



Introducing novel protein functionality using unnatural amino acids

Samuel Charles Reddington

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

School of Biosciences
School of Chemistry
Cardiff University

June 2013

Submission of thesis declaration and statements

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed (candidate) Date

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed (candidate) Date

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed (candidate) Date

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed (candidate) Date

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed (candidate) Date

Acknowledgements

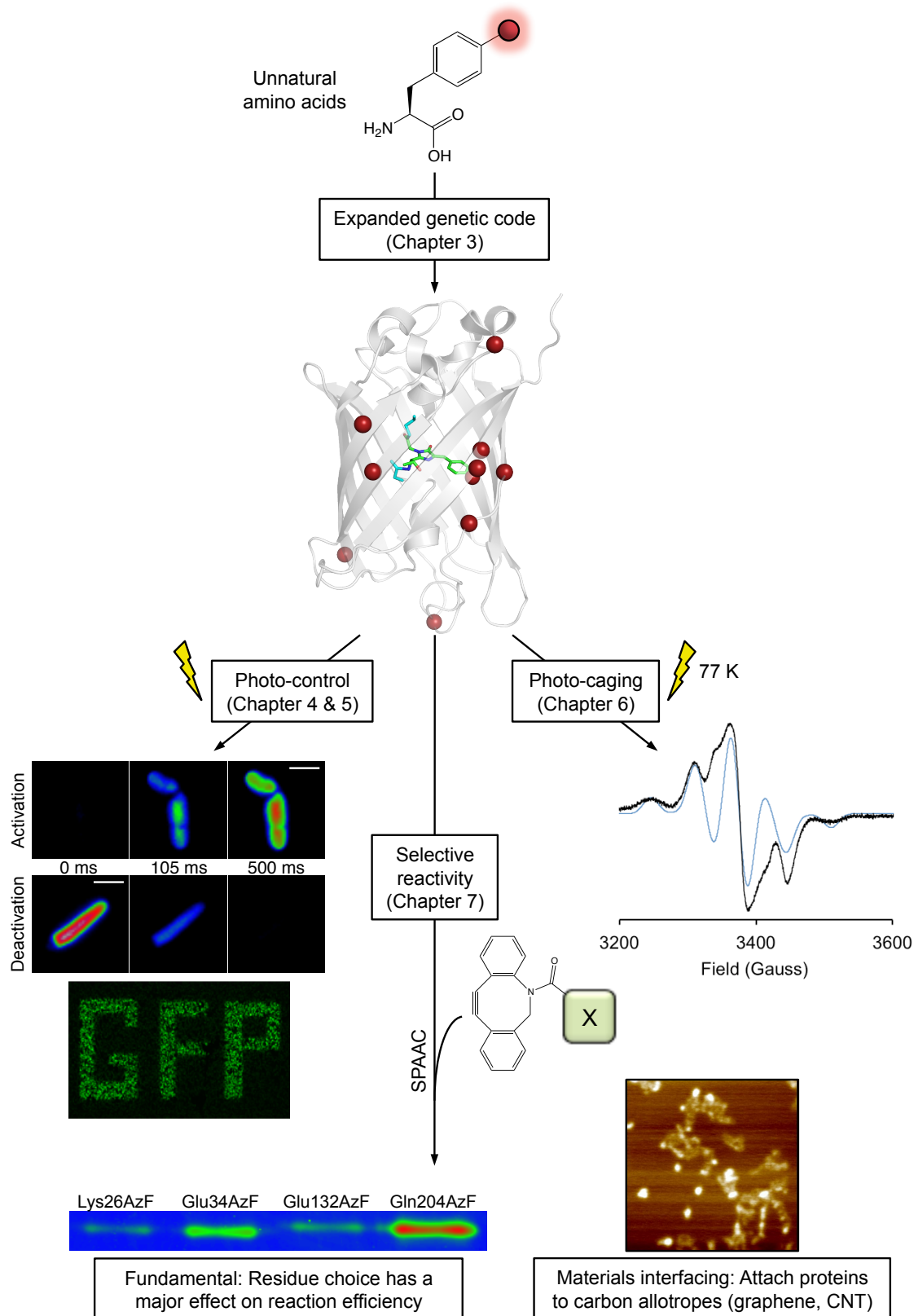
First of all I would like to thank my PhD advisors Dafydd Jones and Eric Tippmann for their continued supervision and guidance during my four years. Thank you for sticking with me and finding me funding – I hope you don't completely regret it. I would also like to recognise the time and effort Dafydd put into proof-reading and improving my thesis.

The two people who were in the lab for the majority of my PhD, James and Andy, deserve a special thanks for making my time as (bearable is the wrong word) fun and enjoyable as it was. I think I speak for James as well when I say, Andy, a little piece of advice for writing up; “Ay, ay, ay, calm down, calm down, calm down...” Huge thanks to Amy Baldwin, Mike Gamble and Matt Edmundson for help when I first started and for continued friendship. I would also like to mention lab members that I've not known for as long but have made my PhD enjoyable; Jonah, Adam, Lisa, Sonali, Athraa and Andriy as well as members of the Tippmann group Josie, Alicia, Zuzka and Sohie. Thank you to my final year undergraduate students Becky, Joanna, Dan, Georgia and Rachel who worked really hard and helped with a lot of the work in this thesis.

I was lucky enough to be surrounded by people who were more than happy to help and guide me during my PhD. A special mention goes to Dr Pete Watson for help with microscopy but mainly for his light-hearted and funny attitude to work and life. Dr Pierre Rizkallah for opening my eyes to the world of protein crystallography – I will never forgive you, but thanks very much for your help. I would also like to acknowledge Drs Damien Murphy, Ian Brewis and Steve Hill for invaluable guidance and assistance with EPR, mass spec and synthesis, respectively. Also, other members of the School who have helped or entertained me; Andy, Rob, Seni, Sarah and of course Joan.

This thesis or the work that went into it would not have been possible without the continued love and support of my family and friends. My parents Corinne and Craig, and little bro James (and more recently Judi) are inspirational and the reason I'm interested in science. There are too many to name personally but cheers to all of the Reddington's and Furneaux's. A special acknowledgement to my loving grandparents Nan French, Grandpa Reddington, (Sir) David and the memory of Marian Reddington and Jack French. And not forgetting my adopted family, the Evans', who have welcomed and supported me like one of their own.

Last but not least is Amy, my lovely fiancé (AKA my diary; the administrator), without whom I would be sat here hungry, dirty... in fact I wouldn't be sat here without Amy. Thank you Aim, you are my motivation and make me smile every day.



Using unnatural amino acids to create protein variants with novel properties. A schematic representation of unnatural amino acid incorporation into the Green Fluorescent Protein (GFP) using an expanded genetic code. GFP variants containing Uaa's were capable of 'Photocontrol', 'Photo-caging' and 'Selective reactivity'.

Abstract

Proteins are inherently limited by the properties of their constituent amino acids. By allowing a protein to sample a wider range of amino acids (in the form of unnatural amino acids; Uaa's) new chemistry can be introduced so leading to the generation of proteins with novel properties. This thesis examines the tolerance and effects of Uaa incorporation into proteins in *Escherichia coli* using an expanded genetic code. Uaa incorporation was used to alter or install new properties in the target proteins, superfolder Green Fluorescent Protein (sfGFP) and cytochrome *b₅₆₂*.

The technical aspects of Uaa incorporation are first discussed (Chapter 3) including the method itself, the orthogonality of the machinery and yield. Generally, Uaa's were well structurally and functionally tolerated and proteins containing Uaa's were purified in high yields. Substitution of residues in sfGFP for unnatural analogues of tyrosine was shown to be a valuable way of altering the properties of the protein. Variants were generated with red-shifted fluorescence and altered excitation spectra (Chapter 4).

The majority of this work focused on the Uaa, *p*-azido-L-phenylalanine (azPhe) as it has a number of properties that would be desirable for use in proteins such as photoreactivity and selective reactivity with alkynes. By incorporating azPhe into key residues of sfGFP, variants were created that could be controlled using light (photocontrol; Chapter 4). Light-dependant fluorescence activation, deactivation and switching were demonstrated *in vitro* and in live cells. The variants generated could be useful as tools in emerging microscopy techniques such as Photo-Activated Localisation Microscopy (PALM). Moreover, as azPhe can be incorporated into any target protein in a range of hosts, the use of Uaa's to control protein activity using light could become a general approach for the study and regulation of biological systems. The molecular bases for the fluorescent changes were investigated by a combination of spectroscopy and X-ray crystallography in Chapter 5.

The photoreactivity of azPhe was exploited for a different purpose in Chapter 6. Proteins were used as an alternative to synthetic cages for studying low temperature phenyl azide photochemistry. Here, two radicals (anilino and triplet phenyl nitrene) were successfully caged and detected on photolysis, with the radical observed dependant on the protein environment. This could be extended to study other photoreactions in a wide range of unique and defined environments.

Finally, in Chapter 7 the selective reactivity of azPhe was used to create proteins capable of site-specific modification (*via* Click chemistry). Protein modification, with non-

natural adducts offers a valuable way of manipulating their properties with a range of biotechnological applications possible. The position of azPhe on the protein surface was shown to have a significant effect on reaction yield and kinetics. There was a surprising inverse relationship between residue surface accessibility and reaction efficiency, which suggested an important role of the surface micro-environment. Modification was used to install proteins with novel properties such as red-shifted fluorescence emission of sfGFP (*via* an attached red fluorescent dye) and the ability to bind to non-biological materials like graphene (*via* an attached pyrene group).

Abbreviations

Amino acids have been described throughout the text using the standard three letter codes. In figures, tables and in the shorthand naming of variants the one letter amino acid code was used. Mutations used to generate variants are denoted as follows: wild type amino acid – residue number – mutant amino acid, in superscript after the protein. For example, sfGFP^{Y66AzF} and cyt *b*₅₆₂^{D5C}.

Abbreviations used throughout the thesis:

λ_{em}	Wavelength of fluorescence emission
λ_{ex}	Wavelength of fluorescence excitation
aaRS	Aminoacyl-tRNA-synthetase
AzF (AzPhe)	<i>p</i> -azido-L-phenylalanine
bp	Base pair
CuAAC	Copper-catalysed azide-alkyne cycloaddition
Cyt <i>b</i> ₅₆₂	Cytochrome <i>b</i> ₅₆₂
DTT	Dithiothreitol
dNTP	Deoxyribonucleotide phosphate
EGFP	Enhanced green fluorescent protein
EPR	Electron paramagnetic resonance
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
K	Kelvin
KNO ₃	Potassium nitrate
LA	Long-axis (cyt <i>b</i> ₅₆₂ variant)
MALDI-TOF	Matrix-assisted laser desorption ionisation-time of flight
NMR	Nuclear magnetic resonance
OD _{xxx}	Optical density (~absorbance) at xxx nm
PAGE	Polyacrylamide gel electrophoresis
PB	Phosphate buffer
PBS	Phosphate-buffered saline
PEG	Polyethylene glycol
QY	Quantum yield
RMSD	Root-mean-square deviation
SA	Short-axis (cyt <i>b</i> ₅₆₂ variant)
SDS	Sodium dodecyl sulphate
sfGFP	Superfolder green fluorescent protein
SPAAC	Strain-promoted azide-alkyne cycloaddition
TAE	Tris, acetate, EDTA
TBE	Tris, borate, EDTA
TEMED	Tetramethylethylenediamine
TLS	Translation – libration – screw
TNG	Tris, NaCl, glycerol
Tris	Tris-hydroxymethyl-aminomethane
tRNA	Transfer-RNA
Uaa	Unnatural amino acid
v/v	Volume per volume
wt (WT)	Wild-type
w/v	Weight per volume

Contents

1. Introduction	1
1.1 Expanding amino acid chemical diversity	1
1.2 Incorporating Uaa's into peptides and proteins	1
1.3 Uaa incorporation within the cell	2
1.4 Engineering an orthogonal tRNA/aaRS pair	7
1.5 Range and uses of unnatural amino acids	11
1.6 <i>p</i> -azidophenylalanine	13
1.6.1 Photoreactivity	13
1.6.2 Selective reactivity via Click chemistry	16
1.6.3 Biophysical probe	19
1.7 Structure, function and engineering of Green Fluorescent Protein	19
1.8 Cytochrome <i>b</i> ₅₆₂	24
1.9 Scope of the project and general aims	28
2. Materials and Methods	30
2.1 Materials	30
2.1.1 Chemicals	30
2.1.2 Bacterial cell strains	30
2.1.3 Bacterial growth media	31
2.2 Molecular Biology	31
2.2.1 DNA purification	31
2.2.2 Polymerase Chain Reaction (PCR)	32
2.2.3 Oligonucleotide primers and DNA sequence analysis	33
2.2.4 Agarose gel electrophoresis	34
2.2.5 DNA quantification	34
2.2.6 Restriction enzyme digestion	34
2.2.7 DNA phosphorylation	34
2.2.8 DNA Ligation	35
2.2.9 Bacterial transformation	35
2.2.10 DNA sequencing	35
2.3 Cloning cytochrome <i>b</i> ₅₆₂ into pBAD	35
2.4 Generation of amber stop codon mutants	36
2.5 Protein production and purification	38
2.5.1 Production of proteins containing Uaa's	38
2.5.2 Cell lysis	38
2.5.3 Protein purification	39
2.5.4 Purification of sfGFP variants	41
2.5.5 Purification of cyt <i>b</i> ₅₆₂ variants	41
2.5.6 Haem titration of cyt <i>b</i> ₅₆₂ variants	41
2.6 Protein analysis	41
2.6.1 SDS-PAGE	41
2.6.2 Protein quantification	42
2.6.3 Fluorescence spectroscopy	42
2.6.4 UV-visible absorption spectroscopy and calculation of extinction coefficients	43
2.6.5 Quantum yield determination	43
2.7 Photolysis of wild-type sfGFP and sfGFP ^{AzF} variants	44
2.8 Microscopy	44
2.9 Mass spectrometry	44
2.10 X-ray crystallography and structural determination	45
2.11 Electron paramagnetic resonance (EPR) spectroscopy	46
2.12 Method and analysis of Click reactions	46
2.12.1 Gel filtration chromatography: Analytical	46
2.12.2 Strain-promoted azide-alkyne cycloaddition (SPAAC) reaction	47

2.12.3	Analysis of SPAAC reactions by fluorescence and UV-visible absorption	48
2.12.4	Fluorescent imaging following SDS-PAGE	48
2.12.5	FRET calculations	48
2.12.6	Reaction kinetics	49
2.13	Non-covalent labelling of carbon allotropes with proteins	50
2.13.1	Synthesis of DBCO-pyrene	50
2.13.2	Atomic force microscopy (AFM)	51
3.	Production and purification of proteins containing unnatural amino acids	52
3.1	Introduction	52
3.2	Results and Discussion	52
3.2.1	Generation of amber stop codon mutations	52
3.2.2	Incorporation of Uaa's into proteins	55
3.2.3	Purification of Uaa-containing proteins	61
3.2.4	Reprogramming the amber stop codon for Uaa incorporation	66
3.2.5	Orthogonal incorporation of Uaa's	68
3.2.6	Efficiency of Uaa incorporation and protein production yields	70
4.	Influencing sfGFP function using unnatural amino acids	73
4.1	Introduction	73
4.1.1	Unnatural amino acids (Uaa's) and their influence on proteins	73
4.1.2	Protein photocontrol	74
4.2	Results	75
4.2.1	Altering the spectral properties of sfGFP by Uaa incorporation	75
4.2.2	Generating photocontrollable sfGFP variants using azPhe	80
4.2.3	Live cell imaging of <i>E. coli</i> using sfGFP ^{Y66AzF} and sfGFP ^{F145AzF}	89
4.3	Discussion	91
4.3.1	A library approach to engineering proteins containing Uaa's	91
4.3.2	Uaa mutations not tolerated by sfGFP	92
4.3.3	Molecular explanation of altered fluorescent properties in sfGFP ^{Uaa}	93
4.3.4	Molecular explanation of photoswitching in sfGFP ^{L44AzF}	96
4.3.5	Molecular explanation of photoswitching in sfGFP ^{T203AzF}	98
4.3.6	Summary	99
5.	Elucidating the photochemical pathways of photocontrollable sfGFP variants	101
5.1	Introduction	101
5.2	Results	101
5.2.1	Investigating the molecular mechanism of sfGFP ^{AzF} variants via incorporation of amino-phenylalanine	101
5.2.2	X-ray crystallography	104
5.2.3	Investigating the molecular mechanism of photodeactivation in sfGFP ^{F145AzF} by mass spectrometry	117
5.3	Discussion	118
5.3.1	Molecular explanation for fluorescence photoactivation in sfGFP ^{Y66AzF}	118
5.3.2	Molecular explanation for fluorescence photodeactivation in sfGFP ^{F145AzF}	120
5.3.3	Photoswitching in sfGFP ^{H148AzF}	121
5.3.4	Effects of azPhe incorporation on the structure of sfGFP	122
5.3.5	Summary	124
6.	Caging reactive chemical intermediates using proteins	125
6.1	Introduction	125
6.2	Results	127
6.2.1	Position of azPhe incorporation	127
6.2.2	Photolysis of sfGFP ^{AzF} variants and EPR measurements	128
6.2.3	Simulation of the observed EPR spectrum	130
6.3	Discussion	133
6.3.1	Characterisation of irradiated sfGFP ^{E34AzF}	133

6.3.2	Characterisation of irradiated sfGFP ^{Y66AzF}	134
6.3.3	AzPhe photolysis in a hydrophobic cavity of a protein.....	138
6.3.4	Proteins as cages to study reactive organic intermediates.....	141
7.	Bio-orthogonal modification of proteins via genetically encoded azide chemistry..	142
7.1	Modification of biomolecules	142
7.1.1	Interfacing proteins with non-biological materials	143
7.2	Results	145
7.2.1	Effect of azPhe position on efficiency of sfGFP modification via Click chemistry.....	145
7.2.2	Effect of azPhe position on the efficiency of cyt <i>b</i> ₅₆₂ modification via Click chemistry.....	153
7.2.3	Modifying sfGFP and cyt <i>b</i> ₅₆₂ with DBCO-pyrene.....	155
7.2.4	Attaching proteins to carbon allotropes.....	157
7.3	Discussion.....	160
7.3.1	Precision of SPAAC based protein modification	160
7.3.2	Effect of azPhe position on SPAAC efficiency.....	161
7.3.3	Effect of azPhe position on the influence of sfGFP SPAAC modification.....	165
7.3.4	SPAAC reaction kinetics	166
7.3.5	Alternative bio-orthogonal chemistries	166
7.3.6	Labelling graphene with pyrene-functionalised proteins	167
8.	Discussion.....	170
8.1	The advantages and disadvantages of genetically encoding Uaa's as a protein engineering technique.....	170
8.2	AzPhe as an addition to the protein engineering toolbox.....	172
8.3	Using SPAAC-based Click chemistry for protein modification	174
8.4	Interfacing proteins with non-biological materials.....	175
8.5	Future work.....	176
8.6	Summary and achievements	178
8.7	Publications associated with this thesis.....	180
9.	References.....	181

1. Introduction

1.1 Expanding amino acid chemical diversity

Proteins are essential for all organisms. They play important roles in virtually every cellular process ranging from metabolism to defence responses to cell signalling and cell shape and motility (1). Despite the huge diversity of structure and function, proteins in all known organisms are composed of the same set of 20 amino acids, with the rare exceptions of selenocysteine (2) and pyrrolysine (3). The natural set of amino acids has relatively restricted chemistry (hydroxyl, carboxyl, sulfhydryl, amine and amide groups) and proteins often require post-translational modifications and cofactors to become fully functional. Expanding the chemical diversity sampled by amino acids beyond that available in nature would allow the engineering of proteins, or even the construction of whole organisms to have new and useful physicochemical properties. Unnatural amino acids (Uaa's) can be rationally designed to have desired properties that will be conferred to a peptide or protein on incorporation. Potential uses of Uaa's include probes for spectroscopy (NMR, X-ray crystallography), imaging (MRI, fluorescence) and post-translational modification.

1.2 Incorporating Uaa's into peptides and proteins

Given the huge potential of introducing new chemistry into proteins, several approaches have been developed to allow incorporation of Uaa's into proteins. The simplest method is incorporation into peptides by solid-phase peptide synthesis (SPPS). However, this fully synthetic approach is limited by the production of simple peptides that can generally only be used *in vitro* rather than in the context of a biological system without the requirement for invasive and non-replicative introduction such as microinjection. Ideally, incorporation would occur during normal cellular protein synthesis. One such approach is through metabolic labelling via substitution of a natural amino acid with a structurally analogous Uaa. Here, a high concentration of Uaa is added to a bacterial strain that is auxotrophic for the natural amino acid. When the natural amino acid has been depleted the Uaa is incorporated in its place throughout the proteome (4). However, this technique does not allow for the target protein or residue position of Uaa incorporation to be controlled resulting in global or partial replacement of the natural amino acid. To more closely mimic natural protein synthesis, Uaa's can be incorporated using chemically aminoacylated tRNAs. With this approach Uaa's are chemically attached to tRNA molecules that either naturally recognise or have been engineered to recognise a stop codon allowing incorporation of an Uaa in response to a specific codon in a

target gene. The main drawback is that unnatural-aminoacylated-tRNAs have to be used with *in vitro* translation systems (5) or introduced transiently and invasively via microinjection (6). Furthermore, these techniques are limited by low yields. The major advance for incorporating Uaa's during normal cellular protein synthesis came by reprogramming the genetic code. Non-endogenous tRNA/aminoacyl-tRNA-synthetase pairs were engineered to recognise a unique codon, normally the rarely used amber stop codon (7). This method was revolutionary because it allowed the production of proteins containing Uaa's to be fully genetically encoded. Uaa incorporation took place autonomously, in parallel with natural protein translation (co-translationally). This approach improved yields and allowed for the incorporation of an Uaa within a defined protein target at specified positions, therefore massively increasing the scope of Uaa-containing proteins as a research and technology tool. Using this method, almost 100 Uaa's (8, 9) with a range of new and useful properties have been introduced into proteins in bacteria (10, 11), yeast (12, 13), mammalian cell lines (14) and recently the first whole organisms; *Caenorhabditis elegans* and *Drosophila melanogaster* (15, 16).

1.3 Uaa incorporation within the cell

The central dogma of biology describes the passage of information from DNA (genes) to proteins, via RNA (17). In *E. coli*, genes are transcribed into complementary messenger-RNA (mRNA) by RNA polymerase II (Figure 1.1A). mRNA is subsequently translated into a chain of amino acids by a combination of transfer-RNA (tRNA), aminoacyl-tRNA-synthetase (aaRS) and ribosome (Figure 1.1B). In mRNA translation, an aaRS recognises and loads a paired tRNA with a specific amino acid (discussed later and in Figure 1.2A). Aminoacylated-tRNAs then enter the ribosome, bind to a specific trinucleotide codon on a transcribed mRNA via an anticodon and the bound amino acid is incorporated into the growing peptide (Figure 1.1C and Figure 1.2B). The machinery of translation is specific for an amino acid with each amino acid having a unique paired tRNA and aaRS, and (at least one) unique codon.

To implement a reprogrammed (or expanded) genetic code approach, a unique codon is selected to code for the incorporation of the Uaa. The genetic code consists of 64 codons of which all are utilised in coding for 20 amino acids and 3 translation stop codons. The most widely used approach is to reprogram the amber stop codon; TAG (or UAG in mRNA) (7). The amber stop codon is the least used codon so the deleterious effects of 'hijacking' the codon are minimised. For example, in *E. coli* K-12 only 326 of 4288 coding sequences (7.6 %) terminate with the amber stop codon (18), as in Figure 1.1D; scenario 1. Furthermore, amber suppression has been shown to not significantly affect cell growth rates due to low usage and also because the 3' untranslated region of a gene is stop codon rich (19). To code for an Uaa, the amber stop codon is introduced into a gene of interest at the desired codon position, as

shown in red in Figure 1.1. Normally, a stop codon placed in the coding region of a gene would terminate translation by release factor 1 (RF-1), producing a truncated protein (Figure 1.1D; scenario 2). However, when the amber stop codon has been reprogrammed to code for an Uaa (as employed here) a full-length protein is produced containing the Uaa instead of a prematurely truncated protein (Figure 1.1D; scenario 3).

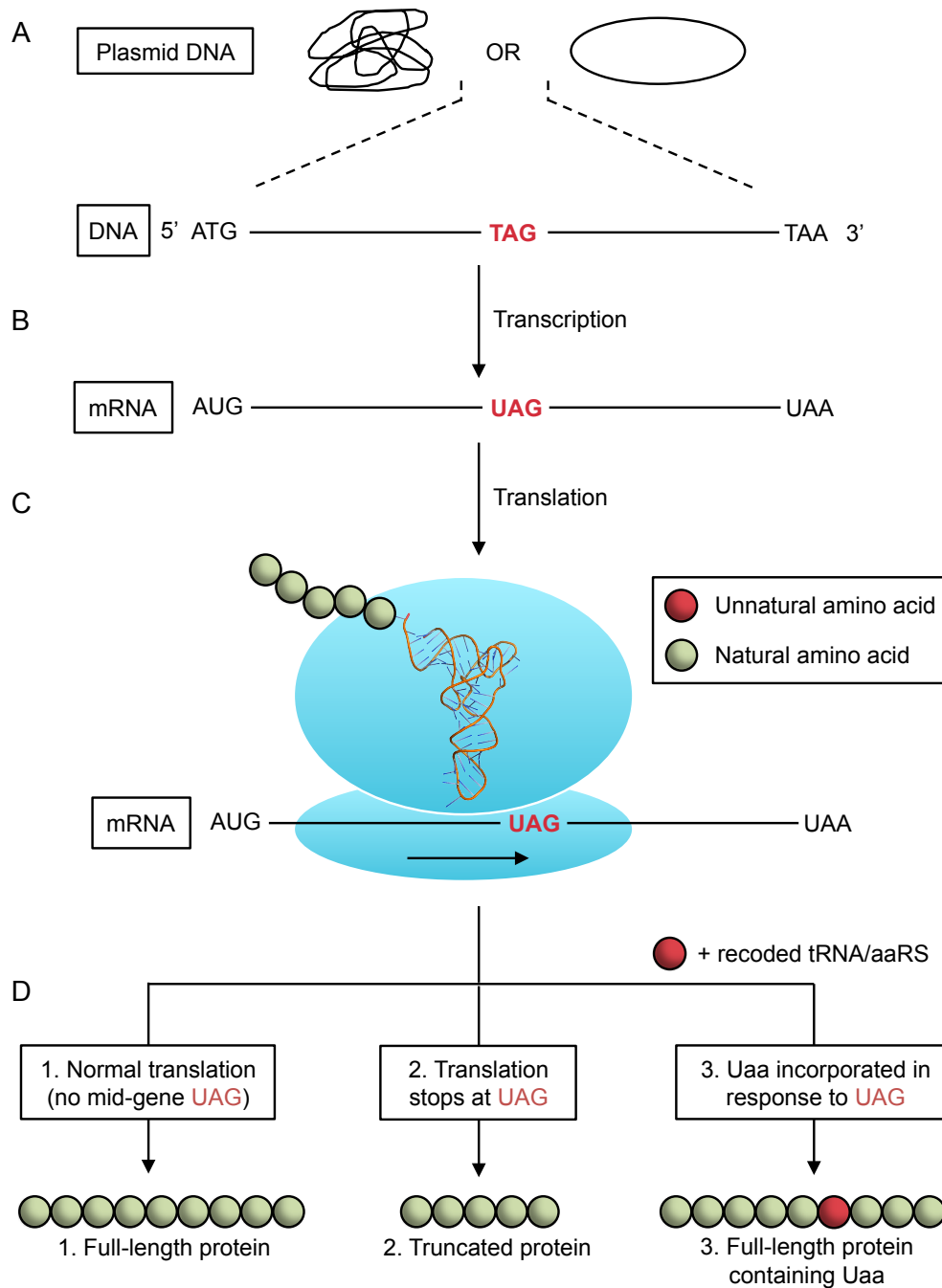


Figure 1.1. Transcription, translation and incorporation of Uaa's via an expanded genetic code. (A) The DNA information coding for proteins is found in genes that are stored on chromosomes or extra-chromosomal elements such as circular plasmids as shown here. The gene shown contains a reprogrammed amber stop codon (red) in the coding region that will code for an Uaa. (B) Double-stranded DNA is transcribed to single-stranded mRNA that matches one (template) strand and is complementary to the other (coding) strand. (C) The ribosome (blue) translocates along mRNA translating codons into amino acids in a growing protein. (D) Translation of a 'normal' gene without a stop codon in the coding region results in full-length protein (scenario 1). When a gene contains a stop codon in the coding region, as in this figure, translation is terminated when the ribosome reaches the stop codon (scenario 2). In a reprogrammed genetic code approach, instead of terminating translation the amber stop codon (UAG) signals for incorporation of an Uaa (red sphere) into the growing peptide (scenario 3).

Although TAG is the most commonly used codon for Uaa incorporation, other codons have been utilised and are necessary if multiple Uaa's are to be incorporated into a single protein or cell. Similarly to the amber stop codon (TAG), the opal stop codon (TGA) has been successfully employed to incorporate an Uaa (20), although with lower efficiency. Lesser-used degenerate codons of some amino acids, such as the least used arginine codon (AGG), have been recoded with minimal negative side effects (21). Novel codons, such as quadruplet codons, are an exciting development as they alleviate the need to reprogram a codon and have been successfully employed to incorporate Uaa's into proteins in *E. coli* (22-24). To use quadruplet codons, the cell requires an additional ribosome, engineered to decode quadruplet codons. The lower efficiencies associated with using alternative stop codons and the additional engineering of a ribosome required for using quadruplet codons makes the amber stop codon (TAG) currently the most amenable codon for Uaa incorporation.

Uaa incorporation during cellular protein synthesis imitates the natural process but uses a new tRNA/aaRS pair specific for that Uaa. aaRS's are enzymes that load tRNA molecules with amino acids via the mechanism described in Figure 1.2A. Uaa-specific aaRS's (aaRS^{Uaa}) work in exactly the same way but are evolved to recognise the desired Uaa, not the original amino acid. Aminoacylated or 'loaded' tRNA molecules enter the ribosome and bind to mRNA via a unique anticodon region that is complementary to the specified codon (Figure 1.2B). When the ribosome reaches the amber stop codon the tRNA specific for the Uaa recognises the amber stop codon via an AUC anticodon and introduce the bound Uaa into the growing peptide (Figure 1.2B). As mentioned, the machinery of translation is specific for an amino acid with each amino acid having a unique paired tRNA and aaRS. For incorporation to be specific for a given Uaa and for the position of incorporation to be controllable, the new tRNA/aaRS pair must not cross-react with any endogenous pairs, *i.e.* it must act orthogonally. This means that the tRNA must only recognise and dock with its paired aaRS, and bind to a specified, unique codon. Furthermore, the introduced aaRS must not acylate any of the endogenous tRNAs with the Uaa and must not acylate the paired tRNA with any natural amino acids. Finally, the Uaa must not be a substrate for endogenous aaRS and must be efficiently transported into the cytoplasm (or synthesised by the organism as in Mehl *et al.* (25)).

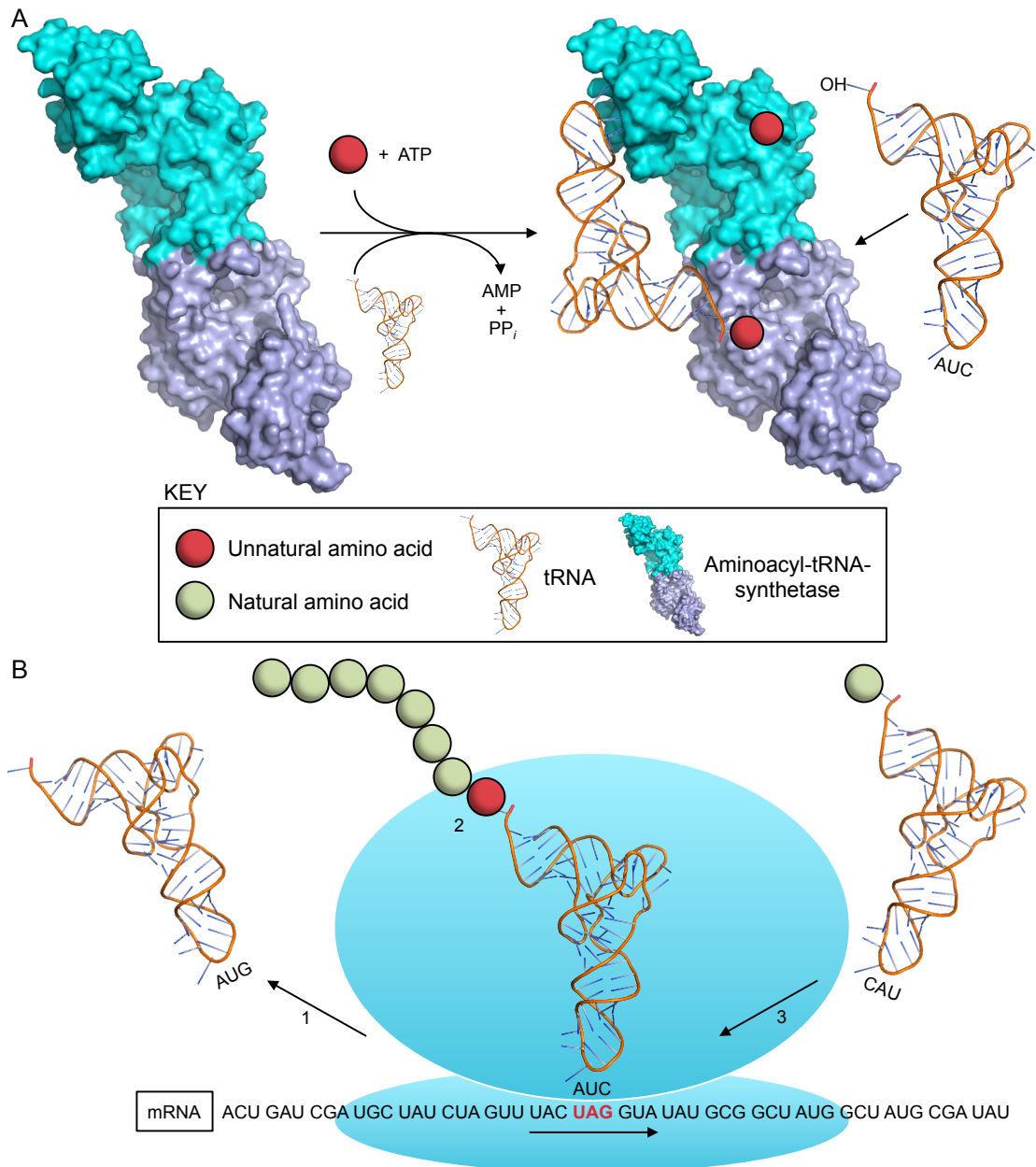


Figure 1.2. Co-translational incorporation of unnatural amino acids (Uaa's). (A) Schematic representation of engineered tRNA_{CUA} (orange) charging with Uaa (red sphere) by an evolved aaRS (cyan and blue). aaRS's catalyse a two-step reaction; firstly, the enzyme binds the amino acid and ATP to catalyse the formation of an aminoacyl-adenylate with loss of PP_i. Secondly, the aminoacyl group is transferred to the terminal OH of a docked tRNA, releasing AMP. Shown is the tyrosyl-RS (TyrRS), which is a homodimer (monomers coloured cyan and blue) and so binds 2 amino acids, ATP's and tRNA's per TyrRS. (B) Subsequent incorporation of the Uaa into a growing peptide within the ribosome (blue). The ribosome moves along the mRNA (direction shown by lower arrow) containing the amber stop codon (bold, red) as with a natural transcript. (1) tRNA molecules carrying natural amino acids enter the ribosome, recognize their cognate codons, present the amino acid for peptide bond formation and leave the ribosome ready to be loaded with another amino acid. (2) When the amber stop codon is reached, instead of acting as a translation termination signal, the charged-tRNA_{CUA} enters the ribosome and delivers the Uaa for incorporation in the growing peptide. (3) Translation continues as normal until a stop codon other than UAG is reached. PDB accessions used: 1U7D, TyrRS; 1J1U, tRNA.

In summary, to incorporate Uaa's into proteins at genetically specified positions (Figures 1.1 and 1.2) cells require:

1. A specified codon, usually the amber stop codon (TAG/UAG),
2. An engineered tRNA that recognises only that codon and only docks with a paired aaRS
3. An aaRS evolved to load only the paired tRNA with only the Uaa of interest
4. The Uaa of interest available for incorporation

1.4 Engineering an orthogonal tRNA/aaRS pair

Uaa's are not normally good substrates for endogenous aaRS's and are outcompeted by the natural amino acid. Thus, specificity for the Uaa needs to be engineered into an aaRS. For a tRNA/aaRS pair to incorporate an Uaa specifically in response to a selected codon it must not cross-react with any endogenous pairs. Rather than extensively engineering an endogenous tRNA/aaRS pair to recognise an Uaa (and not a natural amino acid), a better strategy is to import a pair from a different species. One of the original (and still most commonly used) approaches was to adapt an orthogonal tRNA/aaRS pair from the Archaeon *Methanocaldococcus jannaschii* (*Mj*) for use in *E. coli* (26). The pair that naturally encoded tyrosine incorporation in response to the TAC codon, the tyrosyl-tRNA/tyrosyl-tRNA-synthetase pair (*Mj*-tRNA^{Tyr}/*Mj*TyrRS), was chosen for several reasons. *Mj*-tRNA's have distinct aaRS recognition elements compared to bacterial counterparts (such as the variable arm) so it is less likely to interact with an endogenous *E. coli* aaRS (27). *Mj*TyrRS is ideal because the interaction it makes with the anticodon-loop of *Mj*-tRNA^{Tyr} is relatively weak (28). Altering the specificity of the anticodon loop of *Mj*-tRNA^{Tyr} from tyrosine to the amber stop codon (GUA to CUA) is therefore possible with minor effects on affinity. Additionally, *Mj*TyrRS lacks an editing mechanism that recognises an Uaa as incorrect and results in deacylation (29).

