorescent protein with a high quantum yield ($\Phi = 0.85$), originally discovered from a species of coral (*pectiniidae*) (Ando et al. 2004). In its bright state (termed Dronpa^B), there is a hydrogen bond between the chromophore hydroxyl oxygen and S142 hydroxyl group that is critical for tethering the chromophore to the β -barrel and maintaining its deprotonated state. By contrast, in the dark state (termed Dronpa^D), the protonation of the chromophore hydroxyl moiety results in the loss of hydrogen bond with the S142 and subsequent increase in flexibility to both the chromophore and the β -barrel (Mizuno et al. 2008). The replacement of S142 with other residues (ala, cys, asp, or gly) generated Dronpa^D-like variants, which further highlights the requirement for S142 for the bright fluorescence of Dronpa. A serine is also present at this equivalent position in EosFP a novel FP discovered from scleractinian coral *Lobophyllia hemprichii*, red fluorescent protein from *Discosoma* (DsRed), and yellow-fluorescent protein (zFP538) from *Zoanthus* (Nienhaus et al. 2005; Remington et al. 2005; Tubbs et al. 2005). Within each of these structures, serine forms a key hydrogen bond with the chromophore's hydroxyl group. Thus, as a fairly conserved residue within other species, it raises the question as to how it hasn't replaced H148 within engineered *Aequorea Victoria* GFP-like proteins. This may be due to the way most avGFP variants are generated and the underlying genetic code. Many FP variants like sfGFP are generated by directed evolution; libraries of variants are constructed followed by selection (Cormack et al. 1996; Reddington et al. 2015). During library construction random mutations are normally incorporated into the gene sequence, mostly through single nucleotide changes to relatively few codons. Histidine is encoded by 2 codons: CAU/C and serine has 6 codons: UCN and AGU/C. For His to sample Ser, at least two contiguous nucleotides need to be mutated which is statistically unlikely to occur during random mutagenesis approaches using during directed evolution so has probably never been sampled. It may also be that the sfGFP sequence and structure may promote the effects of S148 that other avGFP variants cannot. Thus, the H148S mutation may not have a general effect. Prof Dafydd Jones' group are currently exploring this by transplanting the H148S mutation into a variety of avGFP variants, including the commonly used enhanced GFP (EGFP).

For sfGFP variants containing valine, alanine and cysteine at position 148 of sfGFP, a bimodal absorbance spectrum was recorded with a significant peak at 400 nm. As valine and alanine are non-polar, they are incapable of hydrogen bonding. As a result, the hydrogen bond originally present between H148 and the chromophore is presumably lost. Excited state proton transfer still occurs as the main emission peak is equivalent to the deprotonated chromophore but now the ground state is dominated by a mixture of CroA and CroB forms. However, deprotonation of the chromophores ground state equilibrium is hindered; causing the neutral, protonated form to predominate, accounting for the major ~ 400 nm absorbance. The key question then arises as to the functional properties of sfGFP^{C148}, and why it has a lower absorbance at ~ 490 nm compared to sfGFP^{A148} and sfGFP^{V148}. Despite the relatively low electronegativity for sulphur (2.58 on the Pauling scale), the side chain of cysteine is weakly polar and thus, can participate in hydrogen bonding (Liu et al. 2019). The crystal structure of sfGFP^{C148.F} revealed a twist in the backbone of β -strand 7, causing residue C148 to face away from the β -barrel. This could be due to the atomic radius of sulphur, as it is greater than that of an oxygen atom. This could therefore result in steric clashes with the chromophore, causing repulsion of the cysteine side chain so that it points outwards away from the β -barrel. In replacement, S147 is now within appropriate distance to form a hydrogen bond with the chromophore. Depending on its associated geometry, it may still be questionable as to if S147 is capable, or as efficient as H148, in stabilising the chromophores phenolate form. Further detailed analysis revealed that one of four chains (chain A) of sfGFP^{C148.F} had an increased distance between S205 and E222, two key residues involved in the ESPT. At 4.0 Å apart, these residues can no longer participate in hydrogen bonding, causing disruption to the proton shuttling network associated with W1, S205, E222 and T65. Thus, in combination with the loss of contribution from residue H148 and T203, the charged phenolate is destabilised. The chromophore would therefore remain in its predominantly, neutral, CroA form in the ground state. However, chains B, C and D can H-bond to key residues of the ESPT (similarly to sfGFP^{WT}) and contain a H-bond with S147. Overall, considering the spectral properties, the structure of sfGFP^{C148} can explain the mixture of chromophore ground states giving rise to 400 and 488 nm absorbance peaks.

4.4 Conclusion

Understanding how specific residues contribute to the structure-function relationship is important to altering protein function through mutation. Residue 148 of sfGFP is located within close proximity to the chromophore and plays an important role stabilising the deprotonated charge of the chromophore. (Zhang et al. 2005). This chapter has shown how replacement of H148 can be used to modulate the function and spectroscopic properties of sfGFP. The production of sfGFP^{C148}, sfGFP^{A148} and sfGFP^{V148} variants suggest the hydrogen bond cannot be formed with the GFP chromophore as is present within sfGFP^{WT}. The resulting absorbance spectra shows a ~ 90 nm blue shift, indicative of the chromophore charged state predominantly present in the neutral form. Whereas the replacement of the imidazole ring of H148 to a short chain hydroxyl containing amino acid, as for sfGFP^{S148}, has given rise to a brighter sfGFP variant with the favourable anionic chromophore state.

The ability to control protein function through a dimerisation-induced approach of sfGFP^{C148} has been demonstrated. The changes in the dominant chromophore ground state are assumed to occur as an indirect result of local interactions. Further structural work is required to deduce the exact mechanism for restoring the more favourable deprotonated chromophore, which will be addressed in Chapter 5.

5 Modulation of sfGFP structure and function by homodimerisation

5.1 Introduction

Protein oligomerisation is central to biology with more proteins existing in complexes than as monomers (Marianayagam et al. 2004; Mei et al. 2005). One of the most important features of protein oligomers is cooperativity between the individual units. Functional cooperativity is achieved through multi-subunit interactions whereby communication between each polypeptide unit can positively or negatively regulate activity, or even result in new properties (see section 1.4 for more information). As protein oligomers provide such useful functional and structural features, there has recently been a push to design and engineer novel systems. Based on similar principles that guide formation of natural proteins oligomers, the development of novel approaches and technologies has been used to construct novel protein assemblies; this encompasses simple de novo design systems (Watkins et al. 2016; Thomas et al. 2018; Hicks et al. 2022) to linking whole functional proteins (Worthy et al. 2019; Johnson et al. 2021; Pope et al. 2021). Designed oligomers have exhibited several useful features such as; extreme stability, designability of 3D topologies and folding pathways. More recently, this have moved beyond simple passive scaffolding (i.e. proteins assemble but still act as independent units) to truly synergistic systems whereby the “sum is greater than the parts” (Gwyther et al. 2019; Worthy et al. 2019). They also show great prospects for several applications as functional devices for biology, biomedicine and synthetic circuits (Zhou et al. 2020).

As briefly mentioned in section 4.1, a recent approach by Worthy *et al.*, 2019 demonstrated how genetically encoded bioorthogonal chemistry can be used to generate artificial protein complexes with enhanced function. Two separate ncAA's (azF and SCO) were incorporated into position 148 of GFP as mutually compatible reaction handles for the construction of protein dimers by strain promoted azide-alkyne cycloaddition (or SPAAC)(Figure 54 a). On dimerisation, 148azF was modified through formation of the triazole link with 148SCO in the cognate monomer. This resulted in a change in backbone and side chain position of residue

148, causing a hole that was then occupied by a water molecule in the dimer (W1) (Figure 54 b). This water molecule forms H-bonds to the chromophore (replacing the H-bond lost on removal of H148), activating the dimer through promotion of the chromophores anionic ground state (CroB). Dimers displayed a positive switch in protein function, with increased brightness (Figure 54 c). Dimerisation also generated a long-range interaction network where three symmetrical water molecules formed an extended inter-monomer water rich hydrogen bond network. This work excellently demonstrates the potential for engineered oligomers to behave as oligomeric biosensing systems. This also highlights a ratiometric approach, meaning it is concentration-independent, dependent on the relative populations of CroA and CroB. However, this work currently requires the use of two separate reaction handles, meaning it is limited to heterodimeric sensing. There are also many hurdles to overcome before ncAA can be used in the biomedicine field; including its effect on protein expression which is especially important during multiple-site incorporation (Gao et al. 2019). Other challenges include low-efficiency incorporation, cross-reaction with endogenous elements and the associated high costs.

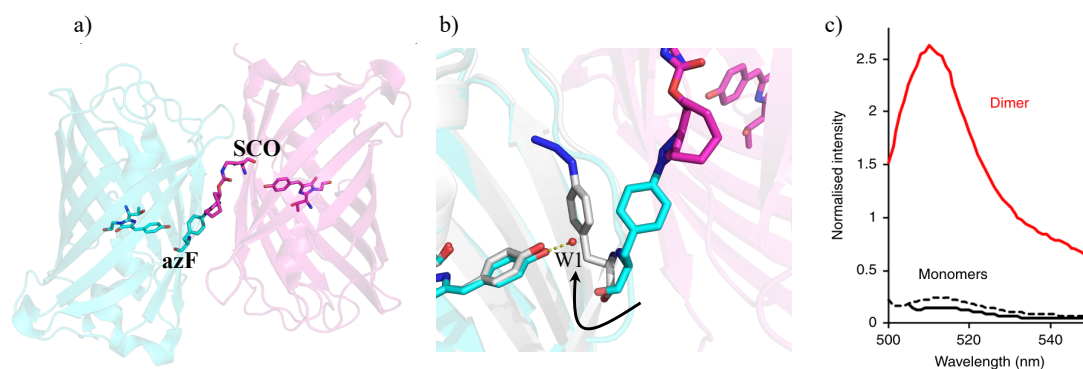


Figure 54. Dimerisation of sfGFP by ncAA at residue 148. a) Dimerisation by ncAA's sfGFP^{148azF} (cyan) and sfGFP^{148SCO} (magenta). b) Structured water molecule within ncAA's linked dimer. SfGFP^{148AzF} monomer in grey. c) Emission spectra of ncAA monomers (black) and dimer (red). PDB ID: 5NHN

Given that majority of oligomeric states are homodimer the monitoring of homodimerisation is a critical event. In this chapter, I will focus on generating sfGFP homodimers that switch function, using chemistry that is available within the natural amino acid repertoire. This will not only bypass the restrictions required with genetically encoded ncAA's, but also remove the use for two separate reaction

handles. Thus, the dimerisation of two identical sfGFP monomer units will allow the monitoring of homodimers.

5.2 Results

5.2.1 Design of mutations to promote dimerisation of sfGFP^{C148}

Throughout Chapter 4, three sfGFP H148X variants were identified (containing valine, alanine, and cysteine) that each had a major absorbance peak at 400 nm, indicative of the neutral chromophore ground state. Of these variants, sfGFP^{C148} demonstrated its ability to dimerise via the formation of disulphide bridges between residue 148 in the corresponding monomer. The use of CuSO₄ promoted the formation of disulphide bridge. However, the reaction was slow and failed to reach completion. Thus, it appears that the H148C mutation alone is not sufficient to promote full dimerisation, at least on useable timescales required to monitor biological events. To encourage the formation of sfGFP^{C148} dimers, additional mutations were designed and incorporated within the protein interface.

As detailed above (section 5.1), the ncAA containing dimer (sfGFP^{148x2}) solved by Worthy *et al.*, (2019) contained an elongated triazole crosslink between residues 148 of sfGFP. With the reaction handle at the same position (residue 148), a similar interface is likely between sfGFP^{C148} proteins. The structure determined for sfGFP^{148x2} was therefore used as a template to select additional residue positions and appropriate residue replacements to encourage and stabilise sfGFP dimerisation. To highlight possible residue targets, the sfGFP^{148x2} structure was analysed using PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>). The results highlighted regions of the protein that are present at the interface including: β -strand 7 containing ncAA's (residues 142 – 149), β -strand 10 (residues 202 – 207) and β -strand 11 (residues 221 and 223) (Figure 55a). As the covalent link between the ncAA's is longer than for a disulphide bridge, the positioning and exact interactions of β -strand 7 at the interface are likely to alter slightly. Thus, investigations into β -strand 10 were next considered. Of the six residues comprising β -strand 10, V206 was targeted as this residue was at the interface in the ncAA linked dimer structure and previous reports of mutants at this residue promote full monomerisation of GFP (Valbuena *et al.* 2020)(Zhang *et al.* 2002)(Ilagan *et al.* 2010). With regards to the latter, V206 could be used to promote dimerisation rather than monomers. Where monomers are

promoted through the introduction of charged residues (e.g. V206K), dimers could instead be promoted through residue 206 by interactions such as hydrophobic interactions, pi stacking, and van der waals.

To determine which residue will be used for replacement at position 206, multiple factors were considered including; side chain polarity, size, hydrophobicity and surrounding residues. Considering the structure of sfGFP^{148x2}, neighbouring residues of V206 lacked side chain polarity, ruling out hydrogen bond capability. However, the short side of V206 highlighted the available space and potential ‘gap’ in the interface for a bulkier residue, possibly to suit an aromatic residue (Figure 55 a). The nearby residue F223 and its potential to form pi-pi attractions then became apparent. The mutation of V206 with aromatics residues was therefore considered; replacing valine for either a phenylalanine, tryptophan or tyrosine.

To model residue replacements, *in silico* mutagenesis was performed on V206 using PyMOL. Of the aromatics (Phe, Tyr and Trp), phenylalanine offers the least bulky side chain, whereas tryptophan is likely too large and would likely lead to steric clashes within the interface. Also, tyrosine and tryptophan are equipped with heteroatoms capable of forming hydrogen bonds with water. This may therefore require the removal of water before interface formation and thus potentially hindering dimerisation. Given this, phenylalanine was selected for incorporation into residue 206, as it would most likely fit between the interface, without disrupting favourable interactions of surrounding residues within the interface. Of the possible conformations suggested in PyMOL, rotamer 2 of V206 (with a suggested 30-43 % probability) was selected for both chain A and B, to avoid residue overlap and the least steric clashes, whilst still offering cross-section interactions. The large side chain of V206F aided extension across the interface contributing to a distance of ~ 3.2 – 3.8 Å to F223 of the opposing chain (Figure 55 b), within the contact limit for previous reporting’s of pi-pi interactions (Chen et al. 2018). It is noted however, that the azide-alkyne triazole crosslink between residues 148 will be replaced with a shorter covalent link by disulphide bridging. The interface will therefore be expected to be slightly altered. It is still likely that V206F can still form attractive forces with F223 or F206 of the opposing chain with its altered position. There were also five possible conformations of V206F available in PyMOL, so the orientation and configuration of its side chain could alter to optimise inter-chain interactions. Based on this evidence, the V206F mutation was decided to incorporate into sfGFP^{C148} to

aid dimerisation and add structural support at the interface. For simplicity, this variant will now be referred to as sfGFP^{C148.F} where F is at position 206.

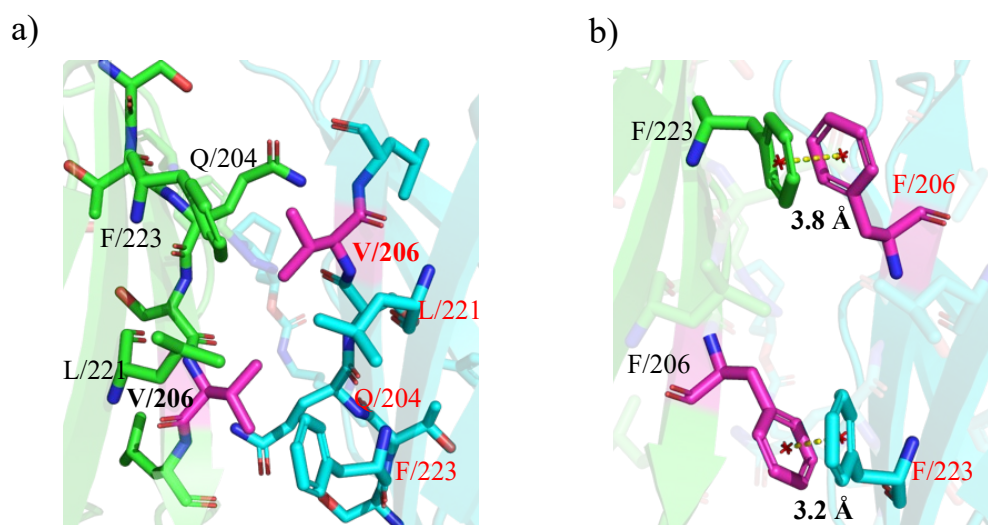


Figure 55. Stick representation of sfGFP^{C148x2} interface. Chain A in green and chain B in cyan. Residue 206 highlighted in magenta. A) Residues at interface. b) Pi-pi interaction network surrounding V206F residue.

5.2.2 Production of sfGFP^{C148.F} monomer

5.2.2.1 Site directed mutagenesis

To incorporate a phenylalanine mutation at residue position 206, the sfGFP^{C148} plasmid was used to perform whole vector PCR with primer pairs listed in Table 3 (Figure 56 a). Agarose gel electrophoresis confirmed amplification of DNA of the approximate size of the pBAD plasmid, containing sfGFP (~ 4.7kb; Figure 56). The DNA was purified, phosphorylated, and ligated. The circularised plasmid was then transformed into *E.coli* and plated onto LB agar plates containing antibiotic and 0.1 % arabinose. The agar plates were visualised under UV light to allow selection of green coloured colonies (Figure 56 c). The final construct and successful incorporation of V206F into sfGFP^{C148} was confirmed via sequencing.

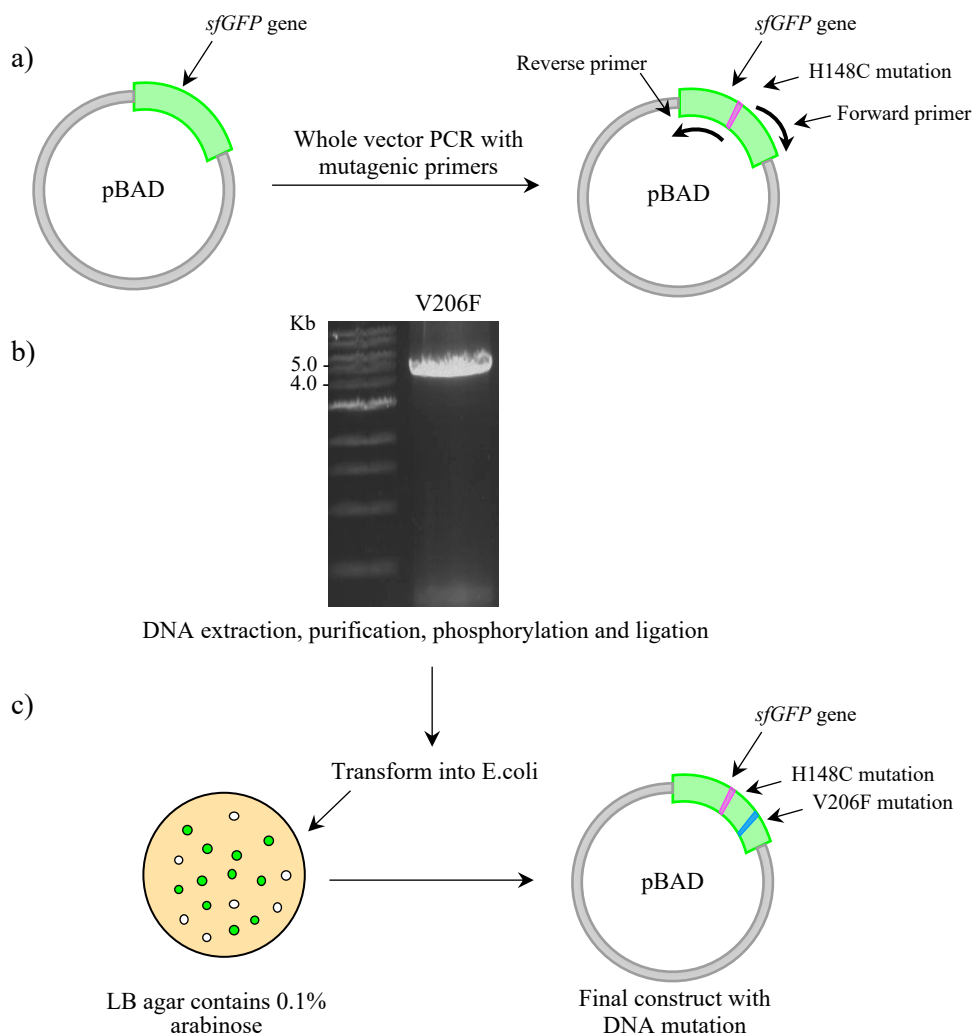


Figure 56. Site directed mutagenesis of sfGFP^{C148} to incorporate phenylalanine mutation at residue position 206. A) Whole vector PCR with mutagenic primers. B) Gel electrophoresis. C) DNA ligation and transformation into *E. coli*.

5.2.2.2 Expression and purification

Before undertaking detailed analysis, sfGFP^{C148.F} variant had to be expressed and purified. This was performed similarly to that for sfGFP^{C148} (section 4.2.1.2) where the protein is expressed and then purified firstly by nickel affinity followed by size exclusion chromatography. A non-reducing gel confirmed the presence of bands of the expected size of monomeric GFP (~ 27 kDa) and dimer (~ 54 kDa) as can be observed in Figure 57a. For characterisation of sfGFP^{C148.F} monomer, the residual dimer was removed by a second purification by SEC carried out with DTT (2 mM). The non-reducing gel confirmed the presence of bands of the expected size of monomeric GFP (~ 27 kDa) (Figure 57b), and the absence of dimer bands. It is

likely that the extremely high protein concentration had encouraged the formation of dimers by C148 disulphide bridging.

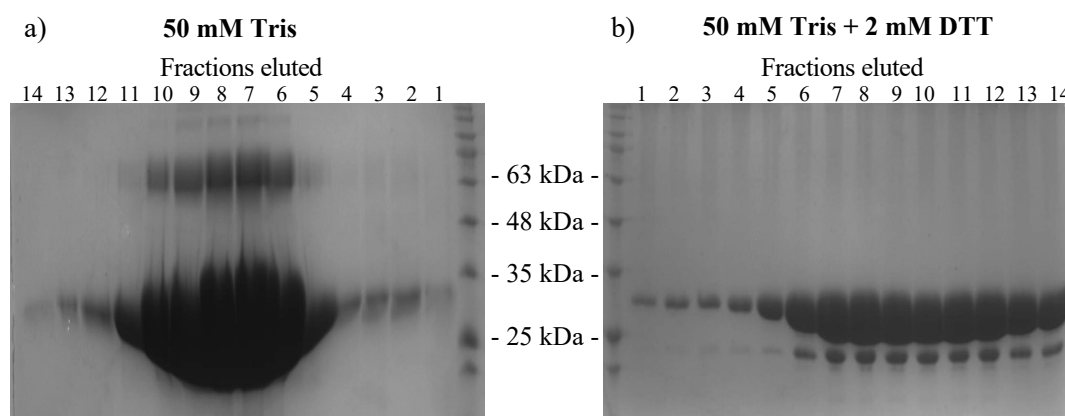


Figure 57. Non-reducing SDS-PAGE analysis of sfGFP^{C148.F} following SEC. Fractions labelled in order of their elution. a) Fractions eluted from SEC in 50 mM Tris, pH 8.0. b) Fractions eluted from SEC in 50 mM Tris, pH 8.0, 2 mM DTT.

5.2.2.3 Spectroscopic Characterisation

The sfGFP^{C148.F} variant was then characterised both with and without DTT, to reduce potential disulphide bridging between C148 residues (Figure 58). The results showed a bimodal absorbance spectrum with peaks at both ~ 400 and 495 nm (Figure 58 a). The molar extinction coefficient was *circa* 13,000 M⁻¹cm⁻¹ at 495 nm regardless of the presence of DTT. However, there were subtle differences in the molar extinction coefficients calculated at 400 nm. This suggests that sfGFP^{C148.F} was susceptible to reducing conditions. It is likely that a minor proportion of sfGFP^{C148.F} had dimerised via residue C148 through the formation of a disulphide bridging. On reduction, the disulphide bonds would break resulting in monomeric sfGFP^{C148.F}. Any potential hydrogen bonding with the chromophore that had occurred as a result of dimerisation (e.g. entrapment of water molecules) may then be lost, causing the destabilisation of the anionic chromophore, and the resulting shift over to CroA. Although, these absorbance changes are minimal and majority of protein is still assumed to be monomeric.

The fluorescence spectra for sfGFP^{C148.F} shows a maximum emission peak at 511 nm upon excitation at 400 nm (Figure 58 b). Compared to sfGFP^{WT} with an optimum excitation wavelength at 485 nm, a 60 – 70 % loss of emission was recorded at the less favourable 400 nm excitation (although this is largely due to the decrease in quantum yield, Table 14).

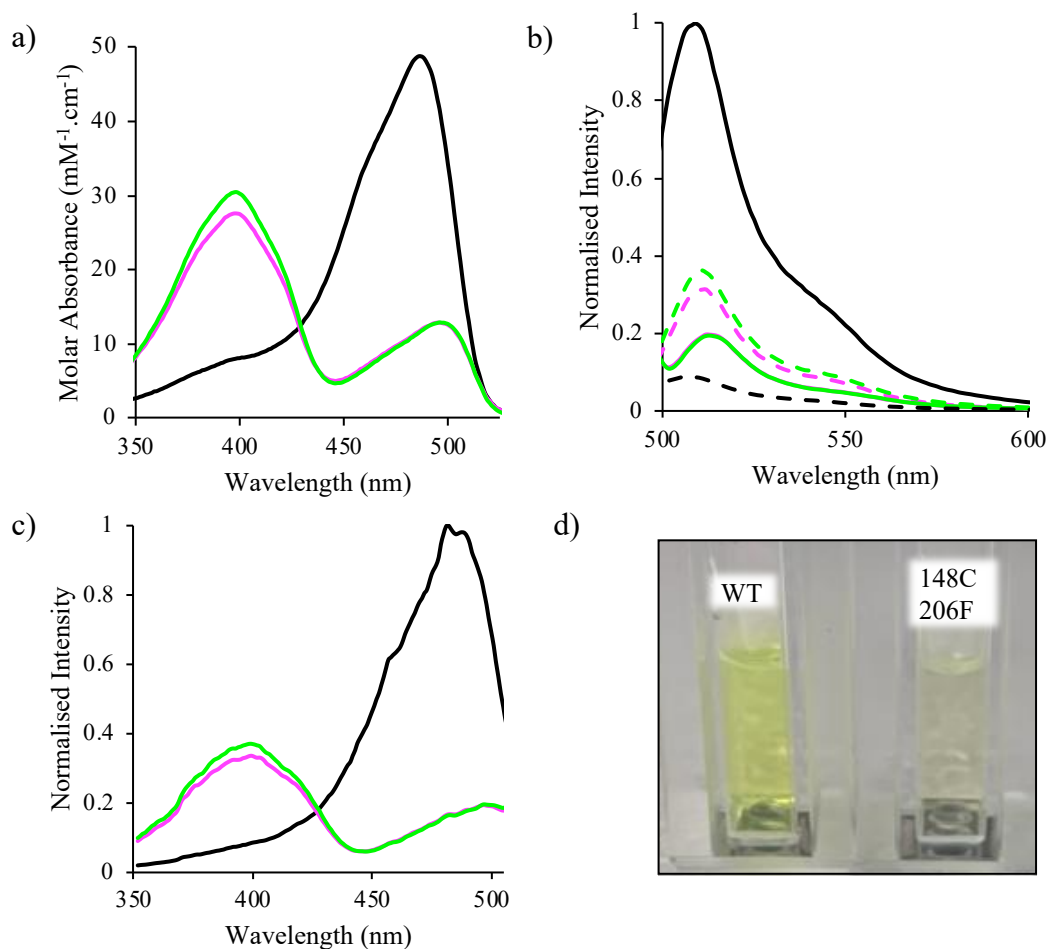


Figure 58. Spectral properties of sfGFP^{C148.F} variant. In all cases, the lines in a-c represent sfGFP^{C148.F} with 1 mM DTT (green line), without DTT (magenta) compared with sfGFP^{WT} (black line). a) Absorbance spectra. b) Fluorescence on excitation at 400nm (dashed lines), 485 and 495 nm (solid lines) for WT and C148.F variant respectively. c) Excitation spectra upon emission at 511 nm.