The initial tRNA developed to recognise the amber stop codon (*Mj*-tRNA_{CUA}) functioned effectively in translation in *E. coli* but was found to be a substrate for endogenous *E. coli* aaRS's (10). *Mj*-tRNA_{CUA} was engineered by directed evolution to be more orthogonal in *E. coli* by reducing interactions with endogenous aaRS's (10). Eleven nucleotides in *Mj*-tRNA^{Tyr}_{CUA} were randomised and variants were put through rounds of negative and positive (double-sieve) selection (Figure 1.3A).

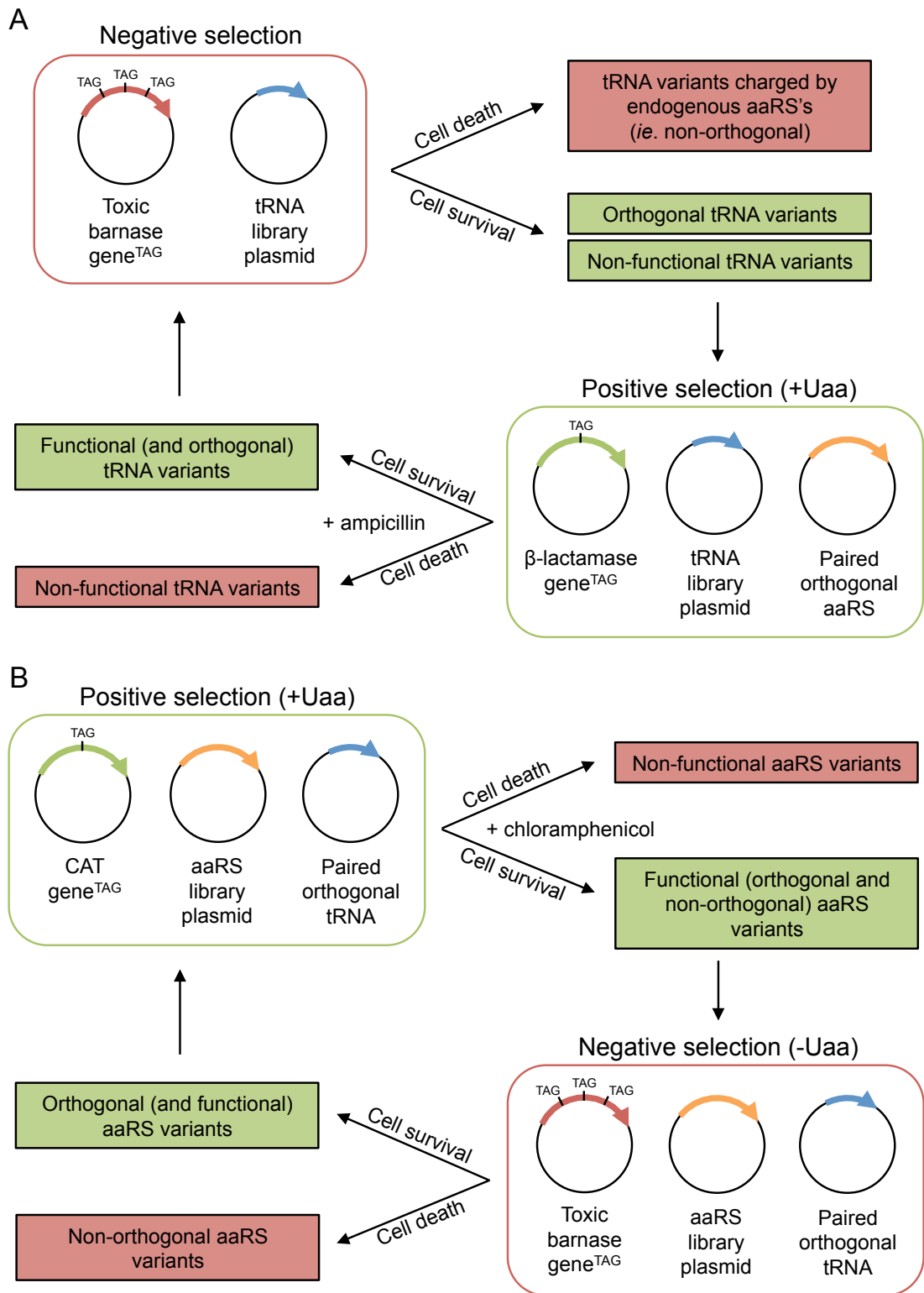


Figure 1.3. Directed evolution of (A) *Mj*-tRNA_{CUA} and (B) *Mj*TyrRS by double-sieve selection. *Mj*-tRNA_{CUA} and *Mj*TyrRS variants generated by random PCR mutagenesis are subjected to rounds of positive and negative selection in *E. coli*, as described in the text. Negative selections are based on suppression of TAG mutations in the gene coding for the toxic barnase protein. These negative selections result in cell death and removal of variants that are non-orthogonal to the *E. coli* host machinery. Positive selections are based on suppression of TAG mutations in the genes coding for antibiotic resistance proteins (β -lactamase and chloramphenicol acetyl transferase (CAT)). Positive selections result in cell survival and propagation of only functional *Mj*-tRNA_{CUA} and *Mj*TyrRS variants.

Negative selection removed any tRNA variants that were charged by endogenous aaRS's (non-orthogonal) by testing variants ability to suppress TAG codons in the gene coding for barnase, a toxic ribonuclease, in the absence of the paired *Mj*TyrRS. Positive selection was performed in the presence of the paired *Mj*TyrRS and tested the variants ability to suppress TAG codons in the β -lactamase gene, therefore conferring ampicillin resistance to functional variants.

Similarly, the *Mj*TyrRS must be evolved to recognise the desired Uaa as a substrate and importantly, not tyrosine or any natural amino acid (Figure 1.3B). A double-sieve selection is again employed with the orthogonal *Mj*-tRNA_{CUA} present in all selection steps (10). First, a library of *Mj*TyrRS mutants is generated by site-directed PCR mutagenesis (via 'doped' oligonucleotide primers containing random base substitutions) focusing on residues that interact with the amino acid (30, 31). Positive selection uses a chloramphenicol acetyl transferase (CAT) gene with a TAG mutation and takes place in the presence of the desired Uaa therefore conferring antibiotic resistance from functional variants. Negative selection uses a barnase gene with TAG mutations and takes place in the absence of Uaa so variants that are able to incorporate natural amino acids are removed. Commonly randomised residues include Tyr32, Leu65, Ala67, His70, Gln109, Asp158 and Leu162 (Figure 1.4) (32). Mutation of Tyr32 and Asp158 to smaller amino acids is common for accommodating larger Uaa substrates (eg. Figure 1.4B-D). Some aaRS^{Uaa}'s have more extreme structural changes such as generation of a cavity in naphthyl-alanineRS by disruption of an α -helix due to a Asp158Pro mutation (Figure 1.4C).

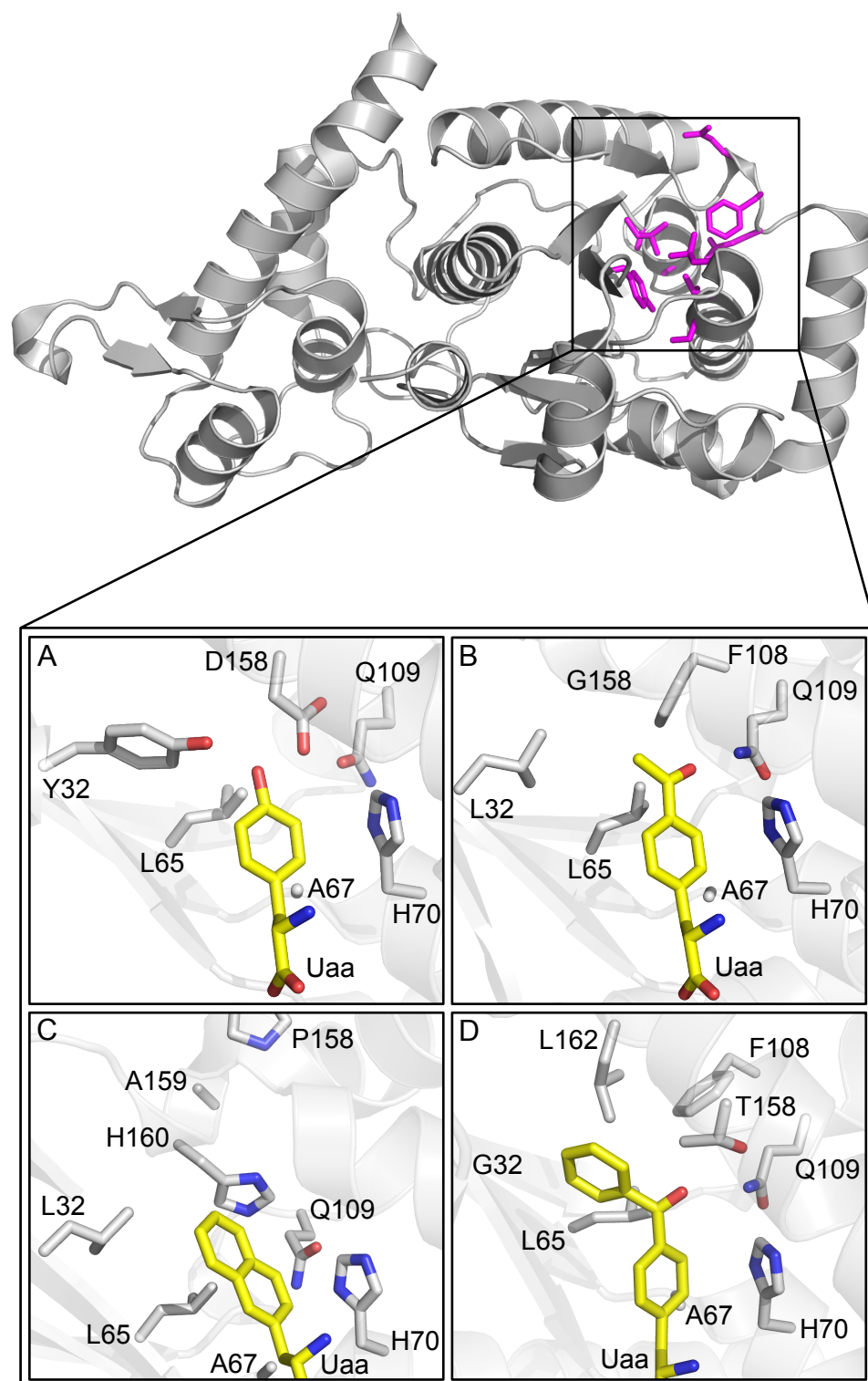


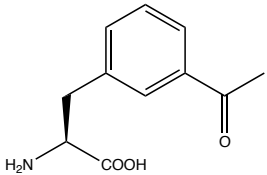
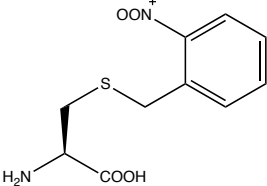
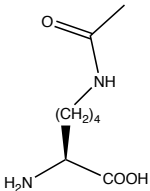
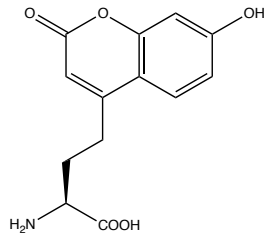
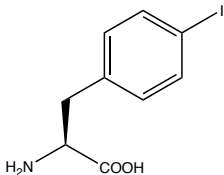
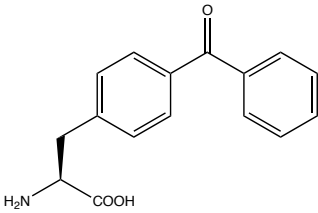
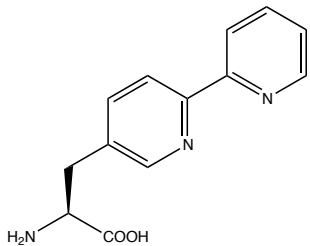
Figure 1.4. Evolution of orthogonal aaRS's for Uaa incorporation. Crystal structures of the amino acid binding region of (A) wild-type *MjTyrRS* and those evolved to recognise (B) acetyl-phenylalanine, (C) naphthyl-alanine and (D) benzoyl-phenylalanine. One monomer of the homodimeric *MjTyrRS* is shown in cyan at the top with the amino acid binding residues coloured pink. The Uaa's in each case are coloured by element with carbons coloured yellow. Residues that are frequently randomised for altering amino acid specificity are shown as sticks and coloured by element. Gly mutations are not displayed by the side chain representation used so labels are shown in the correct position. PDB accessions (top) 2ZP1, (A) 1J1U, (B) 1ZH6, (C) 1ZH0 and (D) 2HGZ.

The same approach is used to engineer aaRS/tRNA pairs for Uaa incorporation in higher organisms. The main difference is the tRNA/aaRS pairs used. *E. coli* tRNA^{Tyr}/TyrRS and tRNA^{Leu}/LeuRS pairs have been engineered to incorporate Uaa's into proteins in *Saccharomyces cerevisiae*. A heterologous *Bacillus stearothermophilus* tRNA^{Tyr}/*E. coli* TyrRS pair has been engineered to allow Uaa incorporation in CHO (Chinese Hamster Ovary) cells (14) and *Bacillus subtilis* tRNA^{Trp}/TrpRS in HEK 293T (Human Embryonic Kidney) cells (20). More recently, pyrrolysyl-tRNA/synthetase (tRNA^{Pyl}/PylRS) pairs from *Methanosarcina* species such as *M. barkeri* and *M. mazei* have been developed for use with *C. elegans* and *D. melanogaster* (15, 16).

1.5 Range and uses of unnatural amino acids

Uaa's are a valuable addition to the proteome because they can be designed to have physical and chemical properties not normally accessible in nature, which are conferred to a protein on incorporation; some examples with their practical uses are shown in Table 1.1. The majority of Uaa's incorporated to date are aromatic, derivatives of phenylalanine, due to wide use of the tRNA^{Tyr}/TyrRS pair as a starting point for evolution of new orthogonal pairs. This is because the recognition mechanism of the tyrosyl hydroxyl group by TyrRS is well known (Figure 1.4), thus simplifying the engineering process by which new Uaa's are recognised. Additionally, the majority of natural amino acids are not aromatic, which means that a major specificity determinant (for the phenyl ring) is already built into the aaRS's. More recent work has focused on the development of aliphatic Uaa's and corresponding tRNA/aaRS pairs to expand the repertoire of Uaa's that can be incorporated into target proteins (9). Substitutions of Uaa's with similar structures to the native residue are better tolerated, therefore unnatural analogues of aliphatic amino acids allow a wider range of protein residues to be probed.

Table 1.1. Structure, properties and potential uses of some interesting Uaa's.

Structure	Name	Property	Use
Selectively reactive groups			
	<i>m</i> -acetyl-phenylalanine	Reactivity with hydrazides and alkoxyamines	Attach fluorescent probes to LamB membrane protein (33)
Photocaged amino acids			
	<i>o</i> -nitrobenzyl-cysteine	Photocage removed at 365 nm	Photoregulation of caspase-3 (13)
Post-translational modification mimics			
	N ^ε -acetyl-lysine	Genetically encodable acetylated lysine	Assess role of histone acetylation on nucleosome state (34)
Biophysical probes			
	7-hydroxy-coumarin-based	Fluorescence	Probe folding and unfolding of superoxide dismutase (35)
	<i>p</i> -iodo-L-phenylalanine	Contains a heavy atom	Phase determination in X-ray crystallography (36)
Photocrosslinking			
	<i>p</i> -benzoyl-L-phenylalanine	Crosslinks with C–H bonds on irradiation	Map protein-protein interactions in SecYEG membrane translocon (37)
Metal chelating			
	Bipyridyl-alanine	Coordinate metal ions (eg. Cu ²⁺ , Fe ^{2+/3+})	Rational design of metalloproteins (38). Engineer proteins capable of photoinduced e ⁻ transfer from a Ru ²⁺ centre (39)

1.6 *p*-azidophenylalanine

p-azidophenylalanine (azPhe; Figure 1.5) is an aromatic Uaa first incorporated into proteins in *E. coli* by Chin and colleagues (11). AzPhe has a 1,3-dipolar azido group (N_3) with four electrons delocalised over three nitrogen atoms that form a π -electron system. Consequently, azPhe can be represented by several resonance structures. AzPhe has a number of useful properties that would be of use in the context of proteins such as photoreactivity, selective reactivity with alkynes and the distinctive azide infra-red absorption spectrum. As discussed in the following sections, these three key characteristics unique to aryl azides are of great potential for introducing new properties into proteins.

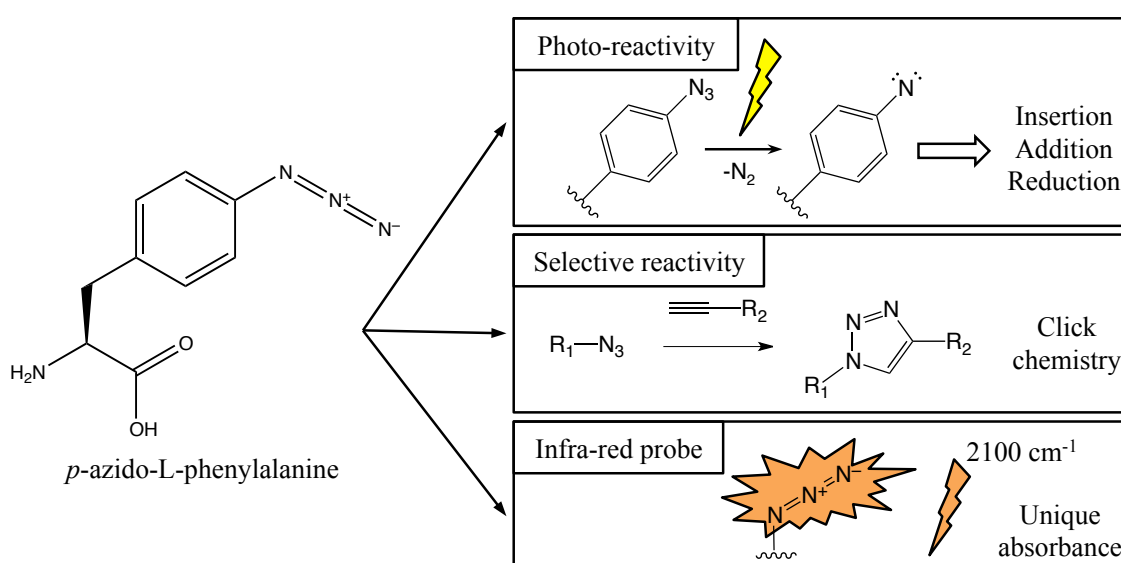


Figure 1.5. Structure and properties of azPhe. The chemical structure of azPhe is shown on the left with unique properties on the right. The azido group is planar and has 4 delocalised electrons. The Phe to azide ($Cp-N\alpha-N\beta$) bond angle is $\sim 120^\circ$.

1.6.1 Photoreactivity

The photochemistry of phenyl azides has been known and exploited for over 90 years (40). Phenyl azides have been used as photocrosslinkers to label antibodies with haptens (41) and to probe protein-peptide interactions (42). Thus, it would be useful to incorporate phenyl azide chemistry into defined positions in a specified target protein to exploit the photochemistry for these and other uses. The reaction mechanism is shown in Figure 1.6. The first step involves activation of the azide (**1**; Figure 1.6) by UV-visible light (or high temperatures), typically at wavelengths below 350 nm. This leads to loss of molecular nitrogen (N_2) and formation of the reactive singlet nitrene species (**2**; Figure 1.6) (43). The singlet nitrene contains six outer-shell electrons and, as such, is a highly reactive electrophile. The

singlet nitrene can follow multiple reaction pathways, the distribution of which depends on the local environment, ring-substituents and temperature. At temperatures above 165 K the singlet nitrene can insert into activated σ bonds, particularly when the phenyl azide is substituted with electron withdrawing groups (44, 45). Singlet nitrenes can also react with carbon double bonds through addition reactions and with nucleophiles such as amines or alcohols (46-48). Additionally, as singlet nitrene is a powerful base it can abstract protons from a variety of sources such as alcohols, amines and acidic C–H bonds to form nitrenium ions (and amines) (45, 49). Although difficult to predict, protonation is more likely in acidic environments or when the phenyl ring is substituted with electron donating groups. The singlet nitrene can also undergo an intramolecular ring expansion to form a cyclic ketenimine (**3**; Figure 1.6), via a short-lived benzazirine intermediate (43). Both the benzazirine and ketenimine can react with proximal nucleophiles to form stable crosslinkages (45, 50, 51). As with singlet nitrene, ketenimines can also insert into σ bonds such as C–H and N–H. At temperatures below 165 K, excitation of the singlet nitrene predominantly produces a radical triplet nitrene state via inter-system crossing (ISC) (52). The triplet nitrene is discussed in more detail in Chapter 6 and has different reactivity (not shown in Figure 1.6) including protonation, hydrogen abstraction and dimerization with another triplet nitrene to form an azo ($-\text{N}=\text{N}-$) linkage (53, 54).

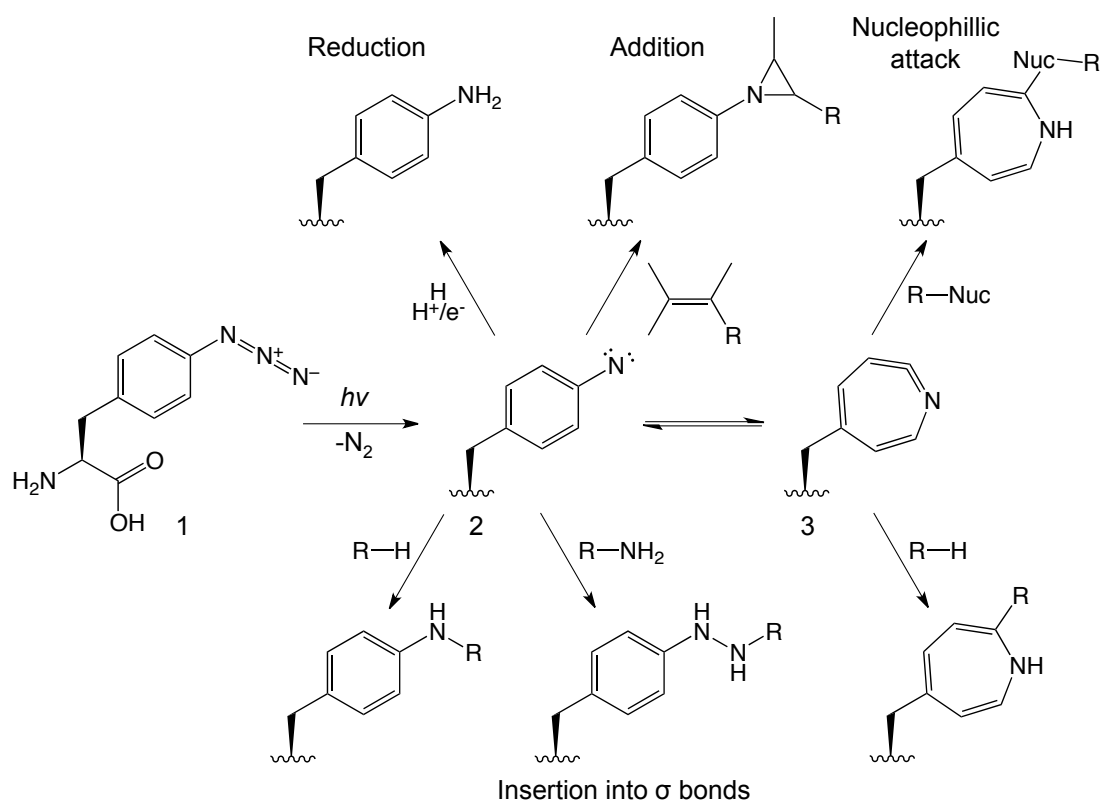


Figure 1.6. Photochemistry of phenyl azides. Upon irradiation with UV light of wavelengths below ~ 350 nm, phenyl azide (**1**) exudes molecular nitrogen (N_2) leaving singlet nitrene (**2**). At temperatures above 165 K, **2** can be protonated to form (a nitrenium ion ($-NH^+$) and) an amine, as indicated. Additionally, **2** can undergo addition reactions into double bonds or insert into activated C–H and N–H bonds. Alternatively, **2** can undergo ring expansion to form a cyclic ketenimine (**3**), via a benzazirine intermediate. Both **3** and the intermediate can be attacked by nucleophiles to form crosslinkages or insert into activated C–H and N–H bonds. The electron configuration shown for **2** is closed-shell, however an open-shelled state also exists.

Although a great deal is known about the photoreactivity of phenyl azides under a range of conditions, our knowledge is not complete. It is important to understand the reaction in order to fully utilise the chemistry and to aid in the design of future experiments. The photochemical pathways of phenyl azides have been studied via a range of different techniques including matrix isolation and time-resolved spectroscopy. One inventive advance in the study of phenyl azides was the use of synthetic host molecules or “cages” that allow greater control over a reaction by defining the environment. Synthetic cages have been designed to provide an inert environment or one with specific chemistry to trap or direct the pathway of reactive chemical intermediates and have provided key insights into the reactions of phenyl azide (55-59). Proteins are interesting targets for caging studies as many have highly defined and studied structures that are extremely diverse, easy to mutate and self-folding and/or assembling. Uaa incorporation via an expanded genetic code (Section 1.2) would be a valuable addition to this technique as it allows for the position of the reactive species of interest to be defined. Phenyl azide chemistry is a good test bed for the use of proteins as cages because the reaction is well studied and the phenyl azide-containing Uaa, azPhe, can be encoded. Furthermore, this could lead to protein caging being used as a general tool for studying other radicals.

As well as for more fundamental studies of photochemistry, the introduction of a photoreactive amino acid would be a highly desirable tool for optogenetics. Optogenetics is the precise control of biological events using a combination of light and genetic engineering. Controlling biological processes using light is a powerful approach because it allows for high spatial and temporal resolution and uses an innocuous effector. AzPhe would be a useful addition to the optogenetic toolbox as it is small, has a specific, controllable photoreaction, and is transferable to any protein. When incorporated at an important protein residue, the photoreaction of azPhe could be used to modulate local protein environment and therefore the overall properties of a protein.

1.6.2 Selective reactivity via Click chemistry

The incorporation of an azido group opens up new possibilities for bioconjugation with high target and site specificity even within complex samples such as cells. Via a set of reactions termed Click chemistry, the azido group can undergo cycloaddition reactions with alkyne (carbon-carbon triple bond) labelled molecules. Click chemistry is a chemical approach developed by the Sharpless group designed to join molecules together quickly and robustly by mimicking nature with the joining of small modular units (60). The original azide-alkyne cycloaddition reaction as described by Huisgen (61) was significantly improved with the addition of a copper catalyst (termed Copper-catalysed Azide-Alkyne Cycloaddition (CuAAC)) and is often referred to as *the* Click reaction (62, 63). CuAAC is stereospecific and

possible under mild conditions thereby alleviating the need for elevated temperatures or pressures and hence making it compatible with natively structured biomolecules. CuAAC provides a powerful tool for joining biomolecules because both azides and alkynes are abiotic and react specifically with each other to form stable triazole linkages (60). CuAAC has been successfully used to modify azide and alkyne functionalised proteins with a range of different chemical adjuncts including fluorescent probes, hydrophobic molecules, long-chain polyethylene glycol (PEG), and purification tags (Figure 1.7) (64-66). Additionally, Click chemistry has been used to modify other macromolecules including RNA (67), lipids (68) and glycans (69). Despite the improvements in reaction rate and stereospecificity, the requirement for copper fundamentally limits CuAAC for use in biological systems. Copper (I) is highly cytotoxic to bacterial and mammalian cells and can elicit damage to biomacromolecules (66, 70, 71). The Bertozzi lab pioneered two alternative reactions for covalent modification in living systems; Staudinger ligation and the copper-free, strain-promoted azide-alkyne cycloaddition (SPAAC). Staudinger ligation utilises the reaction between phosphines and azides and has been successfully used in living animals (72, 73). Staudinger ligation is however limited by ease of oxidation of phosphines in air, low reaction rate and poor water solubility. SPAAC uses the ring strain of a cyclooctyne to activate the alkyne bond hence alleviating the requirement for copper (70). A schematic showing how azPhe can be used for site-specific protein modification using SPAAC can be seen in Figure 1.7.

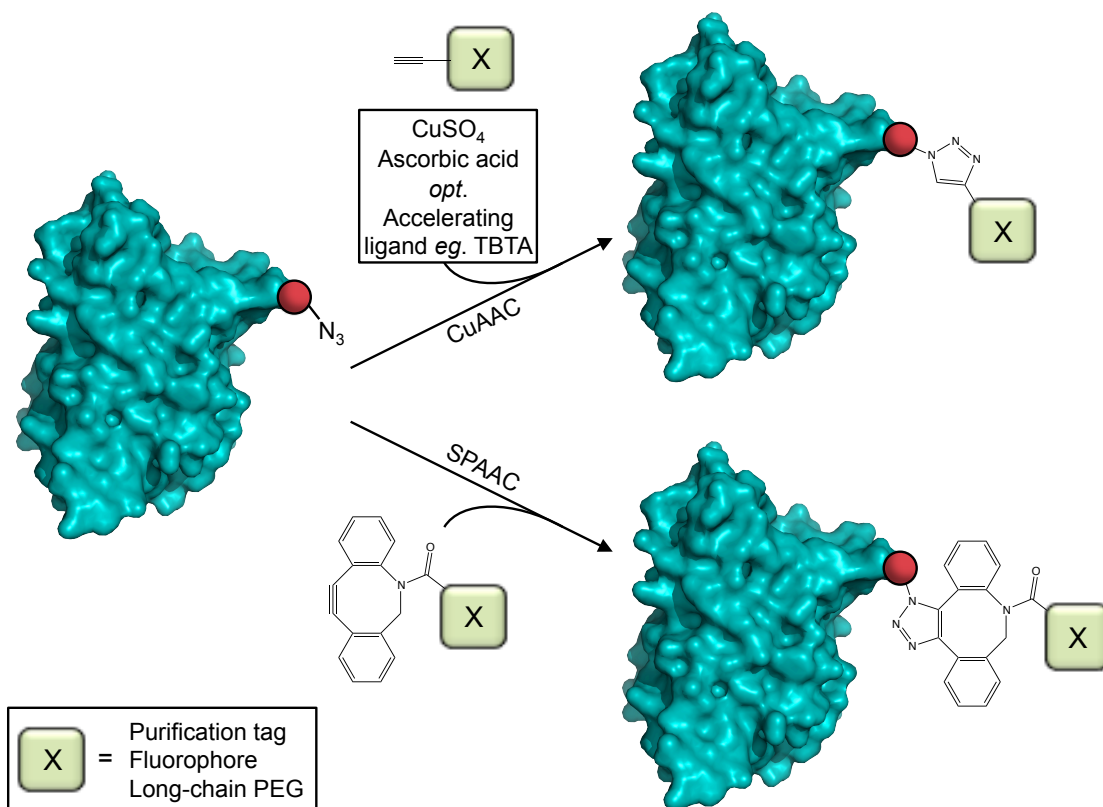


Figure 1.7. Modification of proteins via Click chemistry. Schematic representation of post-translational modification of a protein by copper-catalysed azide-alkyne cycloaddition (CuAAC; top) and strain-promoted azide-alkyne cycloaddition (SPAAC; bottom). Here, azPhe (red sphere) is incorporated into a surface accessible residue of a protein for subsequent modification with an alkyne-functionalised probe. CuAAC requires a Cu(I) catalyst usually generated with Cu(II) (eg. CuSO₄) and a reducing agent (eg. ascorbic acid). SPAAC uses a cyclooctyne-derivatised probe instead of a Cu(I) catalyst. The cyclooctyne group shown is the dibenzylcyclooctyne (DBCO) group used in this study. The protein shown is superfolder GFP; PDB accession 2B3P.

A distortion/interaction model has been proposed to explain the decreased activation energy of strained-promoted cycloaddition compared to strain-free (Cu-catalysed) systems (74, 75). In order for the reaction to take place, the azido and alkyne groups must come into contact and distort to reach a transition state. Strained alkynes have a lower distortion energy (of both the azide and alkyne) required to reach the transition state and hence have lower activation energies than strain-free alkynes (75, 76). Multiple additions to the basic cyclooctyne reactive group have been developed to improve the kinetics and solubility of SPAAC probes including electron-withdrawing groups (77) and additional benzyl rings (78). Currently the most useful functional group is aza-dibenzylcyclooctyne (DBCO) (Figure 1.7). DBCO contains additional ring strain, in the form of two benzyl rings, for favourable kinetics whilst the heterocyclic nitrogen in the cyclooctyne disrupts the hydrophobic surface area (79, 80).

Due to the fast, highly specific and non-toxic modification of biomolecules, SPAAC has found applications in cell surface labelling of glycoproteins (77) and even *in vivo* imaging in Zebrafish (69). In general, an approach for site-specific, orthogonal modification of proteins (and other biomolecules) is valuable. Defined modification allows proteins to be studied in different ways such as with visualisation probes for detection of newly synthesised proteins (BONCAT) (81, 82), or by mimicking natural post-translational modifications of proteins. Additionally, protein modification can produce variants with useful properties like improved therapeutic activity (83, 84) and the ability to bind non-biological materials hence facilitating single molecule studies (85-87).

1.6.3 Biophysical probe

While this dissertation focuses on the photoreactivity and Click-based modification of azPhe, it can also be used as an infra-red (IR) spectroscopy probe due to a unique absorption of approximately 2100 cm^{-1} (88). The spectral properties are also sensitive to electrostatic environment so can be used to probe local changes in protein structure as demonstrated in Ye *et al.* (88). Here, azPhe was incorporated at several sites in rhodopsin to measure changes in the local environments of specific amino acids upon rhodopsin light activation.

1.7 Structure, function and engineering of Green Fluorescent Protein

One of two proteins that are the focus of this thesis is an improved version of the classical Green Fluorescent Protein (GFP). GFP is an intrinsically fluorescent protein first isolated from the jellyfish *Aequorea victoria* (89) and cloned by Prasher *et al.* (90). GFP fluorescence has been exploited in myriad fields of scientific research, often as a way of tracking other proteins. Additionally, GFP is used as a model for the β -barrel protein fold that is found in many membrane-bound proteins. The scientific impact that GFP has had was recognised with the

2008 Nobel Prize in Chemistry (91). Combining the ease of functional analysis with the potential for producing useful research tools, GFP is an excellent target for studying the effects of Uaa incorporation.

GFP is made up of 11 β -strands surrounding a central α -helix in a β -barrel arrangement (Figure 1.8A). The unique fluorescent properties of GFP are centred on a protein-encoded chromophore formed by a chemical rearrangement of contiguous residues during protein folding. Three residues of the central α -helix, Ser65, Tyr66 and Gly67 (in the wild-type protein), undergo cyclisation followed by dehydration and oxidation to produce the mature 4-(*p*-hydroxybenzylidene)imidazolidin-5-one chromophore (Figure 1.8B) (92). The order of the dehydration (step 3) and oxidation (step 4) is disputed and maybe reversed with oxidation preceding dehydration (93). The chromophore has an extended delocalised π -electron system that stretches from the newly formed imidazolinone ring to the phenyl oxygen of Tyr66. Tyr66 can exist in two resonance forms that give rise to two excitation peaks; a major peak at 395 nm corresponding to the neutral phenol ($-\text{OH}$) and a minor peak at 475 nm corresponding to the phenoxide anion ($-\text{O}^-$). When excited at either excitation wavelength, GFP emits green light at ~ 508 nm (94). The single emission peak, regardless of excitation wavelength (and therefore chromophore protonation state), is the result of a phenomenon called excited state proton transfer (ESPT) (92, 95). When the chromophore is excited it becomes more acidic and loses a proton by ESPT, which results in emission from only the anionic chromophore.