5.2.2.4 Comparison of monomeric characteristics of sfGFP^{C148.F} versus sfGFP^{C148}

To determine whether the additional mutation at position 206 influenced the spectral properties and thus function, the absorbance and fluorescence characteristics of sfGFP^{C148} and sfGFP^{C148.F} monomeric proteins were compared. The molar absorbance at 400 nm was 10 % higher for sfGFP^{C148} than sfGFP^{C148.F}, combined with a 18 % lower absorbance at ~ 495 nm suggesting that the additional 206F mutation caused a larger proportion of the sfGFP chromophore to exist in the CroB form (Table 14). It also caused a 3 nm blue shift in CroB's maximum absorbance wavelength from 498 to 495 nm.

Overall, both sfGFP protein variants are spectrally very similar. It was noted however, that the quantum yield at 400 nm was slightly lower for sfGFP^{C148} (32 %) than sfGFP^{C148.F} (39 %).

Table 14. Spectral properties of sfGFP^{C148} variants. [a] Value differs from described in Pédélecq *et al.*, (2006).

Variant	λ_{\max} (nm)	λ_{em} (nm)	ϵ (M ⁻¹ .cm ⁻¹)	Φ (%)	Brightness	
					(M ⁻¹ .cm ⁻¹)	% sfGFP
sfGFP ^{C148} (+DTT)	400	511	34,000	32	10,880	30
	498		10,700	50	5,350	15
sfGFP ^{C148} (-DTT)	400	511	34,000	32	10,880	30
	498		11,100	50	5,550	15
sfGFP ^{C148.F} (+DTT)	400	511	30,600	39	11,934	32
	495		13,000	49	6,370	17
sfGFP ^{C148.F} (-DTT)	400	511	27,600	39	10,764	29
	495		12,800	49	6,272	17
sfGFP ^{WT}	400	511	8,000	N/D	N/D	N/D
	485		49,000 ^[a]	75	36,750	100

5.2.3 Dimerisation of sfGFP^{C148.F}

5.2.3.1 Concentration dependent association

In chapter 4, the dimerisation of sfGFP^{C148} (promoted by addition of CuSO₄) caused a switch from the predominantly neutral chromophore ground state (CroA) to the anionic chromophore (CroB) (section 4.2.2.2). However, with the reaction failing to reach completion, the spectral properties will now be reassessed with the additional V206F mutation; where the main rationale is that the incorporation of V206F mutation should promote formation of the dimer. Firstly, this will be tested without the use of CuSO₄, as similarly described in section 4.2.2.1. The spectra of sfGFP^{C148.F} will be compared when at high (100 μ M) and low protein concentrations (1 μ M) where an increased protein concentration is expected to promote dimerisation.

The emission intensity upon excitation of ~ 400 and 495 nm was recorded for sfGFP^{C148.F} at 1 and 100 μ M. Values were used to calculate the ratio of ~ 495 nm emission over ~ 400 nm emission (CroB/CroA). The relative fluorescence intensity is expected to increase on excitation ~ 490 nm compared to ~ 400 nm, so the CroB/CroA emission ratio should also increase. This is based on prior evidence throughout chapter 4 and comparisons with other sfGFP^{H148X} variants, that dimerisation of sfGFP^{C148} (through formation of disulphide bridges between residues C148) results in changes to the observed spectra.

The results showed a ~ 2.0-fold increase in CroB/CroA ratio (0.8 to 1.6) on increasing sfGFP^{C148.F} concentration (Table 15). This is greater than the ~1.6-fold increase that was recorded for sfGFP^{C148} (1.3 \rightarrow 2.1) (Table 15). This suggests the additional V206F mutation has promoted the switch over to CroB; further promoting the formation of protein dimers through disulphide bridging. The difference in CroB/CroA ratios between sfGFP^{C148} and sfGFP^{C148.F} at 1 μ M (1.3 for sfGFP^{C148} and 0.8 for sfGFP^{C148.F}) and 100 μ M (2.1 for sfGFP^{C148} and 1.6 for sfGFP^{C148.F}) is likely due to the lower quantum yield at 400 nm for sfGFP^{C148} (32%) in comparison to sfGFP^{C148.F} (39%). The overall change in CroB/CroA ratios is greater with V206F mutation, suggesting its role promoting dimerisation.

However, the residual emission recorded when excited at 400 nm suggests that a considerable amount of sfGFP^{C148.F} still contains the neutral chromophore ground state (CroA); indicative of monomeric protein. Thus, the use of CuSO₄ will now be trialled for its use promoting the dimerisation through formation of disulphide bridges.

Table 15. Emission values of sfGFP^{C148} and sfGFP^{C148.F} at 1 and 100 μ M. Emission recorded at 516 nm for excitation at 388 and 483 nm. [a] Gain adjustments for Clariostar plate reader were different for each protein variant.

SfGFP variant	Protein concentration (μ M)	Emission at 516 nm ^[a]		483/388 nm Ratio
		Exc at 388 nm	Exc at 483 nm	
sfGFP ^{C148}	1	1062	1341	1.3
	100	23722	50275	2.1
sfGFP ^{C148C.F}	1	1044	832	0.8
	100	28272	45473	1.6

5.2.3.2 Promoting dimerisation through formation of disulphide bridges using CuSO₄

The use of CuSO₄ was shown to promote the formation of disulphide bridges between C148 residues of sfGFP^{C148} in chapter 4. However, the reaction was slow and failed to reach completion. Thus, dimerisation was retested for sfGFP^{C148.F}, by adding CuSO₄ to the protein.

The results showed that in the presence of CuSO₄, the CroA absorbance ($\lambda_{\max} \sim 400$ nm) decreased, concomitant with an increase in CroB absorbance ($\lambda_{\max} \sim 492$ nm) over time (Figure 59 a). After 2-hours, the major absorbance peak switched from 400 nm ($30,600 \text{ M}^{-1}\text{cm}^{-1}$) to 492 nm with a molar absorbance of $60,000 \text{ M}^{-1}\text{cm}^{-1}$. This suggests the chromophore ground state swapped from a predominantly neutral ground state to the more favourable anionic form. The absorbance changes are also reflected in the fluorescence spectra; with the maximum excitation wavelength switching from 400 to 492 nm (Figure 59 b). After 2 hours, the CroB:CroA emission at 510 nm changed 5.6-fold from 0.6 to 3.3.

To confirm the subsequent spectral properties were recorded as a result of sfGFP^{C148.F} dimerisation, the protein sample was analysed during specific time intervals of CuSO₄ incubation (5, 10, 15, 30, 45 and 60 mins) via a non-reducing SDS-PAGE gel. In the presence of CuSO₄, a higher band corresponding to the size of a sfGFP protein dimer (~ 54 kDa) appeared after just 5 mins (Figure 59 c).

Analysis by ImageJ revealed ~ 75 % of sfGFP^{C148.F} had dimerised by 5 minutes and a further 5% had dimerised by 1 hour. In comparison to sfGFP^{C148} that had dimerised ~5% in 1 hour (Figure 48), this proves the incorporation of V206F mutation encourages the formation of sfGFP^{C148.F} dimers.

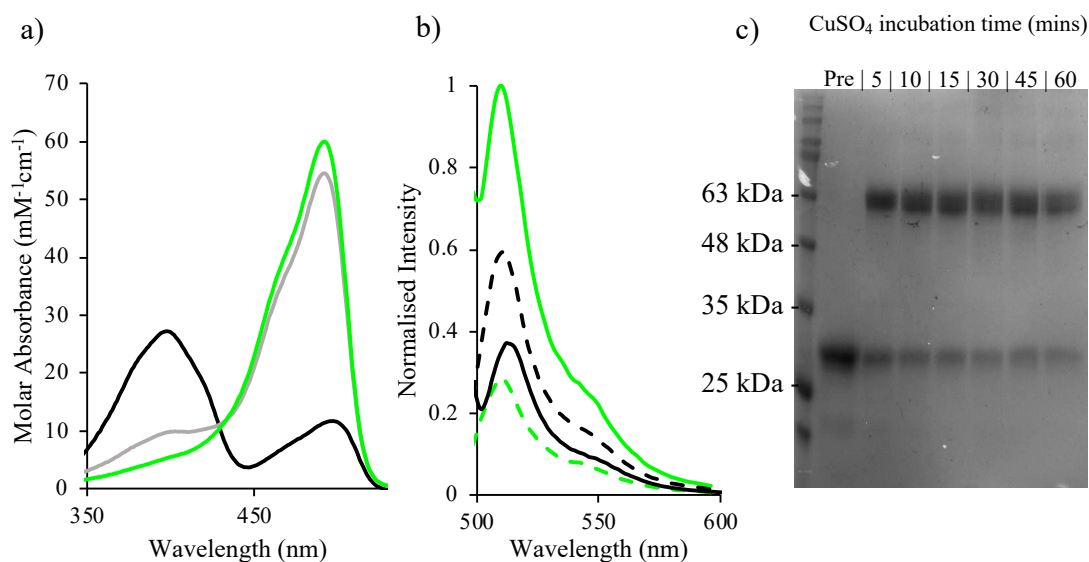


Figure 59. Dimerisation of sfGFP^{C148.F} facilitated by CuSO₄. Addition of 1 mM CuSO₄ for 0 (prior to CuSO₄ addition; black), 15 (grey) and 120 mins (green). a) Molar absorbance. Protein at 10 μM. b) Emission spectra when excited at 400 nm (dashed) and ~ 495 nm (solid). Protein at 0.5 μM. c) Non-reducing SDS-PAGE gel.

5.2.3.3 Purification of sfGFP^{C148} dimers

Before further characterisation could be performed, residual monomer had to be removed from sfGFP^{C148} dimers. Here, the purification of sfGFP^{C148.F} will be described, although the same method was also applied for sfGFP^{C148}.

Purified protein fractions containing sfGFP^{C148.F} were taken and concentrated to a volume of 2 mL, before adding CuSO₄ to a final concentration of 1 mM. The sample was incubated for 2-3 hours before applying to the SEC column. For this purpose, the HiLoadTM 16/600SuperdexTM S75 pg was used, for better resolution and separation of dimer from monomer. Absorbance monitoring was carried out at both 400 and 490 nm. The elution profile shows a major elution peak (peak 1) between 50 – 60 mL and a minor elution peak (peak 2) between 60-70 mL (Figure 60a). For peak 1, there is a ~10-fold difference in absorbance at 490 nm in comparison to 400 nm. As the first elution peak had a predominant CroB absorbance ($\lambda_{\text{max}} \sim 490$ nm),

and therefore mostly likely to have the largest molecular state, it is assumed these protein fractions contained dimeric protein. Whereas the second elution peak of a lower molecular weight is likely to contain the remaining monomeric protein. Non-reducing SDS-PAGE was performed to analyse protein fractions. As expected, primarily dimer bands were visualised for peak 1, and monomer bands from peak 2 (Figure 60b). Purity of the sfGFP^{C148.F} dimeric protein was determined to be 95 %, as calculated with ImageJ.

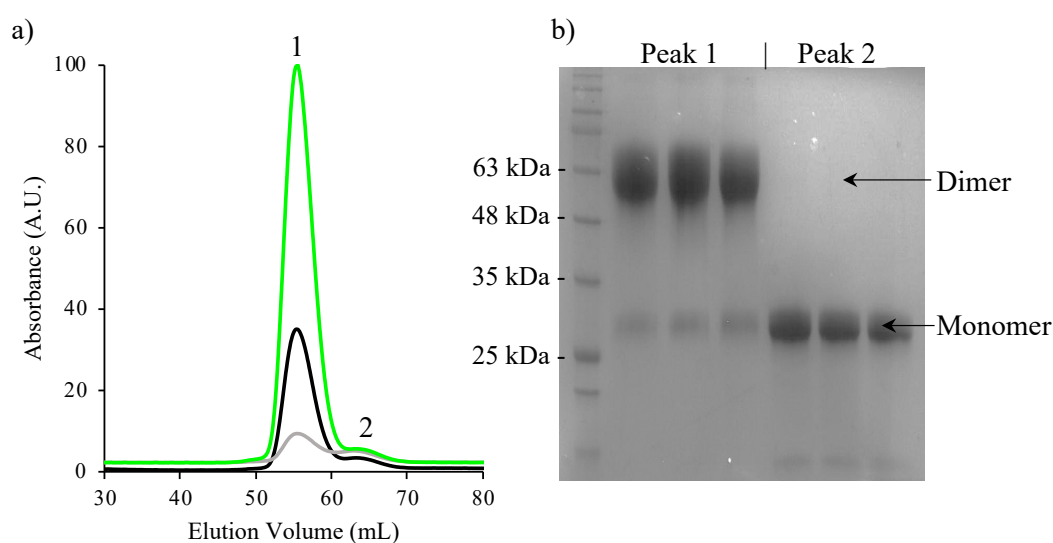


Figure 60. Purification of sfGFP^{C148.F} dimer. A) Elution profile from size exclusion chromatography. Absorbance monitoring at 280, 400 and 490 shown in black, grey and green respectively. B) Non-reducing SDS-PAGE analysis of eluted fractions from SEC.

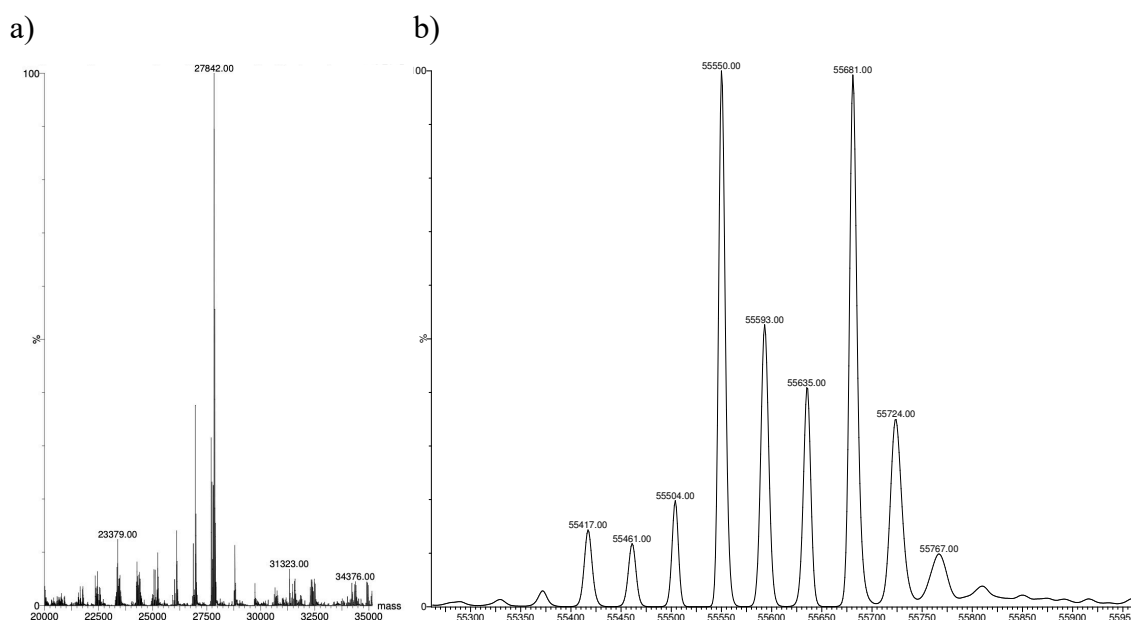
5.2.3.4 Mass spectrometry of sfGFP^{C148.F}

To further confirm the functional switch from CroA absorbance to CroB occurred as a result of covalently linked dimers, mass spectrometry was performed. After purifying dimeric protein from excess CuSO₄ (section 5.2.3.3), the mass was determined by LC/MS-TOF mass spectrometry as described in section 2.5.1. The monomeric mass of sfGFP^{C148.F} was calculated by submitting the amino acid sequence to the online software, ExPASy. The value (27,861 Da) was adjusted to account for chromophore maturation (subtracting 19 Da for H₂O and H) to give an expected mass of 27,842 (Table 16). In its dimeric form, the expected mass is 2x that of the monomer, with a subtraction of 2 hydrogen atoms to account for disulphide bridging (55,682).

Table 16. Masses of sfGFP^{C148.F} in a monomeric and dimeric form.

Variant	Status	Calculated Mw (Da)	Measured Mw (Da)
sfGFP ^{C148.F}	Full	27,842	27,842
sfGFP ^{C148.Fx2}	Full + Full	55,682	55,681
	Full + -NMet	55,551	55,551

Two elution peaks were recorded; a minor (peak 1) and major peak (peak 2). Mass spectrometry confirmed that peak 1 matched the expected mass of 27,842 Da for full length sfGFP^{C148.F}, accounting for cysteine and phenylalanine substitutions, and subsequent chromophore maturation (Figure 61 a). Peak 2, the major elution peak, matched the expected size of sfGFP^{C148.F} dimer (Figure 61 b). Two masses dominated the dimeric form, where either two full length monomers joined (55,681 Da), or a full length joined with a slightly shortened monomer (55,551 Da). As just 131 Da less than full length, the molecular weight of a methionine residue, it is assumed that the initial N-terminal methionine is cleaved. These results confirm that a permanent dimer of sfGFP^{C148.F} persists throughout mass spectrometry analysis; most likely linked via covalent interaction (disulphide bridging between C148 residues), as a non-covalent complex would likely dissociate.

**Figure 61.** Mass spectrometry analysis of sfGFP^{C148.F}. a) Expected masses from peak 1 elution. B) Expected masses from peak 2 elution.

5.2.3.5 Characterisation of sfGFP^{C148} dimers

The spectral properties of pure sfGFP^{C148} dimers were then characterised. The absorbance spectra showed major peaks at ~ 490 nm, with a molar extinction coefficient of 71,300 and 73,800 M⁻¹cm⁻¹ for sfGFP^{C148} and sfGFP^{C148.F} dimers respectively (Figure 62a). This was calculated on a per chromophore basis. This accounts for a ~ 1.5-fold increase in molar absorbance compared to that for sfGFP^{WT}.

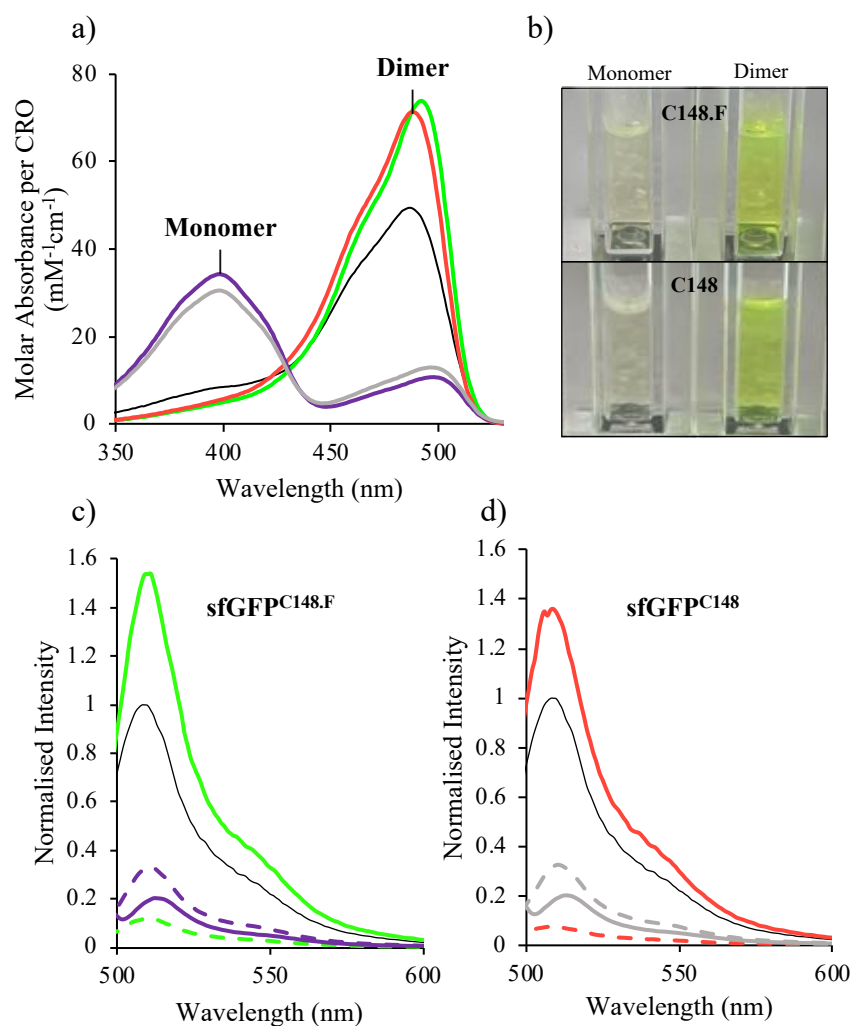


Figure 62. Spectral properties of sfGFP^{C148} and sfGFP^{C148.F} variants. a) Absorbance spectra of sfGFP^{WT} (black), sfGFP^{C148} (purple), sfGFP^{C148.F} (grey), sfGFP^{C148x2} (red), sfGFP^{C148.Fx2} (green). b) Purified protein variants at 10 μM. c) Emission spectra of monomeric and dimeric sfGFP^{C148.F} when excited at 400 nm (dashed) and 495 nm (solid), normalised to sfGFP^{WT}. d) Emission spectra of monomeric and dimeric sfGFP^{C148} when excited at 400 nm (dashed) and 495 nm (solid), normalised to sfGFP^{WT}. Absorbance was recorded at 10 μM. Fluorescence spectra were recorded at 0.5 μM and 0.25 μM for monomers and dimers respectively.

By comparison to their respective monomers, these results highlight the CroA to CroB switch via monitoring of both absorbance and fluorescence spectra. Thus, this poses two possible methods to monitor dimerisation by utilizing a ratiometric system. For example, recording the absorbance of sfGFP^{C148.F} at ~ 492 and 400 nm, the CroB/CroA ratio changes 30-fold; from 0.6 for the monomer, to 14.76 for the dimer. Whereas fluorescence monitoring of 492/400 nm at 510 nm changes from 0.6 (monomer) to 13 (dimer) which equates to a ~ 22-fold change in signal. Note, this is slightly higher for sfGFP^{C148} where a 30-fold switch is recorded from 0.6 to 18 (488/400 nm). This could be due to the slight difference in quantum yields between sfGFP^{C148} and sfGFP^{C148.F} proteins.

By utilising these distinct ratio's, the quantity of dimerised protein can be determined. For example, a 2-hour incubation of 1 mM CuSO₄ with 10 μM of sfGFP^{C148.F} (Figure 59a) resulted in molar absorbance of 5,100 M⁻¹cm⁻¹ and 59,500 M⁻¹cm⁻¹ at 400 and 492 nm respectively. By dividing the CroA absorbance by CroB, a CroB/CroA result is determined as 11.67. Utilizing the calibrated absorbance for purified monomeric and dimeric protein, this equates to 78% dimerised.

These results highlight a significant change in function from a dominant CroA ($\lambda_{\text{max}} = 400$ nm) absorbance for monomeric protein, to CroB ($\lambda_{\text{max}} = 495$ nm) absorbance for protein dimers. The ground state of the chromophore is assumed to have returned to its predominant anionic form, upon which was lost upon H148C mutation.

The quantum yield also improved after dimerisation of sfGFP^{C148} and sfGFP^{C148.F} proteins (~ 1.3 and ~ 1.5 fold respectively) (Table 17). When excited at ~ 495 nm, an ~ 8-fold difference in brightness was recorded between monomer and dimer sfGFP^{C148.F}. Also, when calculated on a per CRO basis, the brightness of sfGFP^{C148.F} increased ~ 1.5-fold higher than sfGFP^{WT}. Thus, the dimer is expected to have an approximate 3-fold higher brightness. This highlights the synergistic properties of this variant, where the 'sum is greater than the parts'.

Table 17. Spectral properties of sfGFP^{C148} variants. [a] Excitation at 400 nm was negligible and so QY was not determined. [b] Value differs from described in Pedelacq et al 205.

Variant	λ_{\max} (nm)	λ_{em} (nm)	ϵ (M ⁻¹ .cm ⁻¹)	Φ (%)	Brightness (M ⁻¹ .cm ⁻¹)
sfGFP ^{C148}	400	511	34,000	32	10,880
	498	511	10,700	50	5,350
sfGFP ^{C148.F}	400	511	30,600	39	11,934
	495	511	13,000	49	6,370
sfGFP ^{C148x2} [a]	488	511	71,300	64	45,632
sfGFP ^{C148.Fx2} [a]	492	511	73,800	72	53,136
sfGFP ^{WT} [a][b]	485	511	49,000	75	36,750

5.2.3.6 Dimerisation rates of sfGFP^{C148} variants

To quantify the impact of mutating residue 206 on the dimerisation rate of sfGFP^{C148.F}, the endpoint fluorescence spectra was recorded using the Clariostar (section 2.4.4). This method was selected for its various advantages including; accurate injection volumes of CuSO₄, dual excitation wavelengths (where wavelengths are read sequentially for each well), rapid emission readings at defined time intervals, and a reduction in sample volume for triplicate readings. Due to the defined bandwidth of the optical filters (8 nm), and minimum distance between excitation and emission wavelengths, the emission of sfGFP variants was monitored at 516 nm upon excitation at 388 and 483 nm, values near the optimum excitation wavelengths. It is worth noting, that due to the modified excitation values, the expected CroB/CroA ratio's were slightly different to those recorded with the fluorimeter, although the same theory applies. Recalibration of monomer and dimer samples was also performed using this method.