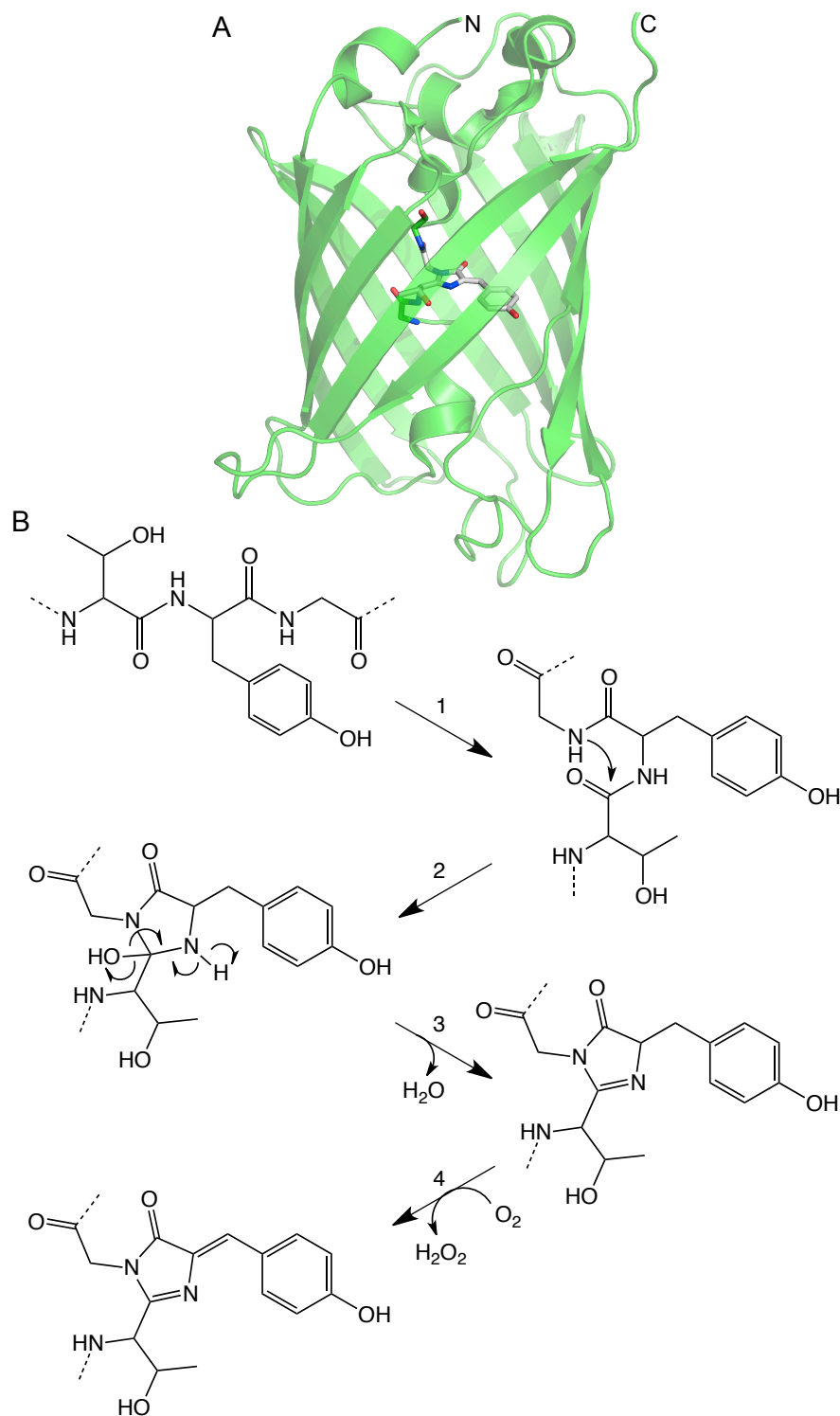


Figure 1.8. GFP structure and chromophore maturation. (A) Three-dimensional structure of GFP showing the β -barrel fold that surrounds a central chromophore (PDB accession 2B3P). The structure is transparent to show the central α -helix that contains the chromophore (coloured by element). (B) Step-by-step maturation of the chromophore. (1) The polypeptide backbone undergoes a series of structural rearrangements that brings the carboxyl C of Thr65 (Ser65 in the wild-type protein) in close proximity to the amino N of Gly67. (2) Nucleophilic attack of Gly67 (N) on Thr65 (C) to form a 5-membered imidazolinone ring, (3) followed by dehydration. (4) Finally, oxidation of the α - β bond of Tyr66 by molecular oxygen extends the conjugated double bond system of Tyr66 to include the imidazolinone ring. The resulting delocalised π -electron system is responsible for green fluorescence.

The protein matrix surrounding the chromophore also plays an important role in defining the fluorescent properties of the protein. Buried polar residues and structural water molecules surround the chromophore that interact to form a complex network of hydrogen bonds (Figure 1.9). The roles of some of the key residues are discussed below. Thr203 lies adjacent to the tyrosyl moiety of the chromophore and participates in a hydrogen bond with the anionic state of the chromophore. Substitution of Thr203 for an aromatic amino acid causes a red shift in the spectral properties due to π - π stacking with the chromophore (96). The Thr203Tyr substitution is the key mutation that gives rise to the yellow fluorescent protein (YFP) variant (97). The side chain of His148 hydrogen bonds with the tyrosyl moiety of the chromophore and together with Thr203 defines the protonation state of the chromophore by promoting ionisation of the phenol hydroxyl group (92, 98). Glu222 is also vital in defining the protonation state of the chromophore. Due to the proximity of the residue with the chromophore, when Glu222 is anionic the chromophore (tyrosyl) cannot exist in a charged state due to electrostatic repulsion (99). Wild-type GFP exists in this state because of a hydrogen bond between Glu222 and Ser65 that stabilises Glu222 in its anionic form. The Ser65Thr mutation in subsequently engineered GFPs, such as enhanced GFP (EGFP), sterically blocks this hydrogen bond and hence keeps Glu222 neutral allowing an anionic chromophore (99, 100). Having a protein that exists solely in the anionic (488 nm) form is useful because it possesses simple, single excitation and emission peaks. Also, the wavelengths are very similar to the widely used organic dye fluorescein, which are commonly used wavelengths in microscope filters and lasers. Glu222 has other matrix contacts and hydrogen bonds including Ser205, which in turns hydrogen bonds with the tyrosyl moiety of the chromophore via a structural water. Additionally, Glu222 and Arg96 have electrostatic roles in chromophore formation by stabilising an intermediate and acting as a general base, respectively (93).

Recent high-resolution structures of GFP (101) have revealed interesting insights into the role of many residues in shuttling protons between the surface of the protein and the chromophore. Proton wires have been shown to have important roles in gathering protons on the surface of GFP from the solvent (collecting antenna), transporting protons to the chromophore (active-site wire) and then providing routes for exit from the protein (101). The collecting antenna is predominantly negatively charged residues with a high proportion of carboxylates that attract and focus protons. Glu5 is thought to be the centre of the antenna and the entrance to the active-site wire. Many residues are implicated in the active-site wire including Ser72, Thr43 and ultimately Glu222-W-Ser205 (101). Proton exits include Asn146 and a gateway created by Thr203 and His148 (101). Furthermore, proton wires are implicated in chromophore maturation (biosynthesis wire) and hence fluorescence.

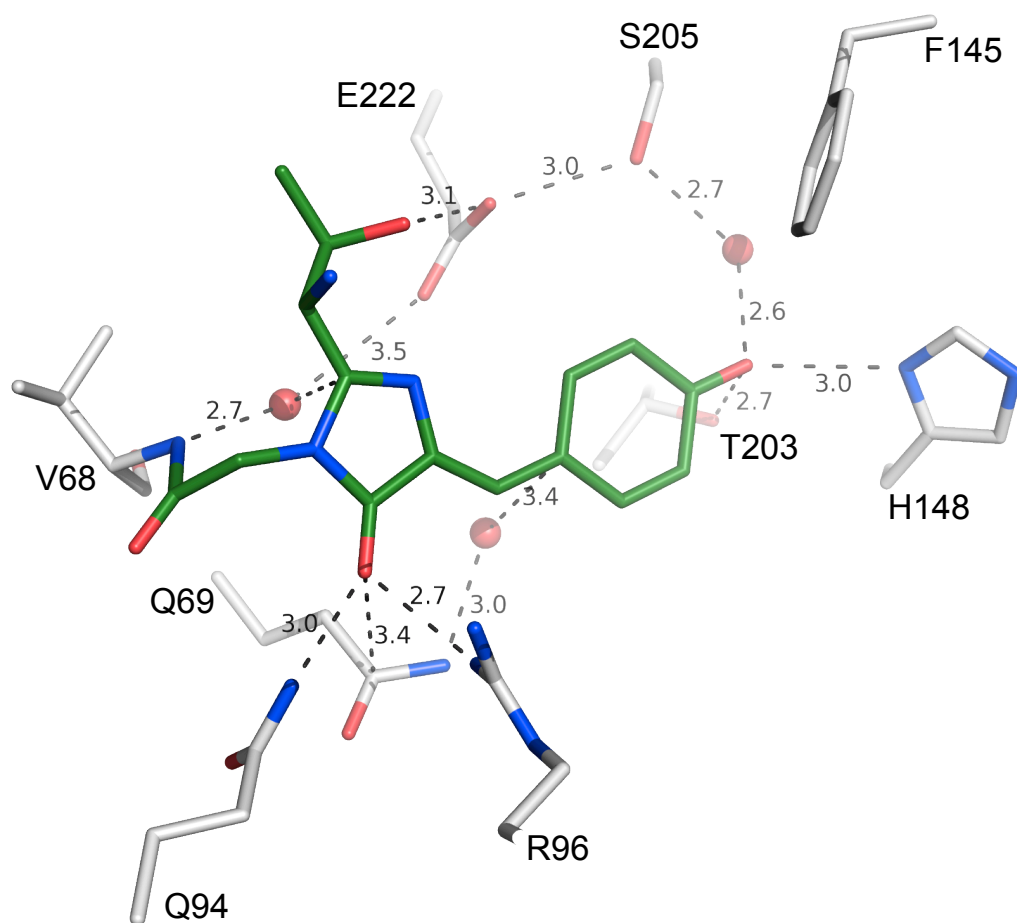


Figure 1.9. Chromophore environment in GFP. Three-dimensional structure of GFP showing interactions between the surrounding protein matrix and the chromophore. Residues are coloured by element with chromophore carbons coloured green. The hydrogen-bonding network is shown by dashed lines with distances given in Angstroms. Only side chains are displayed unless the main chain atoms contribute to the hydrogen-bonding network. Structural water molecules are represented as red spheres.

Many GFP derivatives have been engineered with new and improved properties. One of the first useful GFP variants was enhanced GFP (EGFP). EGFP was engineered to improve production of functional protein at 37 °C which was important to facilitate the use of GFP as a genetically encoded tracer for use in cells (92). EGFP contains two amino acid substitutions; Ser65Thr, which increased fluorescence, improved photostability and resulted in a single excitation peak at 488 nm (102), and Phe64Leu, which improved folding efficiency at 37 °C (103). More recently, the variant termed superfolder GFP (sfGFP) was generated through directed evolution to fold robustly even when fused to poorly folding proteins and to fold efficiently at 37 °C (104). As sfGFP folds more efficiently than other GFP variants more functional protein is produced. Moreover, the improved stability of the scaffold meant sfGFP is more tolerant to point mutations (104). sfGFP was used in this study because it was more likely to tolerate incorporation of Uaa's and because any useful variants generated will benefit from the increased level of protein produced.

1.8 Cytochrome *b*₅₆₂

Cytochrome *b*₅₆₂ (cyt *b*₅₆₂) is a soluble, redox-active protein found in the periplasm of *E. coli* (105) and is thought to act as an electron transport protein although its exact cellular function is unknown (106, 107). It is a small b-type cytochrome (106 amino acid, 12.3 kDa) helix bundle protein composed of four antiparallel α -helices with a left-handed twist (Figure 1.10) (108, 109). Cyt *b*₅₆₂ contains a central haem group that binds non-covalently to the protein via coordination of the iron centre by S_δ of Met7 and N_ε of His102 (Figure 1.10) (108-110). It is this central haem cofactor that confers its electron transfer and redox behaviours with the protein environment tuning these characteristics. Haem also plays a major structural role as the apo-protein is only partially folded. On binding haem, major structural changes occur to form the fully folded, highly stable holo-protein (Figure 1.10B). Haem induced structural changes include full formation of helix 4, increased packing of helix 2 and consequent localisation of the termini (108) as shown in Figure 1.10.

Cyt *b*₅₆₂ is an ideal target for engineering and biotechnological applications due to high pH (pH 2.5 – 10.5) and temperature (< 55 °C) stability, low molecular weight, absence of cysteines and high tolerance to point mutations (107, 108). Cytochromes in general are good model proteins as the helix bundle structure is very common (111) allowing any lessons learnt to be transferred to a wide range of proteins. Moreover, the tuneable redox potential (112, 113), electron transport and potential to link electron transport to secondary processes (114-116) has opened up opportunities to use cyt *b*₅₆₂ as, for example, a transistor component in bio-electronics (117, 118).

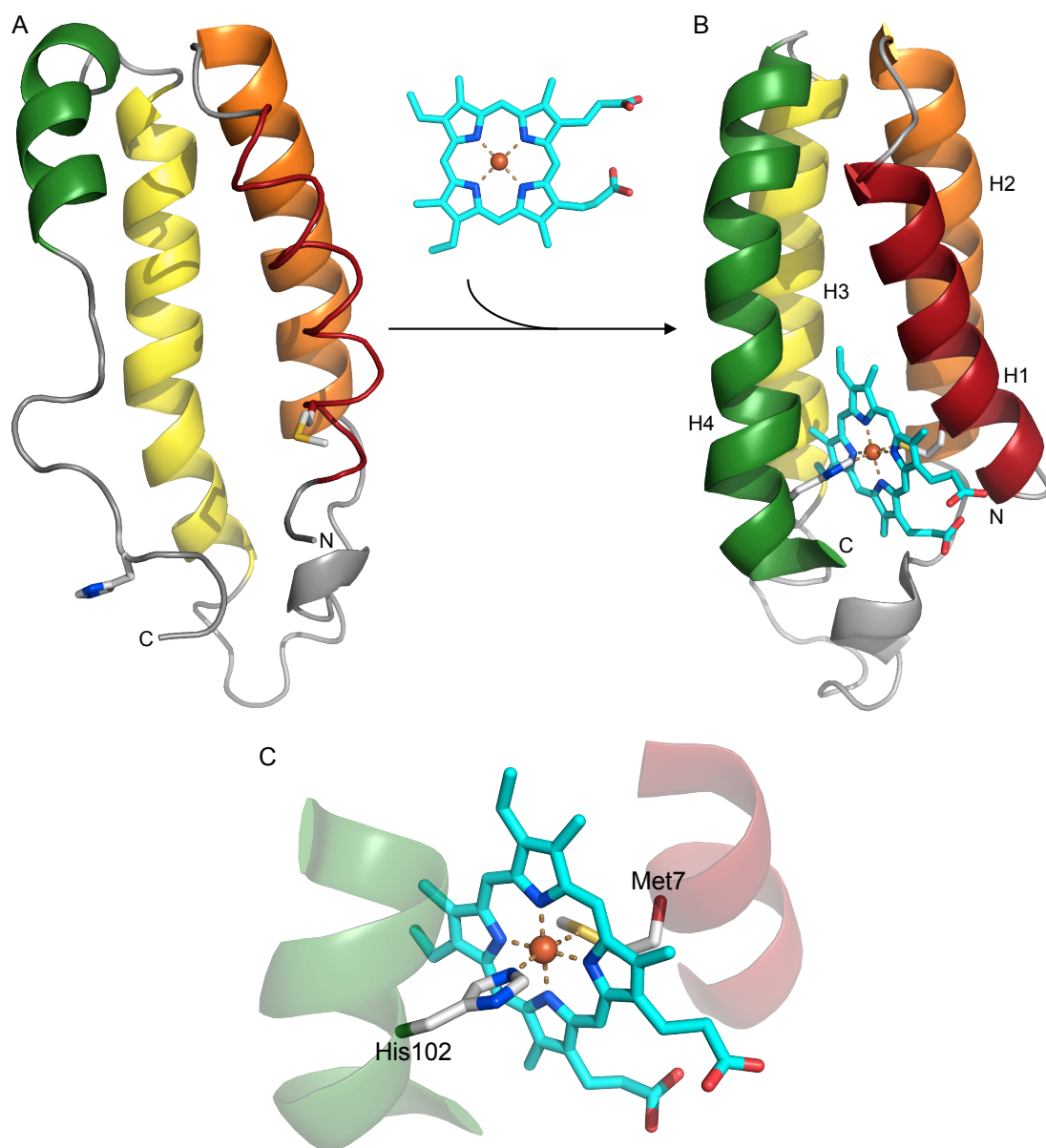


Figure 1.10. Three-dimensional structure of cytochrome b_{562} (cyt b_{562}). The partially structured apo-cyt b_{562} (A) binds haem non-covalently by coordinating the iron centre (dark orange) with Met7 (helix 1) and His102 (helix 4) to form holo-cyt b_{562} (B). α -helices are coloured with a spectral progression (red-orange-yellow-green) and labelled H1 to H4 in (B). Haem, Met7 and His102 are shown as sticks and coloured by element with carbons coloured turquoise, white and white, respectively. Haem binding induces major structural changes in cyt b_{562} including folding of helix 4 (green), increased packing of helix 2 (orange) and consequent localisation of the termini (108). (C) Zoom of haem coordination. Iron is coloured orange and coordinating residues are labelled. PDB accessions; apo-cyt b_{562} , 1APC; holo-cyt b_{562} , 1QPU.

Here, cyt b_{562} was attached to gold surfaces in defined orientations and the conductance through the protein was measured at the single-molecule level using scanning tunnelling microscopy (STM). To facilitate this and other biophysical uses, cyt b_{562} was engineered to introduce surface-exposed cysteine residues so the proteins could be bound to gold and other noble metal ion-based surface and electrodes in defined orientations (97). The ability to make direct electrical contact with two electrodes whilst maintaining native protein structure was revolutionary and opened up myriad possibilities for single molecule studies and creating biophysical devices. The ability to study and analyse proteins at the single molecule level allows for precise details to be learned about mechanism and dynamics by removing the averaging effect of bulk studies. Additionally, single molecule analysis is a prerequisite for exploiting the interesting electron transfer properties of cyt b_{562} for potential integration into an electronic circuit for example. The cysteine mutations introduced are shown in Figure 1.11A along with a schematic of how cyt b_{562} can be bound to a gold surface and connected to an electric circuit via a Pt-Ir tipped electrode as part of a scanning tunnelling microscope (STM) (Figure 1.11B). Two orientations were rationally designed so the haem-to-electrode distances could be changed as a way of altering the conductivity through the protein. Mutation of Asp5 and Lys104 to cysteine created a variant with short haem-electrode distances (both 12 Å) referred to as the short-axis (SA) variant (Figure 1.11A; red). The other variant (long-axis; LA) was generated by replacing Asp21 and Asp50 with cysteine resulting in haem-electrode distances of 20 and 32 Å (Figure 1.11A; blue). The cysteine-noble metal bond mediated by the thiol (-SH) group is effectively covalent in strength and in combination with site-directed mutation can be used to bind proteins to gold in a defined manner. However, the natural abundance of cysteine in proteins reduces its general use (without further engineering).

A combination of the approach taken by Della Pia *et al* (113, 117-120) with the selective reactivity of azPhe (Section 1.6.2) would allow the approach to be transferred to any protein and to other useful materials, such as graphene and carbon nanotubes. The use of azPhe would alleviate the problem of naturally occurring cysteines (or other natural amino acids used) and hence the potential requirement for further engineering or loss of site specificity. AzPhe could be site-specifically incorporated into a target protein and modified with an alkyne-labelled adapter molecule that contains the chemistry to bind to the material of choice. Materials of interest include gold, silver, silicon, quantum dots and the focus of this work, graphene.

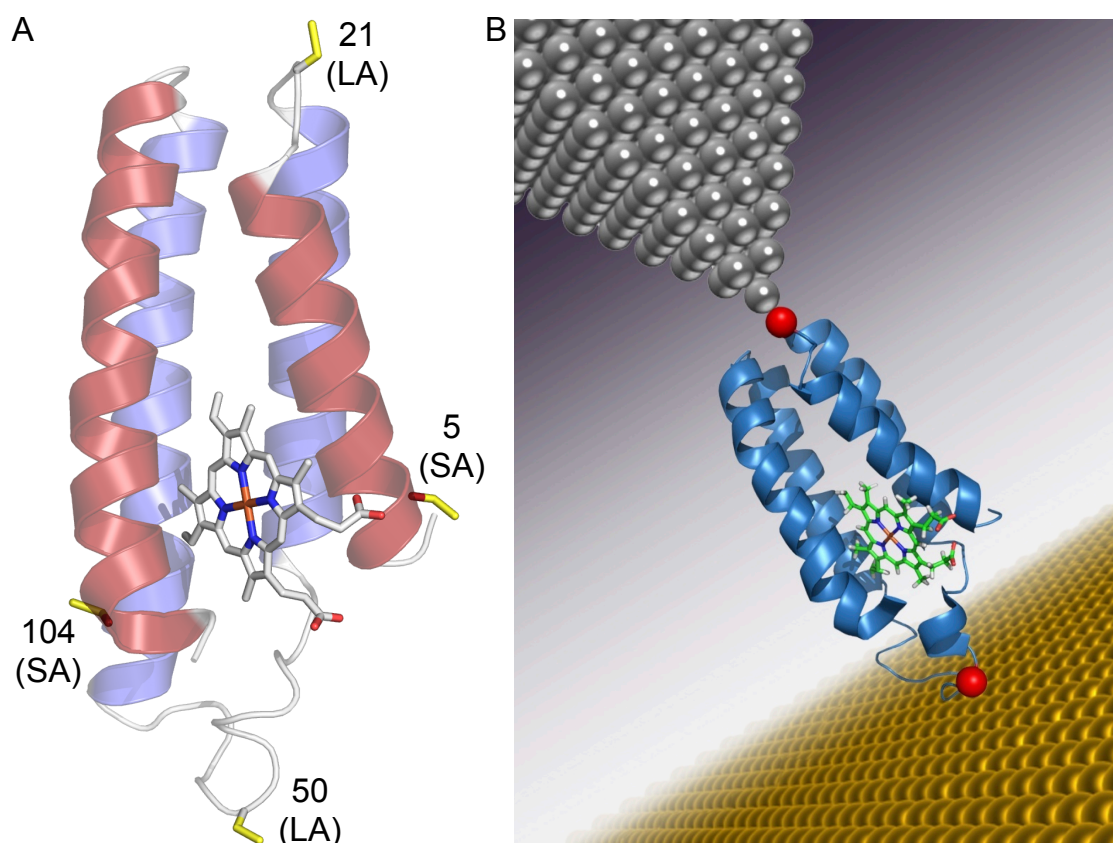


Figure 1.11. Mutation of *cyt b₅₆₂* and use as a single protein transistor. (A) Three-dimensional structure of *cyt b₅₆₂* showing the sites of the cysteine mutations for binding to gold. Mutated residues are shown as yellow sticks and labelled either SA or LA for short and long axis pairs, respectively. Helices 1 and 4 are coloured red and helices 2 and 3 are coloured blue. (B) Schematic of a single *cyt b₅₆₂* (blue cartoon) bound to a gold surface (bottom, gold spheres) and contacted by the Pt-Ir electrode tip (top, silver spheres) of an STM via the introduced cysteine residues (red spheres). Image reproduced with permission of E. A. Della Pia.

1.9 Scope of the project and general aims

Proteins are inherently limited by the properties of their constituting amino acids. By allowing a protein to sample a wider range of amino acids (in the form of Uaa's) new chemistry can be introduced so leading to the generation of proteins with novel properties. This work investigated the tolerance and effects of incorporating Uaa's into proteins using an expanded genetic code approach. More specifically, it focused on two model proteins, sfGFP and cyt *b*₅₆₂, and how Uaa's can be used to influence their properties directly and post-translationally using light and modification with non-biological molecules and materials.

Chapter 3 provides the foundation for the later chapters by describing the experimental basis for Uaa incorporation by amber suppression. This chapter covers the mutation of the genes of interest to protein production and purification. The rest of the chapters focus on the influence of Uaa incorporation in terms of residue position and the influence of subsequent modification. The major focus will be on azPhe (Section 1.6) and its incorporation into sfGFP (Section 1.7) and cyt *b*₅₆₂ (Section 1.8). In **Chapter 4**, other Uaa's were sampled to investigate how they can influence the properties of sfGFP. For azPhe, this work investigated the use of phenyl azide chemistry to generate proteins capable of:

Chapter 4. Modulation via light by exploiting the photochemistry of azPhe (Section 1.6.1).

Using sfGFP as a model, the incorporation of azPhe generated variants that could be activated, deactivated and switch fluorescence excitation in response to light. **Chapter 5** describes the elucidation of the photochemical endpoints reached by the azPhe at each of these residues and hence suggests the pathways taken on photolysis. A combination of spectroscopy and X-ray crystallography points towards reduction to a phenyl amine, insertion into the sfGFP chromophore and a third distinct pathway giving rise to activation, deactivation and fluorescence switching, respectively.

Chapter 6. Caging reactive intermediates. Proteins can be used as 'cages' to trap or direct the pathway of reactive chemical intermediates. sfGFP was used to cage and detect a rare nitrene-derived (anilino) radical following the photolysis of azPhe (Section 1.6.1) at low temperature (77 K). By comparison with a distinct environment in T4 lysozyme (J. Morris (121)) that caged a triplet nitrene, this study showed that the protein matrix influences the radical that is caged.

Chapter 7. Site specific, post-translational modification by Click chemistry (Section 1.6.2).

This part of the study explored the reaction at a fundamental level and investigated the effect of azPhe position (in a protein) on reaction rate and overall yield. Using sfGFP and cyt *b*₅₆₂ as model proteins, a surprising inverse relationship between reaction yield

and residue accessibility was observed, which suggested an important role of the protein surface micro-environment on the reaction. Additionally, residues in defined secondary structures and those with local hydrophobic regions modified to higher extents. The ability to attach useful adjuncts onto proteins opens up myriad possibilities for biotechnology and protein research. This chapter explores how coupling a pyrene group to cyt b_{562} and sfGFP could be used as a general and highly defined way of attaching proteins to novel non-biological materials such as graphene and carbon nanotubes. A pyrene adapter molecule (DBCO-pyrene) was synthesised and attached to proteins with high efficiency. Pyrene-labelled proteins were subsequently seen to bind to the graphene-substitute HOPG.

As well as the more specific outcomes above, more general lessons were learnt regarding the method and influence of Uaa incorporation. The orthogonality of Uaa incorporation was shown to vary for different Uaa's (and therefore aaRS^{Uaa}), but was generally high. Uaa-containing proteins were produced in high yields (average sfGFP^{AzF} 166 mg/L culture *versus* wild-type sfGFP 524 mg/L), which varied substantially depending on the position of incorporation (69 mg/L to 296 mg/L with azPhe) (Chapter 3). On the whole, proteins were shown to be highly structurally and functionally tolerant to the incorporation of Uaa's, even in highly ordered and functionally important regions (*eg.* sfGFP^{Y66AzF}, sfGFP^{H148AzF}; Chapter 4). The only Uaa studied that was poorly tolerated was the large, planar naphthyl-alanine (Chapter 4), which may suggest a limit for the size of Uaa that can be incorporated within a protein core. X-ray crystal structures obtained for the sfGFP^{AzF} variants (Chapter 5) showed only small structural changes (in comparison to wild-type sfGFP) associated with accommodating the Uaa. Uaa's can now be incorporated into a range of organisms including whole animals, with the range of Uaa's available and protein yields ever-increasing. Combined with the ability to alter or install novel properties to a protein (*vide supra*), Uaa's (and particularly azPhe in this work) are an attractive addition to the protein engineering toolbox.

2. Materials and Methods

2.1 Materials

2.1.1 Chemicals

Antibiotics were filter sterilised with a 0.22 μm filter unit (Thermo Fisher Scientific, Leicestershire, UK) as 1000 \times concentrated stock solutions apart from tetracycline, which was prepared as 5 mg/mL solution. The final working concentrations of each antibiotic were; Kanamycin 25 $\mu\text{g/mL}$ (Melford, Suffolk, UK), Chloramphenicol 34 $\mu\text{g/mL}$ (Duchefa Biochemie, Haarlem, Netherlands), Ampicillin 100 $\mu\text{g/mL}$ (Melford) and Tetracycline 25 $\mu\text{g/mL}$ (Fluka Biochemika, Sigma-Aldrich).

Unnatural amino acids (Uaa's) were prepared prior to use by dissolving the appropriate weight of powdered amino acid in ultra-pure water and titrating in 1 M NaOH. Azido-phenylalanine (Bachem, Weil am Rhein, Germany), cyano-phenylalanine (Chem-Impex International, Illinois, USA), trifluoromethyl-phenylalanine (Fluorochem, Derbyshire, UK), naphthyl-alanine (Bachem) and benzoyl-phenylalanine (Bachem) were added to growth media to reach a final concentration of 1 mM. Amino-phenylalanine (Bachem) was used at a final concentration of 10 mM.

Haem (Sigma-Aldrich) was dissolved in 0.5 M NaOH (Sigma-Aldrich) to 10 mM and diluted with dH_2O to working concentrations as required. KNO_3 (BDH AnalaR, Poole, UK) and ascorbic acid (Sigma-Aldrich) were prepared as 100 mM stock solutions, immediately before use. DTT (Melford) and reduced glutathione (Melford) were prepared as 10 mM stock solutions. All reducing and oxidising agents were diluted to working concentrations of 1 mM.

2.1.2 Bacterial cell strains

Escherichia coli cell strains used throughout this work and their genotypes are shown in Table 2.1. Plasmid amplification and propagation for recombinant DNA techniques was performed in *E. coli* DH5 α cells (New England Biolabs (NEB), Hertfordshire, UK). Protein production was performed in *E. coli* TOP10TM cells (Invitrogen, Paisley, UK).

Table 2.1 – Genotypes of *E. coli* strains used

<i>E. coli</i> cell strain	Genotype
DH5 α	F ⁻ <i>fhuA2</i> Δ (<i>argF-lacZ</i>)U169 <i>phoA glnV44</i> Φ 80 <i>lacZ</i> M15 <i>gyrA96</i> <i>recA1 relA1 endA1 thi-1 hsdR17</i>
TOP10 TM	F ⁻ <i>mcrA</i> (<i>mrr-hsdRMS-mcrBC</i>) Φ 80 <i>lacZ</i> M15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ <i>araABC galU galK rpsL endA1 nupG</i>

2.1.3 Bacterial growth media

Lysogeny broth (LB) and LB Agar plates were prepared by dissolving 20 g of granulated LB-Broth (Melford) or 35 g powdered LB Agar (Sigma-Aldrich, Dorset, UK), respectively, in 1 L of ultra-pure water. SOC medium was composed of 2 % (w/v) Bacto Tryptone, 0.5 % (w/v) Bacto Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄. ZYM-5052 autoinduction medium (122) was composed of 1 % (w/v) Tryptone, 0.5 % (w/v) Yeast Extract, 0.5 % (v/v) glycerol, 0.05 % (w/v) glucose, 0.2 % (w/v) lactose, 25 mM Na₂HPO₄, 25 mM KH₂PO₄, 50 mM NH₄Cl, 5 mM Na₂SO₄, 2 mM MgSO₄, 1 × trace metals and 0.05 % (w/v) L-arabinose in ultra-pure water. 1 × trace metal mixture contained 4 μ M CaCl₂, 2 μ M MnCl₂, 2 μ M ZnSO₄, 0.4 μ M CoCl₂, 0.4 μ M CuCl₂, 0.4 μ M NiCl₂, 0.4 μ M Na₂MoO₄, 0.4 μ M H₃BO₃, and 10 μ M FeCl₃ in ultra-pure water. LB, LB agar and SOC media were sterilised by autoclave at 121 °C and supplemented with appropriate antibiotics. For ZYM-5052 autoinduction medium, the appropriate weights of powdered tryptone and yeast extract were dissolved in ultra-pure water and autoclaved. All other constituents, including the appropriate antibiotics, were filter-sterilised with a 0.22 μ m filter unit (Thermo Fisher Scientific) and added to the tryptone and yeast extract mixture.

2.2 Molecular Biology

2.2.1 DNA purification

2.2.1.1 Principle

Briefly, DNA is adsorbed to a silica membrane under high salt concentration, contaminants are washed and DNA is eluted with a low salt buffer as indicated in the manufacturers protocol. All buffers and DNA-binding columns were purchased from Qiagen (West Sussex, UK). The wash buffer (10 mM Tris-HCl, pH 8.5, 80 % (v/v) ethanol) and elution buffer (10 mM Tris-HCl, pH 8.5) are common to all of the below approaches.

2.2.1.2 Recovery from agarose gel

A QIAquick[®] Gel Extraction Kit was used to isolate a specific sized DNA bands after agarose gel electrophoresis. DNA bands were visualized with a UV-transilluminator and excised with a sterile scalpel. The agarose matrix was solubilized in guanidine thiocyanate at 50 °C, releasing the DNA fragment for purification.

2.2.1.3 Recovery from enzymatic reactions

DNA was purified following PCR using a QIAquick[®] PCR Purification Kit. For other enzymatic reactions and when DNA needed to be concentrated, such as preceding a bacterial transformation, a QIAquick[®] MinElute Kit was used. Reaction mixtures were first treated with guanidine hydrochloride and isopropanol to denature enzymes in the reactions, before DNA was bound to the membrane.

2.2.1.4 Plasmid DNA Extraction

Plasmid DNA was isolated from saturated bacterial liquid cultures using QIAprep[®] Miniprep Kit for high-copy number plasmids from 5 – 10 mL cultures or Qiagen MidiPrep kit for low-copy number plasmids from 50 mL cultures. Cells were harvested by centrifugation at $15,400 \times g$ for 4 min. Bacterial cells were lysed under alkaline conditions (including SDS). Lysates were neutralized with 3 M sodium acetate before DNA was bound to the membrane.

2.2.2 Polymerase Chain Reaction (PCR)

Phusion[®] HF polymerase (Finnzymes, Leicestershire, UK) was used to amplify specific genes and for site directed mutagenesis. The reaction mixture is shown in Table 2.2. PCRs were carried out in a Flexigene thermocycler (Techne). The reaction thermocycle used for each polymerase is shown in Table 2.3. The annealing temperature was dictated by the oligonucleotide primer pair used (Table 2.4). When primer annealing temperature was the same as the extension temperature (72 °C), no separate annealing step was used. The extension time was adjusted for the length of fragment to be amplified. Reaction cycles were kept to a minimum (between 25 – 35 cycles).

Table 2.2 - Composition of PCR mixtures used with Phusion[®] polymerase

Component	Concentration
Reaction Buffer ^a	1 ×
dNTPs (each base)	0.2 mM
Forward Primer	0.5 μM
Reverse Primer	0.5 μM
DNA Polymerase	1 U
MgCl ₂	1.5 mM
Template DNA	20 pg – 20 ng
Total volume ^b	50 μL

^a Final concentrations of reaction buffers are as follows; HF Phusion buffer; 10 mM Tris-HCl, 0.1 % (v/v) Triton[®]X-100, pH 8.8.

^b The total reaction volume was made up with ultra pure water

Table 2.3 - PCR cycle conditions

Initial	×25 – 35 reaction cycles			Final
Denaturation	Denaturation	Annealing	Extension	Extension
(98°C)	(98°C)		(72 °C)	(72 °C)
30 sec	8 sec	20 sec	20 sec/kb	5 min

2.2.3 Oligonucleotide primers and DNA sequence analysis

The list of oligonucleotides used in this work is shown in Table 2.4. A suite of sequence, plasmid and PCR analysis programs were used to analyse and manipulate DNA sequences and design oligonucleotide primers. Annealing temperature of primers was determined using the Finnzyme T_m Calculator (www.thermoscientificbio.com/webtools/tmc/) according to modified Breslauer's thermodynamics. OligoAnalyzer (www.idtdna.com/analyzer/applications/) was used to analyse primers for formation of hairpin loops, homo- and hetero-dimers and to confirm annealing temperature. Primers were synthesised by Eurofins MWG Biotech. Amplify 3X[®] v3.1.4 (Bill Engels, University of Wisconsin 2004/5) was used to perform virtual PCRs to check primer specificity. SerialCloner 2-1[®] (F. Perez, SerialBasics 2004-2010) was used for general sequence manipulation including virtual restriction digests, ligations and sequence alignments.

2.2.4 Agarose gel electrophoresis

The analysis and separation of DNA fragments was performed by agarose gel electrophoresis. A range of agarose concentrations, from 0.6 – 1 % (w/v) were used depending on the desired size resolution. Agarose powder (Melford) was dissolved in TAE buffer (40 mM Tris-acetate, 1 mM EDTA; pH 8.8) (Invitrogen, Paisley, UK) to the appropriate concentration and supplemented with 0.3 µg/mL ethidium bromide. Samples were mixed with 6 × loading dye (Final concentrations; 3.3 mM Tris-HCl, 2.5 % (w/v) Ficoll 400, 11 mM EDTA, 0.017 % SDS (w/v), 0.015 % (w/v) Bromophenol Blue; pH 8.0) (NEB) and run against 5 µL of either 100 bp or 1 kbp DNA marker ladders (NEB). Electrophoresis was performed at 90 V until sufficient resolution was achieved as determined visually from the marker ladder. DNA bands were visualised with UV light in a GelDoc-It UV-Transilluminator (Ultra-Violet Products Ltd, Cambridge, UK).

2.2.5 DNA quantification

DNA concentration and purity was determined using absorbance at 230 nm, 260 nm and 280 nm using a NanoDrop[®] ND1000 Spectrophotometer (Thermo Fisher Scientific). The spectrophotometer was blanked against the appropriate buffer and DNA concentration was calculated using the absorbance at 260 nm according to the Beer-Lambert law. Contamination with organic compounds and protein were measured by the absorbance at 230 and 280 nm, respectively. As an alternative to spectrophotometric analysis, 1 µL and 5 µL DNA samples were quantified by agarose gel electrophoresis by comparison to DNA band intensities of known concentration in the DNA marker ladder (NEB) on a UV transilluminator.

2.2.6 Restriction enzyme digestion

Restriction digests were performed with *NcoI*-HF (1 U/µg DNA) and *KpnI*-HF (1 U/µg DNA) (NEB) in 50 µL reactions. Digests were performed in NEB Buffer 4 (50 mM K-acetate, 50 mM Tris-acetate, 10 mM Mg-acetate, 1 mM DTT; pH 7.9). The restriction enzyme was added to the reaction and incubated at 37 °C for 30 min. Restriction enzymes were not heat deactivated as DNA was immediately separated by agarose gel electrophoresis (Section 2.2.4).

2.2.7 DNA phosphorylation

The 5' ends of DNA fragments were phosphorylated using T4 Polynucleotide Kinase (NEB) in 1 × Quick Ligase Buffer (66 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT, 1 mM ATP, 15 % (w/v) PEG 6000; pH 7.6) (NEB) as this provided ATP and allowed subsequent ligation without the need for a buffer change. T4 Polynucleotide Kinase (10 U) was added to DNA and the reaction was incubated at 37 °C for 30 min.

2.2.8 DNA Ligation

Ligation of linear DNA fragments was performed with Quick T4 DNA Ligase (NEB). Ligation reactions were performed in 1 × Quick Ligase Buffer (Section 2.2.7) (NEB) in a total volume of 20 µL at 25 °C for 25 min. Intramolecular ligation (recirculation) was performed with *circa* 100 ng linear plasmid DNA alone. Control reactions were performed that contained no ligase to determine the level of false positives due to the presence of the original parent plasmid. Intermolecular ligations to insert DNA into digested plasmids were performed with 50 ng of vector DNA and 3 molar equivalents of insert DNA. Control reactions were performed that contained no insert DNA to determine the level of false positive due to vector recirculation.