Following addition of CuSO₄ (1 mM), the results showed that for both protein samples, an increase in CroB excitation was concomitant with a decrease in CroA excitation as expected. By dividing the emission values of CroB excitation (λ_{EX} 483 nm) by those of CroA excitation (λ_{EX} 388 nm), the dimerisation can be

monitored more easily as a direct reflection of ratio (483nm:388nm). For pure monomer, the ratio is 0.9 while for pure dimer it is 26.5. Thus, these represent the limits of the ratio in terms of degree of dimerisation. The changes in ratio can then be used to determine the rate of dimerisation. The results were fit to one site binding model using GraphPad. For sfGFP^{C148.F}, the CroB/CroA ratio increased ~ 26-fold after just two hours, from ~ 0.9 to 23 (Figure 63). Whereas, for sfGFP^{C148}, a ~ 2.1-fold increase was observed from ~ 0.9 to 1.9. We can therefore calculate that ~84% of sfGFP^{C148.F} had dimerised compared to ~ 4 % for sfGFP^{C148} after just two hours. These values correlate against prior expectations calculated by absorbance switch's and SDS-PAGE (Figure 59). The second order rate constants were calculated at $2.2 \times 10^{-4} \mu\text{M}^{-1}\text{min}^{-1}$ and $8 \times 10^{-3} \mu\text{M}^{-1}\text{min}^{-1}$ for sfGFP^{C148} and sfGFP^{C148.F} respectively. This is a significant difference (~40-fold) in dimerisation rates, highlighting that addition of V206F mutation contributes towards protein association.

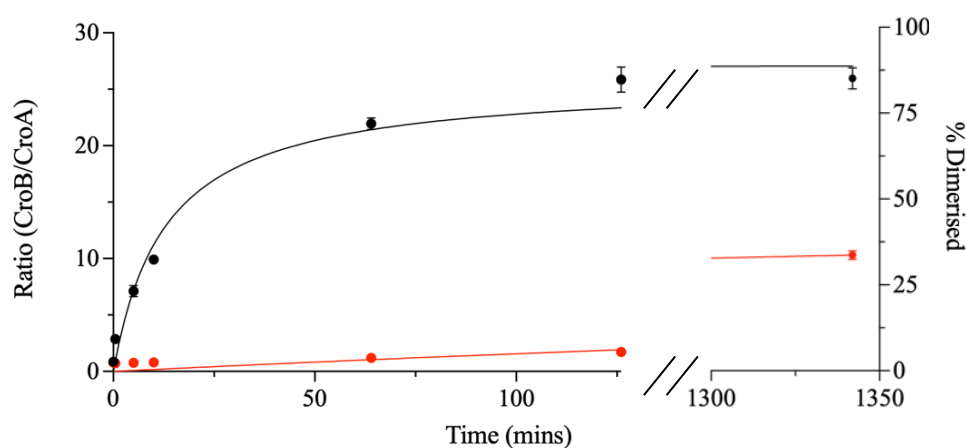


Figure 63. Dimerisation of sfGFP^{C148.F} (shown in black) vs. sfGFP^{C148} (shown in red). Incubation time in 1 mM CuSO₄. Emission ratios of CroB/CroA (left Y axis) and percentage of protein dimerised (right Y axis). Graphpad used to fit 'one site specific binding curve' ($\text{Ratio} = \text{Ratio}_{\text{Max}} * X / (Z + X)$, where X is time and Z is time to achieve half Ratio_{max}).

5.2.3.7 Reversibility of dimerisation

The ability to modulate sfGFP function through covalent modification and dimerisation has been demonstrated in this chapter. As an oxidation-sensitive interaction, there is potential to control monomer-dimer populations through

alteration of external conditions. As already shown above, CuSO₄ can promote dimerisation, presumably through the formation of a disulphide bridge. Conversion of dimer to monomer should thus be possible by adding reducing agents, such as DTT, that traditionally break disulphide bonds. However, the question arises as to whether this conversion of dimer back to the monomer result in the same structure and thus functional features present prior to dimerisation. To address this, the sfGFP^{C148.F} dimer was incubated for 1-hour with DTT (5 mM). The sample was then subjected to size exclusion chromatography, where the SEC buffer also contained 5 mM DTT to ensure conditions were maintained during purification. The results show the major elution peak shifted from 50-60 mL for the sfGFP^{C148.F} dimer to 60-70 mL (Figure 64 a,c). Absorbance monitoring suggests the return of the major 400 nm peak compared to 490 nm for the dimer. Non-reducing gel showed the loss of dimer bands after purification in DTT, suggesting the disulphide bonds had been reduced and caused the subsequent breakdown of sfGFP^{C148.F} dimer (Figure 64 d).

Characterisation was carried out and confirmed that the spectra matched those of the sfGFP^{C148.F} monomer, prior to dimerisation. Thus, we can predict the structure also returns to that prior to dimerisation.

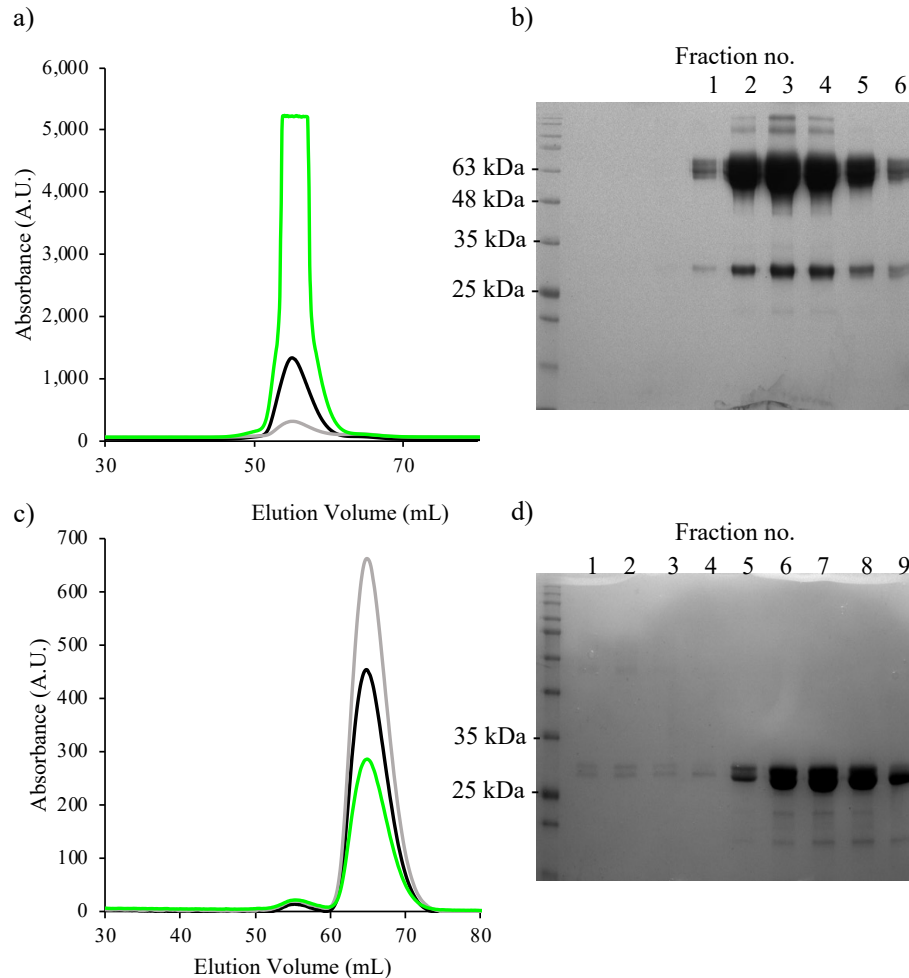


Figure 64. Oligomerisation analysis of sfGFP^{C148.F}. (a-b) Sample pre-treated with 1 mM CuSO₄ for 1 hour before analysis. (c-d) Dimer pre-treated with 5 mM DTT. (a,c) SEC elution profile. Absorbance at 280, 400 and 490 nm shown in black, grey and green respectively. (b, d) Non-reducing SDS-PAGE gel.

5.2.3.8 The impact of metal and salt on dimerisation of sfGFP^{C148.F}

To understand the role of CuSO₄ in mediating dimerisation of sfGFP^{C148.F}, the use of a different copper salt and different transition metal were investigated. To confirm copper ions mediate dimerisation, CuSO₄ was switch for CuCl₂. SfGFP^{C148.F} (~ 5 μM) was mixed with either 1 mM CuSO₄ and CuCl₂. The results showed comparably similar absorbance spectra with the expected decrease in CroA absorbance and subsequent increase in CroB absorbance. The same results were reflected in the fluorescence spectra. It is therefore assumed the copper (II) ion is the key dimerisation mediator. For consistency, future experiments were performed with CuSO₄.

The concentration of metal ion was then investigated to determine its potential impact on reaction rate. The absorbance spectrum was compared for sfGFP^{C148.F} following incubation with 100 μ M and 1 mM CuSO₄. The results showed that after 60 minutes, the CroB/CroA ratio was 4.1 for 100 μ M, while it was 11.7 for 1 mM CuSO₄. This suggests a slower switch from CroA absorbance to CroB is recorded at lower concentrations of CuSO₄ (100 μ M) (Figure 65). A residual peak at 400 nm exists after a 1-hour incubation of sfGFP^{C148.F} in 100 μ M CuSO₄. This isn't present with 1 mM CuSO₄, which just has a major CroB absorbance.

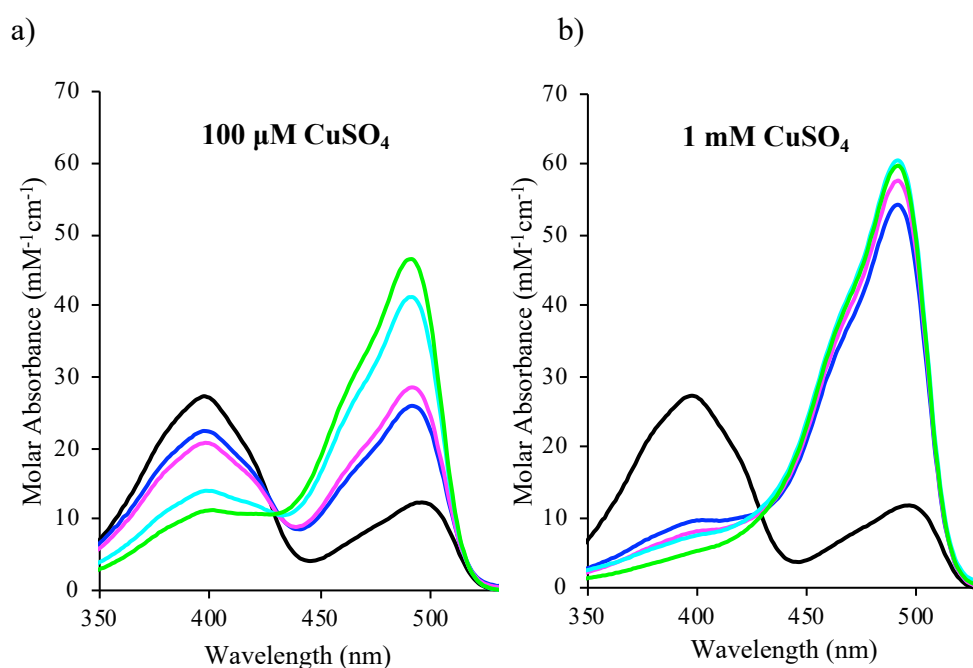


Figure 65. Concentration dependence of CuSO₄-induced dimerisation of sfGFP^{C148.F}. Addition of 1 mM CuSO₄ for 15, 30, 60 and 120 shown in blue, magenta, cyan and green. Pre-treatment sample shown in black.

Another transition metal, manganese (Mn²⁺), was then investigated to determine whether dimerisation was specific for Cu²⁺. The absorbance spectra showed that upon addition of MnSO₄, the 400 nm peak decreased concomitant with an increase at 495 nm (Figure 66). This suggests dimerisation is promoted by Mn²⁺, similarly to Cu²⁺. However, comparison of spectra shows that the absorbance switch (400 nm \rightarrow 495 nm) occurs at a slower rate than for CuSO₄.

Other transition metals were also investigated including iron, cobalt, nickel and zinc. No spectral difference was recorded for sfGFP^{C148.F} as a result of the addition of any of the ions. It is therefore concluded that copper (and partly

manganese) is the most effective transition metal for promoting the formation of sfGFP^{C148.F} dimers.

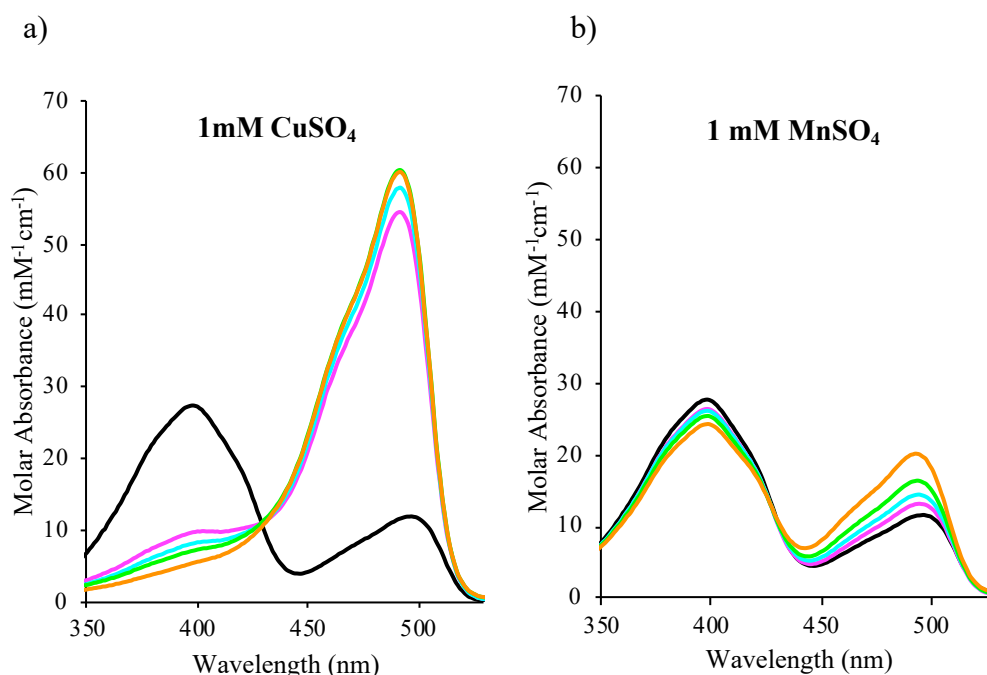


Figure 66. Absorbance spectra of sfGFP^{C148.F}. Incubation times of 15, 30, 60 and 120 min shown in magenta, cyan, green and orange respectively. Pre-treatment sample shown in black. a) Addition of 1 mM CuSO₄ b) Addition of 1mM MnSO₄.

5.2.4 X-ray crystallography of sfGFP^{C148.F} dimer

While the structure of the azide-cyclooctyne click linked dimer acted as the model for designing the current sfGFP^{C148.F} dimer, the nature of the covalent link is very different, especially length. Thus, to understand how dimerisation gives rise to chromophore state switching and enhanced spectral properties, the structure of the sfGFP^{C148.F} dimer was determined. In chapter 4, the structure of the monomer has already been described so changes on the switch from monomer to dimer can also be rationalised.

Protein samples were prepared as similarly described for section 5.2.3.3; with prior incubation with 1mM CuSO₄ followed by SEC to remove excess CuSO₄ and separate any residual monomer from the dimer. Two crystals were harvested and sent to Diamond light source to employ x-ray scattering and diffraction. Both crystal forms generated useful datasets for structure determination. The different crystals had different space groups: orthorhombic (resolution 1.79 Å)(PDB ID: 8BXP) and

triclinic (resolution 1.63 Å). The structures were different from one another, in terms of their protein interfaces', presumably caused by an effect of the crystallographic conditions, with a pH of 7.5 for the orthorhombic, and 4.5 for the triclinic. This is expected to have impacted the structure of sfGFP^{C148.F} dimer, mainly through charge differences implicating residues between the protein interface. Thus, as the protein was used *in vitro* at pH 8.0, the orthorhombic dimer will be used for comparisons against the monomeric structure at the end of chapter 4 (crystal also grew at pH 7.5). The structure of triclinic will then be discussed separately as an alternative structural form of the sfGFP^{C148.F} dimer under acidic conditions (pH 4.5).

5.2.4.1 Structure solution of sfGFP^{C148.F} orthorhombic dimer

5.2.4.1.1 Crystal formation and refinement

Crystals of sfGFP^{C148.Fx2} in the P 2 2 2₁ space group grew in condition G7 on the PACT Premier screen (0.2 M sodium acetate trihydrate, 0.1 M Bis-tris propane, pH 7.5, 20% w/v PEG 3350). The crystals diffracted to the highest resolution of 1.79 Å. Full diffraction data and refinement statistics can be found in Table 18. The structure went through multiple rounds of refinement to reach R-factor and R_{free} values of 20 and 23 % respectively. Analysis of the contents of the crystal cell showed that 2 molecules were present in the unit cell (chain A and B). The structure was submitted to PDBePISA (<https://www.ebi.ac.uk/pdbe/pisa/>) for analysis. Two protein interfaces were identified; where chain A formed a dimer with a symmetry related copy of itself (A-A) and chain B formed a dimer with a symmetry related copy of itself (B-B) (Figure 67). Both interfaces had an interface area of ~ 920 Å² and scored 1.0 for the complex significance score (CSS), the maximum value, implying the interface plays an essential role in complex formation. Whereas, the molecules within the unit cell had an interface area of ~ 420 Å² and a CSS of 0.0, suggesting the interface is insignificant for assembly formation. Thus, further analysis will now be performed on the sfGFP^{C148.F} dimers. For simplicity, the dimers will now be referred to as dimer1(A-A) and dimer2 (B-B).

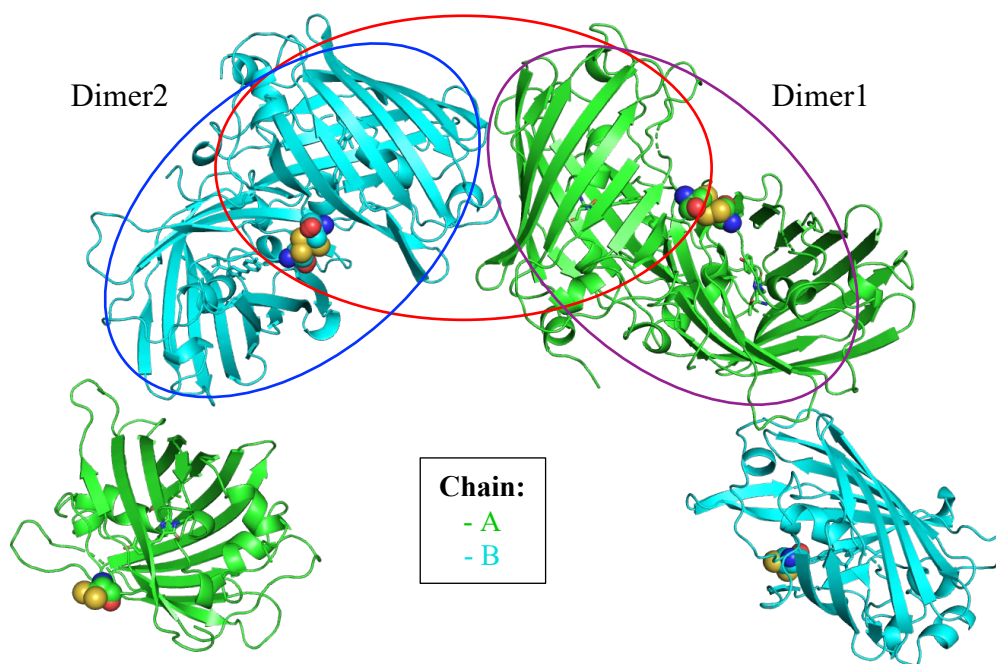


Figure 67. Cartoon representation of molecules making up two dimers found in asymmetric unit. Red oval represents contents of the unit cell, highlighting chain A in green and chain B in cyan. Purple oval shows dimer1 formed by chain A and a symmetry related copy. Blue oval shows dimer2 formed by chain B and a symmetry related copy. Cysteine residues at position 148 are highlighted as spheres.

Table 18. Crystallographic statistics from x-ray diffraction refinement and final bond parameters for crystal structure of orthorhombic sfGFP^{C148.F} dimer. PDB ID: 8BXP.

sfGFP^{C148.F} Orthorhombic Dimer	
Data collection/reduction statistics	
Wavelength (Å)	0.81532
Beamline	I03
Space group	P 2 2 2 ₁
a (Å)	67.80
b (Å)	75.02
c (Å)	125.78
Resolution range (Å)	50.31 – 1.79
Total reflections measured	826,980
Unique reflections	61,413
Completeness (%) (last shell)	99.8 (95.7)
I/σ (last shell)	10.5 (0.4)
R(merge) ^a (%) (last shell)	0.125 (4.902)
B (iso) from Wilson (Å ²)	27.68
Refinement statistics	
R-factor ^b (%)	20.0
R-factor ^c (%)	23.4
RMSD bond lengths (Å)	0.0130
RMSD bond angles (°)	1.795
Ramachandran Plot statistics	
Favored region (%)	96.68
Allowed region (%)	2.65
Disallowed region (%)	0.66

5.2.4.1.2 Structure of dimer1

Upon examination of dimer1, a head-to-tail arrangement was observed where the termini of each monomer lie at opposite ends in the dimer (Figure 68a). The side chain of C148 occupies two alternative conformations (conformer A and B) with a $\sim 75^\circ$ rotation about the β -carbon. The sulphur atoms are within optimum distance (2.0 Å apart) to form disulphide bonds with the corresponding C148 in the other monomer (conformer A with conformer B, and conformer B with conformer A) (Figure 68b). The chromophores point towards the dimer interface with a distance of 17.8 Å apart from one another (Figure 68c). According to PISA (https://www.ebi.ac.uk/pdbe/prot_int/pistart.html), the interface involves 29 residues from each monomer and has a buried surface area of 916.4 Å². There is symmetry between the two monomers, where the same residues from each monomer comprise the interface. There are also five hydrogen bonds present between monomers at the interface involving residues N39, R73, Y200, S202 and Q204 (Figure 68d).

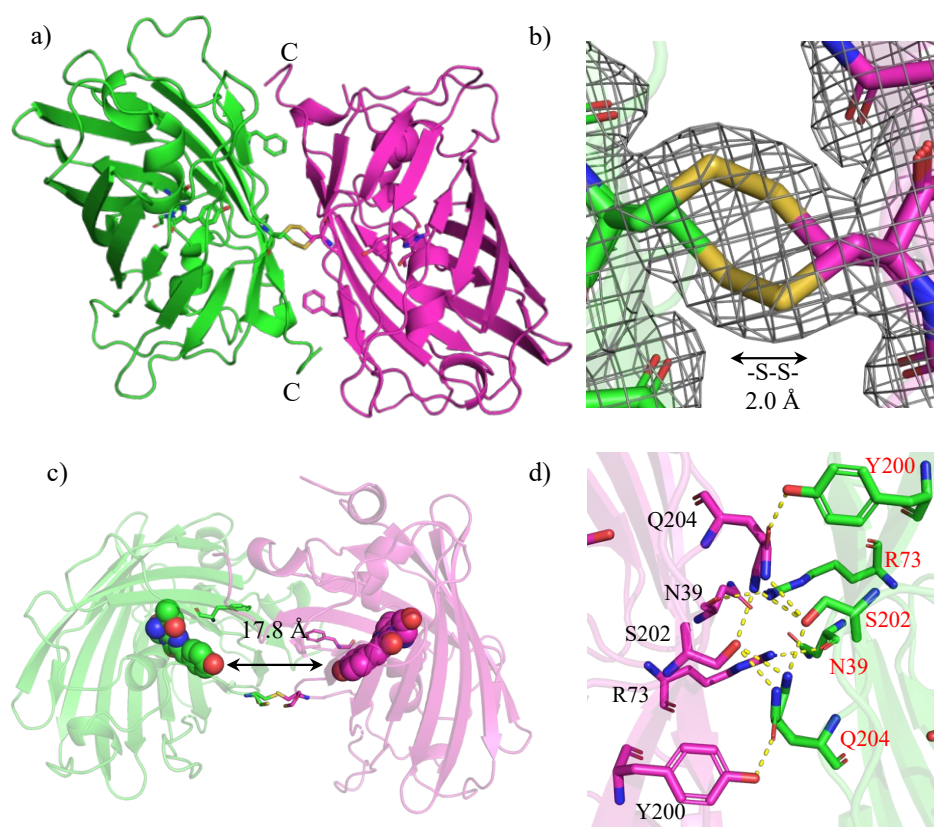


Figure 68. Structural analysis of dimer1. Chain A in green and its symmetry related copy in magenta. a) Head-to-tail orientation of β -barrel displaying C148 and F206 as sticks. b) Disulphide bond between C148 residues, showing 2Fo-2Fc electron density map at 1.0 sigma. c) Distance separating chromophores. d) Hydrogen bonds between monomers present at the interface. Residues labelled red for chain A and black for its symmetry related copy.

The region around residue 206 was then examined to investigate the role of the V206F mutation on promoting protein dimerisation. It was noted, the surrounding interactions and environment of residue F206 were the same within both monomers. According to PISA, F206 has a ~100% buried area percentage as a fraction of the total solvent-accessible surface area. The residues surrounding F206 (between 200 to 209) are also involved in the dimer interface, excluding L201 and T203. The online server PIC (protein interaction calculator) highlighted hydrophobic interactions of F206 with the corresponding residues A227, M233, P75, M78 in the other monomer (Figure 69a). There is also an aromatic-sulphur interaction occurring at a distance of 5.3 Å between F206 and M78, close to the optimum interaction distance (~ 5 Å) (Valley et al. 2012; Gómez-Tamayo et al. 2016). Analysis of residues within a 5 Å contact distance to F206F revealed the T-shaped pi-cation interaction between F206 and H77, with a centroid distance of 4.8 Å (Figure 69b) (Chen et al. 2018). It is therefore likely that F206 plays an important role in promoting dimerisation of sfGFP^{C148,F} through a combination of hydrophobic and pi stacking interactions at the protein interface.