2.2.9 Bacterial transformation

Electro-competent *E. coli* cell aliquots (40 µL), stored at -80 °C, were thawed on ice. Plasmid DNA (20 – 100 ng) was added to an aliquot, gently mixed and transferred to an electroporation cuvette. The cells were subjected to a 4.5 – 5 ms electrical pulse at 12.5 kV/cm field strength using a gene pulser (Bio-Rad laboratories Ltd., UK) with capacitance of 25 µF and resistance of 200 Ω. Transformed cells were recovered by the addition of 360 µL SOC medium and incubation at 37 °C with mixing (200 rpm) for 1 hr. Transformed bacteria were plated on LB agar containing the appropriate selective antibiotic/s and incubated at 37 °C overnight.

2.2.10 DNA sequencing

Sequencing was performed by the Cardiff University Sequencing Core using ABI PRISM[®] BigDye v3.1 Terminator Sequencing technology on the ABI PRISM[®] 3100 DNA Sequencer.

2.3 Cloning cytochrome *b*₅₆₂ into pBAD

The *cybC* gene encoding cytochrome *b*₅₆₂ (cyt *b*₅₆₂) was cloned into the pBAD plasmid by restriction endonuclease digestion followed by ligation. First, *NcoI* and *KpnI* restriction sites were added to the 5' and 3' termini of cyt *b*₅₆₂, respectively, by site directed mutagenesis PCR (Section 2.2.2) using primers SRcNCOF and SRcKPNR (Table 2.4) and pET-cyt *b*₅₆₂ as template. An annealing temperature range of 67 – 70 °C was used. The amplified DNA fragment was then separated by agarose gel electrophoresis (Section 2.2.4) and purified using a QIAquick Gel Extraction Kit (Section 2.2.1.2). Both the purified *cybC* gene (insert) and the pBAD plasmid (vector) were digested by the restriction enzymes *NcoI* and *KpnI* (Section 2.2.6). The pBAD plasmid backbone was separated from the prior fragment present between the *NcoI* and *KpnI* sites within the vector following restriction digestion by agarose gel electrophoresis (Section 2.2.4) followed by QIAquick Gel Extraction Kit cleanup (Section

2.2.1.2). Insert DNA was combined with vector DNA in a 3:1 molar ratio for intermolecular ligation (Section 2.2.8). A MinElute reaction cleanup (Section 2.2.1.3) was used to remove Quick T4 DNA Ligase enzyme and salt from the previous reactions, before *E. coli* DH5 α cells were transformed with the products of the ligation reaction. Transformants were plated on LB agar containing ampicillin to select for colonies containing pBAD. Positive colonies were propagated and pBAD-cyt *b*₅₆₂ derived plasmids sequenced.

2.4 Generation of amber stop codon mutants

Amber stop codon (TAG) mutations were generated by whole plasmid site-directed mutagenesis PCR (Section 2.2.2) using the pBAD plasmid containing the WT gene of interest as a template. pBAD-sfGFP was a kind gift from Ryan Mehl and the gene encoding WT cyt *b*₅₆₂ was cloned into the pBAD plasmid as described in Section 2.3. Oligonucleotide primers were designed with the TAG mutation at the 5' end of the forward primer (underlined in Table 2.4) and have T_m of 72 °C allowing for 2-step PCR. Amplified linear plasmid DNA following PCR was separated by agarose gel electrophoresis (Section 2.2.4) and purified using a QIAquick Gel Extraction Kit (Section 2.2.1.2). Linear DNA was phosphorylated (Section 2.2.7) and recircularised by intramolecular ligation (Section 2.2.8). Reactions were cleaned up using a MinElute Reaction Cleanup Kit (Section 2.2.1.3) and used to transform *E. coli* DH5 α cells. Transformed cells were grown on LB agar containing ampicillin to select for colonies containing circular pBAD plasmid. Positive colonies were picked and grown in LB broth containing ampicillin before plasmid DNA was purified (Section 2.2.1.4) and submitted for sequencing to confirm the presence of the desired mutation with primers SRpBADsQF and SRpBADsQR (Table 2.4).

Table 2.4 – Sequence and use of oligonucleotide primers used for PCR. Restriction sites and TAG mutations introduced by the primers are underlined.

Primer Name	Sequence (5' > 3')	Mutation
sfGFP amber stop codon mutations of surface-exposed residues		
SR26TAGF	<u>TAG</u> TTT AGC GTT CGT GGC GAA GGC	Amber stop codon
SR26TAGR	ATG <u>GCC</u> ATT CAC ATC ACC ATC CAG TTC C	at codon 26
SR34TAGF	<u>TAG</u> GGT GAT GCG ACC AAC GGT AAA CTG A	Amber stop codon
SR34TAGR	GCC TTC GCC ACG AAC GCT AA	at codon 34
SR132TAGF	<u>TAG</u> GAT GGC AAC ATT CTG GGT CAT AAA CTG G	Amber stop codon
SR132TAGR	TTT AAA ATC AAT ACC TTT CAG TTC AAT GCG GTT C	at codon 132
SR204TAGF	<u>TAG</u> AGC GTT CTG AGC AAA GAT CCG AAT G	Amber stop codon
SR204TAGR	GGT GCT CAG ATA ATG ATT ATC CGG CAG CA	at codon 204
sfGFP amber stop codon mutations of internal residues		
SR44TAGF	<u>TAG</u> AAA TTT ATT TGC ACC ACC GGT AAA CTG C	Amber stop codon
SR44TAGR	GGT CAG TTT ACC GTT GGT CGC ATC	at codon 44
SR62TAGF	<u>TAG</u> ACC CTG ACC TAT GGC GTT CAG TGC T	Amber stop codon
SR62TAGR	CAC CAG GGT CGG CCA CG	at codon 62
SR66TAGF	GGG CGT TCA GTG CTT TAG CCG C	Amber stop codon
SR66TAGR	TAG GTC AGG GTG GTC ACC AGG GTC	at codon 66
SR75TAGF	<u>TAG</u> GAT CAT ATG AAA CGC CAT GAT TTC	Amber stop codon
SR75TAGR	ATA GCG GCT AAA GCA CTG AAC GCC	at codon 75
SR119TAGF	<u>TAG</u> GTG AAC CG ATT GAA CTG AAA GG	Amber stop codon
SR119TAGR	GGT ATC GCC TTC AAA TTT AAC TTC CGC	at codon 119
SR145TAGF	<u>TAG</u> AAC AGC CAT AAT GTG TAT ATT ACC GCC G	Amber stop codon
SR145TAGR	ATT ATA TTC CAG TTT ATG ACC CAG AAT GTT GCC	at codon 145
SR148TAGF	<u>TAG</u> AAT GTG TAT ATT ACC GCC GAT AAA CAG AAA CAG AAA AAT GG	Amber stop codon
SR148TAGR	GCT GTT GAA ATT ATA TTC CAG TTT ATG ACC CAG AAT G	at codon 148
SR203TAGF	<u>TAG</u> CAG AGC GTT CTG AGC AAA GAT CCG	Amber stop codon
SR203TAGR	GCT CAG ATA ATG ATT ATC CGG CAG CA	at codon 203
SR222TAGF	<u>TAG</u> TTT GTT ACC GCC GCG GGC	Amber stop codon
SR222TAGR	CAG CAG CAC CAT ATG ATC ACG TTT TTC A	at codon 222
cyt <i>b</i>₅₆₂ amber stop codon and cysteine mutations		
SRc5TAGF	ACC CTC AAC GAC AAT TTA AAA GTG ATC G	Amber stop codon
SRc5TAGR	TTC CAT ATT CTA TTC AAG ATC TGC GG	at codon 5
SRc21TAGF	GCG CAA GTC AAA GAC GCG TTA AC	Amber stop codon
SRc21TAGR	CGC GTT CTA CGC TTT TTC GAT CAC	at codon 21
SRc50TAGF	ACA GCC CGG AAA TGA AAG ATT TC	Amber stop codon
SRc50TAGR	CCG GTG ATT TCT ATT CGA GCT TCG	at codon 50
SRc104TAGF	CCT ATC ACC CGT AGT ATC GTT AAC TCG AGC	Amber stop codon
SRc104TAGR	CGT TGC GGG TCG TTT TCA GTT G	at codon 104
SRc5CYSF	ACC CTC AAC GAC AAT TTA AAA GTG ATC G	Cysteine codon at
SRc5CYSR	TTC CAT ATT ACA TTC AAG ATC T	codon 5
SRc21CYSF ^a	GTG ATC GAA AAA GCG TGT AAC GCG GCG CAA GTC	Cysteine codon at
SRc21CYSR ^a	CTT TGG GAG TTG CTG TTA AAT TTT	codon 21
SRc50CYSF ^a	CCG CCG AAG CTC GAA <u>TGT</u> AAA TCA CCG GAC AGC CCG	Cysteine codon at
SRc50CYSR ^a	CGA CCT ACG CGT TTT TCG TTG C	codon 50
SRc104CYSF	CCT ATC ACC AGT GTT ATC GTT	Cysteine codon at
SRc104CYSR	CGT TGC GGG TCG TTT TCA GTT G	codon 104
Anomalous		
SRcNCOF	TTA ACC ATG GCA GAT CTT GAA GAC AAT ATG GAA ACC	Add <i>Nco</i> I to cyt <i>b</i> ₅₆₂
SRcKPNR	TTG <u>GGT</u> ACC TTA ACG ATA CTT CTG GTG ATA GGC GTT G	Add <i>Kpn</i> I to cyt <i>b</i> ₅₆₂
SRpBADSqF	ATG CCA TAG CAT TTT TAT CC	For sequencing
SRpBADSqR	GAT TTA ATC TGT ATC AGG	inserts in pBAD

^a Oligonucleotide primers were designed by E. Della Pia

2.5 Protein production and purification

2.5.1 Production of proteins containing Uaa's

Two plasmids were required for incorporation of Uaa's into proteins (Table 2.5); pDULE ((136); Tetracycline^R; low copy number) which carries the evolved tRNA_{CUA}/ aaRS pair specific for the desired Uaa and pBAD (Invitrogen; Ampicillin^R; low copy number) which carries the gene of interest with the desired TAG mutation. Both plasmids were added to *E. coli* TOP10 cell aliquots in equal concentrations and cells were transformed as in Section 2.2.9.

Table 2.5. Plasmids used for Uaa incorporation in *E. coli*

Name	Insert 1	Insert 2	Antibiotic resistance
pBAD-sfGFP	sfGFP gene	-	Ampicillin
pBAD-cyt _{b562}	cyt <i>b</i> ₅₆₂ gene	-	Ampicillin
pDULE-cyPheRS	<i>p</i> -cyano-L-phenylalanine aaRS	tRNA _{CUA}	Tetracycline
pDULE-tfmPheRS	<i>p</i> -trifluoromethyl-L-phenylalanine aaRS	tRNA _{CUA}	Tetracycline
pDULE-napAlaRS	naphthyl-alanine aaRS	tRNA _{CUA}	Tetracycline
pDULE-amPheRS	<i>p</i> -amino-L-phenylalanine aaRS	tRNA _{CUA}	Tetracycline

Starter cultures of 5 mL LB broth (containing 25 µg/mL tetracycline and 100 µg/mL ampicillin) were inoculated with a single bacterial colony and incubated with shaking (200 rpm) at 37 °C overnight. Larger expression cultures (10 mL to 1 L) of ZYM-5052 autoinduction medium (Section 2.1.3) were inoculated with a 1/200 dilution from the saturated culture. An hour after inoculation, 1 mM Uaa was added, with the exception of *p*-amino-L-phenylalanine where 10 mM was used, and cultures were incubated at 37 °C for 24-30 hours. Protein production was detected by SDS-PAGE (Section 2.6.1) and fluorescence spectroscopy for sfGFP (Section 2.6.3). When incorporating *p*-azido-L-phenylalanine, cultures were grown in the dark to prevent photolysis of the Uaa by ambient light.

2.5.2 Cell lysis

Three methods, detailed below, were used to lyse cells depending on the volume of expression culture. Following all lysis methods, insoluble material and cell debris was separated by centrifugation (20,000 × g, 30 min, 4 °C). Soluble material (supernatant fraction) was decanted and insoluble material (pellet) was resuspended in the same volume as soluble fractions when required for analysis by SDS-PAGE (Section 2.6.1).

2.5.2.1 French Press

French Press was used for resuspended cell volumes over 5 mL and hence culture volumes over 250 mL. Cells were harvested by centrifugation at $4000 \times g$ for 15 min at 4 °C. Cell pellets were resuspended with a 1 in 50 dilution of original culture size in phosphate buffered saline (PBS; 100 mM sodium phosphate, 300 mM NaCl, pH 8.0) containing 0.1 mg/mL lysozyme (Sigma-Aldrich), 1 mM PMSF (Merck Millipore, MA, US) and ~25 U Benzonase nuclease (Sigma-Aldrich). Cell suspensions were subjected to two rounds of pressure lysis (1010 psig) in a French pressure cell press.

2.5.2.2 Sonication

Sonication was used for resuspended cell volumes of less than 5 mL. Cell pellets were resuspended as with French Press (Section 2.5.2.1) and subjected to 6 to 8 cycles of 10 seconds of high frequency sound waves (20 kHz, 40% amplitude) followed by 20 seconds cooling time on ice.

2.5.2.3 Chemical lysis

BugBuster[®] lysis was used for batch lysis of small culture volumes. Cell pellets were resuspended in BugBuster[®] Protein Extraction Reagent (Novagen) containing 0.2 mg/mL lysozyme and ~25 U Benzonase Nuclease to the required optical density. BugBuster[®] suspensions were incubated at room temperature for 30 min. To ensure full lysis, cell suspensions were subjected to 2 freeze-thaw cycles between -80 °C and room temperature.

2.5.3 Protein purification

All protein purification columns were purchased from GE Healthcare and used with an ÄKTApurifier FPLC. Elution from columns was monitored by absorption at 280 nm and either 485 nm (sfGFP) or 420 nm (cyt *b*₅₆₂) and analysed by SDS-PAGE. Centrifugal concentrator units (Merck Millipore) were used for concentrating samples and for buffer exchange of protein solutions between different purification techniques. Buffer exchange was performed by successive addition of desired buffer and centrifugation at $2,800 \times g$. 10 kDa and 5 kDa MWCO filters were used for sfGFP and cyt *b*₅₆₂ variants, respectively.

2.5.3.1 Nickel affinity chromatography

Hexahistidine-tagged sfGFP variants were purified using a HisTrapTM FF column. Clarified cell lysates were loaded onto the column in PBS (100 mM sodium phosphate, 300 mM NaCl, pH 8.0) with a flow rate of 1 mL/min. The column was washed in PBS with a concentration gradient of 0 to 250 mM imidazole over 20 column volumes (3 mL/min; 40 min).

His SpinTrapTM columns were used for small-scale Ni-affinity purification from small culture volumes and for separation of Click reactions (Section 2.12.3). Purification was performed in a microcentrifuge with each step requiring centrifugation at $100 \times g$ for 30 sec. Columns were first washed with 600 μ L PBS (100 mM sodium phosphate, 300 mM NaCl, pH 8.0). Protein samples were loaded onto the column in PBS containing 15 mM imidazole to reduce non-specific binding. After loading, samples were washed with 3 mL (30 column volumes) PBS containing 40 mM imidazole, before elution with PBS containing 500 mM imidazole.

2.5.3.2 Ammonium sulphate precipitation

For cyt *b*₅₆₂ variants, crude protein samples following lysis were subjected to ammonium sulphate precipitation to concentrate and partially purify samples before gel filtration chromatography. Ammonium sulphate was added to 30 % saturation at 4 °C with constant stirring. Samples were centrifuged at $20,000 \times g$ for 30 min and insoluble proteins (pellet) were removed. Insoluble proteins were resuspended to the same volume as the supernatant for analysis by SDS-PAGE. Then, ammonium sulphate was added to 90 % saturation at 4 °C with constant stirring. Samples were centrifuged at $40,000 \times g$ for 30 min and soluble proteins (supernatant) were removed. Insoluble proteins (pellet) that included cyt *b*₅₆₂ variants were resuspended in PBS (100 mM sodium phosphate, 300 mM NaCl, pH 8.0). Samples were buffer exchanged to remove ammonium sulphate and to concentrate samples.

2.5.3.3 Gel filtration chromatography: preparative

Protein samples were concentrated and loaded onto the column in a 1 mL sample loop to maximise resolution of separation. Samples were applied to a HiLoadTM SuperdexTM 75 pg (preparative grade) gel filtration column (120 mL bed volume) in PBS (100 mM sodium phosphate, 300 mM NaCl, pH 8.0). Samples were run through the column at 0.5 mL/min.

2.5.3.4 Ion exchange chromatography

Proteins were buffer exchanged into Tris-HCl buffer (50 mM, pH 8.5) and applied to a Mono QTM anion exchange column. After washing the column with Tris-HCl buffer (50 mM, pH 8.5) proteins were eluted in Tris-HCl (50 mM, pH 8.5) with a gradient of 0 – 250 mM NaCl over 10 or 20 column volumes.

2.5.4 Purification of sfGFP variants

sfGFP variants were first purified by Ni-affinity chromatography (Section 2.5.3.1) using either a HiTrapTM FF column following large expression cultures (0.1 – 2 L) or HisTrapTM spin columns for smaller cultures (10 – 50 mL). Proteins were generally judged to be >99 % pure following Ni-affinity purification however when further purification was required (for calculating extinction coefficients, quantum yields or crystallisation) proteins were purified by ion exchange chromatography with a Mono QTM column (Section 2.5.3.4).

2.5.5 Purification of cyt *b*₅₆₂ variants

Cyt *b*₅₆₂ variants were first subjected to ammonium sulphate precipitation (Section 2.5.3.2). Samples were then purified by gel filtration chromatography (Section 2.5.3.3). Finally, cyt *b*₅₆₂ variants were buffer exchanged into Tris-HCl buffer (50 mM, pH 8.5) and separated by ion exchange chromatography with a MonoQTM column (Section 2.5.3.4). Proteins were buffer exchanged into PB (100 mM sodium phosphate, pH 8.0), checked for purity by SDS-PAGE and quantified.

2.5.6 Haem titration of cyt *b*₅₆₂ variants

Haem (Sigma-Aldrich) was prepared as a 10 mM solution in 0.5 M NaOH (Sigma-Aldrich) immediately before use. Potassium nitrate (BDH AnalaR, Poole, UK) and ascorbic acid (Sigma-Aldrich) were prepared as 100 mM stock solutions. To ensure cyt *b*₅₆₂ variants were fully saturated with haem (holo-cyt *b*₅₆₂), proteins were incubated with 5 μM haem under reducing conditions (1 mM ascorbic acid) for 48 hrs at 4°C. Free haem was removed by buffer exchange against PB (100 mM sodium phosphate, pH 8).

2.6 Protein analysis

2.6.1 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed as described by Laemmli (123) using the mini-PROTEAN 3 electrophoresis system (Bio-Rad). The stacking and separating gels were composed as shown in Table 2.6. The

concentration of acrylamide varied depending on the molecular weight of the protein to be resolved. Samples were mixed with sample loading buffer (final concentrations; 2 % (w/v) SDS, 0.2 M Tris-HCl pH 6.8, 0.04 % (w/v) bromophenol blue, 8 % (w/v) glycerol, 10 % (v/v) β -mercaptoethanol) and heated to 95 °C for 10 min. When analysing *E. coli* whole cell samples, cells were pelleted by centrifugation ($15,400 \times g$, 10 min) and resuspended in sample loading buffer to an OD₆₀₀ of 10. Samples were run along side broad range (7 to 175 kDa) prestained protein marker (NEB). Gels were stained in 40 % (v/v) methanol, 10 % (v/v) acetic acid, 0.1 % (w/v) R250 Coomassie blue and destained in 40 % (v/v) methanol, 10 % (v/v) acetic acid.

Table 2.6. Constituents of SDS-PAGE separating and stacking gels.

	Separating	Stacking
Acrylamide/bis-Acrylamide ^a	12.5 % (sfGFP) / 15 % (cyt <i>b</i> ₅₆₂) (w/v)	5 % (w/v)
Tris-HCl ^b	0.375 mM	65 mM
SDS	0.1 % (w/v)	0.2 % (w/v)
APS	0.05 % (w/v)	0.1 % (w/v)
TEMED	0.02 % (w/v)	0.02 % (w/v)

^a Acrylamide and N,N'-methylene bis acrylamide in a 37.5:1 (w/w) ratio.

^b pH 8.8 and pH 6.8 solutions were used for separating and stacking gels, respectively.

2.6.2 Protein quantification

Protein concentrations were calculated using the microplate protocol of the Bio-Rad D_C protein assay as per the manufactures instructions. In brief, 5 μ L samples and a protein standard were pipetted into a flat-bottomed 96 well plate (Nunc, Thermo-Fisher) in triplicate. For colour development, 25 μ L of reagent A (alkaline copper tartrate solution) followed by 200 μ L of reagent B (Folin reagent) was added to samples. Samples were incubated at room temperature for 15 - 60 min and absorbencies were measured at 750 nm using a FLUOstar (Omega) spectrophotometer. Protein concentrations of samples were calculated by interpolation of the standard curve produced from the standard protein.

2.6.3 Fluorescence spectroscopy

Excitation and emission spectra were recorded using a Varian Cary Eclipse spectrophotometer and the corresponding Cary Eclipse software. Samples of 750 μ L were analysed in a 5 \times 5 mm QS quartz cuvette (Hellma, Müllheim, Germany). Spectra were recorded at a scan rate of either 120 or 600 nm/min with a slit width of 5 nm. Emission spectra were recorded up to 650 nm from a fixed excitation wavelength at the particular variants

maximum (λ_{ex}). Excitation spectra were recorded by monitoring emission at a set wavelength (λ_{em}) over a range of wavelengths down to 350 nm. For whole cell fluorescence, cultures were centrifuged ($1,500 \times g$, 10 min) and cells were resuspended in TNG buffer (50 mM Tris-HCl, 150 mM NaCl, 10 % (v/v) glycerol, pH 8) to an OD_{600} of between 0.1 and 0.5. For fluorescence measurements in cell lysates, soluble material following lysis was diluted in PBS (100 mM sodium phosphate, 300 mM NaCl, pH 8) to an equivalent OD_{600} of between 0.1 and 1. For more detailed analysis, spectra of pure samples of sfGFP and sfGFP variants were measured at 1 μM in PBS (with the exception of sfGFP^{Y66AzF} which was 10 μM).

2.6.4 UV-visible absorption spectroscopy and calculation of extinction coefficients

UV-visible (UV-vis) absorption spectra were recorded using a JASCO V-660 spectrophotometer in 1 cm pathlength QS quartz cuvettes (Hellma) except for photoswitching data which was recorded in 5×5 mm QS quartz cuvette (Section 2.6.3) Extinction coefficients were calculated by diluting proteins to 12.5 μM (sfGFP) or 5 μM (cyt b_{562}) and recording full absorption spectra from 250 – 650 nm. Extinction coefficients were calculated for cyt b_{562} variants under oxidising and reducing conditions by supplementing the buffer with 1 mM potassium nitrate and 1 mM ascorbic acid, respectively. Proteins were subsequently re-quantified (Section 2.6.2) to confirm the concentration of the diluted sample. The absorbance value at λ_{max} and concentration values were substituted into the Beer-Lambert law (Equation 2.1) to calculate the extinction coefficient. Here, A is the absorbance value, ε is the extinction coefficient ($\text{M}^{-1}\text{cm}^{-1}$), c is concentration (M) and l is pathlength (cm).

$$A = \varepsilon \cdot c \cdot l \quad \text{Equation 2.1}$$

2.6.5 Quantum yield determination

Quantum yields were determined for sfGFP variants by comparison to the reference fluorophore fluorescein with known quantum yield (0.95) and similar λ_{max} (496 nm). sfGFP variants in PBS and fluorescein (in 0.1 M NaOH) were diluted to an A_{485} of 0.05. Emission spectra were recorded after excitation at the λ_{max} of the variant (as in 2.6.3) and the integrated emission spectra of each was calculated between the excitation wavelength and 650 nm and substituted into Equation 2.2.

$$\phi_x = \phi_R \cdot \left(\frac{\text{Int}_x}{\text{Int}_R} \right) \cdot \left(\frac{\eta_x^2}{\eta_R^2} \right) \quad \text{Equation 2.2}$$

Here, Φ is the quantum yield, Int is the integrated emission spectrum and η is the refractive index of the solvent used. The subscripts x and R correspond to the sample being measured and the reference fluorophore, respectively.

2.7 Photolysis of wild-type sfGFP and sfGFP^{AzF} variants

Photolysis of sfGFP variants was performed in fluorimetry cuvettes to allow increased transmission of UV wavelength light. Samples were irradiated for the indicated periods of time with a 302 nm (spectrum from 275 – 380 nm) UVM-57 mid-range UV lamp (6 W; UVP, Cambridge, UK) at a distance of 1 cm and spectra were recorded immediately after. Fluorescence spectra were recorded and therefore photolysis was performed on 1 μ M protein with the exception of sfGFP^{Y66AzF} that was recorded at 10 μ M. Photolysis for UV-visible absorbance spectra were recorded with 10 μ M protein.

2.8 Microscopy

Photoconversion of sfGFP^{Y66AzF} and sfGFP^{F145AzF} variants was visualised in live *E. coli* TOP10 cells using Widefield fluorescence microscopy and Laser Scanning Confocal Microscopy (LSCM). Protein variants were produced in 10 mL cell cultures as detailed in section 2.5.1. After 24 hours, cells were pelleted by centrifugation, culture media was discarded and cells were mounted on slides in 2 % low melting temperature agarose (Invitrogen). Widefield fluorescence images were taken using an Olympus IX-71 with automated stage, shutters and filter wheels (Prior) coupled to a Hamamatsu Orca-AG camera. sfGFP^{Y66AzF} was simultaneously activated and imaged using a 430/24nm excitation filter (Chroma 69308) for the indicated times. sfGFP^{F145AzF} was deactivated and imaging using a 490/20nm excitation filter (Chroma 69302). LSCM was performed using a TCS SP2 AOBS confocal microscope (Leica) fitted with a 20 mW 405nm diode and 458, 476, 488 and 514 nm 100 mW argon ion laser lines, using a 100 \times planApochromat NA 1.4 oil immersion objective. Photoconversion experiments were performed using Leica's FRAP module using flymode allowing line by line visualisation throughout the photoconversion process. Bacteria were photoconverted using the 405 nm laser line and imaged using the 488 nm (sfGFP^{Y66AzF}) and 514 nm (sfGFP^{F145AzF}) lines. Regions of interest containing cells were selected for irradiation with the 405 nm laser using the FRAP module. Full imaging details can be found in Chapter 4.

2.9 Mass spectrometry

Trypsin digest MALDI-TOF (matrix assisted laser desorption/ionisation-time of flight) mass spectrometry was performed to elucidate the photochemical endpoint of sfGFP^{AzF}

irradiation. MALDI-TOF was performed by Central Biotechnology Services (Cardiff University) but a brief description follows. Proteins were excised from SDS-PAGE gels and destained with acetonitrile (CH₃CN) and ammonium bicarbonate (NH₄HCO₃). Samples were dehydrated using CH₃CN and dried in an oven before reduction with 10 mM DTT at 56 °C. Samples were then alkylated with 55 mM iodoacetamide, washed (NH₄HCO₃) and dehydrated again with CH₃CN and heating. Proteins were digested with sequencing grade modified trypsin (Promega) in NH₄HCO₃. Peptides were spotted on 384 well MALDI target plates, mixed with matrix (α -cyano-4-hydroxycinnamic acid) and analysed by a 4800 MALDI TOF/TOF spectrometer (Applied Biosystems).

2.10 X-ray crystallography and structural determination

Full details of the crystallography process and determination of structures are given in Chapter 5. The dark states of sfGFP^{Y66AzF} and sfGFP^{Y145AzF} were concentrated in 50 mM Tris-HCl, 150 mM NaCl, pH 8 (at 25 °C), to 28 and 19 mg/mL, respectively, and used to set up vapour diffusion crystal trays. Initially, in solution irradiated sfGFP^{Y66AzF} was screened for crystal growth by JBScreen Membrane (Jena Bioscience, Germany), K&J and Additive Screen MD1-11 (Molecular Dimensions, Suffolk, UK). Small spiny crystals of sfGFP^{Y66AzF} grew in K&J screen condition A4 (100 mM Tris pH 8.3, 3 M (NH₄)₂SO₄). Screens were then set up around this condition by varying the concentration of (NH₄)₂SO₄ (2 – 3.5 M) and pH (7.9 – 8.9). Protein and buffer were mixed in equal volume in all screens but with varying volumes (0.2, 0.5 or 1 μ L) and equilibrated against 60, 85 or 150 μ L of reservoir buffer at 4 or 18 °C. Also, further protein was produced, concentrated to 20 and 10 mg/mL and used to set up crystal trays. The conditions that produced the best diffracting crystals for each variant can be seen in Chapter 5. At all stages of protein production, purification and crystallization, light exposure was kept to minimum. After harvesting crystals of the dark state, crystal trays of sfGFP^{Y66AzF} and sfGFP^{F145AzF} were irradiated in a Luzchem LZC4 photoreactor (3 \times 8 W Hitachi FL8BL-B (λ_{\max} 350 nm) lamps, 3 \times 8 W LZC-UVB (λ_{\max} 310 nm) lamps) for 30 mins and with a UVM-57 mid-range UV handheld lamp (UVP; 6 W, 1 cm, top down) for 90 min, respectively.

Crystals were transferred to mother liquor supplemented with 13% (w/v) PEG 200 as a cryoprotectant and vitrified in liquid nitrogen. Data were collected at the Diamond Light Source, Harwell, UK. Data were reduced with the XIA2 package (124), space group assignment was done by POINTLESS (125), scaling and merging were completed with SCALA (125) and TRUNCATE (126). Structures were solved by molecular replacement with the previously determined sfGFP structure (PDB accession 2B3P) as a model using PHASER (127). Structures were adjusted manually using COOT (128) and refined by TLS restrained

refinement (sfGFP^{Y66AzF} dark and irradiated) or mixed anisotropic/isotropic restrained refinement (sfGFP^{Y145AzF} irradiated) using RefMac. All non-protein atoms were refined isotropically. The above routines were used as the CCP4 package (126) (www.ccp4.ac.uk). Graphical representations were made with PyMOL Molecular Graphics System, Schrödinger, LLC. Final crystal diffraction and refinement statistics can be found in Chapter 5.

2.11 Electron paramagnetic resonance (EPR) spectroscopy

Purified proteins were snap frozen in liquid nitrogen at 77 K in PBS containing 50% glycerol. Photolysis was performed in quartz–suprasil 4 mm EPR tubes with a 500 W Oriel Instruments UV lamp, incorporating a Hg/Xe arc lamp (250 nm to >2500 nm) with a water filter. The UV output below 280 nm accounted for only 4 – 5% of the total lamp output. Samples were irradiated for 2 min at 77 K in an EPR liquid N₂ finger dewar (Wilmad Glass, NJ, US) and rapidly transferred to the pre-cooled cryostat (10 K) for EPR measurements. The CW-EPR spectra were recorded on a Bruker ESP300e series spectrometer operating at X-band frequencies (9.75 GHz), 100 kHz field modulation, 10 G modulation depth at 10 K in an Oxford liquid helium cryostat.

EPR simulations were performed using Easyspin software (129) and Bruker SIMFONIA software (commercially available) by D. Murphy. Density functional theory (DFT) calculations were performed by J. Platts. All geometries were optimized at B3LYP/6-31+G(d,p) (130-132) in Gaussian03. Hyperfine couplings were calculated at PBE0/EPR-II (133) using ORCA (134). Full details of simulations including can be found in Morris *et al.* (121).

2.12 Method and analysis of Click reactions

2.12.1 Gel filtration chromatography: Analytical

sfGFP^{AzF} variants and wild-type sfGFP were analysed for the presence of any oligomeric species by analytical gel filtration using a SuperdexTM 75 GL column (GE Healthcare). Proteins were loaded at 10 μM and separated at a flow rate of 0.5 mL/min with sodium phosphate buffer (50 mM, 150 mM NaCl; pH 8). Elution volume was monitored by absorption at 280 and 485 nm. The column was calibrated using protein standards of known molecular weight (1.4, 17, 44, 158 and 670 kDa) and used to estimate the molecular mass of samples using Equation 2.3.

$$K_{av} = \frac{V_e - V_0}{V_t - V_0} \quad \text{Equation 2.3}$$

Here, the partition coefficient (K_{av}) is calculated from the elution volume (V_e) of the protein in relation to the void volume (V_0 ; 670 kDa) and total volume (V_t ; 1.4 kDa) of the column. The K_{av} of the standard proteins were plotted against $\log(MW)$ to give a standard curve (Figure 2.1) which was used to calculate the molecular weight of the sfGFP variants.

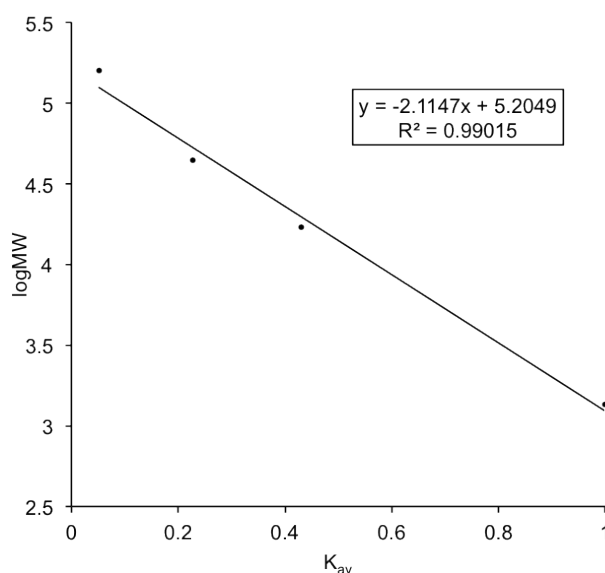


Figure 2.1. Molecular weight standard curve for analytical gel filtration. The partition coefficient (K_{av}) of the protein standards (1.4, 17, 44, 158 kDa) were plotted against the logarithm of their MW.

2.12.2 Strain-promoted azide-alkyne cycloaddition (SPAAC) reaction

SPAAC reactions were performed using either pure protein samples or crude cell lysates (to mimic intracellular conditions more closely) as a source or azPhe-containing protein. The azPhe-containing protein variants (sfGFP and cyt b_{562}) were modified with a dibenzylcyclooctyne (DBCO) functionalised fluorescent probe based on Texas Red referred to as DBCO-585 (Click Chemistry Tools, AZ, US). For pure protein samples, SPAAC was performed using 10 μM protein and various concentrations of DBCO-585 (between 10 and 100 μM) in 100 mM sodium phosphate buffer (pH 8). Proteins were diluted to the desired concentration and the reaction was started by the addition of the DBCO-585. Reactions were incubated at 25 °C in darkness with mixing and allowed to proceed for the indicated times or until the reaction reacted completion. For cell lysate reactions, soluble material following lysis was diluted to an equivalent OD_{600} of 1 with PBS and the reaction initiated on addition of 100 μM of DBCO-585 at 25 °C. The reaction was then left for 20 hours.

2.12.3 Analysis of SPAAC reactions by fluorescence and UV-visible absorption

Reaction mixtures were centrifuged at $15,000 \times g$ for 2 minutes to remove any insoluble material. Unreacted probe was removed from the protein component, and hence the reaction was stopped, by applying the reaction mixture to a His SpinTrap™ Ni-affinity chromatography column (GE Healthcare) and separated using a microcentrifuge (Section 2.5.3.1). Reactions were subsequently analysed by fluorescence and UV-vis absorption spectroscopy at 25 °C as in Sections 2.6.3 and 2.6.4. For fluorescent analysis of sfGFP SPAAC reactions, excitation and emission spectra were measured for the sfGFP protein itself (λ_{ex} 485 nm; λ_{em} 511 nm) and the attached DBCO-585 probe (λ_{ex} 594 nm; λ_{em} 611 nm). Absorption spectra of reactions were measured and used to calculate reaction yields by comparing the absorption at the λ_{max} of the azPhe-containing protein (sfGFP 485nm; cyt b_{562} 417 nm) with DBCO-585 (594 nm). Absorption values were converted to molar concentrations using the calculated extinction coefficients (Section 2.6.4) of the proteins and DBCO-585 ($\epsilon_{594} = 102,000 \text{ M}^{-1}\text{cm}^{-1}$) then used to calculate the molar ratio of protein labelled with DBCO-585.

2.12.4 Fluorescent imaging following SDS-PAGE

SPAAC reactions were also analysed using fluorescent imaging after separation of reaction components by SDS-PAGE (Section 2.6.1). SPAAC reactions were briefly centrifuged and diluted (1/5) in $5 \times$ SDS-PAGE sample buffer (Section 2.6.1) to a final volume of 50 μL . Gels were imaged using a Typhoon 9400 Variable Mode Imager with a 532 nm excitation laser and a 610 (BP 30) nm emission filter. Images were processed and fluorescent bands were quantified using ImageJ software. Gels were subsequently Coomassie stained and destained as in Section 2.6.1 to view total protein.