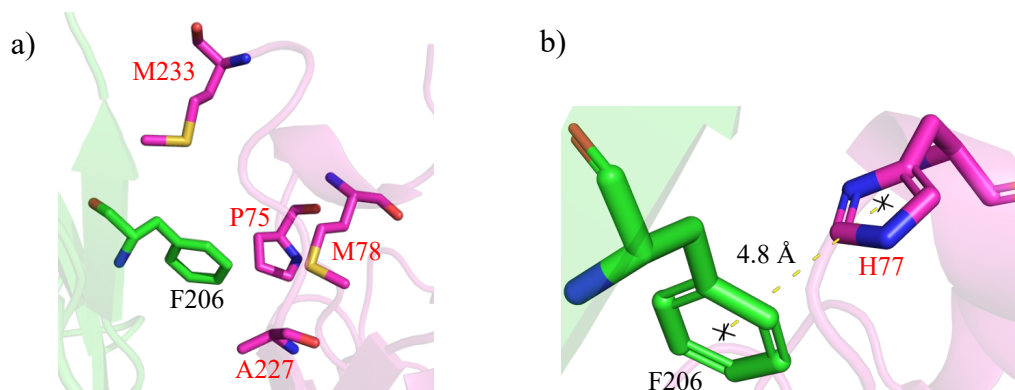


Figure 69. Structural analysis of region surrounding F206 residue in dimer1. Chain A in green and its symmetry related copy in magenta. a) Sticks highlighting to display hydrophobic residues surrounding F206 b) Centroid distance between F206 and H77.

5.2.4.1.3 Structure of dimer1 compared to dimer2

Structural alignment between the dimer1 and dimer2 yielded an RMSD value of 0.481 Å as can be clearly observed by the minor offset of the β -barrel alignments (Figure 70a). According to PISA, the interface surface area of the two chains remains the same ~ 916 Å² despite the minor shift in β -barrels. Dimer2 interface

involves 31 residues whereas dimer1 involves 29. The residues involved in forming 5 hydrogen bonds at the interface (N39, R73, Y200, S202 and Q204) are consistent with dimer1. Alignment of chromophores highlights a 0.3 Å shift in the hydroxyl group from chain A to B which subsequently alters the distance to its partner chromophore from 17.8 Å in dimer1 to 17.5 Å for dimer2 (Figure 70b). A slight offset can also be observed between C148 side chains, although the disulphide bond distance is maintained at 2.0 Å (Figure 70c). The interchain bonds and interactions associated with residue F206 are the same as those for the dimer1.

Overall, the structures of both dimers are similar. The minor shifts in a few residues, and the -CRO-CRO- distance are almost negligible as a comparison of the structures as a whole. Thus, we can conclude their near perfect alignment presents dimers of the same structures.

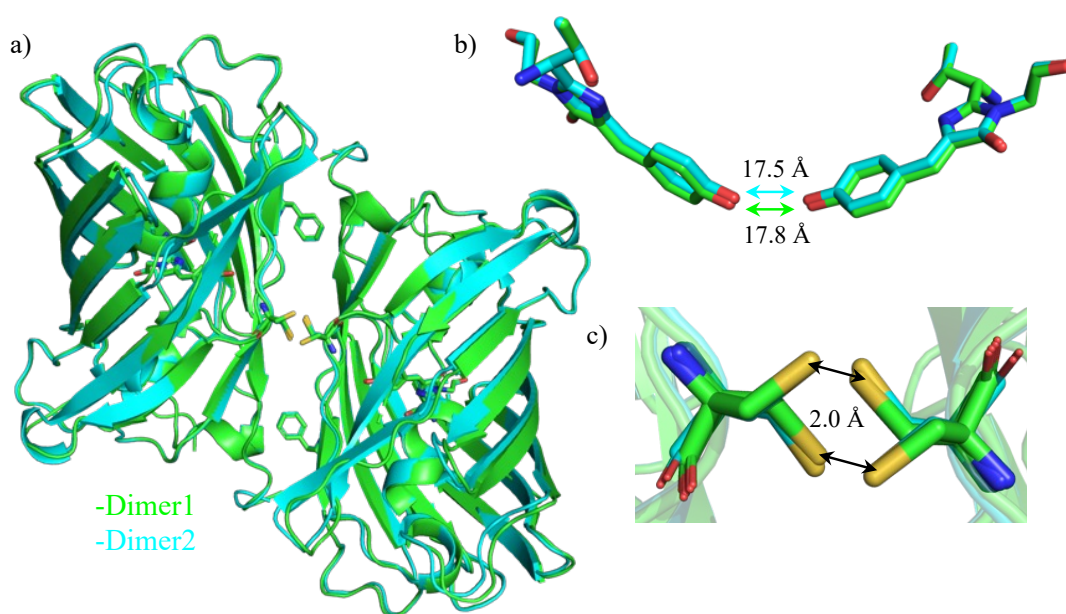


Figure 70. Comparison of orthorhombic dimer1 and dimer2. Dimer1 displayed in green and dimer2 is displayed in cyan. a) Overlay of dimers1 and 2 showing near perfect alignment of overall structure. b) Chromophore alignment. c) Stick representation of C148 residues displaying two alternative conformations in both dimers.

5.2.4.1.4 Comparison of individual monomers within dimer

The structure of each individual monomeric subunits were then compared within the orthorhombic dimer. Structural alignment of chain A and B yielded low RMSD values of 0.156 Å. This suggests that there is very little structural difference between the two chains. Upon visualisation, there are no significant distinctions to

the overall protein structures (Figure 71a). The β -strands housing residues H148C and V206F mutations showed no significant difference except for a minor offset of side chains atoms; the two rotamers of C148 have a slight offset of 0.7 Å between chains (Figure 71b). All other key residues including those involved in the ESPT (i.e. S205 and E222) align near perfectly.

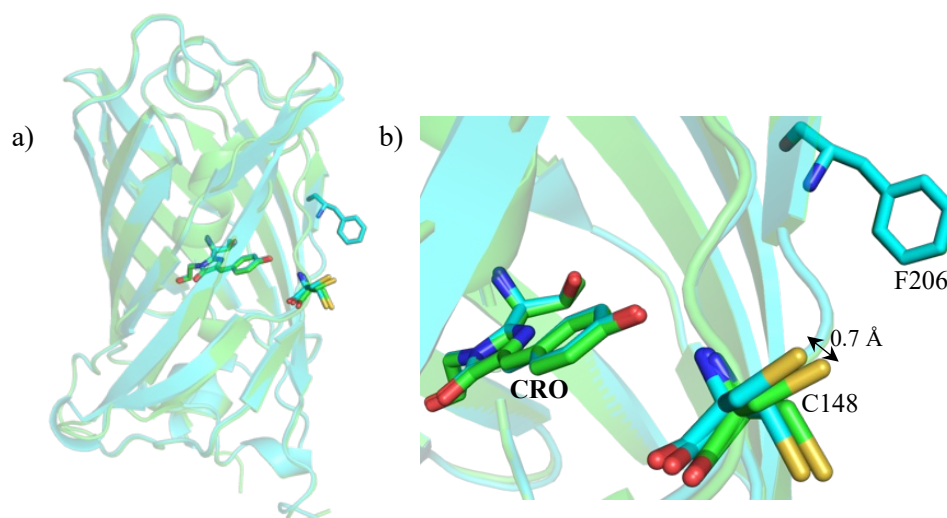


Figure 71. Alignment of chain A and B in orthorhombic dimer A) Cartoon representation of chain A (green) and chain B (cyan). D) The chromophore and residues C148 and F206 represented as sticks.

5.2.4.2 Comparison of sfGFP^{C148.F} dimer (orthorhombic) and monomer

5.2.4.2.1 Structure of dimer vs. monomer

To determine the structural differences that occur as a result of dimerising sfGFP^{C148.F}, the dimer structure was compared against the monomeric structure, described at the end of chapter 4.

Firstly, the sfGFP^{C148.F} subunits within the unit cell of the dimer and monomer structures were compared. Structural alignment of chain A yielded an RMSD score of 0.117 Å, suggesting that the structures are closely matched to one another. Detailed analysis reveals a 1.7 Å difference in β -carbon positions of C148 (Figure 72a). There is a slight offset (0.5 Å) in position of the hydroxyl group for S147 (Figure 72b). However, residue S147 is still of an appropriate distancing to the chromophore to participate in hydrogen bonding within both the dimeric (2.9 Å) and monomeric (2.7 Å) structures. Key residues (T203 and S205) associated with the hydrogen bonding network surrounding the chromophore occupy the same position within both structures. (Figure 72c, d). There is slight difference in position of

residue E222 within monomeric and dimeric structure, where there is rotation of the side chain.

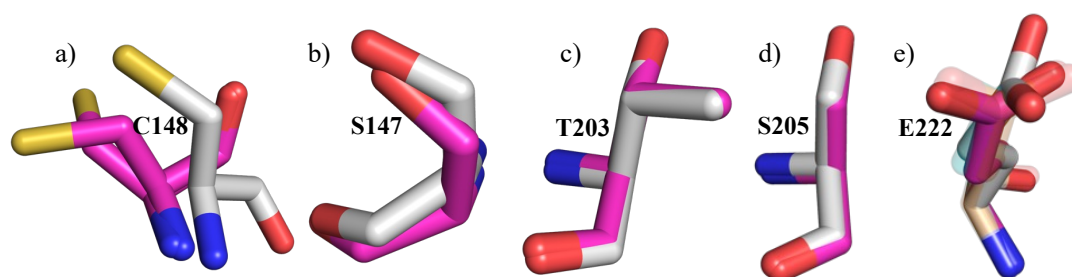


Figure 72. Alignment of chain A within dimer structure (dimer1) and chain A within monomeric structure. Dimer and monomer are shown in magenta and grey respectively. A) Cartoon representation. b) Stick representation of residue H148C. c) Residue S147. D) Residue T203. E) Residue S205. F) Residue E222. Chains B, C and D of monomer shown in background.

5.2.4.2.2 Hydrogen bonding networks

Next, the hydrogen bonding network was compared between the sfGFP^{C148.F} dimer and monomer. The spectral data throughout this chapter demonstrates how dimerisation causes a switch in absorbance from a major 400 nm peak to ~ 495 nm. This suggests the anionic chromophore ground state is stabilised as a result of dimerisation. By analysing the hydrogen bonding network surrounding the chromophore and the key differences between the dimeric and monomeric structures, the exact mechanism inflicting this spectral change will be rationalised.

Analysis of the dimer (dimer1) revealed an extensive hydrogen bond network that extends across the protein interface. Each monomer offers two possible hydrogen bonding networks out of the β -barrel via a water molecule (W1) or S147/OH (Figure 73). The hydrogen bond network via W1 continues with two further water molecules (W2 and W3) and incorporates Y200 from the corresponding monomer (Figure 73). The other route involving S147 extends across the interface connecting the two chromophores via two different water molecules (W1* and W2*) and residues S202 and Q204. By comparison to the structure of dimer2, W2* is missing, so therefore prevents extension of the hydrogen bond network to the partner chromophore. However, hydrogen bonding via W1, W2, W3 and Y200 can still be observed. As well as across the interface, W1 also participates

in hydrogen bonding to S205 and E222 (Figure 73a). Thus, this may suggest that water molecules take on the role, stabilising the charge on the phenolate.

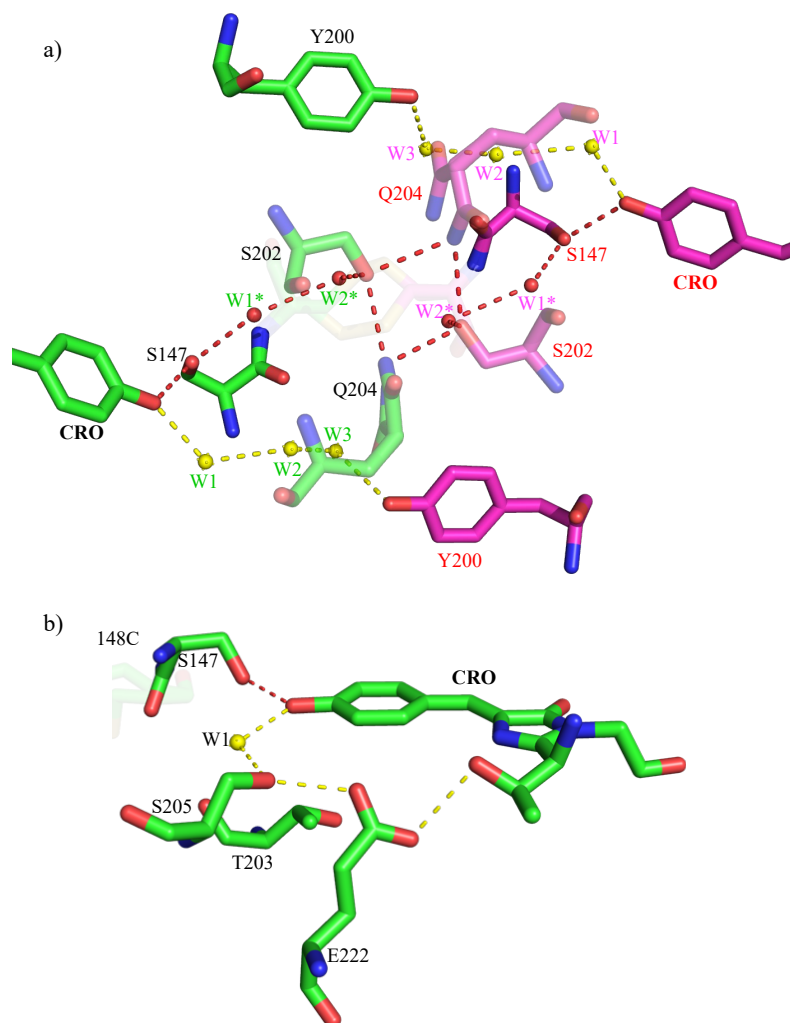


Figure 73. Hydrogen bond network within dimer1 structure. Chain A and its symmetry related copy shown in green and magenta respectively. Hydrogen bonds highlighted in dashed lines. a) Hydrogen bonding networks that extend across the protein interface. Exit routes one and two shown in yellow and red respectively. b) Hydrogen bonding surrounding the chromophore.

Analysis of the monomeric structure showed that the hydrogen bonding network surrounding the chromophore involving residues W1, S205 and E222 (Figure 53), is similar to that within the dimer. The chromophore can hydrogen bond to S147 and W1 within both monomeric and dimeric structures. However, the chromophores cannot hydrogen bond to T203 (for either the dimeric or monomer structure) as does occur within sfGFP^{WT}; where residue T203 plays a critical role stabilising the charge of the phenolate chromophore. Thus, the key differences between the monomer and dimer are likely to be caused as a result of stabilising the

chromophores anionic charge. Thus, the two hydrogen bonding networks that extend across the interface within the dimer are likely to play a key role maintaining the anionic chromophores ground state, giving rise to the ~ 490 nm absorbance peak on dimerisation.

5.2.4.3 Structure solution of sfGFP^{C148.F} triclinic dimer

5.2.4.3.1 Crystal diffraction and refinement

The sfGFP^{C148.Fx2} crystal in the P1 space group was grown in condition A1 on the JCSG Plus screen (0.2 M lithium sulphate, 0.1 M sodium acetate, pH 4.5, 50% PEG 400). The crystal diffracted to the highest resolution of 1.63 Å. Statistics listing the full diffraction and refinement data can be found in Table 19. Analysis of the contents of the crystal cell showed that there were 2 molecules in the unit cell, which together form a dimer. The model went through multiple rounds of refinement to reach R-factor and R_{free} values of 21.7 and 26.8% respectively.

Table 19. Crystallographic statistics from x-ray diffraction refinement and final bond parameters for crystal structure of triclinic sfGFP^{C148.F} dimer.

sfGFP^{C148.F} Triclinic dimer	
Data collection/reduction statistics	
Wavelength (Å)	0.81532
Beamline	i03
Space group	P1
a (Å)	41.69
b (Å)	46.73
c (Å)	65.45
Resolution range (Å)	41.80-1.63
Total reflections measured	194100
Unique reflections	53625
Completeness (%) (last shell)	97.8 (95.1)
I/σ (last shell)	5.4 (0.3)
R(merge) ^a (%) (last shell)	0.125
B (iso) from Wilson (Å ²)	2.035
Refinement statistics	
R-factor ^b (%)	21.7
R-factor ^c (%)	26.8
RMSD bond lengths (Å)	0.0095
RMSD bond angles (°)	1.587
Ramachandran Plot statistics	
Favored region (%)	96.82
Allowed region (%)	2.18
Disallowed region (%)	0

5.2.4.3.2 Overall structure

Upon examination of the sfGFP^{C148.F} triclinic dimer, there is a perpendicular arrangement of monomers (Figure 74a). The two chains are linked via a disulphide bridge between the newly introduced C148 (Figure 74b). The chromophores point towards the dimer interface at a distance of 17.6 Å apart from one another (Figure 74c). The structure was analysed using PISA (https://www.ebi.ac.uk/pdbe/prot_int/pistart.html) which estimated the interface area to be 503.5 Å², and the energy required for dissociation at -2.0 kcal mol⁻¹. There is symmetry between the two monomers, where the same residues from each monomer comprise the interface. Hydrogen bonding at this interface plays an important role, with 10 hydrogen bonds identified between residues N146, S147, N149, Y151, K166, R168, N170 and Y200 (Figure 74d).

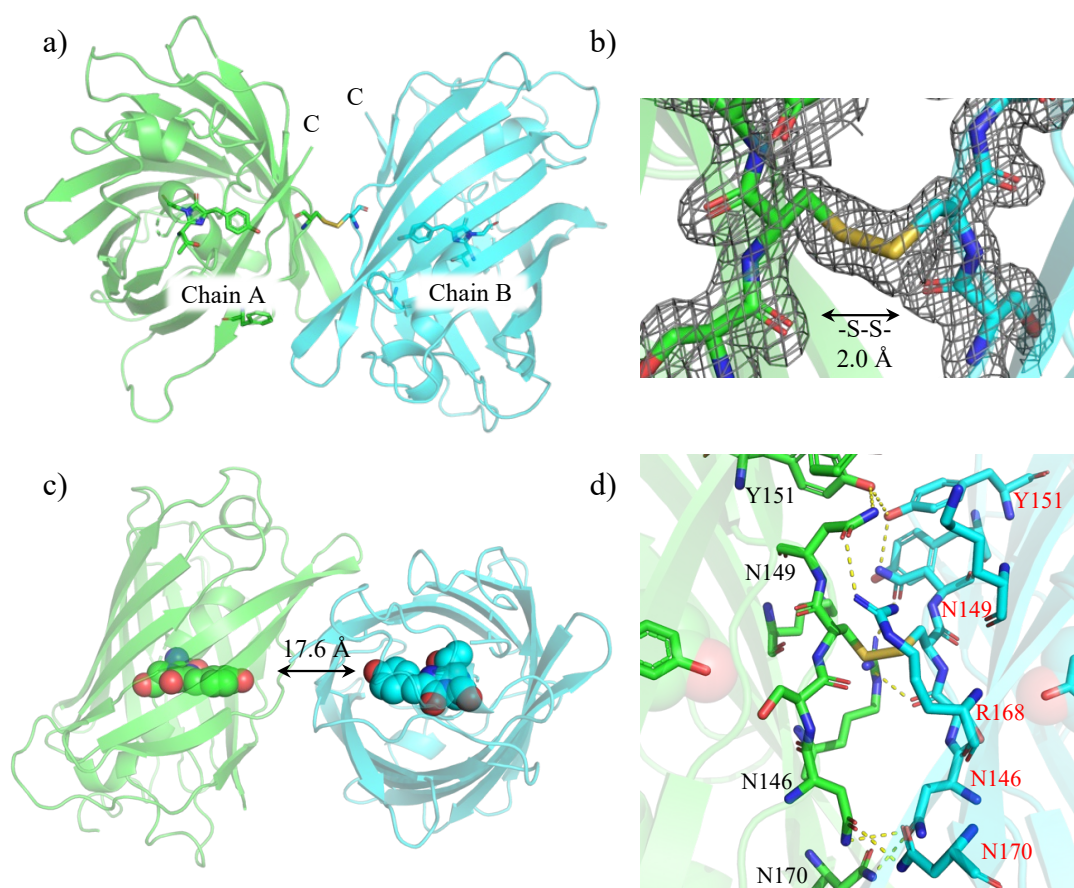


Figure 74. Structural analysis of sfGFP^{C148.Fx2} in the triclinic space group. a) Orientation of monomers showing C148 and F206 residues as sticks. b) Disulphide bond between C148 residues, showing 2Fo-2Fc electron density map at 1.0 sigma. c) Distance highlighted between chromophores. d) Hydrogen bonds between monomers present at the interface shown in dashed yellow lines. Residues labelled black for chain A and red for chain B.

To investigate the structural role of residue F206 promoting protein dimerisation, the exact configuration and interactions with surrounding residues were analysed. According to PISA, F206 has a ~ 40% buried area percentage as a fraction of the total solvent-accessible surface area. It is clear from a top-down view of chain A monomer that although F206 still contributes, it is located at boundary of the protein-protein interface (Figure 75a). Analysis of residues within 5.0 Å of F206 showed that it is mostly surrounded by hydrophobic residues including F223, L221, V219 and G174. The YRB colour scheme was applied in PyMOL to colour code hydrophobicity according to a specific colour scheme; where hydrophobics are yellow (Hagemans et al. 2015). As predicted, the F206 residue resides in a ‘patch’ of yellow proving a clustering of hydrophobic residues within that region (Figure 75b).

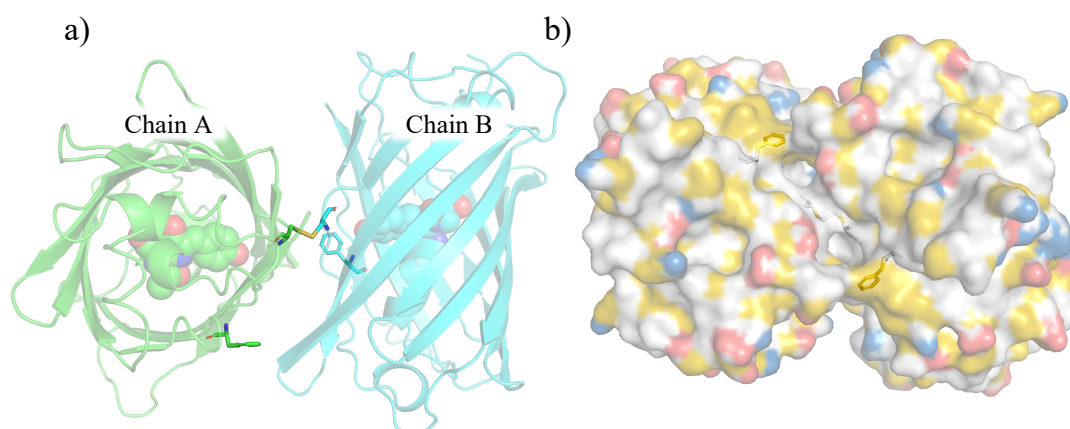


Figure 75. Structure of sfGFP^{C148.Fx2}. a) Top down view of chain A. β -barrel as cartoon, chromophore as spheres, 148 disulphide bridge and F206 residues shown as sticks. b) Surface display using YRB scheme that highlights hydrophobic regions in yellow, nitrogen atoms of LYS and ARG in blue and oxygen atoms of GLN and ASP in red. All other remaining atoms are displayed as white.

5.2.4.3.3 Comparison of monomers in triclinic space group

The monomeric subunits within the triclinic dimer were then be compared. Structural alignment of chain A and B yielded low RMSD scores of 0.156 Å. This suggests that there is very little structural difference between chains. Upon examination, there are no significant distinctions to the overall protein structures (Figure 76a). The β -strands housing residues H148C and V206F mutations showed no significant difference except for a minor offset of side chains atoms; 0.4 Å

between C148 and 1.3 Å of F206 (from a slight rotation) (Figure 76b). Overall, they were comparably similar, so chain A will be used for future comparisons.

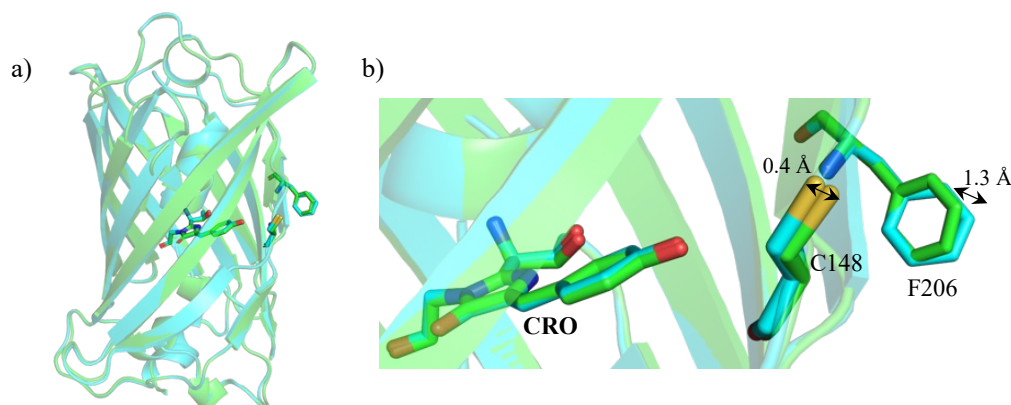


Figure 76. Alignment of chain A and B in triclinc dimer. A) Cartoon representation of chain A (green) and chain B (cyan). b) The chromophore and residues C148 and F206 represented as sticks.

5.2.4.4 Comparison of orthorhombic and triclinic structures

5.2.4.4.1 Dimers in triclinic and orthorhombic space group

As the structure of sfGFP^{C148.F} dimer has been solved in two separate space groups, it is important to compare between the two. Conclusions can then be drawn as to whether both dimeric forms are likely to exist, or whether one form is more favourable than the other. As mentioned (section 5.2.4), the main difference in the growth of sfGFP^{C148.F} crystals was the pH of the crystallographic conditions, 4.5 for triclinic and 7.5 for orthorhombic.

Initial observations show that the structure of the dimers present in the two space groups are largely different from one another in terms of the individual subunit arrangements, as summarised in Table 20. Structural alignment of chain A (OrthoA and TriA) revealed the different relative positions of chain B; where there is ~180° rotation and a large offset in β-barrel positions (Figure 77). Thus, the position of their respective sfGFP^{C148.F} subunit is largely different between space groups. The interface surface area for the triclinic structure (503.5 Å²) is almost half that for the orthorhombic structure (916.4 Å²). This suggests the triclinic structure has a less extensive dimer interface. According to PISA, the triclinic dimer has a CSS of 0.226, in comparison a maximum score of 1.0 for the orthorhombic dimers which implies their interface plays a less essential role in complex formation. There are however 10 potential hydrogen bonds between sfGFP^{C148.F} subunits in the triclinic in comparison

to 5 between the orthorhombic structure. Another key difference is the position of V206F with respect to the dimer interface. Within the orthorhombic structure, V206F is 100 % buried and contributes extensively to the dimer interface. In the triclinic structure, F206 is at the boundary the dimer interface where it is just 40% buried. Given that it drastically improved the rate of dimerisation, it suggests the V206F mutation will play a key role in promoting dimerisation, more evident within the orthorhombic dimer. Residue C148 has an accessible surface area (ASA) of ~109 or 123 Å² for orthorhombic dimers (for dimer1 and dimer2 respectively), double that for the triclinic dimer (59 Å²). According to the experimental data, it may be expected that C148 residues are surface accessible, given that DTT can be used to break the disulphide bridges.

Table 20. Summary of main differences between triclinic and orthorhombic dimers. BSA, buried surface area; ASA, accessible surface area. [a] Two values are given for orthorhombic data where dimer1 and dimer2 values differ.

Considerations	Triclinic	Orthorhombic ^[a]
Crysallisation pH	4.5	7.5
Interface area (Å ²)	~ 503.5	~ 916.4
F206 BSA (%)	40	100
C148 ASA (Å ²)	59	~ 109-123
Interface CSS	0.226	1.00

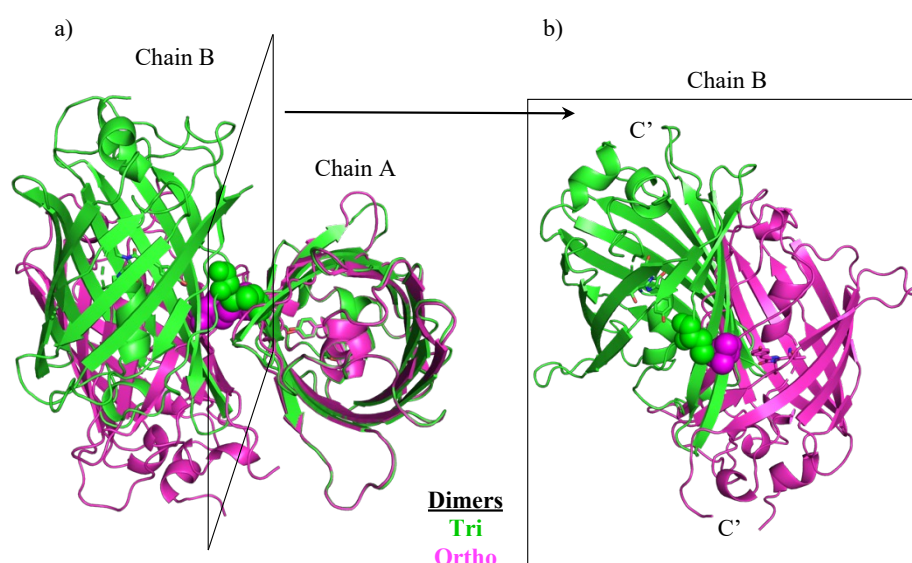


Figure 77. Comparison of triclinic and orthorhombic dimers. a) Structural alignment of orthorhombic chain A and triclinic chain A. Residue C148 shown in spheres. b) Front view showing different alignment of chain B.

5.2.4.4.2 Monomers in triclinic and orthorhombic space group

As the structures in chain A and B are similar to one another within each of the dimers, chain A will now be used to compare the structure of monomers between the two space groups (OrthoA and TriA). Overall, the structures are similar to one another in terms of their global structure (Figure 78a). However, there are subtle yet significant conformational changes around residue 148. In the triclinic space group, C148 occupies one conformation whereas in the orthorhombic there are two conformations. The C148 residues in both space groups are involved in forming disulphide bridges to the respective dimer partners, so the minor offsets do not affect their ability to form the covalent link. There is a difference in the backbone of the β -strand housing residue C148, with a 60° rotation about the $C\alpha$ atoms of C148 (Figure 78b). Whilst C148 rotamers face away from the chromophore in both space groups, C148 protrudes further away from the surrounding protein surface in the orthorhombic structure. As the covalently linked residue, it is assumed that the conformation of C148 would influence the interface and ultimately the structure of the dimer, as explored above.

All side chains of S147 are within the core of the protein, facing towards the chromophore. However, there is a $\sim 60^\circ$ rotational difference around the $C\alpha$ of S147 between triclinic and orthorhombic monomers (Figure 78c). As a result, a hydrogen bond is present between S147 and the chromophore for orthorhombic monomer only. There is also ~ 1.1 Å difference in $C\beta$ positions and complete rotation of the T203 side chain (Figure 78d). This instead causes a hydrogen bond to form between T203 and the chromophore for the triclinic structure only. The F206 residue is similarly positioned in both space groups, with a slight difference in orientation (Figure 78e). Residue E222 also has a difference in orientation between space groups, (Figure 78f).

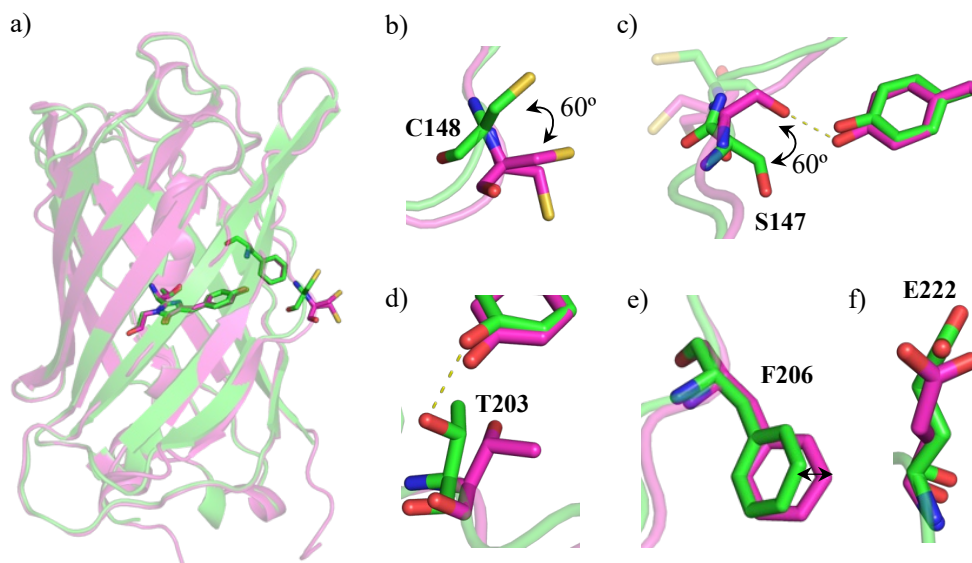


Figure 78. Structural alignment of chain A monomer within triclinic and orthorhombic dimers. Triclinic and orthorhombic structures shown in green and magenta respectively. A) Cartoon representation. B) Difference in rotation of C148. c) Difference in rotation of S147. D) Conformations of T203. E) Residue F206. f) Residue E222.

5.2.4.4.3 Hydrogen bonding

The hydrogen bonding network for the triclinic dimer was then compared against that for the orthorhombic dimer. Analysis of the triclinic dimer showed an extensive hydrogen bonding network surrounding the chromophore, although largely different from that present in the orthorhombic dimer (section 5.2.4.2.2). Where the tyrosyl hydroxyl group of the chromophore hydrogen bonds to S147 and W1 in the orthorhombic structure, this is instead replaced by two structured water molecules (W1 and W2) (Figure 79). The chromophore can also form a hydrogen bond with T203, as is the case for sfGFP^{WT}. The hydrogen bond network can then continue from W1 to S147, or from W2 to C148. The key residues involved in the ESPT can also form hydrogen bonds through W1. Although, unlike the orthorhombic dimer, the hydrogen bond network within the triclinic dimer cannot extend across the protein interface. However, it is possible that the hydrogen bonding network available to the chromophore (from W1, W2 and T203) would still be sufficient to stabilise its deprotonated state. Thus, the dimer structure of sfGFP^{C148.F} solved in the triclinic space group (at pH 4.5) is still expected to have an absorbance spectra indicative of CroB, where the major peak is at ~ 490 nm.

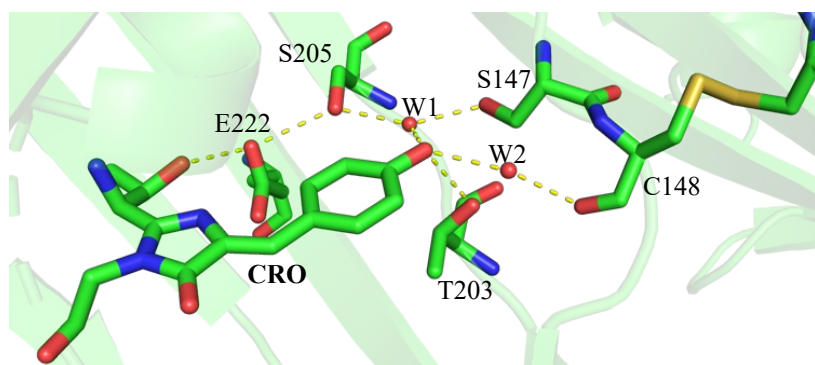


Figure 79. Hydrogen bonding surrounding the chromophore in the triclincic sfGFP^{C148.F} dimer. Key residues and those involved in hydrogen bonding are shown in sticks. Structured water molecules (W1 and W2) are shown as red spheres.

5.2.5 Discussion

This chapter evidenced the role of V206F mutation for stabilising the dimerisation of sfGFP^{C148.F}. The new sfGFP variant evidenced a 40-fold increase in dimerisation rate compared to that for sfGFP^{C148}, and a complete switch from a predominantly neutral chromophore over to the charged phenolate. By comparison of monomeric and dimeric structures of sfGFP^{C148.F}, dimerisation (by the formation of disulphide bridges) is proven to cause the entrapment of water molecules and local conformation changes leading to new hydrogen bond networks that extend across the interface to stabilise the deprotonated state of the chromophore. Thus, with the newly introduced V206F mutation, this complete switch in spectral data occurs at a more measurable rate, so that homodimerisation can now be monitored.

To understand the mechanism of functional switching, the dimeric structure of sfGFP^{C148.F} was solved for comparison to its monomeric structural state detailed in chapter 4. The main difference in protein preparation was the addition of CuSO₄ for the promotion of disulphide bridging between C148 residues. Two structural forms of dimeric sfGFP^{C148.F} were solved, that highlighted the role of differing pH's on the protein interface. However, *in vitro* studies of sfGFP^{C148.F} were consistently carried out at pH 8.0, so the structure solved in the orthorhombic space group (crystal grown at pH 7.5) was chosen for comparison to the monomeric structure. Residue V206F was buried within the protein interface of the orthorhombic dimer and was involved in a variety of hydrophobic and pi-pi interactions. Unlike predictions drawn in section 5.2.1 using the ncAA linked dimer (where F206 was predicted to form pi-pi interactions with F223), the pi-pi interaction occurred

between F206 and H77 instead. For comparison, the dimeric structure of sfGFP^{148ncAAx2} was aligned with the orthorhombic and triclinic dimers, where chain A was overlaid to reveal the different relative positions to chain B. Of the two dimers, the orthorhombic structure more closely matched that of sfGFP^{148ncAAx2} with a similar head-to-tail orientation. According to PISA, the orthorhombic has a greater interface area (916.4 Å²) compared to that for sfGFP^{148ncAAx2} (742 Å²). This is likely due to the elongated triazole link causing the upwards shift and slight rotation of chain B with respect to orthorhombic (Figure 80). In comparison to the triclinic dimer, chain B is largely offset with sfGFP^{148ncAAx2}. Again, this could largely be due to difference in electrostatic charges affecting residues at the protein interface, where the sfGFP^{148ncAAx2} crystal grew at pH 9.0. Disulphide bridging via C148 also requires closer association of protein subunits. Thus, it is not particularly surprising the protein interface of the sfGFP^{C148.F} dimers is slightly altered in comparison to sfGFP^{ncAAx2}. Nevertheless, the sfGFP^{ncAAx2} structure allowed the identification of the key residue mutation that improved the dimerisation capability of sfGFP^{C148}.

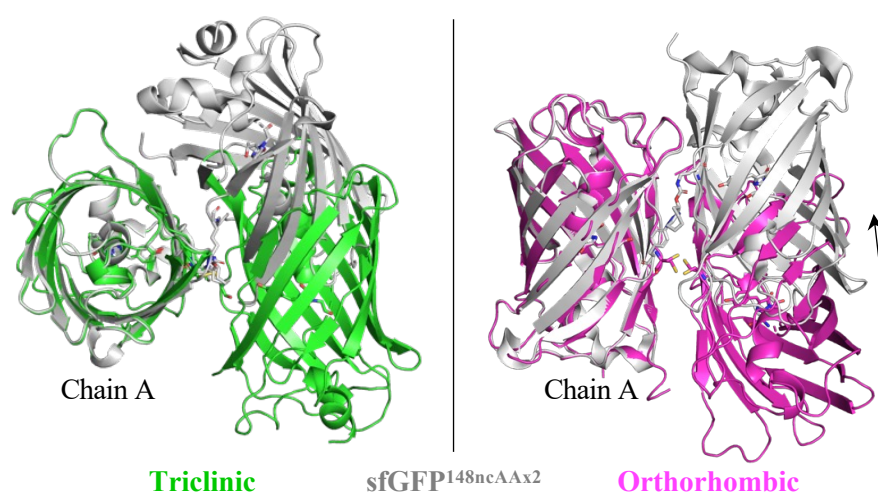


Figure 80. Structural alignment of sfGFP^{148ncAAx2} with triclinic and orthorhombic dimers. Chain A of sfGFP^{148ncAAx2} overlaid with chain A in triclinic (green) and orthorhombic (magenta).

An important consideration for this work was the requirement of Cu²⁺ (or Mn²⁺) for promoting the dimerisation of sfGFP^{C148.F} via C148 disulphide bridging. Cysteine residues with sulphur can exist in various oxidation states (e.g. thiyl radical, sulfenic acid, sulfinyl radical, sulfinic acid, sulfonyl radical and sulfonic acid); among which the thiol and disulphide are best known (Figure 81). The variety of

oxidation states is matched by the number of different redox mechanisms available to sulphur such as nucleophilic attack, electron transfer, hydride transfer, hydrogen radical transfer, and oxygen atom transfer reactions (Giles et al. 2003).

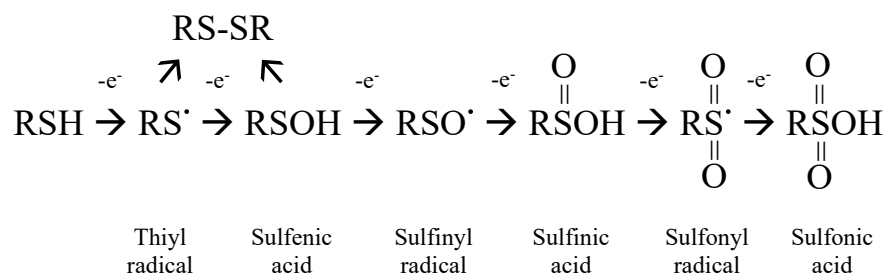


Figure 81. Succession of single-electron transfers giving rise to variety of thiol groups oxidation states. Adapted from Park and Bauerle, (1999).

One such reaction, of both theoretical and practical importance, is the oxidation of thiols to disulphides or sulphur-containing acids, by metal catalysis (Cullis and Trimm 1968). The ability of such changes for the transition metals occur in the order: $\text{Cu} > \text{Mn} > \text{Fe} > \text{Ni} \gg \text{Co}$ (Bagiyan et al. 2003). One example of the single-electron oxidation of cysteine is DAHPS (phenylalanine-sensitive 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase), an enzyme from *Escherichia coli* that undergoes inactivation and destabilisation of its active site cysteines by Cu^{2+} induced oxidation (Giles et al. 2003)(Park and Bauerle 1999). The reaction is initiated by a one-electron transfer from Cu^{2+} , forming Cu^+ and the first cysteinyl radical (Figure 82 a). A hydroxyl radical is then produced by the reaction of Cu^+ and H_2O_2 , regenerating Cu^{2+} (Figure 82 b). The hydroxyl radical reacts with a second cysteine residue, forming another cysteine radical and water (Figure 82 c). The resulting two cysteinyl radicals dimerise to form a disulphide (Figure 82 d). Additional oxidative modifications also occurred for the enzyme as a result of Cu^+ , including attacks on the nonessential cysteine and histidine residues leading to the conformation changes necessary for the repositioning of the cysteine side chain, and subsequent formation of the disulphide.

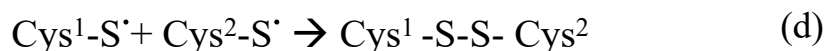
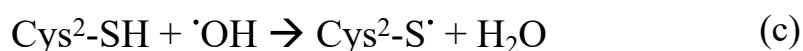
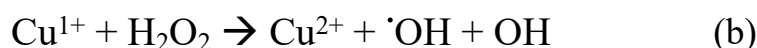
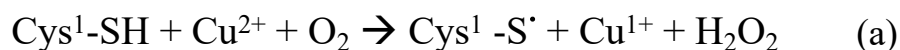


Figure 82. Oxidation of cysteines by metal catalysis. Equation adapted by (Park and Bauerle 1999)

Under mild conditions (pH 7-9) like those used in this study, it may be assumed that thiol compounds can be self-oxidised by oxygen to disulphides. However, Bagiyan *et al.*, (2003) showed that in neutral and alkaline conditions, these reactions are often induced by impurities of variable-valence metals, and the slow kinetics of self-oxidation is indicative of their low reactivity with O₂. Also, the specific reactivity of each cysteine thiol is governed by its microenvironment in the protein such as solvent accessibility, pKa, and polarity of neighbouring residues (Alcock *et al.* 2018).

The oxidation of C148 in sfGFP^{C148.F} to a disulphide is catalysed in accordance to the reported reactivity of transition metals (Cu²⁺ and less effectively Mn²⁺). Thus, it is likely the oxidation of C148 occurs similarly to that described in Figure 82. Also, the trial dimerisation of sfGFP^{C148.F} (section 5.2.3.1) evidenced that dimerisation can partly occur via self-oxidation at higher protein concentrations. The structures of dimeric sfGFP^{C148.F} also prove that the metal ions are not present within the protein or protein interface, so are thus involved in changing the oxidation state of cysteine. By using metal ions to catalyse the reaction, this allows dimerisation to occur at a more measurable rate, reach completion, and allows lower concentrations of proteins to be sampled. The complete spectral switch allows for the calibration of monomeric and dimeric protein, which is critical for the use of sfGFP^{C148.F} to perform as a homodimer biosensor. This will be the focus of Chapter 6.

5.2.6 Conclusion

The design and production of self-assembling protein complexes is currently an area of great interest to protein engineering (Gwyther *et al.* 2019). By using

structural information based on protein oligomers present in nature, protein complexes can be produced that have improved or even novel functional properties. To achieve this, engineered proteins require a favourable interface by which they form a variety of interchain interactions, either non-covalently i.e hydrophobic, electrostatic interactions, hydrogen bonding, or covalently through disulphide bridging. Alternatively, Worthy *et al.*, (2019) demonstrated how ncAA can be utilized for their ability to perform a bioorthogonal reaction, enabling a site specific covalent link between sfGFP monomers (section 5.1). However, not only does the incorporation of ncAA carry its own limitations (i.e. genetic code reprogramming), this dimerisation approach still requires the use of two mutually compatible tags, so cannot monitor homodimerisation. Although effective for heterodimerisation, further engineering would be significantly advantageous to expand this use for homodimers, particularly given that majority of protein complexes are homodimers (Gwyther *et al.* 2019).

Chapter 4 had previously demonstrated the potential of sfGFP^{C148} to form homodimers by the formation of disulphide bridging between C148 residues. By utilizing the structure of sfGFP^{148ncAAx2}, an additional mutation V206F was identified that encouraged and further stabilised the formation of sfGFP^{C148.F} dimers. The spectral properties of the sfGFP^{C148.F} change as the protein exchanges between its monomeric and dimeric state. Thus, by monitoring the spectra, the oligomeric status of the proteins can be determined. The crystal structure of both oligomeric forms was solved that determined the exact mechanism of spectral switching; where symmetrical water molecules are a key factor within the dimer, allowing the formation of continuous transport networks.

6 Design of a homodimer biosensor

6.1 Introduction

Protein-protein interactions (PPIs) refer to the specific, mutually compatible physical contacts that occur between proteins in order to form a multi-protein complex required to carry out a required biological event. PPIs are responsible for a wide variety of biological functions from cell-to-cell interactions to metabolic and developmental control (Rao et al. 2014). It has been revealed that over 80 % of proteins operate as a complex instead of simple monomeric subunits (Berggård et al. 2007). The interactions between proteins can be transient or permanent based on their lifetime and stability (Peng et al. 2017). For example, some protein assemblies are stable because they constitute macromolecular protein complexes and cellular machines (e.g. ATP synthase or RNA polymerase). Transient interactions can both associate and dissociate quickly, such as hormone receptor binding, signal transduction and inhibition of proteases (Wang et al. 2014). PPIs are additionally defined by their biological context and will often depend on cell type, cell cycle phase, development stage, environmental conditions, protein modifications and presence of cofactors. Also, the normal patterns of PPIs can often become disrupted as a result of environmental stress or in disease state (Kuzmanov and Emili 2013).

Such is their importance, the development of approaches to detect and monitor PPIs *in vivo* and *in vitro* has been vital. One key method that is widely used to study biomolecular interactions within cells is FRET. This utilises the non-radiative (dipole-dipole) energy transfer from a fluorescent donor to a compatible acceptor when both fluorophores are situated within ~10 nm (Sekar and Periasamy 2003). When two proteins of interest are labelled with the donor and acceptor tags, FRET only occurs when the two proteins interact with one another. It has proved a powerful and popular approach to monitoring PPIs. However, the dependence on two different fluorophores restricts FRET to hetero-oligomeric events, so is unable to monitor homodimerisation, a significant downfall given that homodimers are the most common oligomeric state (Gwyther et al. 2019). Homo-FRET is possible with identical FPs but is largely an anisotropy-based measurement process rather than a direct change in fluorescence intensity. It also requires specialised spectroscopy and imaging setups, and off pathway processes reduce sensitivity and accuracy (Chan et al. 2010). The accuracy and sensitivity of classical FRET can also pose issues with

several assumptions key to defining energy transfer efficiency (and thus a measure a PPI) potentially leading to misleading results. These include fluorophore dipole alignment, and FP associated binding that can themselves lead to spectral changes (Pope et al. 2021). Alternative methods to FRET have been developed including fragment complementation in split-FPs (Ghosh et al. 1996; Hu et al. 2002) and dimerisation dependent FP exchange (Alford et al. 2012a; Alford et al. 2012b). However, both approaches still require two different FP tags so limit their use for monitoring homodimerisation.

To explore the full repertoire of PPIs, homodimerisation events need to be monitored as well as hetero-oligomerisation events. Worthy *et al.*, (2019) showed how sfGFP can be engineered to undergo a change in its spectral properties on heterodimerisation through switching the charge of the chromophore. Throughout Chapters 4 and 5, we have implemented a similar switch for monitoring homodimerisation. By mutating residue 148 of sfGFP to a cysteine, the resulting neutral chromophore state dominates in the monomer. On dimerisation via the formation of a disulphide bond between C148 residues, the phenolate anion dominates the chromophore ground state exclusively. The engineered dimer retains synergistic properties of the previous dimer constructs with brightness 3.0-fold higher than sfGFP^{WT}. The dynamic range between monomer and dimer is up to 30-fold. Also, by measuring emission on excitation at ~ 400 and 495 nm, the relative populations of monomer and dimer can be estimated, enabling a ratiometric, concentration-independent approach to monitoring PPIs.

In this chapter, the engineered GFP from Chapter 5 (sfGFP^{C148.F}) will be used for attachment to p50, a naturally forming homodimer present within the NF- κ B protein family (see section 1.1.1 for more information). This will be used as a model protein to exemplify the use of GFP^{C148.F} for monitoring homodimerisation. The complex itself is naturally heterogenous with p50; capable of dimerising with itself or RelA, where each possible complex has distinct functional features. Thus, it acts as a good model system to monitor dynamic processes that have implications in terms of effects on transcription. As the primary PPI, p50 is the main driver of protein association, bringing sfGFP^{C148.F} close in space so increasing its effective local concentration and dimerisation propensity. SfGFP^{C148.F} dimerisation is then initiated with CuSO₄ and so effectively reporting the homodimerisation of the

primary PPI. The main assumption is that the rate and extent of GFP^{C148.F} association should be significantly higher in the presence of the primary PPI.

6.2 Results

6.2.1 Design and construction of p50GFP

To determine whether the engineered sfGFP dimerisation system (outlined in chapter 5) can be used to monitor homodimerisation, GFP^{C148.F} was fused to p50. The C-terminus of p50 (or dimerisation domain) was chosen for attachment of GFP^{C148.F} as this ensured FPs were close in space following the primary PPI (Figure 83). A 17 amino acid linker (same sequence used for p50-Venus, Chapter 3) was added between p50 and sfGFP^{C148.F} to allow suitable motion of the FP for dimerisation via the C148 residues.

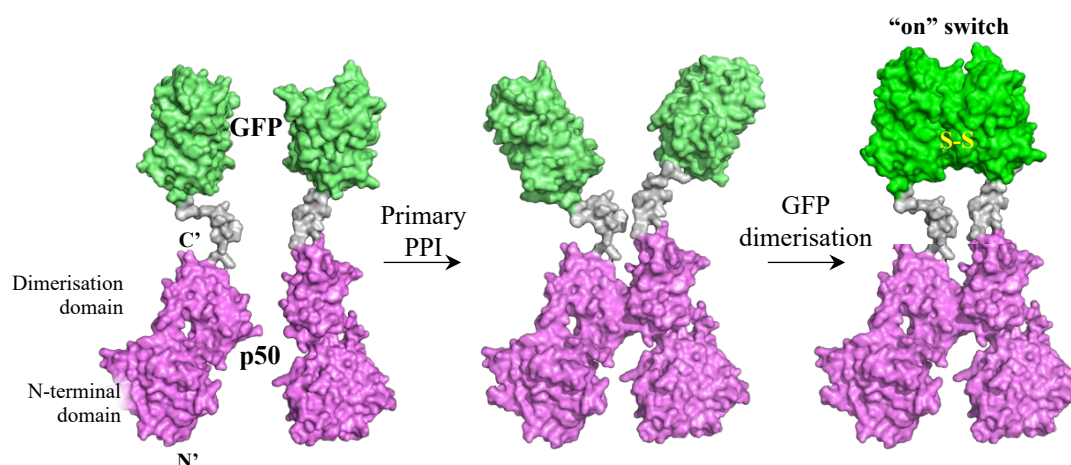


Figure 83. Schematic outline of homodimer PPI process. The purple regions represent p50, grey region is linker segment and green is sfGFP^{C148.F}. The primary PPI, in this case p50 homodimerisation brings sfGFP^{C148.F} into close proximity so promoting formation sfGFP^{C148.Fx2} and thus switching “on” fluorescence emission of the CroB form.

To generate the desired plasmid construct, the sfGFP^{WT} gene was cloned into the pCA24N vector, already containing p50. Suitable restriction sites at the 3' end of p50 were identified and used to design primers for amplification of the sfGFP^{WT} gene (Figure 84 a,b). The restriction endonucleases *SacI* and *HindIII* were used to digest both the vector and gene insert (Figure 84 c). The DNA was checked via an agarose gel to confirm bands of ~ 5.5 and 0.7 kB for vector and sfGFP^{WT} gene respectively (Figure 84d). The complementary ends were then used to ligate the DNA before transformation onto LB agar plates supplemented with antibiotics and

IPTG (0.4 mM). The plate was viewed under UV light, confirming the insert of $sfGFP^{WT}$ gene into the vector by the presence of green colonies. Sequencing results confirmed the correct sequence for the $p50GFP$ gene in the pCA24N plasmid.