2.12.5 FRET calculations

DBCO-585 labeled sfGFP^{Q204AzF} was capable of Förster resonance energy transfer (FRET) from sfGFP (donor) to DBCO-585 (acceptor). First, modified sfGFP^{Q204AzF} was separated from unmodified sfGFP^{Q204AzF} by ion exchange chromatography using a MonoQ column (Section 2.5.3.4). The FRET efficiency of modified sfGFP^{Q204AzF} was determined experimentally using Equation 2.4, where I_A and I_D are the fluorescence intensities of the acceptor and donor, respectively, when excited at the donor's optimal wavelength (485 nm).

$$E = \frac{I_A}{I_A + I_D} \quad \text{Equation 2.4}$$

The Förster distance (R_0) for DBCO-585 labeled sfGFP^{Q204AzF} was determined using Equation 2.5. Here, $J(\lambda)$ is the spectral overlap between the donor emission and the acceptor excitation, Φ_D is the quantum yield of the donor (calculated as 0.65), η is the refractive index of the solvent (1.4 for aqueous solvent) and κ is the assumed orientation factor (2/3 for an undefined orientation).

$$R_0 = 0.211 [\kappa^2 \cdot \eta^{-4} \cdot \Phi_D \cdot J(\lambda)]^{\frac{1}{6}} \quad \text{Equation 2.5}$$

Equation 2.6 was used to calculate the distance between the two chromophores and subsequently, by comparison to the modelling prediction of the inter-chromophore distance, to test the accuracy of the calculation. In Equation 2.6, E equals the observed FRET efficiency, r is the distance between chromophores and R_0 is the Förster distance.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad \text{Equation 2.6}$$

2.12.6 Reaction kinetics

For the study of the reaction rates of sfGFP^{E34AzF} and sfGFP^{Q204AzF}, protein labelling was monitored using FRET between sfGFP and the attached DBCO-585 (as in Section 2.12.5) by fluorescence spectroscopy (2.12.3). The reaction was performed in the fluorimetry cuvette and started by the addition of DBCO-585. The reaction was followed by monitoring DBCO-585 emission (611 nm) after excitation of sfGFP (485 nm) until the reaction reached completion. The endpoint was quantified by absorption as in 2.12.3. Reactions were performed with 10 μ M sfGFP^{AzF} and 20 μ M DBCO-585. Rate constants (k) were calculated by plotting the integrated form of the second order rate equation (Equation 2.7).

$$\frac{1}{(A_0 - B_0)} \left[\ln \frac{(A_0 \cdot B)}{(B_0 \cdot A)} \right] = kt \quad \text{Equation 2.7}$$

Here, A_0 and B_0 are the initial concentrations (in M) of sfGFP and dye **2**, respectively, and A and B are the concentrations of sfGFP and dye **2** at time t (in sec), respectively (135). Rate constants (units $\text{M}^{-1} \cdot \text{sec}^{-1}$) were given by the gradient of the line and were calculated from 3 repeats.

2.13 Non-covalent labelling of carbon allotropes with proteins

Proteins were attached to highly-ordered pyrolytic graphite (HOPG) using the non-covalent π - π stacking interaction between pyrene and the surface. Proteins were modified with the adapter molecule DBCO-pyrene (Section 2.13.1) that contains azide reactivity (DBCO) and HOPG affinity (pyrene) using the SPAAC reaction (2.12.2). 10 μ M sfGFP^{Q204AzF} and cyt *b*₅₆₂^{D5AzF} were reacted with 50 μ M DBCO-pyrene for 16 hours. Unreacted DBCO-pyrene was removed by Ni-affinity purification of sfGFP^{Q204AzF} (Section 2.5.3.1) or size exclusion chromatography of cyt *b*₅₆₂^{D5AzF} with a HiTrap Desalting column (GE Healthcare). Pyrene-labelled proteins were then incubated with HOPG samples under various conditions as detailed in Chapter 7. Labelling of HOPG with proteins was assessed by atomic force microscopy (AFM; Section 2.13.2).

2.13.1 Synthesis of DBCO-pyrene

DBCO-pyrene was synthesised from DBCO-amine (Click Chemistry Tools) and pyrene butanoic acid, succinimidyl ester (AnaSpec, CA, USA) via an S_N2 nucleophilic substitution reaction. The reaction was performed in an argon atmosphere with TEA and reaction progress was monitored by TLC. The reaction was terminated by addition of aqueous HCl (0.1 M, 5 mL) and extracted with EtOAc (3 x 5 ml). Combined organics were washed with brine, dried over MgSO₄ and concentrated *in vacuo*. The crude residue was purified by silica gel column chromatography (eluant-EtOAc) to give DBCO-pyrene (30 mg, 84 %). DBCO-pyrene was analysed by TLC, ¹H-NMR, ¹³C-NMR and ESI-MS, all of which confirmed the expected product (NMR data below). Mass spectrometry was performed by the CITER analytical mass spectrometry facility (School of Chemistry, Cardiff University) and NMR spectrometry was performed by S. Hill (Cardiff University). NMR was performed on a Brücker Avance DPX 400 MHz at 20 °C. NMR spectra were assigned by A. Wood (Cardiff University). The final product was analysed by Further details including yields can be seen in Chapter 7.

¹H NMR (400 MHz, CDCl₃). δ 1.18 (t, 2 H, $J = 7.2$), 1.90 (dq, 1 H, $J_{(AX)(BX)} = 16.6$, $J_{AX} = 7.6$, $J_{AY} = 7.6$, $J_{BX} = 7.6$, $J_{BY} = 7.4$), 1.97 (t, 2 H, $J = 5.0$), 2.01 (d, 2 H, $J = 6.8$), 2.05 (d, 2 H, $J = 6.4$), 2.38 (dq, 1 H, $J_{(AX)(BX)} = 16.8$, $J_{AX} = 7.6$, $J_{AY} = 7.6$, $J_{BX} = 8.0$, $J_{BY} = 7.6$), 3.20 (t, 1 H, $J = 3.2$), 3.29 (m, 1 H), 3.57 (d, 1 H, $J = 14.0$), 4.04 (q, 1 H, $J_{AX} = 14.2$, $J_{AY} = 14.0$), 5.01 (d, 1 H, $J = 13.6$), 5.94 (br t, 1 H, $J = 6.0$), 7.00 (m, 1 H), 7.08 (t, 1 H, $J = 7.4$), 7.17 (m, 1 H), 7.22 (dd, 1 H, $J_{(AX)(BX)} = 2.4$, $J_{AX} = 2.0$, $J_{AY} = 2.0$), 7.56 (d, 1 H, $J = 7.2$), 7.76 (d, 1 H, $J = 7.6$), 7.92 (q, 1 H, $J_{AX} = 16.2$, $J_{AY} = 8.6$), 7.95 (d, 2 H, $J = 2.4$), 8.03 (dd, 2 H, $J_{(AX)(BX)} = 4$, $J_{AX} = 8.4$, $J_{AY} = 8.4$), 8.09 (dd, 2 H, $J_{(AX)(BX)} = 6.6$, $J_{AX} = 5$, $J_{AY} = 7$), 8.18 (d, 1 H, $J = 9.2$)

¹³C NMR (100 MHz, CDCl₃). δ 172.5, 172.3, 151.0, 148.0, 136.0, 132.0, 131.4, 130.9, 129.9, 129.0, 128.8, 128.6, 128.5, 128.3, 127.8, 127.8, 127.5, 127.4, 127.4, 127.3, 126.7, 125.8, 125.6, 125.1, 125.0, 124.9, 124.8, 124.8, 123.5, 122.9, 122.5, 114.8, 107.7, 77.4-76.8 (solvent residual), 55.5, 36.1, 35.2, 34.9, 32.9, 27.4, 21.1, 14.2, 5.61

¹³C NMR (100 MHz, DEPT 90°). δ 131.1, 129.0, 128.6, 128.5, 128.3, 127.8, 127.5, 127.4, 127.4, 127.3, 127.3, 126.7, 125.8, 125.6, 124.9, 124.8, 124.8, 123.5

¹³C NMR (100 MHz, DEPT 135°). δ 132.0, 129.0, 128.6, 128.5, 128.3, 127.8, 127.5, 127.4, 127.4, 127.3, 126.7, 125.9, 125.6, 124.9, 124.8, 124.8, 123.5, 60.4, 55.5, 36.1, 35.2, 34.9, 32.9, 27.4, 21.1, 14.2

2.13.2 Atomic force microscopy (AFM)

Final AFM imaging of HOPG labelled with protein was performed by A. Moskalenko and A. Zaki (School of Physics and Astronomy, Cardiff University). AFM allows accurate measurement of surface topology including height and width measurements. AFM was conducted in air under ambient conditions, measurements were carried out using a Multimode microscope with a Nanoscope III controller (DI Veeco, Santa Barbara, USA) operating using the tapping mode with polysilicon probes.

3. Production and purification of proteins containing unnatural amino acids

3.1 Introduction

The ability to reprogram the genetic code to expand the type of chemistry sampled by proteins has enormous potential in generating new proteins with novel physicochemical properties of both fundamental and technological interest. Wang and colleagues incorporated the first Uaa into a protein using a reprogrammed genetic code approach in 2001 (7). As a relatively young branch of protein engineering the technique of Uaa incorporation during cellular protein synthesis still requires some attention to ensure the approach is performing as designed and producing the intended Uaa-containing proteins in useful yields. Orthogonality is vital for Uaa incorporation via a reprogrammed genetic code to be a useful technique. Incorporation is orthogonal when the requisite machinery (tRNA and aaRS) does not cross-react but still works in parallel with the endogenous protein translation machinery (Section 1.3). This ensures that incorporation is site-specific and that it does not interfere with normal protein synthesis. As discussed in Section 1.4, aaRS/tRNA pairs from other kingdoms (orthologs) are used to afford this orthogonality. This chapter focuses on Uaa incorporation into superfolder GFP (sfGFP; Section 1.7) and cytochrome *b*₅₆₂ (cyt *b*₅₆₂; Section 1.8) in *E. coli* using an engineered tyrosyl tRNA/aaRS from *M. jannaschii* (Section 1.4). The chapter provides a foundation for the later chapters by covering the basis and implementation of gene mutagenesis, and the subsequent production of Uaa-containing proteins in *E. coli*. Many applications for studying and utilizing proteins require the protein of interest in a pure form, isolated from other proteins to allow for more detailed molecular analysis. This chapter also describes the purification of the Uaa-containing proteins used in the subsequent chapters.

3.2 Results and Discussion

3.2.1 Generation of amber stop codon mutations

As discussed in Section 1.3, an existing codon must be reprogrammed to code for an Uaa. The amber stop codon (TAG) is the most commonly reprogrammed codon for Uaa incorporation because it is the least used codon and was thus used in this work. TAG codons were introduced into genes of interest at the desired positions by whole plasmid site-directed mutagenesis (SDM) PCR (Section 2.4). Whole plasmid SDM differs from traditional approaches such as splice-by-overlap PCR mutagenesis because it alleviates the need to clone

the mutated gene back into a host plasmid by restriction digest and ligation. For whole plasmid SDM, plasmids containing the genes of interest are amplified by PCR using oligonucleotide primers containing the TAG mutation (Figure 3.1). An example of SDM to generate a TAG at residues 26, 44, 62 or 75 of *sfGFP* can be seen in Figure 3.1. PCR yielded DNA fragments of ~ 4.8 kb or 4.4 kb that corresponded to the size of pBAD-*sfGFP* or pBAD-*cytb₅₆₂*, respectively (Figure 3.1). DNA fragments were excised, purified and subsequently re-circularised as described in Section 2.4. *E. coli* cells were transformed with the potential plasmids and grown on LB agar with ampicillin to select for cells containing circular pBAD plasmids. Plasmid DNA was isolated from randomly selected colonies (typically 2 for each mutant) and the gene sequenced to confirm the introduction of in-frame TAG codons at the desired positions.

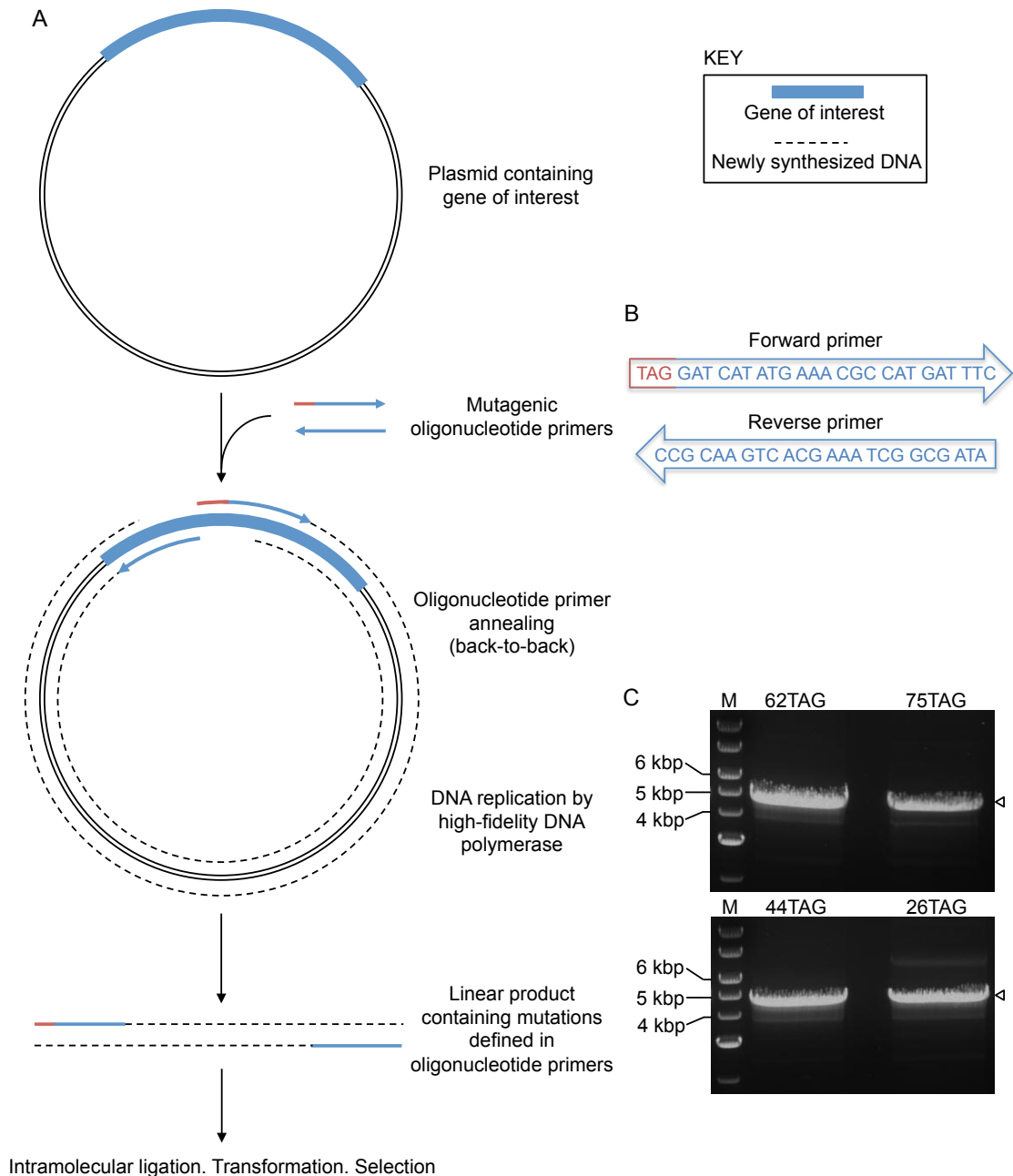


Figure 3.1. Introduction of TAG mutations by whole plasmid site-directed mutagenesis PCR. (A) Schematic showing whole plasmid PCR using mutagenic primers (blue arrows). In the first step, the primers bind back-to-back to the plasmid in the gene of interest (blue block). The TAG mutation (red) was most commonly placed at the 5' end of the forward primer (Table 2.4). The plasmid is then amplified by a high-fidelity DNA polymerase (Phusion[®] DNA polymerase) resulting in a double-stranded linear DNA (dashed lines) containing the TAG mutation (red) as the major DNA component in the reaction mixture. The linear DNA is then re-circularised via a high efficiency intramolecular ligation to regenerate the plasmid. (B) Example of mutagenic primers used in this study. Bases that match with the gene of interest are coloured blue and the TAG mutation is coloured red. (C) Agarose gel electrophoretic analysis of linear DNA following PCR of pBAD-*sfGFP* to introduce TAG codons at residues 26, 44, 62 and 75. Samples were compared to a 1 kb marker ladder (M) and key sizes are indicated. The desired DNA band is indicated by a white arrow (pBAD-*sfGFP* = 4.8 kb).

Uaa's were incorporated into positions throughout sfGFP and *cyt b₅₆₂* for a range of purposes as is discussed in the subsequent chapters. A full list of the mutagenic primers and residues that were mutated in this work can be seen in Table 2.4. Although the specific rationale for selecting the residues used is discussed at the beginning of the appropriate chapters, Figure 3.2 demonstrates the position and purpose of Uaa incorporation, and therefore amber stop codon mutagenesis, in the primary amino acid sequences of sfGFP and *cyt b₅₆₂*. The first main goal of this work was to use Uaa's to alter the properties of a protein, using sfGFP fluorescence as an example (Chapter 4; orange triangles). The second goal was to create sfGFP and *cyt b₅₆₂* variants that were capable of site-specific post-translational modification by Click chemistry with the intention of attaching proteins to non-biological materials like graphene (Chapter 7; blue triangles).

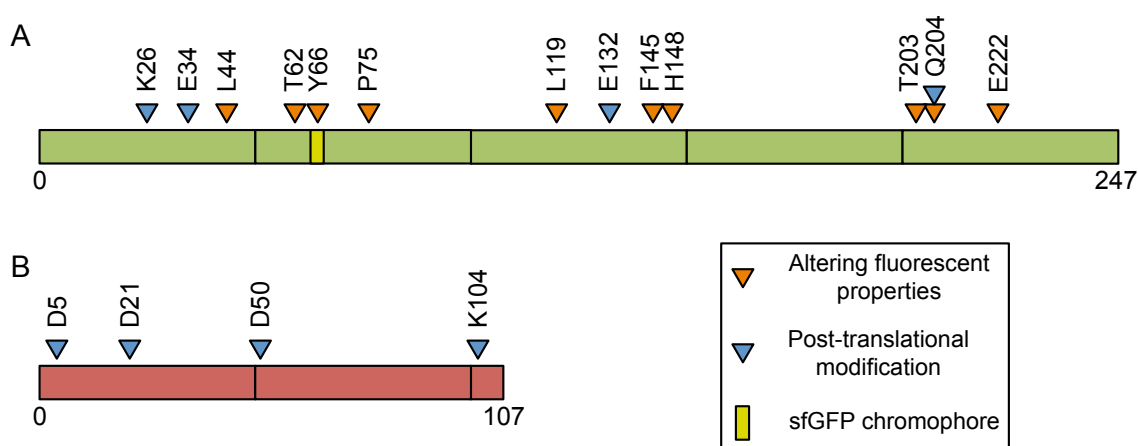


Figure 3.2. Sites of Uaa incorporation in (A) sfGFP and (B) *cyt b₅₆₂*. The positions of Uaa incorporation are shown as triangles on a linear representation of the two proteins (lines represent 50 amino acids). The residue number and native amino acid are shown above each arrow. Orange triangles show incorporation positions used to alter the fluorescent properties of sfGFP (Chapter 4 and 5). Blue triangles show surface-exposed positions in the proteins used to incorporate azPhe for subsequent modification by Click chemistry (Chapter 7). The position of the sfGFP chromophore (Thr65-Tyr66-Gly67) is represented by the yellow block.

3.2.2 Incorporation of Uaa's into proteins

Before *E. coli* can incorporate a desired Uaa into a target protein, the required engineered tRNA/aaRS pair (Section 1.4) needs to be introduced into the cells. This is achieved using a two plasmid system; pBAD and pDULE (Table 2.5; Figure 3.3). pBAD contains the gene of interest under the control of the arabinose-inducible P_{BAD} promoter. pDULE contains the engineered aaRS and paired tRNA under the control of constitutive *lpp* promoters (136). *E. coli* TOP10 cells (Invitrogen) were co-transformed with pBAD and pDULE and used to set up protein production cultures as in Sections 2.2.9 and 2.5.1. TOP10

cells lack the arabinose catabolism gene *araD* resulting in the accumulation of high intracellular arabinose concentrations. Arabinose binds to the AraC repressor protein (dimer), reversing its activity and turning AraC (monomer) into an activator of the P_{BAD} promoter, hence switching on transcription of genes under the control of the P_{BAD} promoter (136, 137). An autoinduction medium (122) was used for protein production that exploits differential levels of sugars to allow very high cell density before induction of P_{BAD} . In addition to arabinose the P_{BAD} promoter requires cAMP (bound to CAP; cAMP binding protein) for activation. Initially, when the medium is rich in glucose, cAMP concentration is low thereby inhibiting expression. Glucose is used as the primary sugar source of the cells and so gets broken down leading to an increase in cAMP concentration and therefore induction of P_{BAD} . Autoinduction medium aids production of Uaa-containing proteins in two ways. Firstly, the breakdown time of the supplied glucose when no energy is expended on target gene induction and protein production allows for a very high cell density. Therefore, when the gene is induced there will be a greater number of cells producing the target protein. Secondly, the gradual breakdown in glucose leads to a steady increase in the induction level of the target gene. This allows more time for the engineered machinery to incorporate the Uaa and so reduces stress on the cells associated with production of truncated protein. The combination of P_{BAD} with TOP10 cells and autoinduction medium results in strong and sustained induction of the gene of interest by a large number of cells, and therefore high levels of protein production.

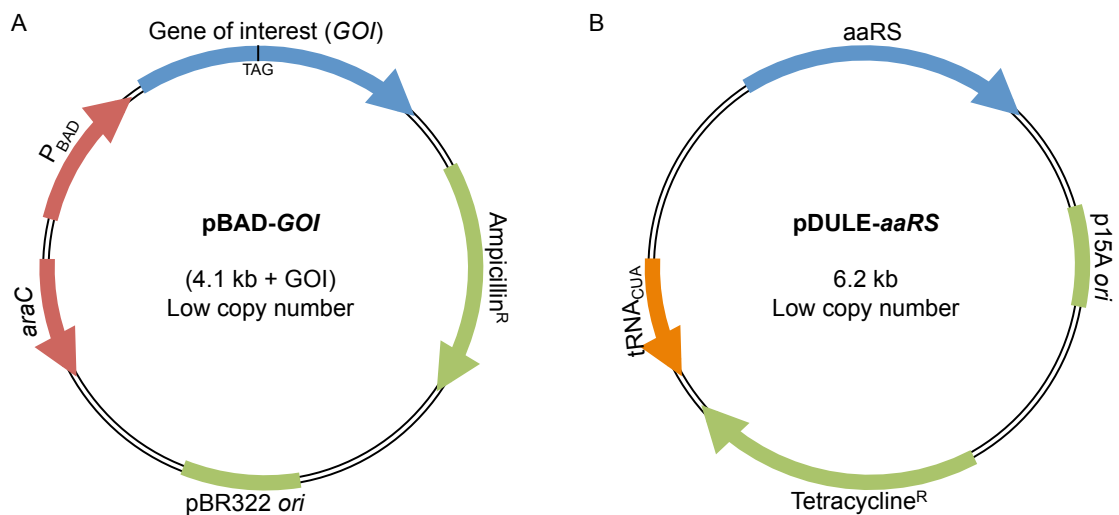


Figure 3.3. Plasmids used for Uaa incorporation. Plasmid maps of (A) pBAD and (B) pDULE showing the key plasmid elements. (A) pBAD carries the gene of interest (*GOI*) containing the amber stop codon (TAG) mutation that codes for the Uaa, under the control of the arabinose inducible P_{BAD} promoter. (B) pDULE contains the machinery for incorporation; one copy of the engineered $tRNA_{CUA}$ and the evolved aaRS specific for the chosen Uaa. Both the $tRNA_{CUA}$ and aaRS genes are located between the constitutive *lpp* promoter and the *rrnC* terminator.

Orthogonality is critical for specific and directed incorporation of Uaa's, and can be assessed by using the original role of the amber stop codon as a signal to stop translation. Parallel protein production cultures are set up; one grown in the presence of the Uaa and one in the absence. In a truly orthogonal system, in the absence of the Uaa the amber stop codon is read as a translation stop signal. This scenario results in protein truncated at the amber stop codon with no full-length protein produced.

Full-length protein should only be produced in the presence of Uaa when the amber stop codon is read as a signal for Uaa incorporation. Figure 3.4 provides an analysis of protein production in *E. coli* cultures of sfGFP and *cyt b₅₆₂* to assess orthogonality of azPhe incorporation. Analysis by SDS-PAGE showed major bands present at ~ 28 kDa for sfGFP (Figure 3.4A) and ~ 12 kDa for *cyt b₅₆₂* (Figure 3.4C) representing full-length protein from cultures grown in the presence of azPhe. As well as full-length protein, truncated protein was observed in the presence of azPhe for sfGFP (Figure 3.4A; Lane I; black arrow). This indicated that the Uaa-incorporation machinery was not efficient enough to fully out-compete RF-1 mediated translation termination (*vide infra*). Protein production levels of *cyt b₅₆₂* did not appear to be as high as sfGFP when analysing SDS-PAGE gels. Although levels were lower, the difference is exaggerated due to the smaller size of *cyt b₅₆₂* resulting in a more indistinct, blurred protein band whilst analysis of culture colour (Figure 3.4D) clearly showed high levels of *cyt b₅₆₂*. The parallel cultures grown in the absence of azPhe showed no full-length protein. Interestingly, for cultures grown in the absence of azPhe an insoluble, over-production band can be seen that corresponds to the predicted mass of the truncated protein (black arrow; Figure 3.4A). sfGFP cultures were also analysed by fluorescence spectroscopy as only the full-length protein is fluorescent. Fluorescence confirmed the SDS-PAGE that full-length protein was produced only from cultures grown in the presence of azPhe (Figure 3.4B). Furthermore, retention of fluorescence coupled with a soluble product strongly suggests that azPhe incorporation at the selected positions is not having a significant effect on protein structure, as discussed later.

Cyt b₅₆₂ is a pink-red colour due to the cofactor haem, which enables the production of full-length *cyt b₅₆₂* to be visualised. Images of cell pellets (Figure 3.4D) following protein production cultures showed functional protein only in cultures grown in the presence of azPhe. As with sfGFP, the production of functional protein suggests that azPhe incorporation is well tolerated. Taken together these results prove the orthogonality of the azPhe incorporation machinery, as was seen for all of the positions sampled (11, 138, 139). Full-length protein is produced only when azPhe is available and no natural amino acid is recognized by the introduced orthogonal tRNA/aaRS pair and incorporated in response to the TAG codon.

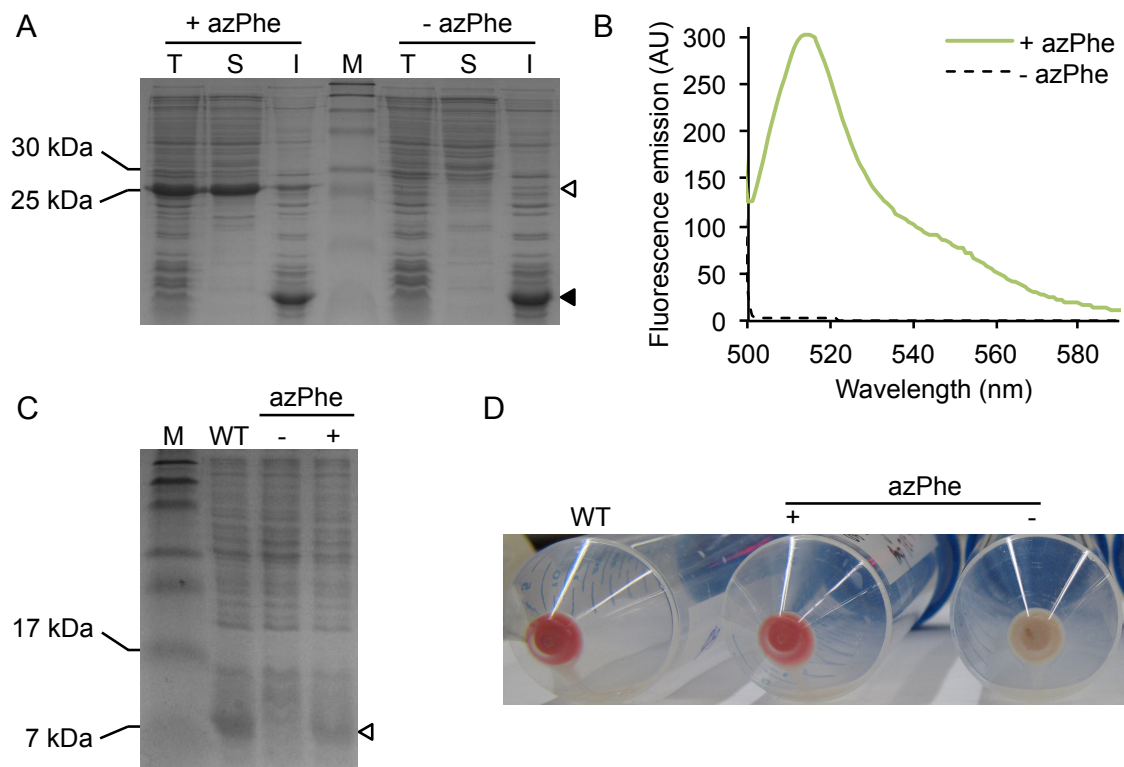


Figure 3.4. Assessment of the orthogonal incorporation of azPhe into sfGFP (A and B) and cyt b_{562} (C and D). SDS-PAGE analysis of (A) sfGFP^{L44azF} and (C) cyt b_{562} ^{D50azF} production. Samples refer to total cell contents (T), and soluble (S) and insoluble (I) fractions following cell lysis and centrifugation. For cyt b_{562} (C) all lanes represent total protein and the variant was compared to wild-type protein (WT). White arrows indicate the predicted full-length protein (~28 kDa for sfGFP and ~12 kDa for cyt b_{562}). The solid arrow in (A) indicates the predicted mass (~5 kDa) of the sfGFP truncated at residue 44. Parallel cultures were grown in the presence (+) and absence (-) of azPhe. (B) Fluorescence emission spectra of sfGFP^{F145azF} cultures grown in the presence (green) and absence (black dashed) of azPhe. Cultures were lysed and soluble material was normalized to an effective OD₆₀₀ = 0.5. Emission spectra were measured after excitation at 485 nm. (D) Cell pellets from cultures producing wild-type (WT) cyt b_{562} , and cyt b_{562} ^{D50azF} grown in the presence (+) and absence (-) of azPhe.

To incorporate different Uaa's, an aaRS (and therefore pDULE plasmid; Table 2.5) is required for that specific Uaa. Each aaRS has been artificially evolved previously to recognize only the Uaa and none of the natural amino acids by random mutagenesis of residues in the amino acid binding pocket of the protein (Section 1.4). Figure 3.5 assesses the orthogonality of the other aaRS enzymes used in this study to incorporate the following Uaa's; *p*-cyano-phenylalanine (cyPhe), *p*-trifluoromethyl-phenylalanine (tfmPhe), naphthyl-alanine (napAla) and *p*-amino-phenylalanine (amPhe). Parallel cultures were grown producing sfGFP^{Uaa} in the presence and absence of the given Uaa and analysed by SDS-PAGE and fluorescence. Generally, Uaa's were incorporated with very high orthogonality as indicated by the production of fluorescent, full-length protein containing Uaa only at the desired residue in the presence, but not absence, of the Uaa (Figure 3.5). cyPhe, tfmPhe and napAla (as well as azPhe; Figure 3.4) were incorporated with high orthogonality with effectively no functional protein detected in the absence of the Uaa. There is a small discrepancy in the production yield of cyPhe-sfGFP; fluorescent analysis indicates a comparable level of protein to other Uaa-sfGFPs however SDS-PAGE analysis shows a lower level. Significant functional protein was produced in the negative controls for amPhe indicating a far less orthogonal aaRS. However, differences in fluorescent and absorption properties of proteins produced in the presence and absence of amPhe (Figure 3.5D and Section 5.2.1) indicate that, although the aaRS is not fully orthogonal, it does incorporate amPhe preferentially over a natural amino acid. This is evident for sfGFP^{Y66TAG} as cultures grown in the presence of amPhe have blue-shifted fluorescence excitation and emission (λ_{ex} 450 nm, λ_{em} 500 nm) compared to sfGFP (λ_{ex} 485 nm, λ_{em} 511 nm).

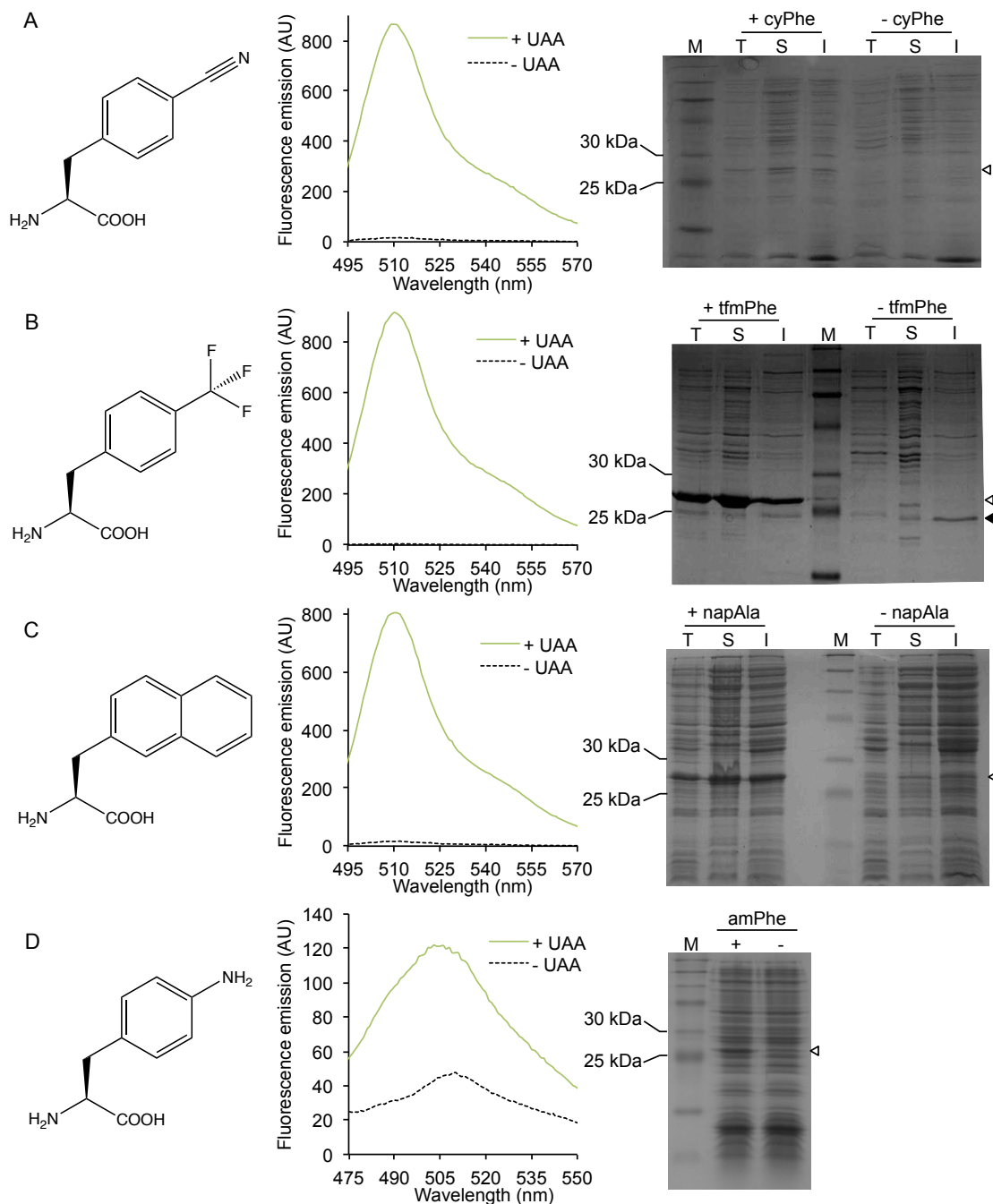


Figure 3.5. Assessment of the orthogonal incorporation of Uaa's into sfGFP. Protein production cultures for sfGFP^{TAG} were set up in the presence (+ Uaa) and absence (- Uaa) of the following Uaa's; **(A)** cyPhe, **(B)** tfmPhe, **(C)** napAla and **(D)** amPhe. For each Uaa **(A-D)**, the chemical structure is shown followed by culture analysis by fluorescence emission then SDS-PAGE. Fluorescence emission spectra of sfGFP cultures grown in the presence (green) and absence (black dashed) of the specific Uaa were recorded after excitation at the λ_{max} of the variant. Cell number was normalized to OD₆₀₀ = 0.5. Lanes on SDS-PAGE gels represent total cell protein (T), and soluble (S) and insoluble (I) proteins following *E. coli* cell lysis, except for amPhe where samples are for total cell protein only. White arrows indicate the predicted full-length protein (~28 kDa) and black arrows indicate protein truncated at the amber stop codon. The sfGFP^{TAG} variants used for fluorimetry were (A) Q204, (B) Q204, (C) Q204, (D) Y66. The sfGFP^{TAG} variants used for SDS-PAGE were (A) P75, (B) Q204, (C) Y66, (D) H148.

3.2.3 Purification of Uaa-containing proteins

In order to characterise variants and to use them for downstream applications, proteins need to be purified. The optimized purification protocol for sfGFP and cyt *b*₅₆₂ variants is described below.

3.2.3.1 Purification of sfGFP variants

All sfGFP variants, including the starting “wild-type” protein, were purified as described below. The sfGFP variants contain a C-terminal hexa-histidine tag sequence allowing for simplified purification by Ni²⁺-affinity (Section 2.5.3.1). Throughout purification, sfGFP could be monitored and tracked using the unique absorbance at 485 nm together with absorbance at 280 nm and fluorescence. After culture lysis, cell lysates (L) were centrifuged at high speed to remove insoluble material (I) from the soluble material (S) containing sfGFP (Figure 3.6A). Soluble material (S) was initially loaded onto a Ni²⁺ HisTrap column using an Äkta FPLC system (Section 2.5.3.1) and flow-through (F) was collected (Figure 3.6B). A gradient of increasing imidazole from 0 to 500 mM was then used to elute proteins that had bound to the column. Two major protein elution peaks were observed during the gradient; one with 280 nm absorption at ~70 mM imidazole and one with 280 and 485 nm absorption at ~145 mM imidazole. Eluent from the Ni²⁺ resin analysed by SDS-PAGE (Figure 3.6C) confirmed the second elution peak (145 - 170 mM imidazole) was sfGFP. Elution fractions containing sfGFP were pooled (between the beginning (B) and end (E) of the peak; Figure 3.6B and C) and, if required, further purified by ion exchange chromatography.

Ni²⁺-affinity pure sfGFP was further purified by anion exchange chromatography using a MonoQ column with an Äkta FPLC system (Section 2.5.3.4). Proteins were eluted from the column using a NaCl gradient from 0 to 300 mM. An intense peak with absorption at 280 and 485 nm was observed at ~180 mM NaCl (Figure 3.6D). Tightly bound proteins or aggregates eluted from the column when NaCl concentration was increased to 1 M. SDS-PAGE confirmed the protein eluted at 180 mM NaCl was sfGFP (Figure 3.6E). Elution fractions from this peak were pooled and exchanged into the required buffer. A final analysis by SDS-PAGE revealed that sfGFP and the sfGFP^{Uaa} variants were pure (Figure 3.6F).

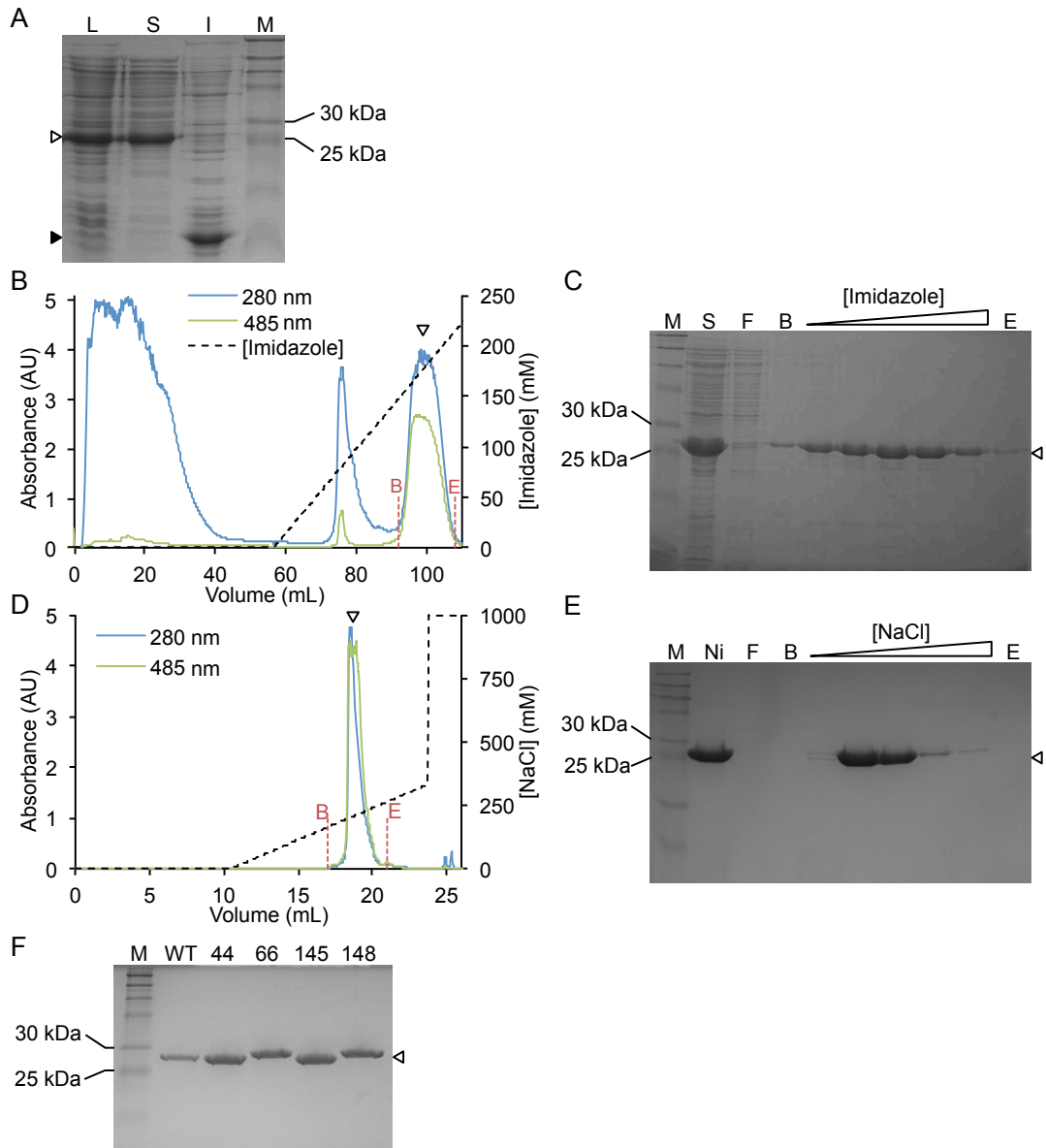


Figure 3.6. Purification of sfGFP variants. The wild-type sfGFP protein is used as an example to demonstrate the general purification process. **(A)** SDS-PAGE analysis of sfGFP over-production. After lysis, lysates (L) were centrifuged to separate soluble (S) and insoluble (I) material. Throughout the figure, white arrows indicate sfGFP (~28 kDa) and SDS-PAGE gels were run against a marker ladder (M) of known molecular weights. Ni^{2+} -affinity purification of sfGFP analysed by **(B)** absorbance of elution (using an Äkta FPLC) and **(C)** SDS-PAGE. **(B)** Chromatogram showing absorbance (280 nm, blue; 485 nm, green) and imidazole concentration (black dashed) over total volume ran through the column. Vertical, red dashed lines (B to E) indicate pooled fractions. **(C)** Soluble material (S) was loaded onto the Ni^{2+} HisTrap column and flow-through (F) was collected. Lanes corresponding to sfGFP elution are between B and E. Anion exchange purification of sfGFP analysed by **(D)** absorbance of elution (using an Äkta FPLC) and **(E)** SDS-PAGE. **(D)** Chromatogram showing absorbance (280 nm, blue; 485 nm, green) and NaCl concentration (black dashed) over total volume ran through the MonoQ column. Vertical, red dashed lines (B to E) indicate pooled fractions. **(E)** Ni^{2+} -affinity pure sfGFP (Ni) was loaded onto the column and flow-through (F) was collected. Lanes corresponding to sfGFP elution on increasing NaCl concentration can be seen between lanes B and E. **(F)** SDS-PAGE analyzing final purity of sfGFP variants. Lane numbers indicate the residue position of the incorporated azPhe. Small differences in protein mobility on SDS-PAGE are due to differences in storage buffer.

3.2.3.2 Purification of cyt *b*₅₆₂ variants

The Ni²⁺-affinity approach used to purify sfGFP is not appropriate for cyt *b*₅₆₂ as the requisite His sequence tag is undesirable for downstream applications such as attaching cyt *b*₅₆₂ to non-biological materials (Chapter 7). Therefore, cyt *b*₅₆₂ and the Uaa-containing variants were purified using more traditional chromatography approaches. Cyt *b*₅₆₂ has a number of distinctive properties that can be exploited during purification such as low molecular weight (~12 kDa), unique absorbance (λ_{max} (ox.) = 417 nm; λ_{max} (red.) = 425 nm) and high aqueous solubility (Section 1.8). Throughout purification cyt *b*₅₆₂ could be monitored by absorbance at 420 nm and visualized by a characteristic red colour. After culture lysis, cell lysates were centrifuged at high speed to remove insoluble material from the soluble material. The insoluble fraction included some cyt *b*₅₆₂ (for both wild-type and azPhe variants), as indicated by a red-coloured pellet. However, a substantial amount was present in the soluble fraction so it was decided to focus only on soluble protein. Initially, soluble material was concentrated and subjected to fractional ammonium sulphate ((NH₄)₂SO₄) precipitation. (NH₄)₂SO₄ was added to soluble lysates to 30 % (w/v) and samples centrifuged at high speed to remove precipitated material. Figure 3.7 shows proteins with low (NH₄)₂SO₄ solubility that were precipitated by 30 % (NH₄)₂SO₄ and removed from samples (Figure 3.7; Lane “30–”). (NH₄)₂SO₄ concentration was then increased to 90 % (w/v) to precipitate the remaining protein (Figure 3.7; Lane “90–”). The precipitated material was resuspended in the minimum possible volume of buffer ready for further purification. Due to the high level of protein some cyt *b*₅₆₂ remained soluble and was lost (Figure 3.7; Lane “90 +”).

The concentrated sample following (NH₄)₂SO₄ precipitation was separated by gel filtration chromatography on a Superdex 75 (10/300 GL) column (Section 2.5.3.3). Figure 3.7B shows the purification chromatogram where elution was monitored by absorption at 280 nm and 420 nm to detect total protein and cyt *b*₅₆₂, respectively. An eluent that absorbed at 280 and 420 nm was observed between 72 - 82 mL elution volume (V_e) for WT and the variants. Additionally, in cyt *b*₅₆₂^{AzF} variants there was a second elution that absorbed at both 280 and 420 nm between 62 - 69 mL V_e . Because these variants were engineered to have surface-exposed cysteine residues the second elution peak observed with a lower V_e (therefore greater hydrodynamic volume) was attributed to the formation of cyt *b*₅₆₂ dimers. Eluent from the column was analysed by SDS-PAGE (Figure 3.7C) and confirmed the elution peak at 72 - 82 mL was cyt *b*₅₆₂. Elution fractions containing cyt *b*₅₆₂ were pooled (between the beginning (B) and end (E) of the peak; Figure 3.7B and C) and further purified by ion exchange chromatography.

After gel filtration, pooled cyt *b*₅₆₂ fractions were subjected to anion exchange chromatography using a MonoQ column (Section 3.2.3.1 and 2.5.3.4). After binding to the column in low salt concentration, proteins were eluted using a gradient of increasing NaCl

concentration from 0 to 300 mM. An intense peak with absorption at 280 and 420 nm was observed at ~170 mM NaCl (Figure 3.7D). Tightly bound proteins or aggregates eluted from the column when NaCl concentration was increased to 1 M. SDS-PAGE confirmed the protein eluted at 170 mM NaCl was cyt b_{562} (Figure 3.7E). Cyt b_{562}^{AzF} variants gave a broader elution peak (absorbed both 280 and 420 nm) that was attributed to cyt b_{562} dimers as confirmed by native PAGE. Elution fractions corresponding to monomeric cyt b_{562} were pooled and exchanged into the required buffer containing ascorbic acid to prevent the formation of dimers.

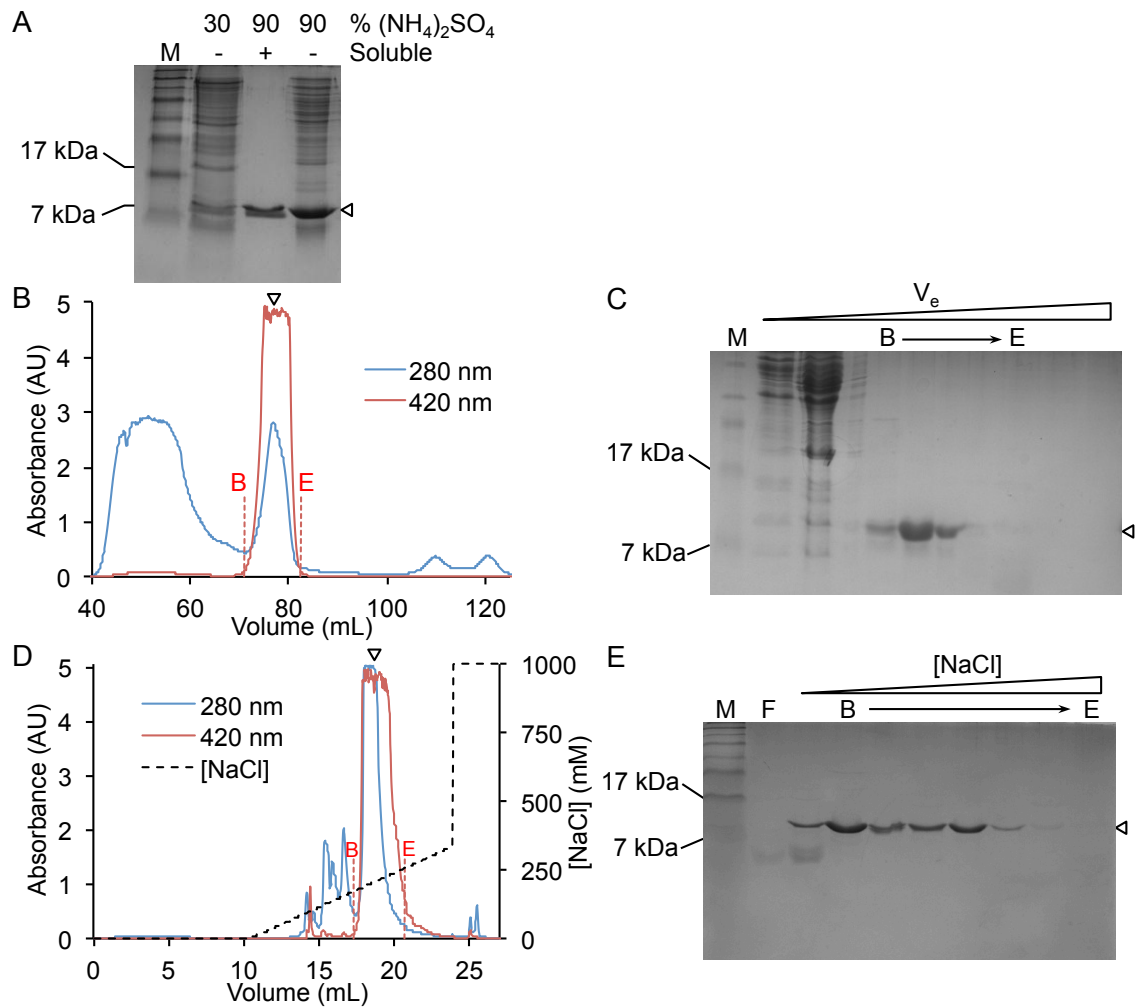


Figure 3.7. Purification of *cyt b₅₆₂* variants. (A) SDS-PAGE analysis of fractional $(\text{NH}_4)_2\text{SO}_4$ precipitation of wild-type *cyt b₅₆₂*. $(\text{NH}_4)_2\text{SO}_4$ was added to soluble lysates to 30 % followed by 90 % (w/v) and samples were centrifuged at high speed to remove precipitated material. Insoluble material (-) was resuspended in an equal volume to the equivalent soluble material (+) for SDS-PAGE analysis. (B and C) Gel filtration of *cyt b₅₆₂* variants. (B) Elution profile chromatogram of *cyt b₅₆₂* measured by absorbance (280 nm, blue; 420 nm, red). Vertical, red dashed lines (B to E) indicate pooled fractions. (C) Fraction analysis by SDS-PAGE. Lanes corresponding to the *cyt b₅₆₂* elution peak are between B and E. (D and E) Anion exchange of *cyt b₅₆₂*. (D) Elution profile chromatogram measured by absorbance (280 nm, blue; 420 nm, red) and NaCl concentration (black dashed) over total volume ran through the MonoQ column. Vertical, red dashed lines (B to E) indicate pooled fractions. (E) Fraction analysis by SDS-PAGE. Lanes corresponding to *cyt b₅₆₂* elution from the increasing [NaCl] gradient can be seen between lanes B and E. Throughout the figure, white arrows indicate the predicted size of *cyt b₅₆₂* (~12 kDa) and SDS-PAGE gels were run with a marker ladder (M) of known molecular weights.

3.2.4 Reprogramming the amber stop codon for Uaa incorporation

The first, and still most commonly recoded codon for genetic incorporation of Uaa's is the amber stop codon (TAG) (26). Amber stop codon suppression has been used to produce proteins containing Uaa's in various cell types (from *E. coli* to *S. cerevisiae* to HEK and HeLa cells) and even whole organisms (*C. elegans*, *D. melanogaster*) (Section 1.4). As well as having the advantage over solid-phase peptide synthesis (SPPS) of being *in vivo* and autonomous, this approach does not require invasive external manipulation by microinjection associated with the use of chemically aminoacylated-tRNA's. Its use in *E. coli* in combination with improved culture conditions, as demonstrated here (Section 3.2.2), means that significant and useful quantities of Uaa-containing proteins can now be produced. However, as with recoding any codon there are negative effects. The amber stop codon is the least used codon however 7.6 % of genes in *E. coli* (18) still require it for termination. Therefore, for cells expressing the reprogrammed tRNA/aaRS pair, the native proteins encoded by TAG-terminated genes may be extended and contain the Uaa. This could be an issue if the technique is to be used for *in vivo* or cell lysate applications, as the unique Uaa may be incorporated into undesirable proteins and the extended proteins may have altered activity. Despite this, amber stop codon recoding has been shown to not significantly effect cell growth rates (19), partly because non-coding regions are stop codon-rich so an alternative termination signal would soon be reached. Therefore, full-length native proteins are still produced, just now containing C-terminal extensions, which often do not detrimentally affect protein function (in comparison to a mid-protein Uaa) as shown by C-terminal purification tags. Additionally, the level of these endogenous proteins (now containing Uaa's) would be far lower than an over-produced protein of interest.

A similar situation arises from the recoding of rare, degenerate amino acid-coding codons; even though the codons are rare, they are not absent. The alternative approach is to 'create' new codons by extending the codon to four nucleotides (quadruplet codon). Here, translational machinery is evolved to recognize a completely unused codon such as AGGU or CGGU and incorporate the desired Uaa in response (22, 23). However the benefits of not interfering with normal cellular translation are counteracted by the requirement for additional machinery and low yields. The extra machinery required in comparison to the amber stop codon (Figure 3.3) is an engineered ribosome (O-ribosome, ribo-X) (24) capable of decoding the quadruplet codons that is itself a burden on the mutant cells. Also, if the first three bases are recognised as a standard triplet codon then translation will be frame-shifted resulting in production of junk protein and hence lower yields. Reported protein production yields from quadruplet codons are up to 40 mg/L culture for the incorporation of cysteine (140) and 2 mg/L for the Uaa homoglutamine (141). Therefore, although not perfect, the combination of

protein yields and accessibility makes the amber stop codon currently the best choice for Uaa incorporation.

There are other potential concerns that need to be addressed with Uaa incorporation by amber stop codon suppression. Firstly, amber stop codon suppression near the C-terminus of a protein can be a problem; truncated protein can retain the function of the full-length protein. For example, the cyt $b_{562}^{K104TAG}$ variant (TAG at codon 104 of 107) retains the ability to bind haem (Figure 3.8A). This would be problematic if the ability to control functional protein production by the presence of the Uaa in the growth medium was key and for the +/- control experiment for orthogonality (Section 3.2.2). Normally, the amber stop codon signals translation termination via binding of release factor 1 (RF-1) in the A-site of the ribosome (142, 143). As RF-1 is still present, amber suppression is in competition with translation termination, which could result in a small amount of truncated protein being produced even in the presence of Uaa (Figure 3.4A, Figure 3.5B and Figure 3.8B) due to the reduced efficiency of the reprogrammed tRNA/aaRS compared to endogenous pairs. This mixture of full-length and truncated protein could interfere with subsequent analysis such as quantification and enzymatic analysis of the full-length, functional protein.

Secondly, a more general issue with amino acid mutation as a whole is the potentially negative effect of the residue change on protein structure, folding and function. sfGFP^{E222azF} for example is produced in high yields but the protein is insoluble, most likely because it is partially/fully unfolded (Figure 3.8B), and is therefore non-fluorescent. This is probably due to the position of Glu222 in sfGFP facing towards the chromophore at the core of the protein in a tightly packed region. Replacing Glu with the much larger azPhe may result in mis-packing of the protein core and an incorrectly folded protein. Alternatively, as Glu222 has an important role in chromophore maturation, replacing it may preclude this role leading again to different packing at the core of the protein and incorrect folding. It is therefore important to check the solubility (*eg.* by SDS-PAGE; Figure 3.8B) and function (*eg.* fluorimetry for sfGFP; Figure 3.5E) of the protein after incorporation of an Uaa, particularly for internal residues where changes to residue size and polarity can have a profound effect.

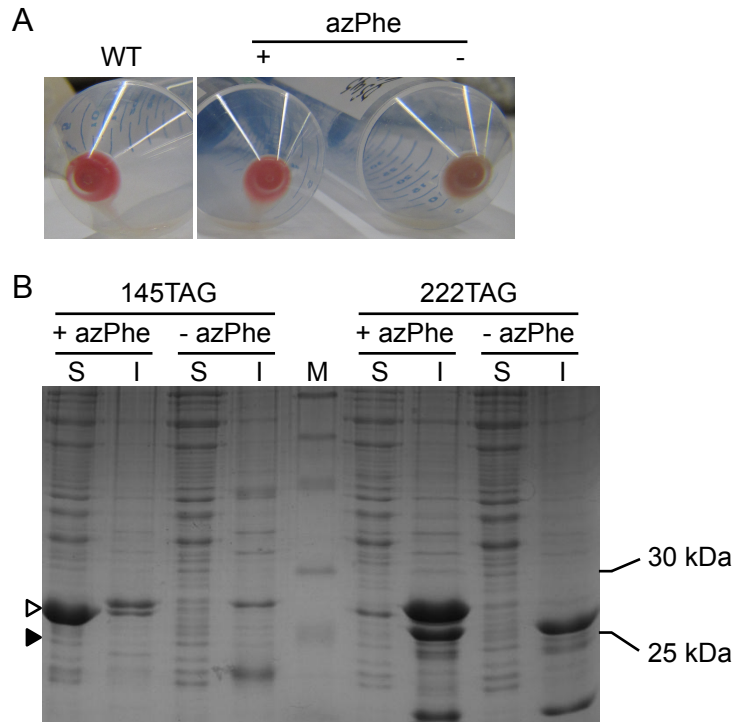


Figure 3.8. Issues with stop codon suppression and Uaa incorporation. (A) Truncated protein retaining function. Cell pellets from cultures producing wild-type (WT) *cyt b₅₆₂*, and *cyt b₅₆₂^{K10450azF}* grown in the presence (+) and absence (-) of azPhe. Red colour is a result of *cyt b₅₆₂*-bound haem (holo-*cyt b₅₆₂*). (B) Production of insoluble protein of interest. SDS-PAGE analysis of soluble sfGFP^{F145azF} (left) and insoluble sfGFP^{E222azF} (right) production. Samples were made from cultures after cell lysis that represent soluble (S) and insoluble (I) proteins. Parallel cultures were grown in the presence (+) and absence (-) of azPhe. White and black arrows indicate the predicted full-length protein (~28 kDa) and truncated protein, respectively.

3.2.5 Orthogonal incorporation of Uaa's

As discussed in Sections 1.3 and 3.1, for Uaa incorporation via a reprogrammed genetic code to be a useful and general technique the Uaa-incorporating machinery must be orthogonal to the endogenous machinery. The major factor for orthogonality seems to be the aminoacyl-tRNA-synthetase (aaRS) responsible for loading the tRNA with the Uaa (Figure 3.5). As shown in Figure 3.5D, some aaRS^{Uaa}'s appear to have poor amino acid specificity so produce full-length protein in the absence of Uaa. With regards to the former, the aaRS^{Uaa} may be recognizing and loading a natural amino acid in response to the amber stop codon. It is important to ensure that when the Uaa *is* present, only the Uaa is incorporated in response to the amber stop codon. Differences in fluorescent and absorption properties of sfGFP variants produced in the presence and absence of Uaa indicate that even poor aaRS^{Uaa}'s do incorporate the Uaa preferentially over a natural amino acid. The best demonstration of this is the incorporation of amPhe at residue 66 of sfGFP (Figure 3.5D; discussed in Section 5.2.1). After chromophore maturation residue 66 (Tyr in wild-type) forms the central moiety of the

chromophore and is key in defining the precise spectral properties of the protein (Section 1.7). The protein produced in the absence of amPhe has identical spectral properties to wild-type sfGFP ($\lambda_{\text{abs/exc}} = 485$; $\lambda_{\text{em}} = 511$), whereas the protein produced in the presence of amPhe has significantly shifted properties ($\lambda_{\text{abs/exc}} = 446$; $\lambda_{\text{em}} = 500$) with no sign of multiple peaks (therefore a single species). Together these results indicate that the majority of protein produced must contain the Uaa and not a natural amino acid. In the absence of Uaa, the misacylated natural amino acid is likely to be tyrosine as this was the original substrate for the aaRS prior to directed evolution. This competition for amino acid can be pushed in favour of the Uaa (amPhe in this case) by increasing the concentration of Uaa in the growth medium and thus putatively increasing intracellular concentration (Section 2.5.1). Overloading the cell with Uaa together with the superior tRNA loading efficiencies of endogenous aaRS's for their cognate amino acid should drive the engineered aaRS to utilize the Uaa.

It is important to understand why some of the current artificially evolved aaRS^{Uaa}'s are not as orthogonal or efficient as natural aaRS's, as this has important implications for future protein engineering aimed at generating better aaRS^{Uaa}'s. One major factor is evolutionary time; the natural aaRS's have evolved over billions of years to be perfectly adapted to recognize only the desired amino acid, whereas the current approach utilizes a library of $\sim 10^9$ mutants (7). Another discernable factor is the aaRS residues that are targeted for directed evolution. The most commonly used approach is to randomly mutate 6 active site residues that make direct contact with the amino acid. By focusing on these few residues the aaRS library will never sample the important effects of second (or third) shell mutations or more dramatic changes in protein architecture such as moving, adding or removing domains or even dimerization. An excellent example is the comparison between tyrosylRS and phenylRS. The amino acids differ by a single hydroxyl group however the enzymes fall into different aaRS structural classes (tyrosylRS, class I; phenylRS, class II) and hence evolved from completely different lineages with unrelated catalytic domains (144). This highlights the disparity (and evolution time) required to distinguish two very similar structures. Two even more similar amino acids are Tyr and amPhe differing by a hydroxyl vs amino group (Figure 3.5D). It would be naive to imagine that the current approach of selecting active site residues could generate an aaRS that could absolutely distinguish amPhe from Tyr. Although there are currently some specific limitations, a reprogrammed genetic code is still a very useful tool that is becoming increasingly used to engineer proteins.

3.2.6 Efficiency of Uaa incorporation and protein production yields

Although the protein yields are high with the current approach, this required a great deal of optimization from the original system (7, 136, 145). Optimization included varying the copy number and, induction time and strength of the machinery and gene of interest as well as improvements to tRNA sequence and aaRS's. Initial protein production yields were reported as 2 mg/L culture for *o*-methyl-tyrosine (in DHFR; (7)) and azPhe (in GST; (11)). These figures may be misleadingly high and not represent all residues in a protein as amber suppression has been shown to be highly context dependant (146). The use of pDULE, pBAD and arabinose autoinduction medium (Figure 3.3) (136, 139) generates high yields of protein. Protein production yields were calculated in mg of purified protein per litre of growth medium for some azPhe-containing variants. sfGFP^{AzF} variants ranged from 69 mg/L (sfGFP^{K26AzF}) up to 296 mg/L (sfGFP^{Q204AzF}) with an average of 166 mg/L (of 7 variants), as compared to 524 mg/L for wild-type sfGFP (Table 3.1). Therefore under these conditions Uaa incorporation gave an average production yield of 32 % of the “natural” over-production of the wild-type protein. A comparison of the suppression efficiency of the various amber stop codon mutations (Figure 3.2) with azPhe is shown in Table 3.1.

Table 3.1. Suppression efficiencies of amber stop codon mutations in sfGFP and cyt *b*₅₆₂. Suppression efficiencies are given relative to other amber stop codon positions in the same protein (– < + < ++ < +++ < ++++). Protein production yields are given (as mg protein/litre culture) where calculated and shown as a percentage of the wild-type protein yield.

Amber stop codon position ^a		Relative suppression efficiency	Soluble?	Protein yield mg/L	% wild-type
sfGFP					
26	Lys	+	+	69	13
34	Glu	++	+	132	25
44	Leu	++	+	-	-
62	Thr	++	–	-	-
66	Tyr	+++	+	205	39
75	Pro	+	+	-	-
119	Leu	++	+	-	-
132	Glu	++	+	106	20
145	Phe	+++	+	144	27
148	His	+++	+	210	40
203	Thr	+	+/-	-	-
204	Gln	++++	+	296	56
222	Glu	+++	–	-	-
cyt <i>b</i> ₅₆₂					
5	Glu	++	+	7	-
21	Glu	+	+	4	-
50	Glu	++	+	6	-
104	Lys	+	+	4	-

^a See Figure 3.2.

One limiting factor for production yield is the incorporation machinery and in particular the aaRS^{Uaa}. The development of improved aaRS^{Uaa}'s would benefit protein production yields, as well as orthogonality (Section 3.2.5). Kinetic analysis of aaRS^{Uaa}'s indicated low k_{cat} and K_m for the PP_i exchange step resulting in a k_{cat}/K_m ratio 10³ lower than for the wild-type enzyme (147). This decreased efficiency (and orthogonality) creates the requirement for large excesses of the Uaa (148). Both of these facets of aaRS^{Uaa} function could be remedied through a more stringent protein strategy as discussed in Section 3.2.5 and Chapter 8.

The inefficiency of Uaa incorporation is highlighted when multiple Uaa's are incorporated into a single protein. When a gene contains two amber stop codons the machinery must suppress both (in competition with RF-1) in order to produce a full-length protein and subsequently the protein yield is exponentially lowered (Figure 3.9). The production level is similar to endogenous proteins rather than an over-production (Figure 3.9; 1+). Enough protein is produced to visualize via Coomassie stained SDS-PAGE, but only when compared to the negative control grown in the absence of Uaa (Figure 3.9; 2+ vs 2-).

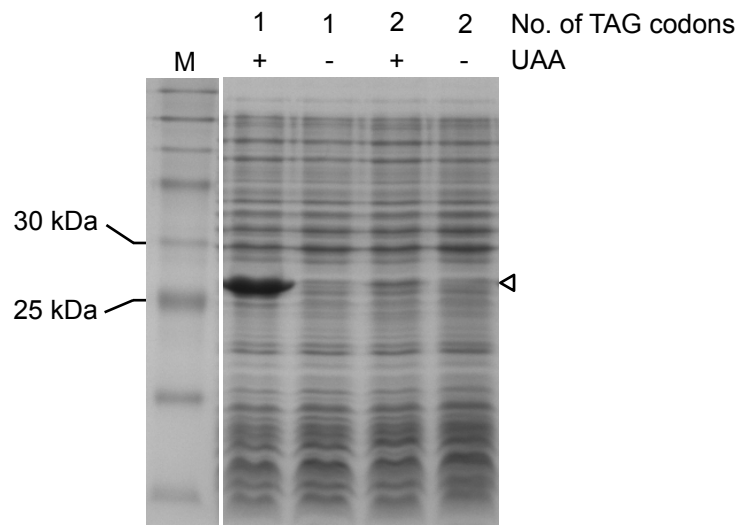


Figure 3.9. Incorporating multiple Uaa's into a single protein. SDS-PAGE analysis of protein production cultures of sfGFP containing one (left) and two (right) azPhe residues. Samples represent total protein from cultures where cell number was normalized to $OD_{600} = 10$. Parallel cultures were grown in the presence (+) and absence (-) of azPhe. The white arrow indicates the predicted full-length protein (~28 kDa). The sfGFP^{TAG} variants used were T203 (1 Uaa) and Y66/T203 (2 Uaa's).

This level of protein should still be useful for many applications but would need to be improved if Uaa technology is to be of wider applicability. One particular direction of incorporating multiple Uaa's is to encode two different Uaa's into the same protein such as the complementary reactive azido and alkyne groups, as in (149). As discussed, one factor limiting protein yields is the competition between amber suppression with translation termination by RF-1. RF-1 knockout in *E. coli* is lethal (150) so alternative strategies are required (8). One promising, yet arduous strategy is the replacement of all amber stop codons in the genome allowing deletion of RF-1. Implemented by Multiplex Automated Genome Engineering (MAGE), this technique creates a blank codon to code for Uaa's (151, 152). The technique could be expanded to degenerate amino acid-coding codons leaving all but one codon for each amino acid in use (eg. five of the six codons for arginine). Therefore, this could potentially expand the genetic code for a huge number of extra Uaa's. Nonetheless, all techniques to bypass problems with reprogramming codons also require high quality aaRS^{Uaa}'s that could be generated by extensive, whole-protein evolution.

4. Influencing sfGFP function using unnatural amino acids

4.1 Introduction

4.1.1 Unnatural amino acids (Uaa's) and their influence on proteins

Traditional protein engineering approaches such as amino acid substitution, deletion and insertion have proved very useful but are inherently limited by the chemistry of the natural amino acids. Uaa's have the capacity to go beyond the natural repertoire and have physicochemical properties not normally accessible in nature. In conjunction with an expanded genetic code approach (Section 1.3) to site-specifically introduce these new properties, Uaa's offer a powerful way of altering and/or studying the properties of a protein. The Uaa's used in this study are aromatic, phenylalanine derivatives and were chosen because their properties are not found in the natural amino acids. The chemical structures of the target Uaa's can be seen in Table 1.1 and Figure 3.5. The properties of azPhe are discussed in Section 1.6.

p-Cyano-phenylalanine (cyPhe) has a highly electron-rich, polar cyano bond ($\delta^+ \text{C} \equiv \text{N} \delta^-$). Polarity of the cyano group is between methylene and amide groups which allows cyPhe to be tolerated in both hydrophobic and hydrophilic environments (153). CyPhe has unique infrared absorption at 2200 cm^{-1} and is weakly fluorescent (154, 155). The fluorescence properties of cyPhe complement Trp and Tyr and can be used as a Förster resonance energy transfer (FRET) acceptor, which has been utilized to study protein unfolding (154, 155) and could be used to alter the spectral properties of a fluorescent protein (FP).

p-Trifluoromethyl-phenylalanine (tfmPhe) contains highly polarized carbon-fluorine bonds ($\delta^+ \text{C}-\text{F} \delta^-$), due to the electronegative fluorine atoms. TfmPhe is one of the strongest electron withdrawing groups which can cause adjacent groups to become more electrophilic (156). Additionally, recent evidence suggests fluorinated amino acids could be used to stabilise protein folding (157). Naphthyl-alanine (napAla) has a large, planar, conjugated naphthalene group. The conjugated π -system and resultant absorption and fluorescence properties could be used to alter the electrochemistry of a protein into which it is incorporated. NapAla is also highly hydrophobic, and so could have an effect on the hydrophobic packing of a protein including protein folding (32, 158). Additionally, cyPhe, tfmPhe and azPhe can act as hydrogen bond acceptors in contrast to napAla and Phe, which cannot hydrogen bond, and Tyr that can be both a hydrogen bond donor and acceptor. Taken together, tfmPhe, cyPhe and napAla are interesting candidates for making subtle changes to the electrochemistry and/or structure and stability of a target protein.

By introducing the above properties into a receptive residue in a protein the properties of that region or even the function of the whole protein can be altered. As discussed in Section 1.7, sfGFP is an ideal target protein to test this approach. sfGFP is highly stable, tolerant to point mutations (104) and the intrinsic fluorescence can be easily visualized and measured. Additionally, the importance of fluorescent proteins (FPs) to cell biology and microscopy means there is potential for producing useful research tools.

4.1.2 Protein photocontrol

One recent revolution in the study of biological systems was the emergence of approaches to control protein function, and therefore biological processes, using light. This approach, known as optogenetics (159, 160), provides control with high spatial and temporal resolution in response to an innocuous stimulus; light. Optogenetics has created a new paradigm for neuroscience allowing neurons in specific regions of the brain to be switched on or off. The ability to control individual neurons with high temporal control (20 Hz) has facilitated the dissection of many different neuron subtypes (161, 162). Optogenetics approaches currently rely on a limited number of cofactor-dependent light sensitive proteins such as opsin channels (163) and LOV domains (160), which may restrict general use of the approach. Uaa's such as *p*-azido-phenylalanine (azPhe; Section 1.6) would be a useful addition to the optogenetic toolbox as they are generally small, have specific, controllable photoreactions, and are transferable to any protein. When incorporated at a protein residue important for structure or function, the photoreactions (Figure 1.6) could be used to modulate local protein environment and therefore the overall properties of a protein.

In this study sfGFP was used as the preliminary target to evaluate the use of azPhe as a small optogenetic probe because of the ease of visualisation of function (fluorescence). Moreover, photocontrollable FPs have recently become desirable as molecular highlighters for use in high spatial, temporal and molecular resolution cell imaging, such as PALM (Photo-activated Localisation Microscopy) (164, 165). Thus, considerable effort over the last few years has been devoted to engineering useful photo-controllable FPs (166) but there are still some limitations including the robustness and oligomeric state of the base protein scaffold. Also the energy input required for photoconversion can be high, which can cause damage to bio-macromolecules and cells. Ideally, photocontrol would be engineered into a stable, monomeric, fast-folding FP such as sfGFP. The model photocontrollable protein would have fast photoconversion kinetics and high magnitude changes on the application of relatively low energy.

This chapter explores the use of Uaa's to alter the fluorescent properties of sfGFP. Residues were selected for Uaa incorporation by both rational design using X-ray crystal structures, and using a directed evolution approach to generate interesting variants that would not have otherwise been sampled. Firstly, mutation to the various phenylalanine derivative-Uaa's discussed above (Section 4.1.1) generated sfGFP variants with altered spectral properties. Additionally, incorporation of the photoreactive Uaa azPhe (Section 4.1.2) allowed the properties of azPhe-variants to be controlled post-translationally with UV-blue light.