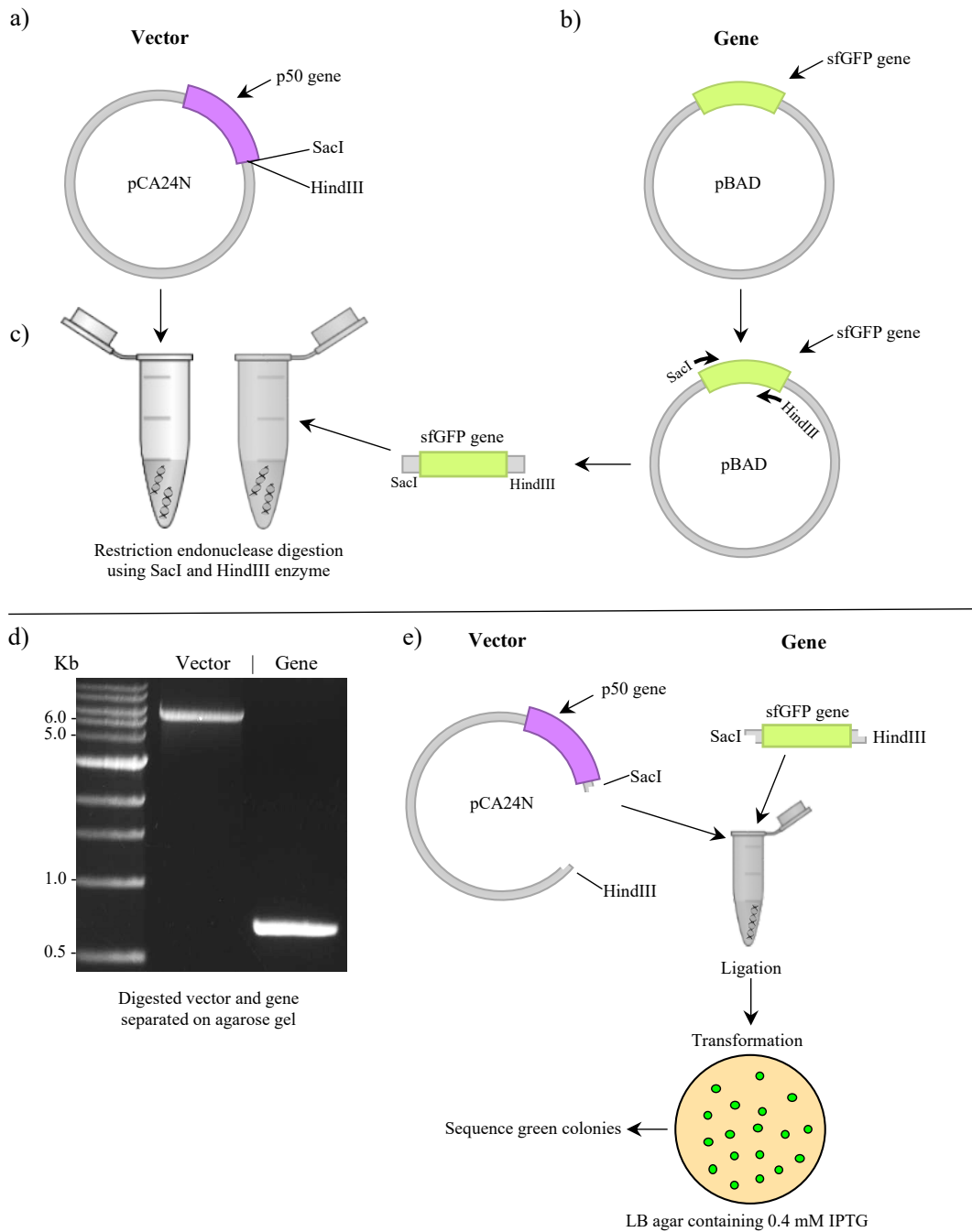


Figure 84. Construction of pCA24N vector containing $p50GFP$ fusion. A) The vector pCA24N with *SacI* and *HindIII* restriction sites at the 3' end of $p50$ gene (purple). b) The pBAD vector map used in PCR with primers to amplify the $sfGFP$ gene (green) with *SacI* and *HindIII* restriction sites. C) Restriction endonuclease digestion of both vector and gene. d) Agarose gel showing digested vector (~ 5.5 kb) and $sfGFP$ gene (~ 0.7 kb). E) $SfGFP$ gene inserted into the pCA24N vector between *SacI* and *HindIII* sites by ligation, followed by transformation into competent cells.

For use as a homodimerisation sensor, H148C and V206F mutations were required within sfGFP. Whole vector PCR was performed with mutagenic primers on the pCA24N vector (Figure 85a). DNA amplification was confirmed by the presence of bands the approximate size of the vector (~ 6.2 kb) (Figure 85b). The DNA was purified, phosphorylated and ligated before transformation into *E.coli* competent cells. The plate was viewed under UV light to select green colonies. Successful mutagenesis was confirmed via sequencing.

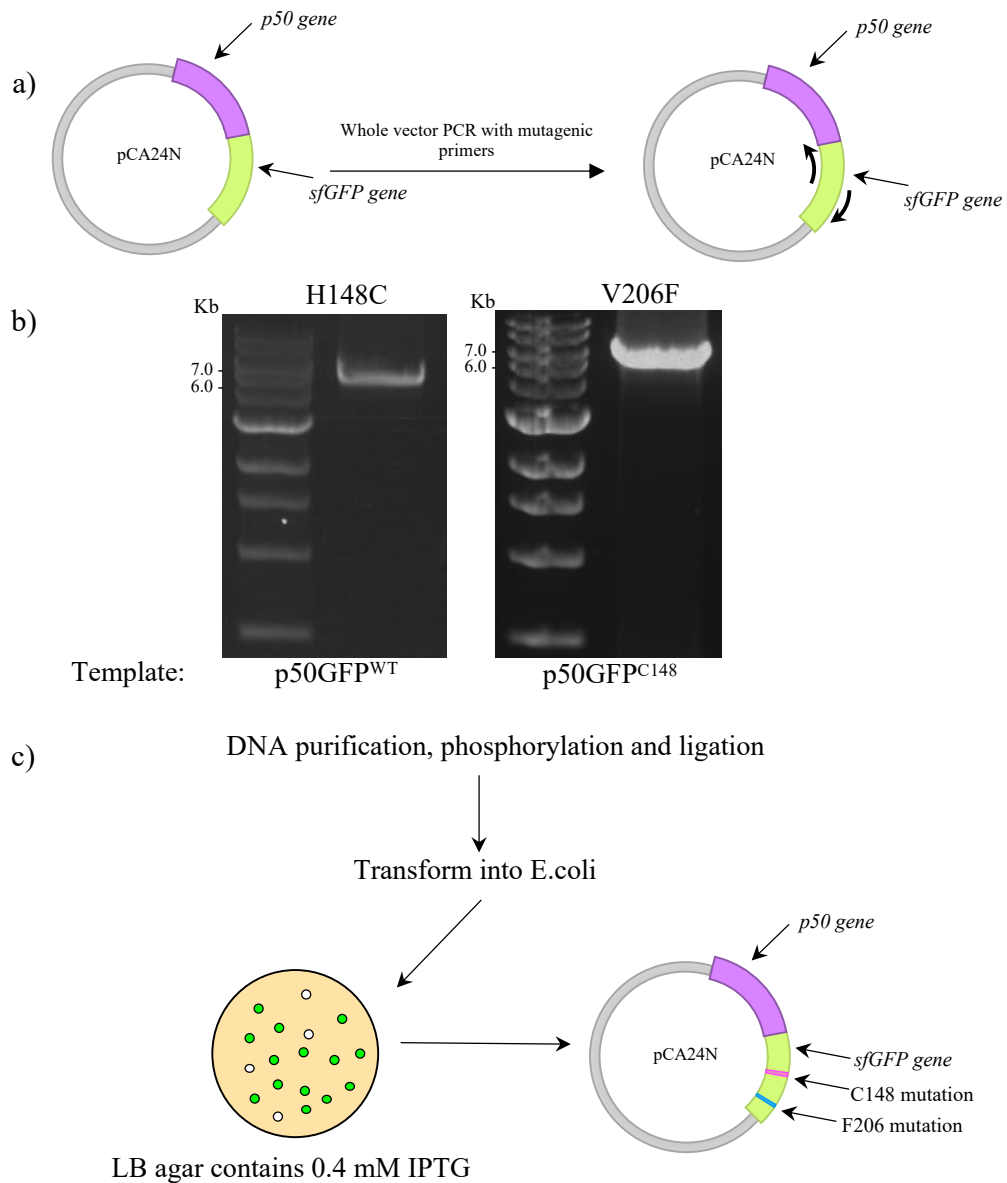


Figure 85. SDM of p50GFP to incorporate sfGFP H148C and V206F mutations. A) Whole vector PCR with mutagenic primers. B) Gel electrophoresis. C) DNA ligation and transformation into *E.coli*.

6.2.2 Expression and purification of p50GFP^{C148.F}

Prior to analysis, p50GFP^{C148.F} had to be expressed and purified. P50GFP^{C148.F} contains an N-terminal His tag to allow binding to a nickel column and allow non-target proteins to flow through. However, elution and brief storage in high concentrations of imidazole had proven to increase protein degradation. Thus, two purifications were carried out in tandem, where the SEC column (HiLoadTM 16/600SuperdexTM S200 pg) was connected after contaminating proteins had eluted from the nickel column. A short interval (10 mL) with 500 mM imidazole was applied to elute the protein from the nickel column straight onto the SEC column (Figure 86a). The original buffer (50 mM Tris, pH 8.0) was then used to flow the protein through the SEC column, allowing separation from imidazole. Protein was analysed via an SDS-PAGE gel, which showed bands the approximate size of p50GFP^{C148.F} (~ 65 kDa) (Figure 86 b). Successful purification was confirmed with ~ 95 % purity using ImageJ. It should be noted however that after a brief period of time (as short as 24 hours), p50GFP spontaneously cleaved into p50 (~39 kDa) and sfGFP (~26 kDa) (Figure 86c). It was therefore important to use p50GFP^{C148.F} for downstream analysis on the same day as purification.

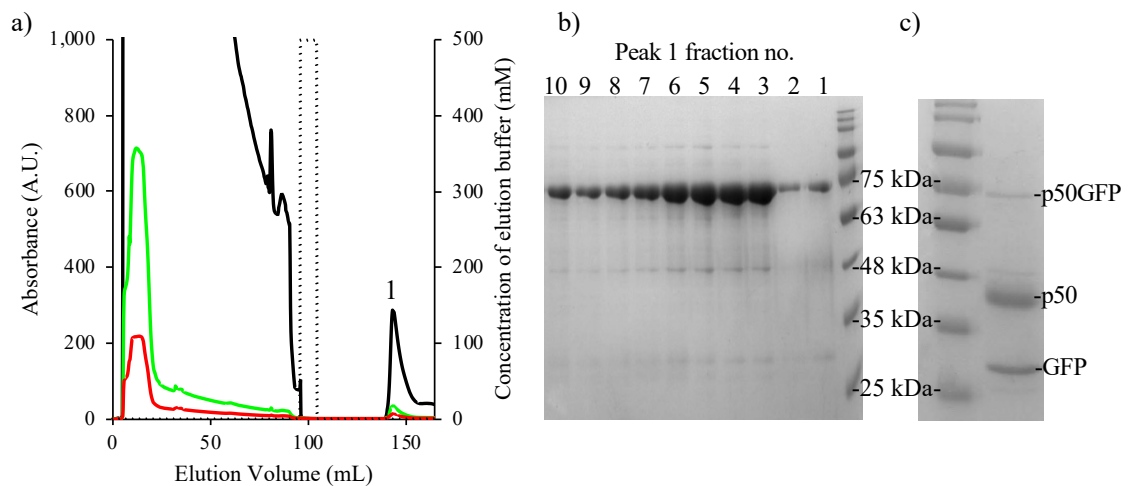


Figure 86. Purification of p50GFP^{C148.F}. a) Elution profile from nickel affinity chromatography and size exclusion chromatography. Absorbance at 280, 400 and 495 shown in black, green and red respectively. Dotted line shows concentration of imidazole. b) SDS-PAGE gel showing purified protein following SEC. c) SDS-PAGE gel of proteins 24 hours after purification.

6.2.3 Dimerisation rates of p50GFP^{C148.F}

To determine whether homodimerisation of p50GFP^{C148.F} can be monitored and whether this is driven by the primary PPI, the 483:388 emission ratio was recorded as previously performed in chapter 5. The fluorescence was recorded upon excitation at both 388 and 483 nm following addition of 1 mM CuSO₄ at specific time intervals (0.5, 5, 10, 15, 30, 60 and 120 mins). The results were used to determine the 483:388 emission ratio and plotted against a ‘one site-specific binding’ fit using GraphPad software. The results showed a rapid increase in emission ratio during the initial 5 minutes for each of the concentrations of protein tested (0.1, 0.25, 0.5, 1.0 and 2 μM) (Figure 87a). Upon conversion to % dimerised, the results showed that for increased protein concentrations, a higher percentage dimerised was recorded; at 0.1 μM, 20% was dimerised compared to ~ 65 % at 2.0 μM (Figure 87b). At each concentration, the majority of dimerisation occurred with the first 10 min.

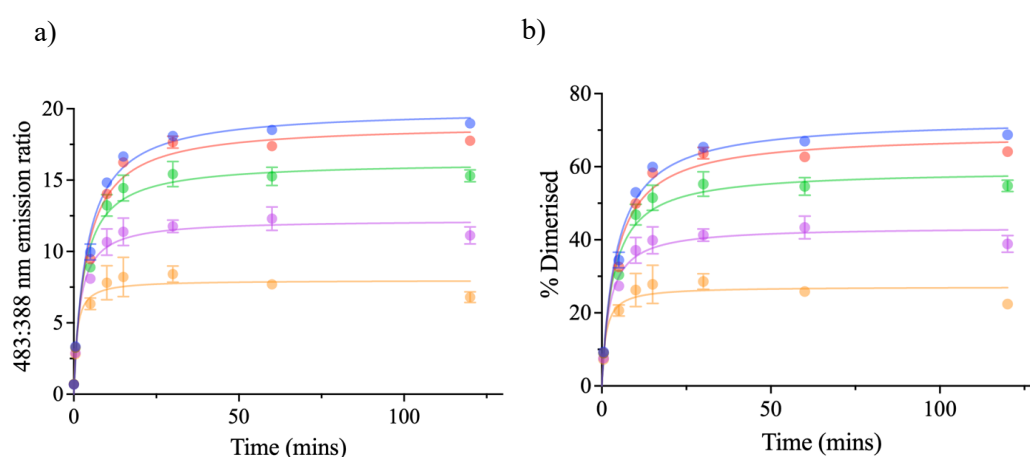


Figure 87. Dimerisation rates of p50GFP^{C148.F}. Data plotted using Graphpad. Non-linear curve plotted as one site-specific saturation curve with 95% confidence level ($\text{Ratio} = \text{Ratio}_{\text{Max}} * X / (Z + X)$, where X is time and Z is time to achieve half $\text{Ratio}_{\text{max}}$). Mean points plotted with standard deviation (SD) error bars. Varying protein concentrations were as followed; 0.1 μM in yellow, 0.25 μM in purple, 0.5 μM in green, 1.0 μM in red and 2.0 μM in blue. a) Raw results of CroB values divided by CroA values. Data recorded on plate reader (CroA = 388-8/516-8, CroB = 483-8/516-8). b) Percentage dimerised calculated from 483:388 nm emission ratios.

For this system to effectively report on homodimerisation, the rate and extent of sfGFP^{C148.F} association should be significantly lower in the absence of the primary PPI. The dimerisation rates were therefore compared for p50GFP^{C148.F} against sfGFP^{C148.F}. The experiment was carried out similarly to that above using a variety of protein concentrations. The results showed that the emission ratio increased rapidly for sfGFP^{C148.F} when fused to p50. For each of the reactions (0.1, 0.25 and 0.5 μM), p50GFP^{C148.F} had reached completion after ~ 5 mins. In comparison, the emission ratio for sfGFP^{C148.F} was still increasing after 120 mins at 0.5 μM . The second order rate constant was determined as $0.34 \pm 0.1 \mu\text{M}^{-1}\text{min}^{-1}$ for p50GFP^{C148.F} and $0.008 \pm 0.004 \mu\text{M}^{-1}\text{min}^{-1}$ sfGFP^{C148.F} respectively. Thus, the rate of association is ~ 40 -fold faster when attached to p50. This suggests that p50 is the primary driver of PPI bringing sfGFP^{C148.F} close in space so increasing its effective local concentration and dimerisation propensity. Thus, this system can effectively monitor p50 homodimerisation.

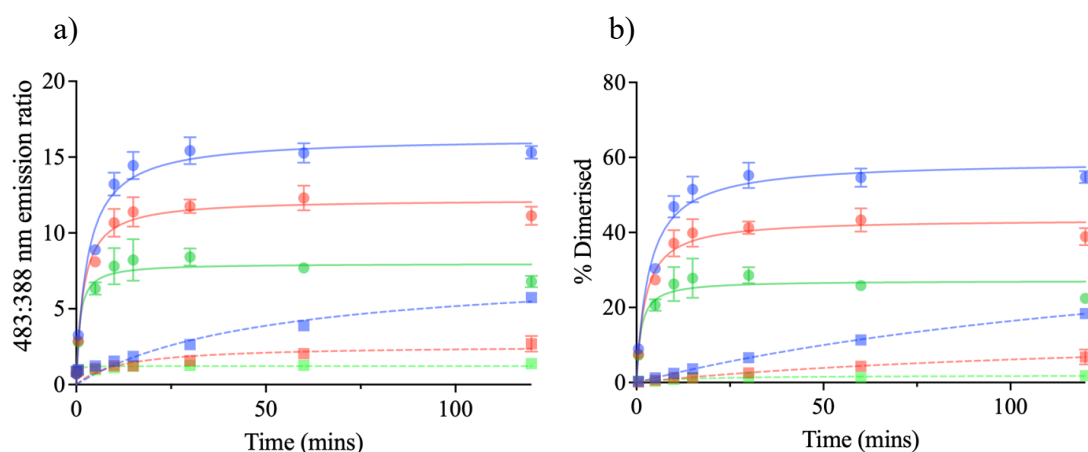
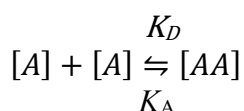


Figure 88. Dimerisation rates of p50GFP^{C148.F} (solid lines) vs GFP^{C148.F} (dashed lines). Data plotted using Graphpad. Non-linear curve plotted as one site-specific saturation curve with 95% confidence level ($\text{Ratio} = \text{Ratio}_{\text{Max}} * X / (Z + X)$, where X is time and Z is time to achieve half $\text{Ratio}_{\text{max}}$). Mean points plotted with standard deviation (SD) error bars. Varying protein concentrations were as followed; 0.1 μM in green, 0.25 μM in red, 0.5 μM in blue. a) Raw results of CroB values divided by CroA values. Data recorded on plate reader (CroA = 388-8/516-8, CroB = 483-8/516-8). b) Percentage dimerised calculated from 483:388 nm emission ratios.

6.2.4 Calculation of dissociation constant

A commonly used term to quantitatively define PPIs is the dissociation constant or K_D and its less commonly used reciprocal, the association constant or K_A . Both K_D and K_A are related by the reversible dissociation and association of the individual complex components as outlined in Equation 12.

Equation 12.



To determine the equilibrium constant, normally K_D , we need to know the proportion of free and complexed components at a particular concentration as indicated by Equation 13.

Equation 13.

$$K_D = \frac{[A][A]}{[AA]}$$

The reciprocal of K_D is the K_A (i.e. $1/K_D$). Dissociation constants is used to quantitatively describe affinity, with the lower the K_D , the higher the affinity. Although the association of sfGFP^{C148.F} is considered irreversible (unless with the use of DTT), the change in spectral properties can be used to report a pseudo or apparent K_D of the primary PPI, especially when early timepoints (sub 30 minute incubations) for complexation are monitored for each concentration, where the main driver for signal change is the primary PPI event and not the secondary sfGFP association.

The dimerisation values (30 mins for p50GFP^{C148.F} mins for sfGFP^{C148.F}) for each of the concentrations sampled (0.1, 0.25, 0.5, 1.0 and 2 μ M) were taken and plotted using GraphPad (Figure 89). The pseudo K_D of p50 was calculated as $0.276 \pm 0.024 \mu$ M. This value is within the low micromolar range as previously reported (Sengchanthalangsy et al. 1999; Kohl et al. 2019). In comparison, sfGFP^{C148.F} in the absence p50 had a much lower affinity; the pseudo K_D of sfGFP^{C148.F} dropped ~ 22 -fold to $6.114 \pm 0.111 \mu$ M. This highlights the primary PPI event (i.e. p50 homodimerisation) is driving association. Also, it confirms the differences in the

observed kinetics of association as the equilibrium constants are function of the on and off rates.

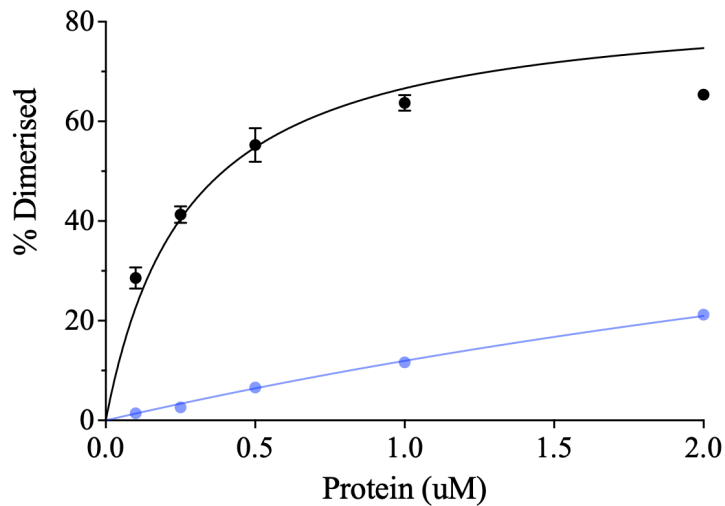


Figure 89. Affinity of sfGFP^{C148.F} alone (blue) or when fused to p50 (black). Mean values with 95% confidence level plotted with SEM error bars.

6.2.5 Measuring homodimerisation in *E. coli*.

Now the system has been proven to work *in vitro*, it was then important to determine whether it could effectively report on homodimerisation within a cell. As a first approach, *E. coli* was used as it required no further cloning. However, one of the main limitations to this approach was the instability of p50GFP^{C148.F} following *E. coli* lysis. Samples of cells expressing p50GFP^{C148.F} were taken both pre and post induction, and analysed by SDS-PAGE. The gel showed an increase in band intensity at sizes corresponding to p50GFP (65 kDa), p50 (39 kDa) and sfGFP^{C148.F} (26 kDa) following addition of IPTG (Figure 90). Of the three components, p50GFP^{C148.F} represented just 6 % (according to ImageJ). Thus, monitoring the spectral switch of cell lysate following expression of p50GFP^{C148.F} would inaccurately report on homodimerisation, as the majority of sfGFP^{C148.F} had cleaved from p50.

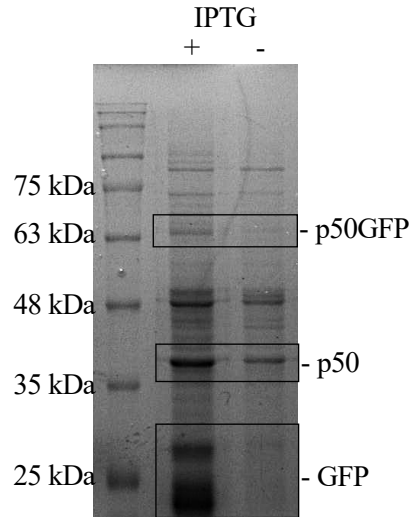


Figure 90. Protein expression test of p50GFP^{C148.F} in *E. coli*. Soluble pre and post induction of cells with IPTG. Bands normalised to an OD₆₀₀ of 0.7.

Thus, the homodimerisation system was trialled within whole *E. coli* cells by adding CuSO₄ (1 mM) to the *E. coli* cell culture, following expression of p50GFP^{C148.F}. However, the 483:388 nm emission ratio remained at ~ 0.9 after 60 mins of CuSO₄ incubation, indicative of monomeric protein (Table 21). Thus, it is presumed that copper sulphate cannot enter *E. coli* and initiate the dimerisation and the formation of the disulphide bond between sfGFP^{C148.F} for effective reporting of p50.

Table 21. The 483:388 nm emission ratios following incubation in 1 mM CuSO₄ of whole cells expressing sfGFP^{C148.F} and p50GFP^{C148.F}. The OD was measured at 600 nm.

Variant	OD	Time (mins)				
		0 (w/o)	5	10	30	60
sfGFP ^{C148.F}	1	0.9	0.9	0.9	0.9	0.9
	0.5	0.9	0.9	0.9	0.9	0.9
p50-GFP ^{C148.F}	1	0.9	0.9	0.9	0.9	0.9
	0.5	0.9	0.9	0.9	0.8	0.9

6.3 Discussion

Homooligomers are the most commonly observed oligomers, at least in the PDB, with symmetrical homodimers being the most prevalent structural state (Gwyther et al. 2019). With current methods to detect this PPI limited to predominantly monitoring hetero-oligomerisation processes, there is a need to develop alternative and compatible approaches to monitor homo-oligomerisation. To overcome current limitations, we have demonstrated how a single engineered fluorescent protein can undergo a spectral switch upon dimerisation. When fused to a protein of interest, this system can be used to monitor and detect homodimerisation *in vitro* (and soon potentially *in vivo*). The primary PPI increased the rate of sfGFP dimerisation by ~ 40 -fold and accurately reported a dissociation constant of p50 of $0.276 \pm 0.024 \mu\text{M}$.

The additional benefit of the new system developed in this thesis is the ability to monitor simultaneously linked and/or competing homo- and hetero-oligomerisation events. This can be achieved through changes in FRET. For example, Chapter 3 described how FRET was used to monitor p50 interaction with Bcl3. Using the current p50GFP^{C148.F} combined with a suitable FP labelled Bcl3, both self-association of p50 and complexation with Bcl3 can be monitored as a dynamic process providing information on the assembly mechanism. For example, if Bcl3 binds to monomeric p50 (fused to sfGFP^{C148.F}), FRET will occur predominantly on excitation at 400 nm; this will switch to 490 nm if the homodimer of p50 is required for assembly. As p50 can form heterodimers as well as homodimers, the approach can be used to determine which cellular events drive which oligomerisation process. For example, by tagging RelA (also known as p65) with a FRET compatible FP both homo-dimerisation and hetero-dimerisation events can be monitored. Thus, the approach developed here can be applied more broadly to monitor the full repertoire of PPIs, with suitable linker sequences being used.

The chapter has focused on the use of sfGFP^{C148.F} as the main dimerisation monitoring system. I have shown that the K_D for this association is $6.114 \pm 0.111 \mu\text{M}$ (Figure 89). While this is suitable for the p50 system used here due to the higher affinity of the PPI, many PPIs, especially transient interactions are in the low μM range. This would mean that the GFP^{C148.F} dimerisation aspect could be as dominant as the target PPI. One way to overcome this problem is revert to using the sfGFP^{C148}

variant (no V206F mutation). This original variant had lower binding kinetics and thus affinity (see section 5.2.3.6) so should be useful for assessing lower affinity PPIs.

One of the main limitations within this chapter was the spontaneous cleavage of p50GFP into its component parts (p50 and sfGFP) in *E. coli*. The linker (17 amino acids) used between p50 and GFP was the same as that used between p50Venus, in chapter 3. Although similar cleavage events occurred between p50-Venus, it seemed to occur significantly more with p50GFP. The main difference between protein preparation and storage was the buffer used; 50 mM Tris, pH 8.0 for p50GFP and 20 mM HEPES, 300 mM NaCl, pH 8.0 for p50Venus. It was clear throughout initial expression trials, that p50 was more stable within HEPES. However, this buffer caused the immediate precipitation of CuSO₄, the compound required for disulphide bond formation. Thus, Tris was used throughout this chapter, as already proven stable for sfGFP (throughout chapters 4 and 5). It is therefore predicted that the use of Tris caused the quicker degradation of p50. This is also likely the case given the SDS-PAGE analysis of component proteins following cleavage, as sfGFP has a greater band intensity than p50. As a result, the spontaneous cleavage of the fusion protein meant that monitoring homodimerisation of the primary PPI could not be performed with the bacterial cell lysate, as the sfGFP signal was primarily contributed by free unbound protein (Figure 90). The instability of p50GFP also required that downstream analysis to be performed on the same day as purification. This severely restricted use of the protein and required extensive forward planning of downstream experiments.

The requirement for copper could be considered either a major problem or a route to explore other sensing events. Copper is an essential micronutrient that serves as a cofactor for key enzymes involved in respiration and redox defence such as cytochrome c oxidase, NADH dehydrogenase, and superoxide dismutase (Rensing and Grass 2003) However, excess levels of copper are toxic to aerobic bacteria due to the formation of reactive oxygen species (ROS) (Grey and Steck 2001)(Nies 1999). Also, high levels can cause the inactivation of enzymes involved in the TCA cycle and pentose phosphate cycle by displacement of iron from iron-sulphur proteins (Decembrino et al. 2021). Thus, bacteria restrict the amount of copper that is available to react with intracellular targets by; controlling copper import, secreting excess copper and increasing cells buffering capacity (Giachino and Waldron 2020).

As well as the inherent resistance and regulatory mechanisms, copper exists in a chemical equilibrium where Cu(II) is favoured in oxidising conditions such as the periplasm and Cu(I) prevalent during anoxic growth within the cytoplasm (Rensing and Grass 2003)(Giachino and Waldron 2020). Given that the GFP^{C148.F} dimerisation is copper dependent, it could act as a genetically encoded biosensor for copper levels inside of cells. The p50 would have to be replaced with an alternative biological element as the monomer kinetics are too slow to measure copper levels in real time. One such dimerisation element is the leucine zipper component of some DNA binding proteins. The leucine zipper is a short peptide sequence that homo-dimerises with high affinity so essentially forming a permanent dimer. By fusing the leucine zipper sequence to the terminus of GFP^{C148.F}, a potential copper sensing construct can be developed. This work is currently being pursued in the Jones' lab.

Throughout this project, copper (CuSO₄) has been used *in vitro* to catalyse the oxidation of C148 sidechain resulting in the formation of a disulphide bond between two GFP subunits. This covalent bond is critical for the spectral switch of CroA to CroB, and the effective monitoring of the PPI. Currently, the redox compartmentalisation of bacterial cells prevents the *in vivo* use of copper as an oxidising agent for GFP^{C148.F}. This explains why dimerisation could not be monitored inside bacterial whole cells. The same limitation could also be the case for the use of this homodimerisation biosensor for live cell imaging within mammalian cells. Under non-stressed conditions, disulphide bond formation is primarily restricted to the oxidative environment of the endoplasmic reticulum (ER) in eukaryotic cells (Cumming et al. 2004). Whereas, the majority of protein cysteine residues remain protonated within the reducing environment of the cytoplasm. However, recent cell imaging work using the p50GFP^{C148.F} construct suggests this may not be such a significant issue in mammalian cells as it is in *E. coli* which some encouraging imaging results (Zitti, Watson and Jones, unpublished data). However, ideally no copper should be required and a purely non-covalent GFP homo-dimer complex should be sought. This should be achievable given that the structure of the GFP homo-dimer complex is known. The structure can be used to guide future engineering effort to optimise the non-covalent homo-dimer interface interactions through mutations. The leucine zipper construct mentioned above can also be used in this engineering endeavour. GFP variants are initially selected with a dominant CroA form (e.g. H148V or H148A). The GFP is then engineered or directly evolved in the

context of the leucine zipper dimerisation domain; those that switch to the CroB form will be selected for as potential homodimerisation sensing variants and explored further. Again, this work is currently under way in the Jones' lab.

6.4 Conclusion

Within this chapter, p50, a member of the NF- κ B protein family was used to demonstrate the use of our engineered sfGFP variant for monitoring homodimerisation. By fusing sfGFP^{C148.F} to the dimerisation domain of p50, a ~ 40 fold increase in dimerisation rate was recorded. As well as kinetics, a pseudo K_D was calculated as $0.276 \pm 0.024 \mu\text{M}$, within the expected range for p50 homodimers. Thus, the spectral switch of sfGFP induced by a disulphide bond between residues C148, can be used to directly monitor homodimerisation of p50. This offers an alternative approach to current methods (e.g. homo-FRET) for detecting homodimers.

7 Discussion

7.1 General overview

Protein oligomerisation offers new structural and functional properties beyond those available to a simple monomeric system. It plays a key role throughout many cellular processes; from the formation of large complex structures (cytoskeleton and collagen) and multienzyme complexes (RNA polymerase and pyruvate dehydrogenase), to small functional dimers (cell receptors and transcription factors) (Gwyther et al. 2019). With such importance, protein oligomerisation is often considered a phenomenon crucial in triggering a variety of physiological pathways. Thus, identifying PPIs that comprise protein oligomers is crucial to underpin a vast number of diverse biological processes.

High expression of Bcl3 has been detected in many types of cancers including breast cancer, nasopharyngeal carcinoma, colorectal, cervical and ovarian cancers (Cogswell et al. 2000)(Choi et al. 2010)(Ma et al. 2020)(Urban et al. 2016)(Zhao et al. 2016)(Zou et al. 2018). When complexed to the p50 homodimer, Turnham *et al.*, (2020) evidenced the involvement of Bcl3 as an upstream regulator of the two principal forms of breast cancer cell motility, involving collective and single cell migration. Thus, investing the PPI between Bcl3 and its binding partner is critical for improving our understanding of biological role of the protein complex. This thesis was therefore broken up into two main sections; firstly to understand the PPI between the Bcl3-p50 complex, including how the newly designed drugs effect these interactions. Second was to address the challenges in monitoring PPIs, particularly homo-dimerisation events.

The Bcl3-p50 complex was successfully formed by the co-expression of Bcl3 and p50 proteins in *E.coli*. Characterisation by SDS-PAGE and SEC-DLS confirmed the stoichiometric ratio, where one Bcl3 molecule was bound to a p50 homodimer. However, monitoring of the PPI by FRET highlighted a major challenge throughout chapter 3 and subsequently impacted the investigation of drug interactions with Bcl3. *In silico* modelling proved how assumptions that are necessary for FRET calculation, are not always true. For example, FPs can be either freely rotating or closely associated affecting the distance between chromophore pairs. Thus, this chapter has proven how modelling the dynamics of FP fused protein complex can

help predict these aspects to improve accuracy of FRET, for improved reporting of PPI.

Another limitation of FRET that brings us to the second aspect of this work; is that hetero-FRET will inefficiently report on homo-dimerisation events and will suffer from a complex signal arising from the distribution of fluorophores. Whilst homo-FRET has potential, monitored by changes in fluorescence polarisation, it is not able to discern affinity and suffers from major challenges in tractability of detecting the signal change. Thus, we need suitable tools that can detect homo-oligomerisation and potentially couple that through to any associated hetero-dimerisation events that comprise a functional complex i.e. fluorescence changes on homo-oligomerisation enables FRET on hetero-oligomerisation. We engineered a FP (sfGFP) that will undergo a switch in its spectral properties upon dimerisation. By tagging the FP to p50, a naturally forming homodimer, the fluorescence was monitored and used to confirm its oligomeric state. As the most common type of protein assembly in nature, the detection of homodimers within cells would improve our understanding of many molecular interactions. With that knowledge, new targets for drug design can be identified that were not previously known so providing new opportunities for treatments and therapies for cancer.

7.2 Usefulness of *in silico* modelling

In silico modelling has achieved structure prediction accuracy at such a level that is being used in drug design, virtual screening, protein engineering, and site directed mutagenesis applications (Gupta et al. 2014). It has also aided researchers to reveal the sequence to structure relationship, unveiling the novel pathway of protein folding and characterising the active and catalytic sites of proteins. Thus, a wide range of computational software has been developed and improved over the years to achieve these goals such as: GROMACS, Rosetta, GAMESS, alphafold and MOE. These techniques have been utilized on their own or through combination of multiple software. The more recent development of AlphaFold has a novel machine learning approach that employs a deep learning algorithm to predict protein structures, even in cases where no similar structure is known (Jumper et al. 2021). GROMACS was used by Watkins *et al.*, (2015) for successful *de novo* protein design of oxygen-activating oxidoreductases. Whereas, a

combination of *in silico* techniques were utilized for the drug design of JS6, the compound used in this study that was designed by Prof Andrea Brancale (School of Pharmacy, Cardiff University) to bind to a unique pocket within ANK repeat 6 and 7 of Bcl3 and block interactions with p50 (Soukupova et al. 2021).

Throughout this work, MD simulations were performed to relax the proteins and allows the proteins to sample dynamics. In order to carry out a MD simulation, a forcefield must first be selected which governs the way in which all atoms, bonds and angles of the system are treated. The AMBER forcefield (AMBER99sb) was selected for this project as one of the most widely set of parameters used in biomolecular simulations (Hornak et al. 2006). As a non-standard amino acid, the chromophores within the fluorescent proteins sfGFP and Venus required new additional parameters to be created. This involved adding each of the atom types, bonds, and bond lengths, angles and dihedrals, (as discussed in chapter 3). For the most part this was considered successful, particularly for determining the relative interchromophore distances (r) and the impact of protein dynamics on the accuracy of FRET. However, several issues became apparent after the MD simulations were performed whereby the chromophores loses the correct conformation and planarity was not conserved. This was attributed to incorrect parameterisation and is an area of improvement for future work. This was not considered a significant problem for chapter 3, as the overall difference in chromophore distances (as a result of a loss in chromophore planarity) would be minimal (~ 1 Å). However, this was considered within chapter 4, where the exact position of the chromophore was critical for confirming the hydrogen bond capability of the chromophores hydroxyl group to residue 148. Thus, GROMACS was only used for energy minimisation, to relieve any residue clashes (particularly as hydrogens are now added to the structure at this point) and change the potential energy of the system; where a constant kinetic energy will be maintained and as such a reduction in the total energy. The results were therefore a prediction based on available conformations in PyMOL, and subsequent relaxation by EM. Across all models, the chromophore was of the correct conformation and did not lose its planarity following EM. The hypothesis determined by EM models did fit experimental data whereby the hydrogen bond predicted between the chromophore and residues C148 and S148 of sfGFP gave rise to the 490 nm absorbance peak, indicative of the anionic charged state. Thus, although there were limitations to the MD simulations in this work, overall the

models were successful in drawing predictions and rationalising outcomes to explain experimental data. It is also worth mentioning that forcefields are continuously being developed and improved, including parameters for the molecular weight of atoms, bond distances, angles and dihedrals. This requires dedicated time, focus, and often involves collaboration of many individuals. The AMBER forcefield has developed since 1995 (Bayly et al. 1995), has over 60 authors and is currently on the 22nd iteration (Case Ross C Walker and Roitberg Kenneth Merz Pengfei Li 2010).

7.3 Significance of small structural perturbations on local conditions

Residue H148 plays a critical role defining the fluorescent properties of GFP, through formation of a H-bond with the tyrosyl group of the chromophore. Throughout chapters 4, mutation of this residue demonstrated how changes to the H-bond network surrounding the chromophore can be used to influence protein function. The replacement of H148 with non-polar residues (V148 and A148) resulted in the switch from a major absorbance peak at ~ 485 nm over to 400 nm, indicative of a neutral chromophore state. Thus, the negative charge of the phenolic group is no longer stabilised by H148. Interestingly, replacement of H148 with cysteine, a polar residue, resulted in similar spectral properties as H148V and H148A. Structural analysis revealed that the side chain of C148 faces away from the chromophore and a subsequent loss in H-bond. However, dimerisation of sfGFP^{C148.F} via disulphide bridging at residue C148 resulted in the return of the major absorbance peak at ~ 490 nm (chapter 5). This is due to the presence of long-range H-bond networks across the interface. Comparison of monomer and dimeric structures highlighted the key structural differences, where residues provided by the dimer interface are deemed critical for stabilising the negative charge of the chromophore.

7.4 Future work

Throughout chapter 4, a novel sfGFP variant with a H148S mutation was identified that was 50% brighter than wild type. The hypothesis is that serine can

form a hydrogen bond with the chromophore's hydroxyl group, and samples less conformational flux than H148. To confirm this, the crystal structure of sfGFP^{S148} is required. The structural impacts upon residue replacement can then be deciphered by comparison to sfGFP^{WT}. From analysing other FP structures in the PDB it appears that within other species, it appears serine is surprisingly well conserved at this residue position. However, for avGFP-like proteins (where H148 is conserved), H148S is statistically unlikely to have occurred during random mutagenesis approaches as at least two contiguous nucleotides need to be mutated. As we have evidenced the improved brightness within sfGFP, the question now arises as to if H148S would enhance brightness of other avGFP-like proteins (e.g. EGFP). One of the most commonly used marker proteins is EGFP, as its expression can be easily monitored by flow cytometry or confocal microscopy (Cooray et al. 2012). If replacement of a single amino acid could improve its brightness further, this would significantly improve its use in live cell imaging allowing more accurate and reliable monitoring. Thus, work is being continued by Prof Dafydd Jones' lab to determine whether this mutation is advantageous to other avGFP-like proteins.

One of the main discoveries of chapter 5 was that incorporation of V206F mutation promoted the formation of sfGFP^{C148.F} dimers; both by increasing the rate of reaction 40-fold and allowing full dimerisation to be reached. However, the transition metals copper and manganese (although less effective) were still required to catalyse the oxidation of C148 thiols. For an effective homodimer sensor, ideally no copper should be required and a purely non-covalent interface should be sought. The sfGFP^{C148.F} structure can therefore now be used as a template for future engineering. One option is the incorporation of additional mutations at the protein interface. Alternatively, a separate approach involves the addition of a leucine zipper to the terminus of sfGFP^{C148.F}. The leucine zipper is a short peptide sequence that homo-dimerises with high affinity so essentially forming a permanent dimer. This would ensure sfGFP molecules are within close proximity so should enhance the dimerisation capability. On the contrary, aside from a mechanism of detecting homodimers (as evidenced in chapter 6), the dependence of metals for dimerisation could be exploited for use as a metal sensor, or more specifically a copper sensor.

References

- Alcock, L.J., Perkins, M. V and Chalker, J.M. 2018. Chemical methods for mapping cysteine oxidation. *Chemical Society Reviews* 47(1), pp. 231–268. doi: 10.1039/c7cs00607a.
- Alford, S.C., Abdelfattah, A.S., Ding, Y. and Campbell, R.E. 2012a. A Fluorogenic Red Fluorescent Protein Heterodimer. *Chemistry & Biology* 19, pp. 353–360. doi: 10.1016/j.chembiol.2012.01.006.
- Alford, S.C., Ding, Y., Simmen, T. and Campbell, R.E. 2012b. Dimerization-Dependent Green and Yellow Fluorescent Proteins Scheme. *ACS. Synth. Biol.* 1, pp. 569–575. doi: 10.1021/sb300050j.
- Ando, R., Mizuno, H. and Miyawaki, A. 2004. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* 306(5700), pp. 1370–1373. doi: 10.1126/SCIENCE.1102506.
- Arpino, J.A.J., Rizkallah, P.J. and Jones, D.D. 2012. Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 Å Resolution Reveals Alternative Conformations for Glu222. *PLoS ONE* 7(10), p. 47132. doi: 10.1371/journal.pone.0047132.
- Auhim, H.S. et al. 2021. Stalling chromophore synthesis of the fluorescent protein Venus reveals the molecular basis of the final oxidation step. *Chemical Science* 12(22), pp. 7735–7745. doi: 10.1039/d0sc06693a.
- Bagiyan, G.A., Koroleva, I.K., Soroka, N. V and Ufimtsev, A. V 2003. Oxidation of thiol compounds by molecular oxygen in aqueous solutions. *Russian Chemical Bulletin* 52(5), pp. 1135–1141. doi: 10.1023/A:1024761324710.
- Bajar, B.T., Wang, E.S., Zhang, S., Lin, M.Z. and Chu, J. 2016. A guide to fluorescent protein FRET pairs. *Sensors* 16(9). doi: 10.3390/s16091488.
- Barondeau, D.P., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. 2002. Structural Chemistry of a Green Fluorescent Protein Zn Biosensor. *J. AM. CHEM. SOC* 124, p. 45. doi: 10.1021/ja0176954.
- Barondeau, D.P., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. 2005. Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* 44(6), pp. 1960–1970. doi: 10.1021/bi0479205.
- Barondeau, D.P., Putnam, C.D., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. 2003. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences of the United States of America* 100(21), pp. 12111–12116. doi: 10.1073/pnas.2133463100.
- Bayly, C.I. et al. 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* 117(19), pp. 5179–5197. doi: 10.1021/ja00124a002.
- Berggård, T., Linse, S. and James, P. 2007. Methods for the detection and analysis of protein-protein

- interactions. *Proteomics* 7(16), pp. 2833–2842. doi: 10.1002/PMIC.200700131.
- Bonizzi, G. and Karin, M. 2004. The two NF- κ B activation pathways and their role in innate and adaptive immunity. *Trends in Immunology* 25(6), pp. 280–288. doi: 10.1016/j.it.2004.03.008.
- Branchini, B.R., Nemser, A.R. and Zimmer, M. 1998. A computational analysis of the unique protein-induced tight turn that results in posttranslational chromophore formation in green fluorescent protein. *Journal of the American Chemical Society* 120(1), pp. 1–6. doi: 10.1021/ja973019j.
- Brasier, A.R., Lu, M., Hai, T., Lu, Y. and Boldogh, I. 2001. NF- κ B-inducible BCL-3 Expression is an Autoregulatory Loop Controlling Nuclear p50/NF- κ B1 Residence. *Journal of Biological Chemistry* 276(34), pp. 32080–32093. doi: 10.1074/jbc.M102949200.
- Bundy, B.C. and Swartz, J.R. 2011. Efficient disulfide bond formation in virus-like particles. *Journal of Biotechnology* 154, pp. 230–239. doi: 10.1016/j.jbiotec.2011.04.011.
- Bundy, D.L. and McKeithan, T.W. 1997. Diverse effects of BCL3 phosphorylation on its modulation of NF- κ B p52 homodimer binding to DNA. *Journal of Biological Chemistry* 272(52), pp. 33132–33139. doi: 10.1074/jbc.272.52.33132.
- Campanini, B. et al. 2013. Role of histidine 148 in stability and dynamics of a highly fluorescent GFP variant. doi: 10.1016/j.bbapap.2013.01.014.
- Cannon, M.B. 2006. Re-engineering redox-sensitive green fluorescent protein for improved response rate. *Protein Science* 15(1), pp. 45–57. doi: 10.1110/ps.051734306.
- Carmody, R.J., Ruan, Q., Palmer, S., Hilliard, B. and Chen, Y.H. 2007. Negative regulation of toll-like receptor signaling by NF- κ B p50 ubiquitination blockade. *Science* 317(5838), pp. 675–678. doi: 10.1126/science.1142953.
- Case Ross C Walker, D.A. and Roitberg Kenneth Merz Pengfei Li, A.M. 2010. Amber 2020 Reference Manual Principal contributors to the current codes.
- Chan, F.T.S., Kaminski, C.F. and Kaminski Schierle, G.S. 2010. HomoFRET Fluorescence Anisotropy Imaging as a Tool to Study Molecular Self-Assembly in Live Cells. *ChemPhysChem* 12, pp. 500–509. doi: 10.1002/cphc.201000833.
- Chan, K.Y., Hei, Y., Kwong, H. and Szeto, D.M. 2022. Green Fluorescent Protein: Its Development, Protein Engineering, and Applications in Protein Research. *Journal of Young Investigators* 25(3), pp. 19–32. doi: 10.22186/25.3.1.1.
- Chen, F.E., Huang, D. Bin, Chen, Y.Q. and Ghosh, G. 1998. Crystal structure of p50/p65 heterodimer of transcription factor NF- κ B bound to DNA. *Nature* 391, pp. 410–412. doi: 10.1038/34956.
- Chen, T., Li, M. and Liu, J. 2018. π - π Stacking Interaction: A Nondestructive and Facile Means in Material Engineering for Bioapplications. *Crystal Growth and Design* 18(5), pp. 2765–2783. doi: 10.1021/acs.cgd.7b01503.
- Chen, Y., Wu, J. and Ghosh, G. 2003. κ B-Ras binds to the unique insert within the ankyrin repeat

domain of I κ B β and regulates cytoplasmic retention of I κ B β -NF- κ B complexes. *Journal of Biological Chemistry* 278(25), pp. 23101–23106. doi: 10.1074/jbc.M301021200.

Choi, H.J. et al. 2010. Bcl3-dependent stabilization of CtBP1 is crucial for the inhibition of apoptosis and tumor progression in breast cancer. *Biochemical and Biophysical Research Communications* 400(3), pp. 396–402. doi: 10.1016/j.bbrc.2010.08.084.

Cogswell, P.C., Guttridge, D.C., Funkhouser, W.K. and Baldwin, A.S. 2000. Selective activation of NF- κ B subunits in human breast cancer: Potential roles for NF- κ B2/p52 and for Bcl-3. *Oncogene* 19(9), pp. 1123–1131. doi: 10.1038/sj.onc.1203412.

Concetti, J. and Wilson, C.L. 2018. NFKB1 and Cancer: Friend or Foe? *Cells* 7(133), pp. 1–16. doi: 10.3390/cells7090133.

Cooray, S., Howe, S.J. and Thrasher, A.J. 2012. Retrovirus and lentivirus vector design and methods of cell conditioning. In: *Methods in Enzymology.*, pp. 29–57. doi: 10.1016/B978-0-12-386509-0.00003-X.

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

Cramer, P., Larson, C.J., Verdine, G.L. and Müller, C.W. 1997. Structure of the human NF- κ B p52 homodimer-DNA complex at 2.1 Å resolution. *The EMBO Journal* 16(23), pp. 7078–7090. doi: 10.1093/EMBOJ/16.23.7078.

Cubitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A. and Tsien, R.Y. 1995. Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences* 20(11), pp. 448–455. doi: 10.1016/S0968-0004(00)89099-4.

Cullis, C.F. and Trimm, D.L. 1968. Homogeneous catalysis of the oxidation of thiols by metal ions. *Discussions of the Faraday Society* 46, pp. 144–149. doi: 10.1039/DF9684600144.

Cumming, R.C., Andon, N.L., Haynes, P.A., Park, M., Fischer, W.H. and Schubert, D. 2004. Protein disulfide bond formation in the cytoplasm during oxidative stress. *Journal of Biological Chemistry* 279(21), pp. 21749–21758. doi: 10.1074/jbc.M312267200.

Decembrino, D., Girhard, M. and Urlacher, V.B. 2021. Use of Copper as a Trigger for the in Vivo Activity of E. coli Laccase CueO: A Simple Tool for Biosynthetic Purposes. *ChemBioChem* 22(8), pp. 1470–1479. doi: 10.1002/cbic.202000775.

Duff, J.R., Grubbs, M.R. and Howell, J.E. 2011. Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. *J. Vis. Exp* (55), p. 2796. doi: 10.3791/2796.

ExpASy ProtParam tool. 2008. <https://web.expasy.org/protparam/> [Accessed 12 December 2017]

Gao, W., Cho, E., Liu, Y. and Lu, Y. 2019. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Frontiers in Pharmacology* 10(611). doi: 10.3389/fphar.2019.00611.

- Gao, X., Dong, X., Li, X., Liu, Z. and Liu, H. 2020. Prediction of disulfide bond engineering sites using a machine learning method. *Scientific Reports* 10(1), p. 10330. doi: 10.1038/s41598-020-67230-z.
- Garratt, R.C., Fonseca Valadares, N., Fernando, J. and Bachega, R. 2013. Order of Reaction. *Encyclopedia of Biophysics. Springer, Berlin* G. C. K. Roberts, pp. 1807–1808. doi: 10.1007/978-3-642-16712-6.
- Ghosh, G., Van Duyne, G., Ghosh, S. and Sigler, P.B. 1995. Structure of nf- κ b p50 homodimer bound to a κ b site. *Nature* 373(6512), pp. 303–310. doi: 10.1038/373303a0.
- Ghosh, I., Hamilton, A.D. and Regan, L. 1996. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *A Proc. Natl. Acad. Sci. U.S.A* 21(2), pp. 5658–5659. doi: 10.1021/ja994421w.
- Giachino, A. and Waldron, K.J. 2020. Copper tolerance in bacteria requires the activation of multiple accessory pathways. *Molecular Microbiology* 114, pp. 377–390. doi: 10.1111/mmi.14522.
- Giles, N.M., Watts, A.B., Giles, G.I., Fry, F.H., Littlechild, J.A. and Jacob, C. 2003. Review Metal and Redox Modulation of Cysteine Protein Function. *Chemistry & Biology* 10, pp. 677–693. doi: 10.1016/S.
- Goedhart, J. et al. 2012. Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nature Communications* 3(751). doi: 10.1038/ncomms1738.
- Gómez-Tamayo, J.C., Cordoní, A., Olivella, M., Mayol, E., Fourmy, D. and Pardo, L. 2016. Analysis of the interactions of sulfur-containing amino acids in membrane proteins. *Protein Science* 25, pp. 1517–1524. doi: 10.1002/pro.2955.
- Goodsell, D.S. and Olson, A.J. 2000. Structural symmetry and protein function. *Annual Review of Biophysics and Biomolecular Structure* 29, pp. 105–153. doi: 10.1146/annurev.biophys.29.1.105.
- Grey, B. and Steck, T.R. 2001. Concentrations of Copper Thought to Be Toxic to Escherichia coli Can Induce the Viable but Nonculturable Condition. *Applied and Environmental Microbiology* 67(11), pp. 5325–5327. doi: 10.1128/aem.67.11.5325-5327.2001.
- Gupta, C.L., Akhtar, S. and Bajpai, P. 2014. In Silico protein modeling: Possibilities and limitations. *EXCLI Journal* 13, pp. 513–515.
- Gwyther, R.E.A., Dafydd Jones, D. and Worthy, H.L. 2019. Better together: Building protein oligomers naturally and by design. *Biochemical Society Transactions* 47(6), pp. 1773–1780. doi: 10.1042/BST20190283.
- Hagemans, D., van Belzen, I.A.E.M., Luengo, T.M. and Rüdiger, S.G.D. 2015. A script to highlight hydrophobicity and charge on protein surfaces. *Frontiers in Molecular Biosciences* 2(56), pp. 1–11. doi: 10.3389/fmolb.2015.00056.
- Hanson, G.T. et al. 2002. Green fluorescent protein variants as ratiometric dual emission ph sensors. 1. Structural characterization and preliminary application. *Biochemistry* 41(52), pp. 15477–15488.

doi: 10.1021/bi026609p.