4.2 Results

4.2.1 Altering the spectral properties of sfGFP by Uaa incorporation

Due to the molecular complexity of protein structure, folding and function sometimes a fully rational approach to protein engineering is naïve. It can be difficult to predict the full effects of a mutation due to the complex and cooperative network of interactions that constitute a functional protein. Directed evolution methods were developed to alleviate such problems by randomly generating molecular diversity and selecting protein variants with desirable properties (167-169). These approaches therefore remove the bias of rational design and allow effectively every residue in a protein to be subject to mutagenesis and selection. By sampling every residue in a protein it is possible uncover influential residues that would have otherwise been overlooked because, for example, they are distal from the active site. Directed evolution has been used to randomly sample amino acid substitution (167), deletion (170) and recently the introduction of Uaa's (17).

In this section a directed evolution approach called Trinucleotide exchange (TriNEx) (171, 172) was employed (by A. Baldwin and M. Edmundson, unpublished data) to randomly introduce amber stop codons (TAG) throughout the *sfGFP* gene. Initially the *sfGFP*^{TAG} library was screened for in-frame TAG codons with a positive/negative selection in the presence or absence of the desired Uaa, respectively, based on fluorescence of *E. coli* colonies (17). Colonies that were fluorescent in the absence of the Uaa were removed from the library as they did not contain an in-frame TAG codon and hence could produce full-length, functional protein. Clones that were fluorescent only in the presence of the Uaa were selected and the resulting library of *sfGFP*^{TAG} variants was further analysed. The library highlighted, among others, residues Leu44 and Leu119 of sfGFP for further analysis (Figure 4.1). Leu44 and Leu119 are residues in the β -barrel with side chains that face the interior of the protein. Both residues are surrounded by hydrophobic residues that contribute to the packing of the protein core (Figure 4.1C and D). Leu44 is level with the plane of the chromophore and faces the former Thr65 (-CH₃) moiety at a distance of approximately 3.5 Å. Leu119 is further from the

chromophore with the nearest contact the former Gly67 (C=O) at a distance of ~ 6.7 Å and the C-terminal residue Val68 at 3.6 Å.

In addition to these randomly identified residues, His148 and Thr203 were rationally chosen for mutagenesis. Both residues reside in β -strands and face the tyrosyl moiety of the chromophore in close proximity (Figure 4.1A and B). The side chains of both His148 and Thr203 hydrogen bond with the tyrosyl hydroxyl group of the chromophore, which promotes ionisation to the anionic form (Section 1.7; (92, 98)). The side chain of His148 is known to be flexible and can rotate away from the chromophore. Additionally, when substituted for larger amino acids, Thr203 can π - π stack with the chromophore resulting in red-shifted fluorescence, such as in YFP (Thr203Tyr) (96) (97).

The Uaa's (azPhe, tfmPhe, cyPhe and napAla) as well as Tyr were incorporated into the targeted residues discussed above, as in Section 2.5.1 and 3.2.2. The resulting sfGFP^{Uaa} variants were analysed by fluorescence spectroscopy (Figure 4.2).

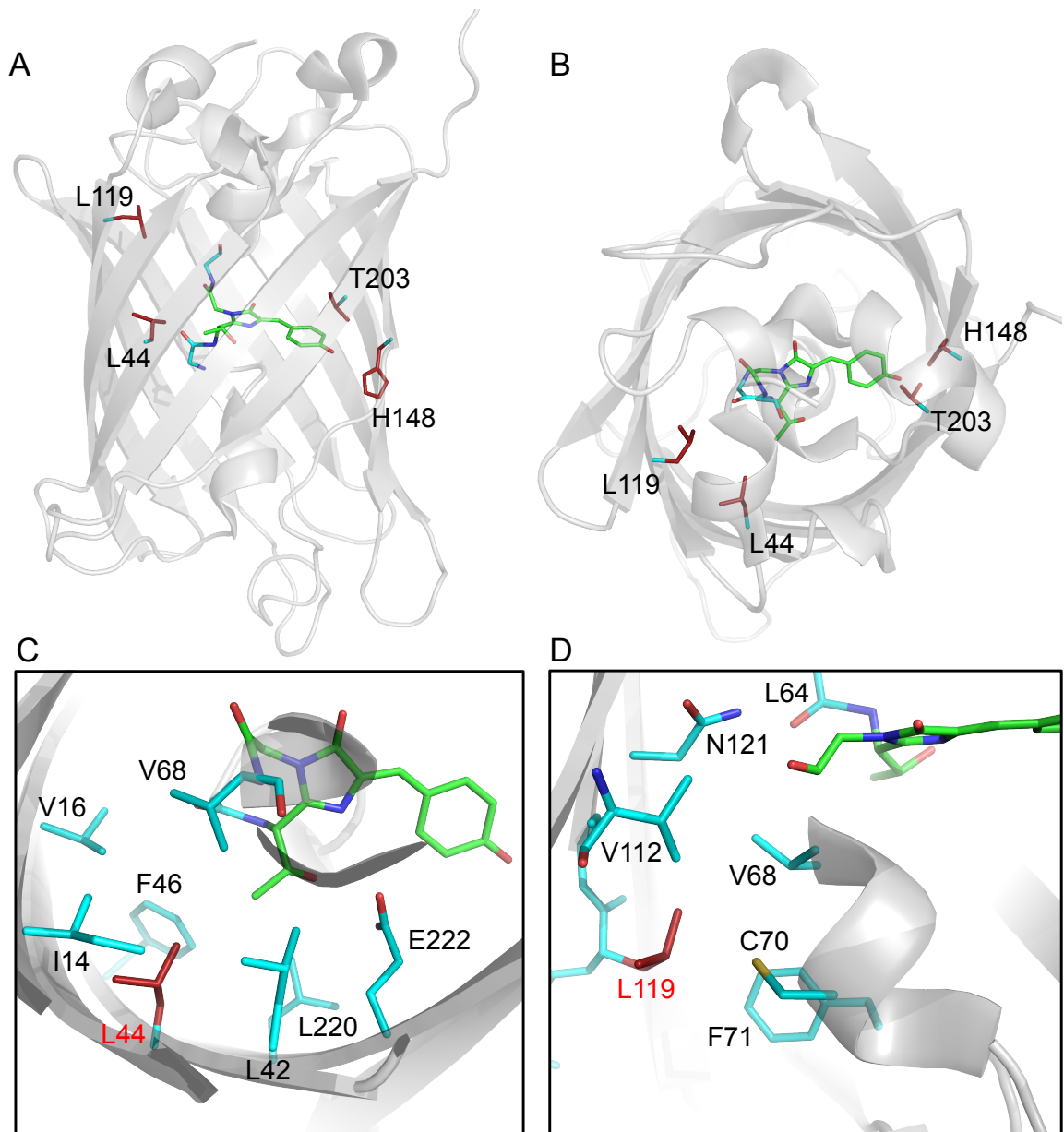


Figure 4.1. Residues of sfGFP mutated to different Uaa's. Three-dimensional representation of sfGFP from two views; (A) side-on and (B) top-down. The 4 residues mutated to Uaa's are labeled and shown as red sticks. The chromophore is shown by stick representation with carbon atoms coloured green. Local environments surrounding (C) Leu44 and (D) Leu119. The environment of His148 and Thr203 can be seen in **Figure 1.9**. PDB accession 2B3P.

4.2.1.1 Uaa incorporation at Leu44

Mutation of Leu44 had a significant effect on the excitation spectrum of sfGFP (Table 4.1), affecting the proportion of two major excitation peaks at 394 and 485 nm that represent the neutral and anionic forms of the chromophore, respectively (Section 1.7). Wild-type sfGFP absorbs and is excited at 485 nm (black dashed trace; Figure 4.2A). On substitution of Leu44 with Uaa's the ratio of the neutral peak (in relation to anionic peak) increased with the pattern cyPhe > tfmPhe > azPhe > napAla (Figure 4.2A, Table 4.1). Interestingly, the variant that was most shifted to the neutral (394 nm) form, and therefore deviated furthest from wild-type sfGFP ($\lambda_{\text{anionic}}/\lambda_{\text{neutral}} = 8.33$), was replacement with Tyr ($\lambda_{\text{anionic}}/\lambda_{\text{neutral}} = 0.16$). Also, the anionic excitation peak was blue-shifted by between 2 and 5 nm (to 480-483 nm) on substitution of Leu44 (Table 4.1). Excitation at either λ_{ex} generated the same wavelength emission spectra with changes in intensity matching the relative intensity of the λ_{ex} . The λ_{em} between the different amino acids was effectively the same with a 3 nm blue-shift in cyPhe44 and tfmPhe44 (508 nm) compared to wild-type (511 nm). Although napAla44 was produced in very low yield, fluorescence spectra indicated a mainly anionic chromophore with λ_{ex} and λ_{em} of 481 and 508 nm, respectively (Table 4.1).

Table 4.1. Spectral properties of sfGFP^{Uaa} variants. Excitation and emission spectra were recorded as described in Figure 4.2.

Residue	Amino acid ^a	λ_{ex}				λ_{em} (nm)
		λ_{neutral} (nm)	λ_{anionic} (nm)	Ratio		
				neutral:anionic	anionic/neutral	
44	cyPhe	394	480	0.64 : 1	1.56	508
	tfmPhe	394	480	0.58 : 1	1.72	508
	azPhe	394	483	0.45 : 1	2.22	511
	napAla	n/a	481	n/a	n/a	508
	Tyr	394	482	1 : 0.16	0.16	509
148	cyPhe	400	486	0.19 : 1	5.26	510
	tfmPhe	400	487	0.08 : 1	12.50	509
	azPhe	400	500	1 : 0.66	0.66	514
	napAla	396 (broad)	486	1 : 1	1	510
	Tyr	400	500	1 : 0.60	0.60	514
203 ^{b,c}	cyPhe	-	511	-	-	525
	tfmPhe	-	508	-	-	520
	Tyr	-	511	-	-	525
119	cyPhe	400	482	0.06 : 1	16.67	511
	tfmPhe	400	484	0.09 : 1	11.11	511
	azPhe	400	483	0.12 : 1	8.33	510
	napAla	n/a	482	n/a	n/a	511
	Tyr	400	482	0.08 : 1	12.50	509
Wild-type	-	400 (shoulder)	485	0.12 : 1	8.33	511

^a sfGFP^{hapA} variants were generally produced in low yields and as such were difficult to analyse. Therefore only reliable data was included in this table with data not available indicated by n/a.

^b sfGFP^{203AzF} was light sensitive and so not included in this analysis (discussed in Section 4.2.2).

^c sfGFP^{203Uaa} variants showed no neutral excitation peaks.

4.2.1.2 Uaa incorporation at His148

Similarly to Leu44, mutation of His148 had a significant effect on the fluorescence excitation spectrum of sfGFP. However, His148 variants displayed neutral and anionic excitation peaks at 400 and 485 nm, respectively (*versus* 394 and 480 nm for Leu44) (Figure 4.2B, Table 4.1). The excitation spectra of sfGFP^{148cyF} and sfGFP^{148tfmF} were most similar to wild-type sfGFP with the anionic (485 nm) form the major species. The ratios of the anionic to neutral chromophore forms are given for all of the variants in Table 4.1. TfmPhe148 completely removed any neutral (400 nm) excitation whereas cyPhe148 resulted in an increase in this species in comparison to wild-type sfGFP. The sfGFP^{148napA} variant was difficult to analyse because of the low level of soluble protein produced. However the ratio of excitation peaks indicates an equal proportion of the neutral and anionic forms. In contrast, replacement of His148 with either azPhe or Tyr generated variants in which the neutral chromophore (400 nm) was the predominant form, with ($\lambda_{\text{anionic}}/\lambda_{\text{neutral}}$) ratios of 0.66 and 0.60 (Table 4.1). Also, the anionic excitation peak of these variants was red-shifted by 15 nm (to 500 nm). The emission spectra generated from excitation at both λ_{ex} were virtually the same with small variation between 509–514 nm.

4.2.1.3 Uaa incorporation at Thr203

Mutation of Thr203 to Uaa's and Tyr resulted in large red-shifts in excitation and emission spectra of variants (Figure 4.2C, Table 4.1). Excitation spectra were red-shifted from wild-type sfGFP by 23 nm (to 508 nm) for tfmPhe203, and 26 nm (to 511 nm) for cyPhe203 and Tyr203. The emission spectra matched the red-shifted pattern seen for the excitation with λ_{em} of 520 nm for tfmPhe203, and 525 nm for Tyr203 and cyPhe203. Reliable fluorescence spectra could not be recorded for napAla203 because very little soluble, functional protein was produced. AzPhe203 was extremely sensitive to ambient light and is discussed in the following section (4.2.2; Figure 4.4C).

4.2.1.4 Uaa incorporation at Leu119

Substituting Leu119 for Uaa's had a subtle effect on fluorescence (Figure 4.2D, Table 4.1) resulting in small shifts in λ_{ex} and λ_{em} of between 3 and 5 nm from wild-type sfGFP. Although the properties of the sfGFP^{L119Uaa} variants were not significantly altered, Leu119 acts as a good control to show the shifts in fluorescence of the other residues (Figure 4.4A-C) were a result of Uaa position in the protein and not simply due to the presence of the Uaa. Additionally, all of the Uaa's incorporated at Leu119 were tolerated and produced fluorescent proteins (Table 4.1).

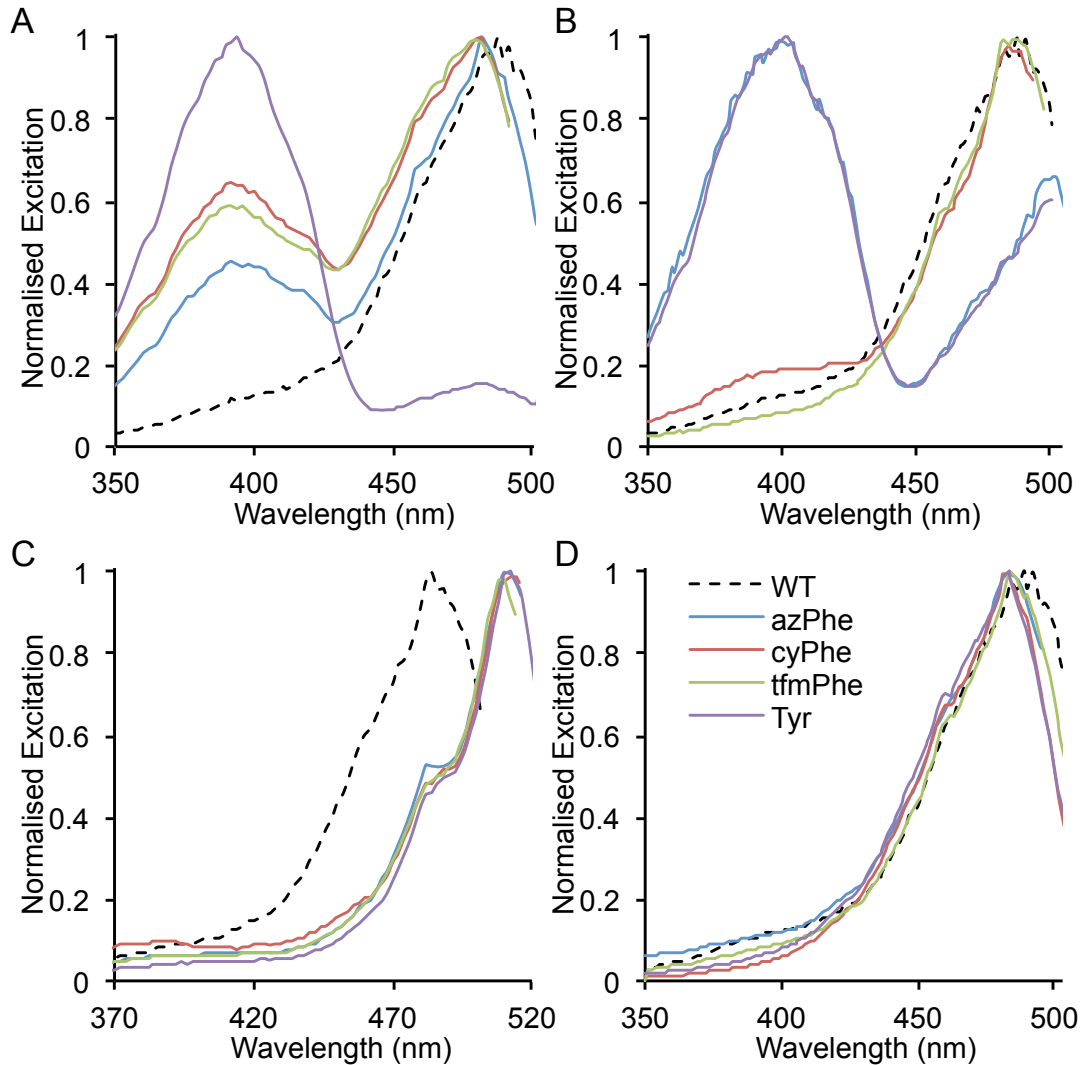


Figure 4.2. Fluorescent properties of sfGFP^{Uaa} variants. Excitation spectra for Uaa's incorporated at residue (A) Leu44 (B) His148, (C) Thr203 and (D) Leu119. Fluorescence spectra were recorded for each residue following the incorporation of azPhe (blue), tfmPhe (red), cyPhe (green), napAla (not included) and Tyr (purple). The spectrum of wild-type sfGFP (black dashed) is shown as a comparison for each. Fluorescence was normalized to 1 for comparison of λ_{max} . Emission spectra were recorded following excitation of the sample at the respective λ_{exc} and excitation spectra by monitoring emission at the λ_{em} . Emission λ_{em} and intensity followed the same pattern as λ_{exc} for all variants, as described in the text, so are not shown. Spectra were measured on soluble cell lysates from protein production cultures that were diluted to equivalent OD₆₀₀ between 0.1 – 0.5.

4.2.2 Generating photocontrollable sfGFP variants using azPhe

One property that is absent in natural amino acids but highly desirable to the protein engineer is photoreactivity. To test the hypothesis that a photoreactive Uaa could be used to modulate the properties of a protein, the photosensitive azPhe (Figure 1.6) was incorporated into selected sfGFP residues. In addition to Leu44, His148 and Thr203 studied in Section 4.2.1, two other rationally designed positions were investigated; Tyr66 and Phe145. Following

maturation, Tyr66 is the central residue of the sfGFP chromophore and is crucial in defining fluorescence (Figure 1.9). Phe145 is part of the β -barrel and points towards the tyrosyl region of the chromophore in an amphipathic environment formed by Pro58, Val61, Thr62, Ile167 and His169 (Figure 1.9).

AzPhe was incorporated into these 5 residues and the spectroscopic properties of the resulting sfGFP^{AzF} variants were analysed before, during and after photolysis using UV light (302 nm, 6 W). Proteins were studied in whole cells, cell lysates and as pure protein by fluorescence and absorption spectroscopy. The sfGFP^{AzF} variants displayed one of three distinct light-dependent events: activation (sfGFP^{Y66AzF}), deactivation (sfGFP^{F145AzF}) or switching (sfGFP^{L44AzF}, sfGFP^{H148AzF} and sfGFP^{T203AzF}) (Figure 4.3 and Figure 4.4). Wild-type sfGFP was used as a negative control and was photolysed under the same conditions as the sfGFP^{AzF} variants and showed no change in fluorescence intensity or wavelength (*vide infra*; 4.2.2.4).

4.2.2.1 sfGFP photoactivation: azPhe incorporation at residue 66

sfGFP^{Y66AzF} had negligible fluorescence before irradiation but became highly fluorescent after exposure to UV light (Figure 4.3B,C). The spectral properties of the photoactivated (irradiated) protein were different to sfGFP, with relatively broad excitation and emission peaks that were blue shifted to 446 and 500 nm, respectively (Figure 4.3A-C, Table 4.2). Along with sfGFP^{F145AzF} and sfGFP^{H148AzF}, the extinction coefficient (ϵ), quantum yield (QY) and brightness of sfGFP^{Y66AzF} was calculated before and after irradiation. Extinction coefficient (formerly molar absorption coefficient) is a measure of how much energy a chromophore absorbs per molar concentration and is measured in $M^{-1}.cm^{-1}$. QY is a measure of the conversion efficiency of absorbed to emitted energy (photons emitted / photons absorbed) and in combination with extinction coefficient is important because it determines the brightness of a fluorescent protein (brightness = $\epsilon \times QY$). Brightness is a key characteristic for whether a fluorescent protein is useful, and for this needs to have a suitable combination of extinction coefficient and QY. If, for example, a given fluorescent protein has a very low QY (low conversion of absorption energy to emission) then regardless of how well it absorbs light (extinction coefficient) it does not give a practical level of emission (brightness).

Prior to irradiation, a non-radiative major absorbance peak at 390 nm was observed with an extinction coefficient of $\sim 40,000 M^{-1}.cm^{-1}$. This converted to a 446 nm peak on the application of UV light. The extinction coefficient at λ_{446} for the non-irradiated form was $2,300 M^{-1}.cm^{-1}$, which increased to $24,600 M^{-1}.cm^{-1}$ on irradiation. The QY of photoactivated sfGFP^{Y66AzF} was calculated as 0.16 giving a brightness of $3,939 M^{-1}.cm^{-1}$ (Table 4.2). This is a brightness of only 11 % of sfGFP, however sfGFP is a very bright variant (160 % of EGFP

(104)) so even 11 % should give a useful level of fluorescence. The difference in brightness pre and post irradiation revealed a large ~13 fold gain in brightness.

Table 4.2. Spectral properties of the sfGFP^{AzF} variants.

Variant	λ_{\max} (nm)	λ_{em} (nm)	State	ϵ (M ⁻¹ cm ⁻¹)	Quantum yield	Brightness		Fold change
						M ⁻¹ cm ⁻¹	% sfGFP	
sfGFP ^{Y66AzF}	446	500	Dark	2300	0.13	299	1	13
			UV	24600	0.16	3939	11	
sfGFP ^{F145AzF}	490	511	Dark	47700	0.76	36279	99	≥53 ^a
			UV	11400	0.06	686	2	
sfGFP ^{H148AzF}	400	511	Dark	34200	0.69	23630	64	11
			UV	24000	0.30	7207	20	
	490	511	Dark	15800	0.32	5046	14	
			UV	34300	0.51	17474	48	
sfGFP	485	511	-	49000 ^b	0.75	36777	100	-

^a Value for fold change is approximate due to low fluorescence intensity in the light state.

^b ϵ calculated during this study for sfGFP, which differs from the previously reported value of 83,300 M⁻¹cm⁻¹ (104).

4.2.2.2 sfGFP photodeactivation: azPhe incorporation at residue 145

sfGFP^{F145AzF} was produced as a highly fluorescent protein with spectral properties similar to sfGFP; $\lambda_{\text{ex}} = 490$ nm and $\lambda_{\text{em}} = 511$ nm (Table 4.2). The similarity and high level of fluorescence (99 % brightness of sfGFP) indicates that incorporation of an additional azido group at the *para* position of the phenyl side chain of Phe145 had little effect on chromophore environment. The non-irradiated sfGFP^{F145AzF} had an extinction coefficient and QY of 47,700 M⁻¹.cm⁻¹ and 0.76, respectively, which was comparable to sfGFP (Table 4.2). On exposure to low intensity UV light, sfGFP^{F145AzF} rapidly loses fluorescence (Figure 4.3E,F) and overall brightness (Table 4.2). After irradiation, the ~490 nm peak in both the absorbance and excitation spectra decreases in intensity (to 11,400 M⁻¹.cm⁻¹) with the emergence of a non-radiative secondary peak at 386 nm in the absorbance spectrum (Figure 4.3D-F). Irradiation of sfGFP^{F145AzF} deactivates the protein to an effectively non-fluorescent form with a fold change of at least 53.

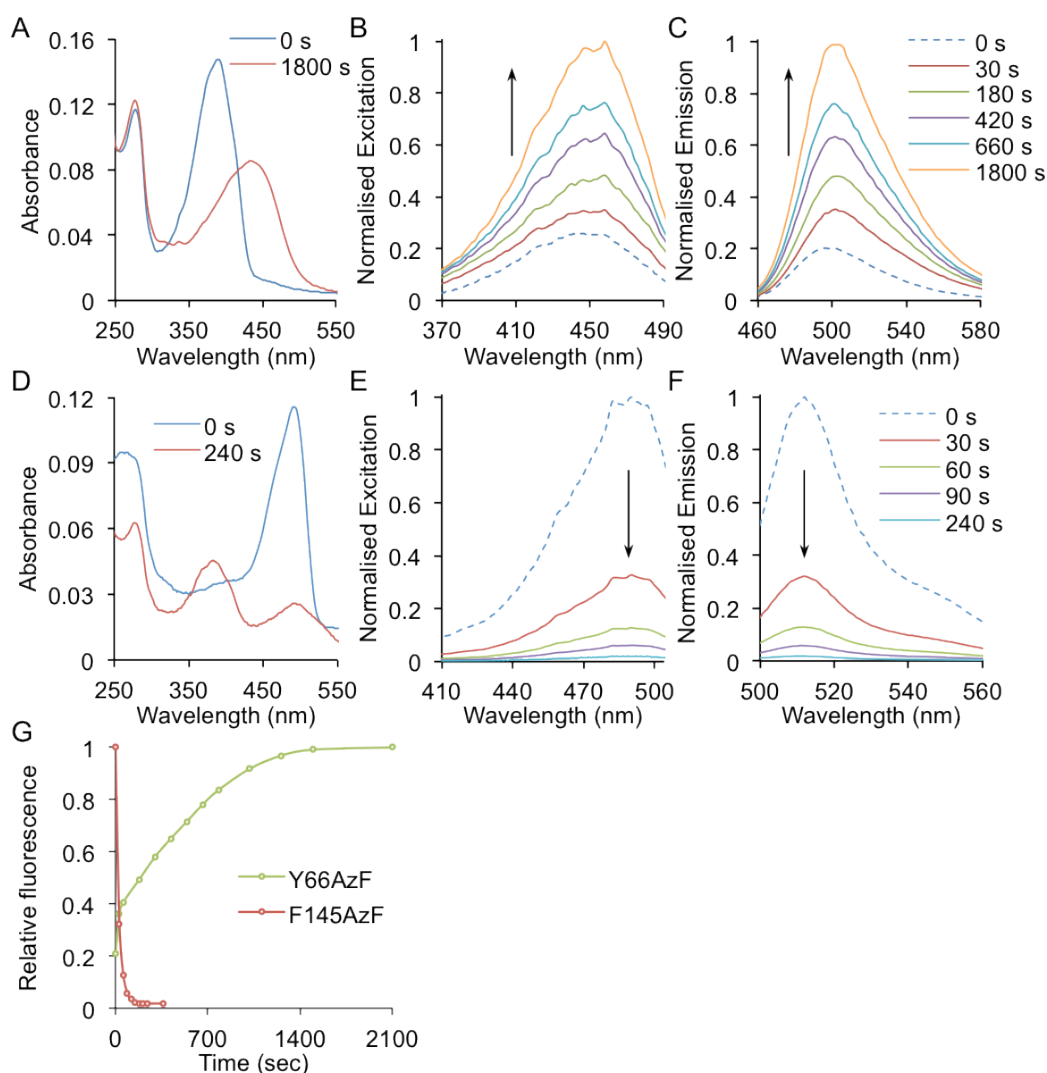


Figure 4.3. Spectral characterization of a photo-activatable and -deactivatable sfGFP variants. Photoconversion was performed by irradiation of purified protein (302 nm handheld UV lamp, 6 W) for the indicated amount of time. sfGFP^{Y66AzF} spectra are shown on the top row (A-C) and sfGFP^{F145AzF} on the bottom (D-F). UV-vis absorbance spectra of (A) sfGFP^{Y66AzF} and (D) sfGFP^{F145AzF} measured before (blue) and after (red) exposure with UV light. (B) Excitation spectra (emission monitored at 498 nm) of sfGFP^{Y66AzF} after 0 s (dashed blue), 30 s (red), 180 s (green), 420 s (purple), 660 s (cyan), 1800 s (orange) UV exposure. (E) Excitation spectra (emission monitored at 510 nm) of sfGFP^{F145AzF} after 0 s (dashed blue), 30 s (red), 60 s (green), 90 s (purple), 240 s (cyan) UV exposure. Fluorescence emission spectra of (C) sfGFP^{Y66AzF} (excitation at 446 nm) and (F) sfGFP^{F145AzF} (excitation at 490 nm) were measured with the same UV exposure time points as excitation spectra (B and E, respectively). Spectra were recorded with pure protein on 1 μ M for sfGFP^{F145AzF} fluorescence and 10 μ M for all others. (G) Rate of photo-activation and deactivation of sfGFP^{Y66AzF} (green) and sfGFP^{F145AzF} (red), respectively, determined by plotting the emission intensity against UV irradiation time.

4.2.2.3 sfGFP photoswitching: azPhe incorporation at residues 44, 148 and 203

Three variants, sfGFP^{L44AzF}, sfGFP^{H148AzF} and sfGFP^{T203AzF} displayed photoswitching with respect to maxima in their excitation (and absorbance) spectra and/or emission spectra. Prior to irradiation, sfGFP^{H148AzF} showed major and minor excitation peaks at 400 and 500 nm, respectively (Figure 4.4A,B, Table 4.2). On exposure to UV light, the 500 nm peak increased in intensity and blue shifted by 10 nm while the 400 nm peak decreased by ~85 % of its original intensity, with an isobestic point at 435 nm (Figure 4.4B). The λ_{em} remains at 511 nm, irrespective of excitation wavelength (400 nm or 500 nm) but intensity changed in line with the changes in the two λ_{ex} (Figure 4.4C,D; Table 4.2). Absorbance spectra pre- and post-irradiation mirrored changes in the excitation spectra, although the switch was less pronounced than the fluorescence excitation switch (Figure 4.4A). The 400 nm peak decreased in extinction coefficient from 34200 to 24000 M⁻¹.cm⁻¹ (70 % of original), while the extinction coefficient measured at 490 nm increased from 15800 to 34300 M⁻¹.cm⁻¹ (217 % of original). The more distinct fluorescent changes are a result of a change in QY of the two absorbance peaks; on irradiation the 400 nm peak decreased to 43 % of the original value, whereas the 490 nm peak increased by 159 % (Table 4.2). This ratiometric switching of excitation peaks means that the overall switching ratio is amplified (decrease in 400 nm × increase in 490 nm) giving an overall switch of 11 fold.

sfGFP^{T203AzF} proved to be very light-sensitive and rapidly reacted even on exposure to ambient light. In the non-irradiated state sfGFP^{T203AzF} had red-shifted fluorescence and absorbance spectra (λ_{max} 511 nm; λ_{em} 525 nm) in comparison to wild-type sfGFP (Figure 4.4E-G). On irradiation (UV lamp), sfGFP^{T203AzF} fluorescence blue-shifted by 8 nm and increased in intensity by 2.5-fold (Figure 4.4F,G). Absorbance matched the shift in wavelength (517 to 509 nm), however the peak intensity dropped to ~20 % of the original with the emergence of a non-radiative secondary peak at 393 nm (Figure 4.4E). The photoswitching behaviour in ambient light was investigated by visibly monitoring the colour of a cell lysate sample of an *E. coli* culture producing sfGFP^{T203AzF} over time (Figure 4.4H). Initially the cell lysate was red but rapidly lost its red colour over a 30 min period.

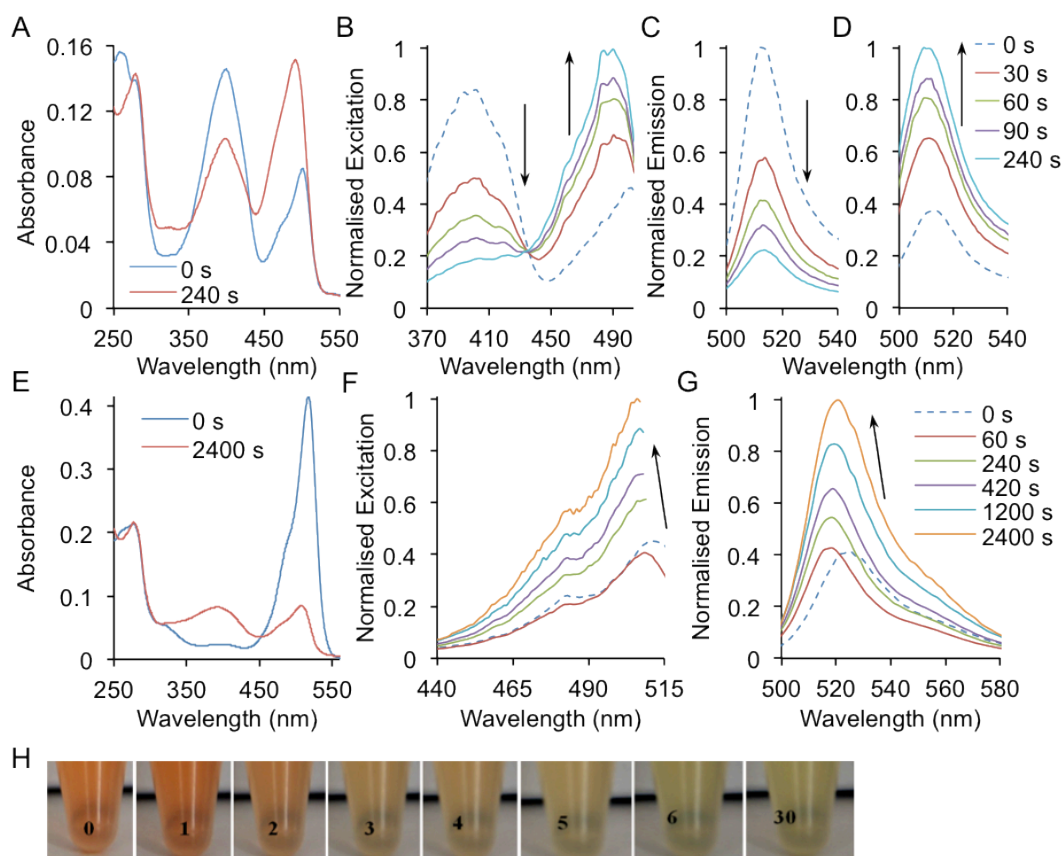


Figure 4.4. Spectral characterization of photo-switching sfGFP variants. Photoconversion was performed by irradiation of purified protein (302 nm handheld UV lamp, 6 W) for the indicated amount of time. sfGFP^{H148AzF} spectra are shown on the top row (A-D) and sfGFP^{T203AzF} on the bottom (E-G). UV-vis absorbance spectra of (A) sfGFP^{H148AzF} and (E) sfGFP^{T203AzF} measured before (blue) and after (red) exposure with UV light. (B) Excitation spectra (emission monitored at 511 nm) of sfGFP^{H148AzF} after 0 s (dashed blue), 30 s (red), 60 s (green), 90 s (purple), 240 s (cyan) UV exposure. (F) Excitation spectra (emission monitored at 525 nm) of azPhe145 after 0 s (dashed blue), 60 s (red), 240 s (green), 420 s (purple), 1200 s (cyan), 2400 s (orange) UV exposure. Fluorescence emission spectra of (C) sfGFP^{H148AzF} (excitation at 400 nm), (D) sfGFP^{H148AzF} (excitation at 490 nm) and (G) sfGFP^{T203AzF} (excitation at 485 nm) were measured with the same UV exposure time points as excitation spectra (B and F, respectively). Spectra were recorded on 1 and 10 μ M pure protein for fluorescence and absorbance, respectively. (H) Ambient light photoswitching of sfGFP^{T203AzF}. Images of a cell lysate sample left in ambient room light for the indicated amount of time (in min).

sfGFP^{L44AzF} photoswitching was more complex, being dependant on sample preparation and conditions. Prior to irradiation, sfGFP^{L44AzF} had major and minor excitation peaks at 394 and 484 nm, respectively, (Figure 4.5A-C). In whole cell samples sfGFP^{L44AzF} displayed classic photoswitching with respect to its excitation and absorbance spectra, whereby the 484 nm peak increased in intensity and the 394 nm peak decreased with an isobestic point at 433 nm (Figure 4.5A). However, when the same experiment was performed with cell lysates the intensity of the 484 nm peak remained constant while the 394 nm peak decreased to effectively zero (Figure 4.5B). Furthermore when purified, both excitation peaks decreased on irradiation with the 485 nm peak decreasing to ~60 % of its original value and the 394 nm peak decreasing to essentially zero (Figure 4.5C). Interestingly, continued irradiation of the sample resulted in a small increase in the 484 nm peak with a further small decrease in the 384 nm peak. No changes were observed in the wavelength of emission and intensity changes followed the same pattern as the respective excitation spectra.

This environmental effect on sfGFP^{L44AzF} photoswitching was further investigated using purified sfGFP^{L44AzF} protein by altering the pH of the buffer and the redox conditions (Figure 4.5D-E). Changing pH had no effect on the overall pattern of photoswitching with both excitation peaks decreasing on irradiation (as with pure protein samples) but caused the excitation peaks to decrease at slightly different rates on irradiation. Redox potential however had a significant effect on photoswitching. All redox agents tested caused a decrease in both excitation peaks (Figure 4.5D-1) however reducing agents resulted in a secondary increase in the 484 nm excitation peak (Figure 4.5D-2; as with no additive). The extent of this secondary increase varied depending on the redox agent with stronger reducing agents giving rise to larger increases with the pattern DTT > reduced glutathione > ascorbic acid (all 1 mM). This increase is shown in Figure 4.5E with the excitation spectra at the start (un-irradiated) and endpoint of irradiation. Promoting oxidising conditions with hydrogen peroxide (H₂O₂) resulted in a large decrease to both excitation peaks post irradiation, with no secondary increase (Figure 4.5E). Photoswitching of sfGFP^{L44AzF} was also monitored by UV-vis absorption, which showed a ratiometric switch (decrease in 394 nm, increase in 484 nm) for all redox agents as with the fluorescence excitation spectra of whole cell samples (Figure 4.5F).