Hanson, G.T., Aggeler, R., Oglesbee, D., Cannon, M., Capaldi, R.A., Tsien, R.Y. and Remington, S.J. 2004. Investigating Mitochondrial Redox Potential with Redox-sensitive Green Fluorescent Protein Indicators. *Journal of Biological Chemistry* 279(13), pp. 13044–13053. doi: 10.1074/jbc.M312846200.

Hartley, A., Worthy, H., Reddington, S., Rizkallah, P. and Jones, D. 2016. Molecular basis for functional switching of GFP by two disparate non-native post-translational modifications of a phenyl azide reaction handle. *Chemical Science* 7(10), pp. 6484–6491. doi: 10.1039/c6sc00944a.

Hastings, J.W. and Morin, J.G. 1969. Comparative biochemistry of calcium-activated photoproteins from the ctenophore, Mnemiopsis and the coelenterates Aequorea, Obelia, Pelagia and Renilla. *Biological Bulletin* 173(3), p. 402.

Hay, R.T. and Nicholson, J. 1993. DNA binding alters the protease susceptibility of the p50 subunit of NF- κ B. *Nucleic Acids Research* 21(19), pp. 4592–4598. doi: 10.1093/nar/21.19.4592.

Hayden, M.S. and Ghosh, S. 2004. Signaling to NF- κ B. *Genes and Development* 18(18), pp. 2195–2224. doi: 10.1101/gad.1228704.

Hellenkamp, B. et al. 2018. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nature Methods* 15, pp. 669–676. doi: 10.1038/s41592-018-0085-0.

Hicks, D.R. et al. 2022. De novo design of protein homodimers containing tunable symmetric protein pockets. *Proceedings of the National Academy of Sciences* 119(30). doi: 10.1073/pnas.

Hoesel, B. and Schmid, J.A. 2013. The complexity of NF- κ B signaling in inflammation and cancer. *Molecular Cancer* 12(86). doi: 10.1186/1476-4598-12-86.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. 2006. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics* 65(3), pp. 712–725. doi: 10.1002/prot.21123.

Hu, C.D., Chinenov, Y. and Kerppola, T.K. 2002. Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular Cell* 9(4), pp. 789–798. doi: 10.1016/S1097-2765(02)00496-3.

Hubbard, R.E. and Kamran Haider, M. 2010. *Hydrogen Bonds in Proteins: Role and Strength*. doi: 10.1002/9780470015902.a0003011.pub2.

Huynh, K. and Partch, C.L. 2016. Analysis of protein stability and ligand interactions by thermal shift assay. *Current protocols in protein science* 79, pp. 28.9.1-28.9.14. doi: 10.1002/0471140864.ps2809s79.

Ilagan, R.P., Rhoades, E., Gruber, D.F., Kao, H.T., Pieribone, V.A. and Regan, L. 2010. A new bright green-emitting fluorescent protein - Engineered monomeric and dimeric forms. *FEBS Journal* 277(8), pp. 1967–1978. doi: 10.1111/j.1742-4658.2010.07618.x.

- Johnson, R.L., Blaber, H.G., Evans, T., Worthy, H.L., Pope, J.R. and Jones, D.D. 2021. Designed Artificial Protein Heterodimers With Coupled Functions Constructed Using Bio-Orthogonal Chemistry. *Frontiers in Chemistry* 9, p. 629. doi: 10.3389/FCHEM.2021.733550/BIBTEX.
- Jumper, J. et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, p. 583. doi: 10.1038/s41586-021-03819-2.
- Kerrigan, J.J., Xie, Q., Ames, R.S. and Lu, Q. 2011. Production of protein complexes via co-expression. *Protein Expression and Purification* 75(1), pp. 1–14. doi: 10.1016/j.pep.2010.07.015.
- Kneen, M., Farinas, J., Li, Y. and Verkman, A.S. 1998. Green Fluorescent Protein as a Noninvasive Intracellular pH Indicator. *Biophysical Journal* 74, pp. 1591–1599. doi: 10.1016/S0006-3495(98)77870-1.
- Kobayashi, N. and Arai, R. 2017. Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Current Opinion in Biotechnology* 46, pp. 57–65. doi: 10.1016/j.copbio.2017.01.001.
- Kohl, B. et al. 2019. Comparison of backbone dynamics of the p50 dimerization domain of NFκB in the homodimeric transcription factor NFκB1 and in its heterodimeric complex with RelA (p65). *Protein Science* 28(12), pp. 2064–2072. doi: 10.1002/pro.3736.
- Kuzmanov, U. and Emili, A. 2013. Protein-protein interaction networks: Probing disease mechanisms using model systems. *Genome Medicine* 5(37). doi: 10.1186/gm441.
- Lehrman, S.R. 2017. Protein structure. In: *Fundamentals of Protein Biotechnology.*, pp. 9–38. doi: 10.4324/9780203967621-8.
- Lindman, S., Hernandez-Garcia, A., Szczepankiewicz, O., Frohm, B. and Linse, S. 2010. In vivo protein stabilization based on fragment complementation and a split GFP system. *Proceedings of the National Academy of Sciences of the United States of America* 107(46), pp. 19826–19831. doi: 10.1073/pnas.1005689107.
- Liu, Z.Y. et al. 2019. Sulfur-Based Intramolecular Hydrogen-Bond: Excited-State Hydrogen-Bond On/Off Switch with Dual Room-Temperature Phosphorescence. *Journal of the American Chemical Society* 141(25), pp. 9885–9894. doi: 10.1021/jacs.9b02765.
- Ma, F.F. et al. 2020. Dysregulated NF-κB signal promotes the hub gene PCLAF expression to facilitate nasopharyngeal carcinoma proliferation and metastasis. *Biomedicine and Pharmacotherapy* 125. doi: 10.1016/j.biopha.2020.109905.
- Malek, S., Huang, D. Bin, Huxford, T., Ghosh, S. and Ghosh, G. 2003. X-ray crystal structure of an IκBβ·NF-κB p65 homodimer complex. *Journal of Biological Chemistry* 278(25), pp. 23094–23100. doi: 10.1074/jbc.M301022200.
- Manavalan, B., Basith, S., Choi, Y.M., Lee, G. and Choi, S. 2010. Structure-function relationship of cytoplasmic and nuclear IκB proteins: An in silico analysis. *PLoS ONE* 5(12). doi: 10.1371/journal.pone.0015782.

- Marianayagam, N.J., Sunde, M. and Matthews, J.M. 2004. The power of two: Protein dimerization in biology. *Trends in Biochemical Sciences* 29(11), pp. 618–625. doi: 10.1016/j.tibs.2004.09.006.
- Martin, E.W. et al. 2019. Assaying Homodimers of NF- κ B in Live Single Cells. *Frontiers in Immunology* 10(2609). doi: 10.3389/fimmu.2019.02609.
- Massoumi, R., Chmielarska, K., Hennecke, K., Pfeifer, A. and Fässler, R. 2006. Cxcl12 Inhibits Tumor Cell Proliferation by Blocking Bcl-3-Dependent NF- κ B Signaling. *Cell* 125, pp. 665–677. doi: 10.1016/j.cell.2006.03.041.
- Matthews, J.R., Nicholson, J., Jaffray, E., Kelly, S.M., Price, N.C. and Hay, R.T. 1995. Conformational changes induced by DNA binding of NF- κ B. *Nucleic Acids Research* 23(17), pp. 3393–3402. doi: 10.1093/nar/23.17.3393.
- Mei, G., Di Venere, A., Rosato, N. and Finazzi-Agrò, A. 2005. The importance of being dimeric. *The FEBS Journal* 272(1), pp. 16–27. doi: 10.1111/J.1432-1033.2004.04407.X.
- Mena, M.A., Treynor, T.P., Mayo, S.L. and Daugherty, P.S. 2006. Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nature Biotechnology* 24(12), pp. 1569–1571. doi: 10.1038/nbt1264.
- Michel, F., Soler-Lopez, M., Petosa, C., Cramer, P., Siebenlist, U. and Müller, C.W. 2001. Crystal structure of the ankyrin repeat domain of Bcl-3: A unique member of the I κ B protein family. *EMBO Journal* 20(22), pp. 6180–6190. doi: 10.1093/emboj/20.22.6180.
- Miyawaki, A. and Tsien, R.Y. 2000. Monitoring protein conformations and interactions by fluorescence resonance energy transfer between mutants of green fluorescent protein. *Methods in Enzymology* 327, pp. 472–500. doi: 10.1016/S0076-6879(00)27297-2.
- Mizuno, H. et al. 2008. Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* 105(27), pp. 9227–9232. doi: 10.1073/pnas.0709599105.
- Moorthy, A.K., Huang, D.-B., Ya-Fan Wang, V., Vu, D. and Ghosh, G. 2007. X-ray Structure of a NF- κ B p50/RelB/DNA Complex Reveals Assembly of Multiple Dimers on Tandem κ B Sites. doi: 10.1016/j.jmb.2007.08.039.
- Morin, J.G. and Hastings, J.W. 1971. Energy transfer in a bioluminescent system. *Journal of Cellular Physiology* 77(3), pp. 313–318. doi: 10.1002/JCP.1040770305.
- Morise, H., Shimomura, O., Johnson, F.H. and Winant, J. 1974. Intermolecular energy transfer in the bioluminescent system of aequorea. *Biochemistry* 13(12), pp. 2656–2662. doi: 10.1021/bi00709a028.
- Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K. and Miyawaki, A. 2002. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology* 20(1), pp. 87–90. doi: 10.1038/nbt0102-87.
- Nicholson, E.M. and Scholtz, J.M. 1996. Conformational stability of the Escherichia coli HPr protein: Test of the linear extrapolation method and a thermodynamic characterization of cold denaturation.

- Biochemistry* 35, pp. 11369–11378. doi: 10.1021/bi960863y.
- Nienhaus, K., Nienhaus, G.U., Wiedenmann, J. and Nar, H. 2005. Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. *Proceedings of the National Academy of Sciences of the United States of America* 102(26), pp. 9156–9159. doi: 10.1073/pnas.0501874102.
- Nies, D.H. 1999. Microbial heavy-metal resistance. *Applied Microbial Biotechnology* 51, pp. 730–750.
- Nijnik, A., Mott, R., Kwiatkowski, D.P. and Udalova, I.A. 2003. Comparing the fine specificity of DNA binding by NF- κ B p50 and p52 using principal coordinates analysis. *Nucleic Acids Research* 31(5), pp. 1497–1501. doi: 10.1093/nar/gkg231.
- Oeckinghaus, A. and Ghosh, S. 2009. The NF- κ B Family of Transcription Factors and Its Regulation. *Cold Spring Harbor Perspect Biol* 1. doi: 10.1101/cshperspect.a000034.
- Ormo 1996. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* 273(5280), pp. 1392–1395.
- Pahl, H.L. 1999. Activators and target genes of Rel/NF- κ B transcription factors. *Oncogene* 18(49), pp. 6853–6866. doi: 10.1038/sj.onc.1203239.
- Paketurytė, V., Zubrienė, A., Ladbury, J.E. and Matulis, D. 2019. Intrinsic Thermodynamics of Protein-Ligand Binding by Isothermal Titration Calorimetry as Aid to Drug Design. In: *Methods in Molecular Biology*. Humana Press, pp. 61–74. doi: 10.1007/978-1-4939-9179-2_5.
- Pang, H. et al. 2004. Crystal structure of human pirin: An iron-binding nuclear protein and transcription cofactor. *Journal of Biological Chemistry* 279(2), pp. 1491–1498. doi: 10.1074/jbc.M310022200.
- Park, O.K. and Bauerle, R. 1999. Metal-catalyzed oxidation of phenylalanine-sensitive 3-deoxy-D-arabino- heptulosonate-7-phosphate synthase from *Escherichia coli*: Inactivation and destabilization by oxidation of active-site cysteines. *Journal of Bacteriology* 181(5), pp. 1636–1642. doi: 10.1128/jb.181.5.1636-1642.1999.
- Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 24(1), pp. 79–88. doi: 10.1038/nbt1172.
- Peng, X., Wang, J., Peng, W., Wu, F.X. and Pan, Y. 2017. Protein-protein interactions: detection, reliability assessment and applications. *Briefings in bioinformatics* 18(5), pp. 798–819. doi: 10.1093/bib/bbw066.
- Perkins, N.D. 2007. Integrating cell-signalling pathways with NF- κ B and IKK function. *Molecular Cell Biology* 8, pp. 49–62. doi: 10.1038/nrm2083.
- Piñero, J., Furlong, L.I. and Sanz, F. 2018. In silico models in drug development: where we are. *Current Opinion in Pharmacology* 42, pp. 111–121. doi: 10.1016/j.coph.2018.08.007.

- Pope, J.R. et al. 2021. Association of Fluorescent Protein Pairs and Its Significant Impact on Fluorescence and Energy Transfer. *Advanced Science* 8(1). doi: 10.1002/adv.202003167.
- Rao, V.S., Srinivas, K., Sujini, G.N. and Kumar, G.N.S. 2014. Protein-Protein Interaction Detection: Methods and Analysis. *International Journal of Proteomics* 2014, pp. 1–12. doi: 10.1155/2014/147648.
- Reddington, S.C., Baldwin, A.J., Thompson, R., Brancale, A., Tippmann, E.M. and Jones, D.D. 2015. Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chemical Science* 6(2), pp. 1159–1166. doi: 10.1039/c4sc02827a.
- Reddington, S.C., Rizkallah, P.J., Watson, P.D., Pearson, R., Tippmann, E.M. and Jones, D.D. 2013. Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence. *Angewandte Chemie - International Edition* 52(23), pp. 5974–5977. doi: 10.1002/anie.201301490.
- Reddington, S.C., Tippmann, E.M. and Dafydd Jones, D. 2012. Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry. *Chem. Commun* 48, pp. 8419–8421. doi: 10.1039/c2cc31887c.
- Remington, S.J. 2006. Fluorescent proteins: maturation, photochemistry and photophysics. *Current Opinion in Structural Biology* 16(6), pp. 714–721. doi: 10.1016/j.sbi.2006.10.001.
- Remington, S.J. 2011. Green fluorescent protein: A perspective. *Protein Science* 20(9), pp. 1509–1519. doi: 10.1002/pro.684.
- Remington, S.J., Wachter, R.M., Yarbrough, D.K., Branchaud, B., Anderson, D.C., Kallio, K. and Lukyanov, K.A. 2005. zFP538, a yellow-fluorescent protein from *Zoanthus*, contains a novel three-ring chromophore. *Biochemistry* 44(1), pp. 202–212. doi: 10.1021/bi048383r.
- Rensing, C. and Grass, G. 2003. *Escherichia coli* mechanisms of copper homeostasis in a changing environment. *FEMS Microbiology Reviews* 27, pp. 197–213. doi: 10.1016/S0168-6445(03)00049-4.
- Rupesh, K.R., Smith, A. and Boehmer, P.E. 2015. Ligand induced stabilization of the melting temperature of the HSV-1 single-strand DNA binding protein using the thermal shift assay. *Biochemical and Biophysical Research Communications* 454(4), pp. 604–608. doi: 10.1016/j.bbrc.2014.10.145.
- Schuster, M., Annemann, M., Plaza-Sirvent, C. and Schmitz, I. 2013. Atypical I κ B proteins-nuclear modulators of NF- κ B signaling. *Cell Communication and Signalling* 11(23), pp. 1–11.
- Scott, A.J. et al. 2021. Constructing ion channels from water-soluble α -helical barrels. *Nature Chemistry* 13, pp. 643–650. doi: 10.1038/s41557-021-00688-0.
- Seifert, M.H.J., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B. and Holak, T.A. 2003. Backbone dynamics of green fluorescent protein and the effect of Histidine 148 substitution. *Biochemistry* 42(9), pp. 2500–2512. doi: 10.1021/bi026481b.
- Sekar, R.B. and Periasamy, A. 2003. Fluorescence resonance energy transfer (FRET) microscopy

- imaging of live cell protein localizations. *Journal of Cell Biology* 160(5), pp. 629–633. doi: 10.1083/jcb.200210140.
- Sen, R. and Baltimore, D. 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. *Cell* 46(5), pp. 705–716. doi: 10.1016/0092-8674(86)90346-6.
- Sengchanthalangsy, L.L., Datta, S., Huang, D. Bin, Anderson, E., Braswell, E.H. and Ghosh, G. 1999. Characterization of the dimer interlace of transcription factor NF κ B p50 homodimer. *Journal of Molecular Biology* 289(4), pp. 1029–1040. doi: 10.1006/jmbi.1999.2823.
- Shekhawat, S.S. and Ghosh, I. 2011. Split-protein systems: Beyond binary protein-protein interactions. *Current Opinion in Chemical Biology* 15(6), pp. 789–797. doi: 10.1016/j.cbpa.2011.10.014.
- Shimomura, O., Johnson, F.H. and Saiga, Y. 1962. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology* 59(3), pp. 223–239. doi: 10.1002/JCP.1030590302.
- Shrestha, D., Jenei, A., Nagy, P., Vereb, G. and Szöllösi, J. 2015. Understanding FRET as a Research Tool for Cellular Studies. *Int. J. Mol. Sci* 16, pp. 6718–6756. doi: 10.3390/ijms16046718.
- Siebenlist, U., Franzoso, G. and Brown, K. 1994. Structure, regulation and function of NF- κ B. *Annual Review of Cell Biology* 10, pp. 405–455. doi: 10.1146/annurev.cb.10.110194.002201.
- Sinha, S., Mitra, N., Kumar, G., Bajaj, K. and Surolia, A. 2005. Unfolding studies on soybean agglutinin and Concanavalin A tetramers: A comparative account. *Biophysical Journal* 88(2), pp. 1300–1310. doi: 10.1529/biophysj.104.051052.
- Sinha, S. and Surolia, A. 2005. Oligomerization endows enormous stability to soybean agglutinin: A comparison of the stability of monomer and tetramer of soybean agglutinin. *Biophysical Journal* 88(6), pp. 4243–4251. doi: 10.1529/biophysj.105.061309.
- Soukupova, J. et al. 2021. The discovery of a novel antimetastatic Bel3 inhibitor. *Molecular Cancer Therapeutics* 20(5), pp. 775–786. doi: 10.1158/1535-7163.MCT-20-0283.
- Su, H. and Xu, Y. 2018. Application of ITC-based characterization of thermodynamic and kinetic association of ligands with proteins in drug design. *Frontiers in Pharmacology* 9. doi: 10.3389/fphar.2018.01133.
- Thomas, F. et al. 2018. De Novo-Designed α -Helical Barrels as Receptors for Small Molecules. *ACS Synthetic Biology* 7(7), pp. 1808–1816. doi: 10.1021/acssynbio.8b00225.
- Tsien, R.Y. 1998. THE GREEN FLUORESCENT PROTEIN. *Annu. Rev. Biochem* 67, pp. 509–553.
- Tubbs, J.L., Tainer, J.A. and Getzoff, E.D. 2005. Crystallographic structures of Discosoma red fluorescent protein with immature and mature chromophores: Linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. *Biochemistry* 44(29), pp. 9833–9840. doi: 10.1021/bi0472907.

- Turnham, D.J., Yang, W.W., Davies, J., Varnava, A., Ridley, A.J., Conlan, R.S. and Clarkson, R.W.E. 2020. Bcl-3 promotes multi-modal tumour cell migration via NF- κ B1 mediated regulation of Cdc42. *Carcinogenesis* 41(10), pp. 1432–1443. doi: 10.1093/carcin/bgaa005.
- Udalova, I.A., Mott, R., Field, D. and Kwiatkowski, D. 2002. Quantitative prediction of NF-B DNA-protein interactions. *PNAS* 99, pp. 8167–8172.
- Urban, B.C. et al. 2016. BCL-3 expression promotes colorectal tumorigenesis through activation of AKT signalling. *Gut* 65, pp. 1151–1164. doi: 10.1136/gutjnl-2014-308270.
- Valbuena, F.M., Fitzgerald, I., Strack, R.L., Andruska, N., Smith, L. and Glick, B.S. 2020. A photostable monomeric superfolder green fluorescent protein. *Traffic* 21(8), pp. 534–544. doi: 10.1111/tra.12737.
- Valley, C.C., Cembran, A., Perlmutter, J.D., Lewis, A.K., Labello, N.P., Gao, J. and Sachs, J.N. 2012. The methionine-aromatic motif plays a unique role in stabilizing protein structure. *Journal of Biological Chemistry* 287(42), pp. 34979–34991. doi: 10.1074/jbc.M112.374504.
- Wakefield, A., Soukupova, J., Montagne, A., Ranger, J., French, R., Muller, W.J. and Clarkson, R.W.E. 2013. Bcl3 Selectively Promotes Metastasis of ERBB2-Driven Mammary Tumors. *Cancer Research* 73(2), pp. 745–755. doi: 10.1158/0008-5472.CAN-12-1321.
- Wall, K.P., Dillon, R. and Knowles, M.K. 2014. Fluorescence Quantum Yield Measurements of Fluorescent Proteins: A Laboratory Experiment for a Biochemistry or Molecular Biophysics Laboratory Course. *Biochemistry and Molecular Biology Education* 43(1), pp. 52–59. doi: 10.1002/bmb.20837.
- Wang, J., Peng, X., Peng, W. and Wu, F.X. 2014. Dynamic protein interaction network construction and applications. *Proteomics* 14(4–5), pp. 338–352. doi: 10.1002/PMIC.201300257.
- Wang, V.Y.-F., Li, Y., Kim, D., Zhong, X., Du, Q., Ghassemian, M. and Ghosh Correspondence, G. 2017. Bcl3 Phosphorylation by Akt, Erk2, and IKK Is Required for Its Transcriptional Activity. *Molecular Cell* 67, pp. 484–497. doi: 10.1016/j.molcel.2017.06.011.
- Watkins, D.W. et al. 2016. A suite of de novo c-type cytochromes for functional oxidoreductase engineering. *Biochimica et Biophysica Acta - Bioenergetics* 1857(5), pp. 493–502. doi: 10.1016/j.bbabi.2015.11.003.
- Wehr, M.C. et al. 2006. Monitoring regulated protein-protein interactions using split TEV. *Nature Methods* 3(12), pp. 985–993. doi: 10.1038/nmeth967.
- Wood, T.I., Barondeau, D.P., Hitomi, C., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. 2005. Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. *Biochemistry* 44(49), pp. 16211–16220. doi: 10.1021/bi051388j.
- Worthy, H.L. et al. 2019. Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry. *Communications Chemistry* 2(1), pp. 1–12. doi: 10.1038/s42004-019-0185-5.

- Wulczyn, F.G., Naumann, M. and Scheidereit C 1992. Candidate proto-oncogene bcl-3 encodes a subunit-specific inhibitor of transcription factor NF-kB. *Nature* 358, pp. 597–599. doi: 10.1038/246170a0.
- Yang, F., Moss, L.G. and Phillips, G.N. 1996a. The Molecular Structure of Green Fluorescent Protein. *Nature Biotechnology* 14(10), pp. 1246–1251. doi: 10.1038/nbt1096-1246.
- Yang, T.T., Cheng, L. and Kain, S.R. 1996b. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Research* 24(22), pp. 4592–4593. doi: 10.1093/nar/24.22.4592.
- Yang, W.W. 2018. *Understanding the Role of Bcl-3 in Triple-Negative Breast Cancer*. PhD. Cardiff University.
- Yao, S. et al. 2022. De novo design and directed folding of disulfide-bridged peptide heterodimers. *Nature communications* 13(1539), pp. 1–10. doi: 10.1038/s41467-022-29210-x.
- Yu, Y., Wan, Y. and Huang, C. 2009. The Biological Functions of NF-kB1 and its Potential as an Anti-Cancer Target. *Current Cancer Drug Targets* 9(4), pp. 566–571. doi: 10.2174/156800909788486759.
- Zhang, jin, Campbell, R.E., Ting, A.Y. and Tsien, R.Y. 2002. Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology* 3(12), pp. 906–918. doi: 10.1038/nrm976.
- Zhang, M.-Y., Harhaj, E.W., Bell, L., Sun, S.-C. and Miller, B.A. 1998. Bcl-3 Expression and Nuclear Translocation Are Induced by Granulocyte-Macrophage Colony-Stimulating Factor and Erythropoietin in Proliferating Human Erythroid Precursors. *Blood* 92(4), p. 1225.
- Zhang, R., Nguyen, M.T. and Ceulemans, A. 2005. A concerted mechanism of proton transfer in green fluorescent protein. A theoretical study. *Chemical Physics Letters* 404(4–6), pp. 250–256. doi: 10.1016/j.cplett.2005.01.094.
- Zhao, H., Wang, W., Zhao, Q., Hu, G., Deng, K. and Liu, Y. 2016. BCL3 exerts an oncogenic function by regulating STAT3 in human cervical cancer. *OncoTargets and Therapy* 9, pp. 6619–6629. doi: 10.2147/OTT.S118184.
- Zhou, W., Šmidlehner, T. and Jerala, R. 2020. Synthetic biology principles for the design of protein with novel structures and functions. *FEBS Letters* 594(14), pp. 2199–2212. doi: 10.1002/1873-3468.13796.
- Zou, Y., Uddin, M.M., Padmanabhan, S., Zhu, Y., Bu, P., Vancura, A. and Vancurova, I. 2018. The proto-oncogene Bcl3 induces immune checkpoint PD-L1 expression, mediating proliferation of ovarian cancer cells. *Journal of Biological Chemistry* 293(40), pp. 15483–15496. doi: 10.1074/jbc.RA118.004084.