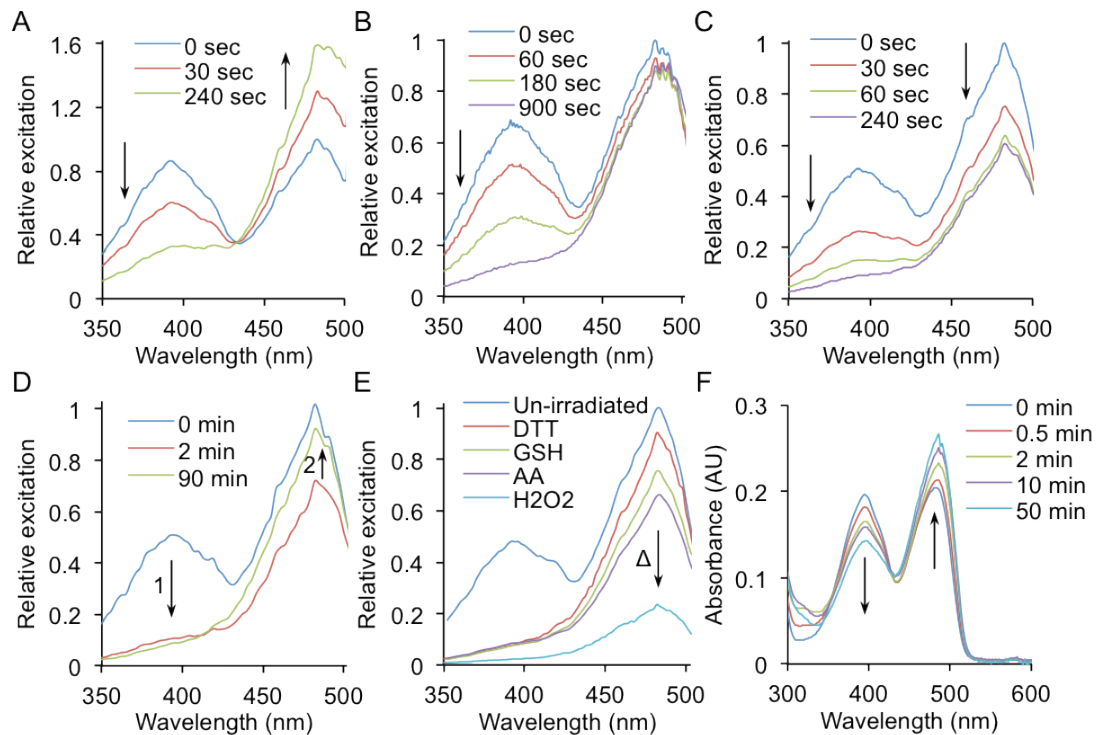


Figure 4.5. Spectral characterization of sfGFP^{L44AzF} photoswitching. Photoconversion was performed by irradiation (302 nm handheld UV lamp, 6 W) of (A) whole cells, (B) cell lysates or (C) pure protein producing sfGFP^{L44AzF} for the indicated amount of time. Excitation spectra were recorded by monitoring emission at 511 nm. Spectra of pure sfGFP^{L44cyF} were recorded with 1 μ M protein and an OD₆₀₀ of 1 for whole cell and lysates. (D) Photoconversion of 1 μ M sfGFP^{L44AzF} in 1 mM DTT showing the spectrum before (0 min), midway (2 min) and after (90 min) photolysis. The spectra show (1) the initial decrease in both excitation peaks followed by (2) an increase in the 484 nm peak. (E) Photoconversion of 1 μ M sfGFP^{L44AzF} under varying reducing and oxidising conditions. Fluorescence excitation (emission monitored at 511 nm) readings were taken before irradiation (un-irradiated) and then after photolysis in 1mM DTT, reduced glutathione (GSH), ascorbic acid (AA) and hydrogen peroxide (H2O2). The endpoint of photolysis was 90 mins. The 484 nm excitation peak decreased to different levels as indicated by Δ . (F) UV-vis absorption spectra during photolysis of 10 μ M sfGFP^{L44AzF} in PB.

4.2.2.4 Irradiation of wild-type sfGFP: a control for azPhe photocontrol

Importantly, under the same irradiation conditions, wild-type sfGFP fluorescence was unchanged in terms of intensity and spectral properties (Figure 4.6). This shows that the photoreactive properties of the sfGFP^{AzF} variants are a result of the incorporation of azPhe. Additionally, as a control for the sfGFP^{L44AzF} photoswitching experiments (Section 4.2.2.3) wild-type sfGFP photolysis was performed with the range of pH and redox conditions. No changes were observed under any of these conditions showing that azPhe44 was responsible for the fluorescence changes in sfGFP^{L44AzF}.

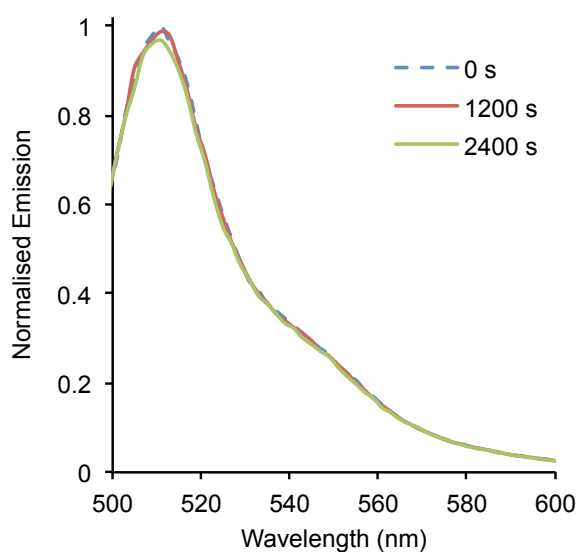


Figure 4.6. Effect of UV exposure on wild-type sfGFP fluorescence. Irradiation was performed as in Figure 4.3 and Figure 4.4 for the indicated amount of time. Emission spectra (excitation at 485 nm) of sfGFP after 0 s (dashed blue), 1200 s (red), 2400 s (green) UV exposure.

4.2.3 Live cell imaging of *E. coli* using sfGFP^{Y66AzF} and sfGFP^{F145AzF}

If the variants generated here are to be useful tools for cell biology (*vide infra*) they must exhibit their photoreactive changes inside cells. This is also an important prerequisite if azPhe is to be used as a general method of photocontrolling other target proteins *in vivo*. The photoconversion characteristics of sfGFP^{Y66AzF} (activation) and sfGFP^{F145AzF} (deactivation) *in vivo* were demonstrated through live cell imaging of *E. coli*. Initially widefield fluorescence microscopy was used to assess photoconversion. The technique uses a white light source (usually from a mercury lamp) with optical filters, which allowed a range of excitation wavelengths to be assessed. Widefield microscopy revealed that sfGFP^{Y66AzF} could be simultaneously activated and imaged with a 430 nm excitation filter (Figure 4.7A). Fluorescent activation was rapid with maximum intensity reached after 0.4 sec (Figure 4.7A). Laser scanning confocal microscopy (LSCM) was then used to gain better control over the wavelength, power and time of excitation light on the sample. LSCM uses a laser to give more precise control over excitation wavelength, power and area (versus whole field of view for widefield). LSCM also has the advantage of spatial resolution due to a “pinhole” that eliminates out-of-focus and scattered light. LSCM allowed separate activation and visualisation of sfGFP^{Y66AzF} at 405 nm (0.05 mW delivering 0.285 W/cm² at the sample) and 488 nm (0.21 mW delivering 1.195 W/cm²), respectively (Figure 4.7C). Gain in fluorescent signal on photoactivation was ~20 fold with a $t_{1/2}$ of 5.0 sec at 0.05 mW power (Figure 4.7C). Once activated, very little loss of fluorescence was observed during imaging (Figure 4.7C).

Widefield imaging of sfGFP^{F145AzF} revealed rapid deactivation whilst imaging using a 490 nm excitation filter with cells becoming essentially non-fluorescent after ~0.5 sec (Figure 4.7B). Using LSCM, sfGFP^{F145AzF} was imaged at 514 nm (0.02 mW delivering 0.114 W/cm²) without deactivation, with deactivation initiated with a 405 nm laser at low energy output (0.63 mW delivering 3.584 W/cm²). The $t_{1/2}$ for fluorescence deactivation was 9.0 sec at 0.63 mW power input (Figure 4.7D). Fluorescence of individual cells producing either sfGFP^{Y66AzF} or sfGFP^{F145AzF} could be modulated on the application of a defined light source (Figure 4.7E-G). By defining regions of interest to be irradiated with 405 nm light it was possible to activate or deactivate single cells within a population of cells. With ever improving accuracy and resolution of microscopes it should be possible to expand this technique to examine a small sub-cellular population of proteins. Increasing intensity of the photoconversion laser (405 nm; ≤1.61 mW delivering 9.159 W/cm²), full photo-activation or –deactivation was achieved at high speed in a single step (400Hz scan rate, 512×512 image; pixel dwell time of ~5μs).

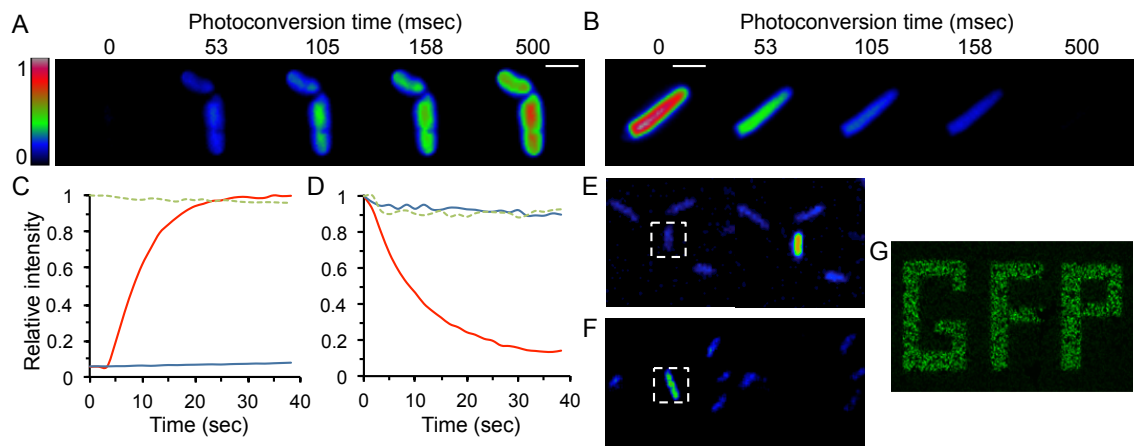


Figure 4.7. Photoconversion of sfGFP^{AzF} variants in live *E. coli* cells. Widefield microscopy images of cells producing (A) photoactivating sfGFP^{Y66AzF} or (B) photodeactivating sfGFP^{F145AzF}. Cells were photoconverted using a 430 nm and 488 nm filter, respectively, for the indicated amount of time. Images are false-coloured (scale shown on the far left). Scale bars = 2 μ m. The rates of photoconversion were monitored by confocal microscopy. Cells producing (C) sfGFP^{Y66AzF} or (D) sfGFP^{F145AzF} were photoconverted with a 405 nm laser (red trace). As a control, cells were not irradiated with 405 nm and only visualised pre-photoconversion (blue trace) at 488 nm (sfGFP^{Y66AzF}) or 514 nm (sfGFP^{F145AzF}). Cells producing wild-type sfGFP (green dashed) were irradiated at 405 nm under the same conditions as for photoconversion experiments. Photoconversion was quantified by monitoring individual cell fluorescence and plotting the value against time using Leica LCS software (sfGFP^{Y66AzF} n = 65, sfGFP^{F145AzF} n = 100). Spatial resolution of (E) sfGFP^{Y66AzF} photoactivation and (F) sfGFP^{F145AzF} photodeactivation using the FRAP module of the Leica confocal software. Regions of interest were selected for irradiation at 405 nm with the full field of view imaged with either a 488 nm (sfGFP^{Y66AzF}) or 514 nm (sfGFP^{F145AzF}) laser. (G) “Writing” in a layer of cells producing sfGFP^{Y66AzF}. The “GFP” region of interest was selected for irradiation at 405 nm as in (C) and (E).

4.3 Discussion

The incorporation of Uaa's into proteins offers a useful way of altering the properties of the protein. sfGFP was used as a model protein to assess whether incorporation of Uaa's could modify protein fluorescence. Uaa's can be designed to have altered chemistry such as polarity, H-bonding ability, electron-withdrawing ability and size. The introduction of such novel properties led to the generation of sfGFP variants with red-shifted fluorescence and altered excitation spectra, dependant on the Uaa used. The other aspect explored in this chapter is the use of a photoreactive Uaa, azPhe, to create whole proteins that can be controlled using light. Photocontrol is a particularly useful property to introduce into proteins given that control of protein and thus cellular activity can be achieved using a relatively benign, highly controllable effector. Again sfGFP was the target and azPhe incorporation generated variants capable of photo –activation, –deactivation and –switching.

4.3.1 A library approach to engineering proteins containing Uaa's

Proteins are complex molecules and despite our ever-increasing knowledge of structure, function and folding we are not in a position to engineer proteins in a fully rational way. Two strategies were employed for selection of residues for Uaa incorporation; rational design and a random library approach. Library approaches are used extensively in protein engineering as they allow many residues or mutations to be sampled without bias. Trinucleotide exchange (TriNEx) was used to randomly introduce amber stop codons into the *sfGFP* gene. TriNEx generated variants with interesting and useful properties that may have never been found with a fully rational approach. Leu44 is an internal residue facing towards the chromophore, however it doesn't make a direct interaction with the chromophore (Figure 4.1C) and has no mention in literature of GFP function. Mutations surrounding the tyrosyl group of the chromophore are known to influence protonation state, including those used to generate EGFP (eg. Ser65Thr; (96)). However, residue 44 has not been previously identified as one such mutation, especially given its remote location with respect to the tyrosyl group of the chromophore. Leu119 is even less of an obvious choice for mutation as it is further still from the chromophore. Both residues face the opposite side of the chromophore to the tyrosyl moiety, which is where the majority of residues are found that are recognized as being important (Figure 4.1C,D). Another benefit of a library approach is finding residues that were tolerant to Uaa mutagenesis that would have been overlooked as being too detrimental to protein structure. One such example is Pro75 of sfGFP, which is at the end of the central α -helix of sfGFP and was sampled in a TriNEx library by A. Baldwin and K. Streichert (unpublished data). The steric restraints of proline residues mean they often install important structural features, such as disruption of secondary structures, and as such are disregarded for

mutation. However because of the random library approach used, Pro75 was shown to be tolerant with mutation to various Uaa's generating highly fluorescent variants with similar properties to wild-type sfGFP (A. Baldwin and K. Streichert, unpublished results).

4.3.2 Uaa mutations not tolerated by sfGFP

It is important to discuss mutations that were not tolerated for future engineering attempts. In this context, lack of tolerance entails substitution of a residue with a particular Uaa resulting in a non-functional (non-fluorescent) protein. The implications of such a mutation are that it interferes with protein folding, final structure or the mechanism of fluorescence. For example, a mutation may be structurally tolerated (*ie.* it will generate a correctly folded protein) but may knock out function.

The Uaa studied here that was least tolerated was napAla. This is most likely because it is large, hydrophobic amino acid that could disrupt protein packing and/or folding. None of the positions in sfGFP tested in this chapter (Leu44, Tyr66, Leu119, His148, Thr203) produced useful levels of soluble, functional protein containing napAla. However, napAla was successfully tolerated at Gln204 (Chapter 3, Figure 3.5) probably because this residue is surface exposed and has fewer packing constraints.

The photocontrol approach presented here (Section 4.1.2) requires incorporation of azPhe into a protein of interest. Therefore for this approach to be of general utility azPhe must be well tolerated by a protein, which was generally found to be the case (Figure 4.2 to Figure 4.5). However some of the residues tested were not functionally tolerated including Thr62 and Glu222. Incorporation of azPhe at both residues resulted in the production of high levels of insoluble and therefore most likely incorrectly folded protein (Figure 4.8A). Thr62 and Glu222 are in the core of the protein in a tightly packed and highly ordered region (Figure 4.8B and C). Both form part of the complex H-bonding network in the core of the protein that includes structural waters and mediates proton transfer with the chromophore (92, 101). Glu222 is also important in proton transfer to and from the chromophore in ground-excited state cycling and in defining the charge state of the chromophore (Section 1.7, (92)). Moreover, Thr62 (along with Glu222) have roles in chromophore maturation (via a conserved H-bond with His181 (101)), hence mutations may prevent the chromophore forming in the first place. Considering the important structural (and functional) position of these residues it is not surprising that mutation to an aromatic amino acid with a very different structure resulted in non-functional protein. The production of (mainly) insoluble protein suggests that azPhe incorporation is affecting protein folding or final structure, possibly by preventing chromophore maturation or altering the core protein packing.

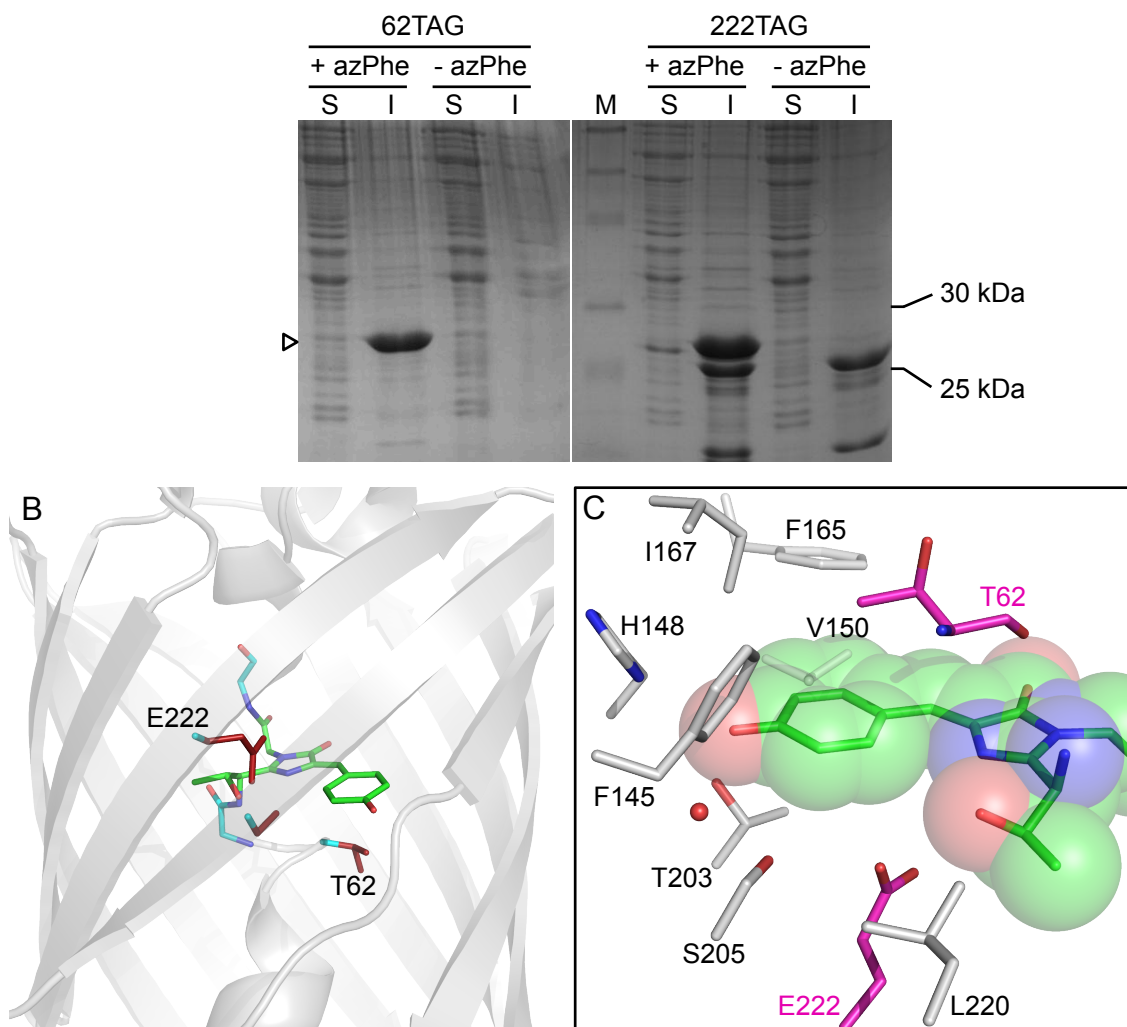


Figure 4.8. Residues not tolerant to azPhe incorporation. (A) Production of insoluble protein. SDS-PAGE analysis of soluble (S) and insoluble (I) fractions of sfGFP^{T62AzF} (left; 62TAG) and sfGFP^{E222AzF} (right; 222TAG) production. Parallel cultures were grown in the presence (+) and absence (-) of azPhe. The predicted size (~28 kDa) of full-length protein is indicated by the white arrow. (B) 3-dimensional structure of sfGFP (PDB accession 2B3P) showing residues that were not tolerant to azPhe as red sticks. (C) Chromophore environment showing surrounding residues as sticks. Thr62 and Glu222 are coloured by element with carbons coloured pink. The space occupied by the chromophore is displayed as spheres and shows the packing of neighbouring residues around the chromophore.

4.3.3 Molecular explanation of altered fluorescent properties in sfGFP^{Uaa}

Some of the variants of sfGFP studied displayed interesting and potentially useful changes in fluorescence. Given that ultimately these effects on function would ideally be applied in general to the proteome, it is important to understand the molecular basis of such changes for future rational engineering efforts.

4.3.3.1 Leu44

In several cases, the Uaa substituted sfGFP variants resulted two distinct excitation peaks (~400 nm and ~490 nm; Figure 4.2A and Table 4.1). An excellent example of such fluorescent changes are the mutations centred on Leu44. Classically, in GFP-like proteins two excitation peaks correspond to the protonation state of the ground state chromophore tyrosyl moiety with ~400 and 485 nm arising from the neutral and anionic forms, respectively. On substitution of Leu44 the proportion of the neutral peak (in relation to anionic peak) increased with the pattern Tyr > cyPhe > tfmPhe > azPhe > Leu (Table 4.1). Tyr44 exists effectively fully in the neutral form (394 nm) and Leu44 fully anionic (485 nm) (Figure 4.2A). Leu44 points towards the core of the protein (at 3.5 Å from the chromophore) and is surrounded by other hydrophobic residues that contribute to packing. Mutation of the small, hydrophobic amino acid to large aromatic Uaa's with variable polarity may have changed the hydrophobic packing at the core of the protein and resulted in a different structural arrangement. Additionally, the excitation and emission peaks were blue-shifted by a small amount in sfGFP^{L44cyF}, sfGFP^{L44tfmF} and sfGFP^{L44Y}, which can be indicative of changes in the residue packing and contacts around the chromophore. The exact mechanism by which a change in protein packing could affect the protonation state of the distal chromophore is currently unknown but could be the result of disruption or alteration to the H-bonding network at the core of GFP-like proteins that is key for function (92). A change in protein packing (caused by mutation of Leu44) could cause shifts and therefore alter the roles of important nearby residues, such as Thr65 or Glu222 that are part of the H-bonding network and are involved in proton cycling and the charge state of the chromophore (92). Alternatively (or in combination with structural packing) the introduction of a new chemical group capable of H-bonding might extend or modify this H-bonding network. Such changes could restore the H-bond between Thr65 and Glu222 (as in wt-GFP with Ser65), which could lead to deprotonation of Glu222 as its charged form would now be stabilized by the H-bond (Section 1.7, (99)). The charge on this anionic Glu222 would prevent ionization of the chromophore and hence promote the neutral 400 nm form. The large differences in excitation spectra were caused by relatively similar structured Uaa's with quite different ring substituents (*eg.* cyano *versus* trifluoromethyl group). This suggests that the explanation is not a change in structural packing due to the size of the Uaa alone but an alteration (or extension) of the H-bonding network.

Alternatively, disruption of the proton wires that feed the chromophore with protons from the solvent (and recycle protons from the chromophore to exit back to the solvent) (101). Specifically, Thr43 forms part for the proton entrance wire near the surface of the protein that precedes the internal wire for delivery to the chromophore (101). Mutation of the adjacent Leu44 may cause a shift in this region (*vide supra*) that could promote or even knockout the proton wire. The differences observed between Uaa's of similar size but with differing polarity

indicates some sort of charge effect. Mutating Leu44 to charged amino acids as well as solving an X-ray crystal structure of an sfGFP^{L44Uaa} variant with shifted excitation (eg. sfGFP^{L44cyF}, sfGFP^{L44Y}) could provide the answer.

4.3.3.2 His148

As with Leu44, mutation of His148 resulted in changes in excitation spectra in terms of the λ_{ex} observed (Figure 4.2B). In contrast to Leu44, His148 is a dynamic residue on the β -barrel of sfGFP that can point inward and H-bond with the tyrosyl moiety of the chromophore (Sections 1.7 and 4.2.1). His148 is well-documented to play a major role in the protonation state of the chromophore (173) and promotes ionisation to form the anionic (~485 nm) form (92). Therefore, the observed changes in excitation spectra are unsurprising, however the particular pattern is more difficult to explain. The proportion of the neutral peak (in relation to the anionic peak) increased with the pattern Tyr = azPhe > cyPhe > His > tfmPhe at residue 148. H-bonding and the presence of proximal polar groups increases the ionisation and hence deprotonation of the chromophore (100). TfmPhe contains the most polar group (Section 4.1.1) and the sfGFP^{L148tfmF} chromophore exists solely in the anionic form. Conversely, the hydroxyl group of Tyr and azido the group of azPhe are both polar however they have the reverse effect on excitation. The relatively long azido group may prevent azPhe from occupying the same conformation in the protein, removing the H-bond and neighbouring polar group from the chromophore.

Alternatively, the substitution to slightly larger amino acids may affect the packing around the chromophore, shifting residues that make important interactions with the chromophore or are involved with the proton shuttling network. Knocking-out (or even promoting) such an interaction would thereby alter the fluorescent properties of the protein. The different H-bonding abilities could even result in the Uaa's contributing to and therefore altering the H-bonding network around the chromophore. Again, changes to this important feature of sfGFP activity will manifest as changes in fluorescence. Both of these theories are supported by the red-shifted (by 15 nm to 500 nm) excitation wavelength of the anionic form of sfGFP^{H148AzF} and sfGFP^{H148Y}, which is indicative of a change in the environment of the chromophore. Shifts to longer wavelengths can be the result of new H-bonds with the chromophore reducing the energy for excitation. In a similar vain, His148 makes up part of an exit route for protons from the chromophore (101) so mutating it and making it (potentially) less dynamic could affect the ratio of chromophore protonation.

4.3.3.3 Thr203

The effects of Thr203 mutations (red-shifted fluorescence; Figure 4.2C) are easier to rationalise. A single Thr203Tyr mutation is responsible for the red-shifted fluorescent properties of yellow fluorescent protein (YFP) (97). In fact, mutation of Thr203 to any aromatic residue gives rise to red-shifted proteins (96). Thr203 faces the chromophore and mutation to the larger Tyr results in the residue lying parallel and in-plane with Tyr66 of the chromophore. This stacking of aromatic groups results in the delocalisation of π -electrons over both Tyr203 and Tyr66 (and hence the whole chromophore). This phenomena known as π -stacking results in an effective expansion of the delocalised electron system lowering the energy (longer λ) required to excite an electron, hence the observed red-shifted characteristics. Small differences were observed between sfGFP^{T203Uaa} variants that can also be explained by the level of delocalisation. Any groups on the phenyl ring at residue 203 would have an effect on the electronics of the ring and therefore the whole chromophore. Electron-donating groups (-OH, -NH₂) increase electron density to the whole chromophore and red-shift fluorescence whereas electron-withdrawing groups (-NO₂, -CO) cause blue-shifts. For example, trifluoromethyl was the most electron-withdrawing group tested due to the highly electronegative fluorine atoms and sfGFP^{T203fmF} had the shortest λ_{exc} and λ_{em} (Figure 4.2C).

4.3.4 Molecular explanation of photoswitching in sfGFP^{L44AzF}

As well as trying to understand the more traditional substitution mutations above, it is important to elucidate the molecular mechanism responsible for the photoswitching properties of the sfGFP^{AzF} variants. This will aid in the placement of azPhe in future rational engineering attempts with different residues or even different proteins. The photoconversion mechanisms of sfGFP^{Y66AzF}, sfGFP^{F145AzF} and sfGFP^{H148AzF} were investigated by X-ray crystallography and are discussed in greater detail in Chapter 5.

Substituting Leu44 for azPhe and subsequent photolysis resulted in changes in the proportion of the two excitation peaks (394 and 484 nm) with emission intensity changes mirroring the changes in excitation. The two excitation peaks correspond to the protonation state of the chromophore tyrosyl moiety with 394 and 484 nm likely to arise from the neutral and anionic forms, respectively. Photoconversion characteristics were very sensitive to environment with differing conditions (reducing and oxidising agents as well as whole cell and cell lysate samples) giving different photoconversion patterns. In whole cell samples, sfGFP^{L44AzF} is located in the cytosol which is reducing, at pH 7.2 – 7.8 and contains many salts, ions and macromolecules such as proteins (174). Under these conditions, a ratiometric switch was observed in the excitation spectra (decrease in 394 nm, increase in 484 nm) on photolysis. Cell lysates contain all of the same components but were diluted (in PBS) reducing

their effect on sfGFP^{L44AzF} photoswitching. Here, photolysis resulted in a decrease in the 394 nm peak while the 484 nm peak remained constant. Conversely, photolysis of samples of pure sfGFP^{L44AzF} showed decreases in both excitation peaks followed by variable increases in the 484 nm peak dependant on the addition of redox agents. There was correlation between the strength of reducing agent (*ie* redox potential) and photoconversion pattern. The stronger reducing agents (DTT > reduced glutathione > ascorbic acid) resulted in larger secondary increases in the 484 nm peak and hence switching more similar to the ratiometric increase in the 484 nm peak observed for whole cell samples. This can be rationalized as the cytosol is also reducing (*vide supra*). Further, when the protein is under oxidizing conditions (with the addition of 1 mM H₂O₂) no secondary increase in the 484 nm peak was observed. Together, this shows a key role of redox potential on sfGFP^{L44AzF} photoswitching and indicates that reducing conditions are required for ratiometric switching of sfGFP^{L44AzF} excitation spectra. Interestingly, the UV-vis absorption spectra during photolysis (Figure 4.5F) showed a ratiometric switch regardless of redox agent. This suggests that photolysis converts the sfGFP chromophore from the neutral to the anionic form under all conditions but that the anionic (484 nm) form is no longer fluorescent under certain conditions (*eg.* oxidizing conditions).

Leu44 points towards the core of the protein (3.5 Å from the former Thr65 (-CH₃) moiety of the chromophore; Figure 4.1) and is surrounded by other hydrophobic residues that contribute to packing. Mutation of the small, hydrophobic amino acid to azPhe may have changed the hydrophobic packing at the core of the protein and resulted in a different structural arrangement. The fluorescent changes on photolysis could be caused by the redox potential affecting the photochemical pathway taken by azPhe, with the chemical endpoint having different effects on fluorescence. For example, reducing conditions may promote reduction to an amine whereas oxidizing conditions might allow ring expansion and crosslinking, which are less favorable for fluorescence hence the deactivation. The alternative is that the photochemical pathway is unaffected and instead the endpoint form is more sensitive to redox potential than the wild-type protein. However, the difference between whole cell samples (ratiometric) and pure protein under reducing conditions (decrease in both peaks with a secondary increase in 484 nm) suggests that there are more factors at work than simply the redox potential. The cytosol is a complex environment and the various components (*eg.* salts, other stabilizing macromolecules) may be responsible for the difference observed, which is currently being investigated.

This interesting response to differing redox conditions opens up the possibility of using sfGFP^{L44AzF} (or other sfGFP^{L44Uaa} variants) as redox sensors, which are useful for monitoring oxidative stress in, for example, disease states. Leu44 mutations could also be useful in combination with existing redox sensors such as rxYFP¹⁴⁹₂₀₂ (175) to increase or modulate redox fluorescent switching.

4.3.5 Molecular explanation of photoswitching in sfGFP^{T203AzF}

The red-shifted fluorescent and absorption of sfGFP^{T203AzF} are most likely due to the well characterised π - π stacking interactions of residue 203 with the tyrosyl moiety of the chromophore (Section 4.3.3.3, (97)). The underlying photochemical mechanism responsible for photoswitching in sfGFP^{T203AzF} is currently unknown but probably also revolves around the π - π interaction. On irradiation, sfGFP^{T203AzF} absorbance, excitation and emission are blue-shifted by 8 nm and fluorescence increased to 250 % of the original. The increase in fluorescence suggests the removal of a quencher, which is likely to be the electron-rich azido group. Azido groups are used to quench small fluorescent probes (176) and could act in a similar manner in sfGFP^{T203AzF} as the chromophore is delocalised with residue 203. On sfGFP^{T203AzF} irradiation, azPhe203 loses N₂ and no longer acts as a fluorescent quencher, hence the increase in fluorescence.

Despite the observed blue-shift on photoconversion, the irradiated state of sfGFP^{T203AzF} is still red-shifted in comparison to wild-type sfGFP (517 nm *versus* 511 nm) suggesting that the π - π interaction is intact, but altered. Blue-shifts (higher energies) are indicative of a decrease in electron density of a delocalised system. A decrease in chromophore delocalisation could also be the result of poorer stacking and therefore a poorer interaction of residue 203 with the chromophore. This could be due to conversion of azPhe to a form that may not have such a strong π - π stacking interaction as a phenyl ring (*eg.* ring expanded ketenimine). Alternatively the photochemical transformation may result in a shift of residue 203 so that it can no longer stack with the chromophore. A photochemical endpoint that is no longer (or only partially) interacting with the chromophore is supported by the non-radiative secondary absorbance peak at 393 nm that could be the photoconverted form of azPhe203. Changes to the *para* ring-substituent of azPhe203 caused by irradiation could provide an alternative explanation for the decrease in delocalisation. A more electron-withdrawing or less electron-donating group than the azido group would contribute less electron density and hence decrease the conjugation with the chromophore. The azido group is weakly electron withdrawing (177, 178), however most of the potential photochemical endpoints are electron donating perhaps detracting from this explanation.

One potential endpoint of azPhe photochemistry could be probed by incorporation of the Uaa amPhe at residue 203. The spectral properties of the resulting protein could then be compared to the irradiated state of sfGFP^{T203AzF} to confirm or rule-out photoreduction to a phenyl amine. Ideally an X-ray crystal structure of sfGFP^{T203AzF} would elucidate the endpoint of residue 203 and its interaction with the chromophore.

4.3.6 Summary

This chapter has demonstrated how Uaa's can be used to alter the properties of a protein beyond what is possible with the 20 natural amino acids, using sfGFP fluorescence as an example. A combination of rational design and a library approach was used to generate sfGFP variants with altered spectral properties (Section 4.2.1) including significantly red-shifted fluorescence (sfGFP^{T203CyF}) that is useful for whole-animal imaging due to the increased tissue penetration of longer wavelengths of light. Section 4.2.2 demonstrated how a specific chemistry that is absent in natural amino acids, photoreactivity, could be used to generate useful proteins. AzPhe was incorporated into defined residues of sfGFP which itself altered the spectral properties of the protein (sfGFP^{Y66AzF} and sfGFP^{H148AzF}). The phenyl azide photoreactivity of azPhe was conferred to the whole protein generating variants that could be activated, deactivated and switched in response to light.

Photocontrollable FPs are desirable for super resolution microscopy via novel techniques such as PALM (Section 4.1.2). A range of photocontrollable FPs exist that can be grouped into three broad classes; (i) irreversible off-to-on, (ii) irreversible fluorescent state conversion (eg. green-to-red), or (iii) reversible fluorescent state conversion. PA-GFP (Class i) was one of the first photocontrollable FPs developed and irreversibly activates on application of high-intensity violet light due to decarboxylation of Glu222 (165). Kaede (Class ii), a naturally occurring photoreactive protein irreversibly converts from green to red fluorescence on irradiation due to a break in the protein backbone (179). However, Kaede is a tetramer so is unsuitable for protein tagging. Dronpa (Class iii) can be cycled between a green fluorescent form (by irradiation at 400 nm) and a non-fluorescent form (by irradiation at 470-510 nm) due to *cis-trans* isomerisation of the chromophore (180, 181). Unfortunately, in practice Dronpa is not as bright as in principle possibly due to overlap between the inactivation and excitation (488 nm) wavelengths (182). Although not perfect, sfGFP^{Y66AzF} and sfGFP^{F145AzF} have many benefits over existing photoswitching FPs (above) such as low (or at least comparable) energy of photoconversion (0.285 W/cm², 405 nm *versus* 30 kW/cm², 488 nm for PA-GFP (183) or <2.5 W/cm², 405 nm for PATagRFP (184)), being predominantly monomeric and exhibiting high switching ratios (Table 4.2). These characteristics pave the way for using sfGFP^{Y66AzF} to examine high-speed cellular events through photoactivation experiments using low intensity blue light to modulate fluorescence. Photodeactivable sfGFP^{F145AzF} could be useful as an alternative for fluorescent recovery after photobleaching (FRAP)-type experiments. Here, fluorescence is switched off by intense, prolonged irradiation and diffusion of fluorescent proteins (and hence recovery of fluorescence) into the bleached area is measured allowing the dynamics of a system to be examined (185). The advantage of sfGFP^{F145AzF} is sensitivity; the protein can be deactivated by relatively very low energy (3.584 W/cm², 405 nm *versus* 69-640 kW/cm², 488 nm for EGFP (186)). The sensitivity allows quicker 'bleaching' and therefore

more high-speed information can be gained as well as being less harmful to a target cell or organelle. Conversely the sensitivity makes sample preparation and imaging more difficult.

Controllably photoreactive proteins could be useful for studying important cellular processes such as cell signalling or as an alternative to a conditional knockout/knock-in mutation. For example by creating a photo-activatable or –deactivatable signalling receptor (or intermediate) specific cellular processes (*eg.* differentiation) could be modulated by light. An approach like this could afford total user control over cellular function for studying a specific protein, cellular process and/or signalling pathway (187, 188). Using light to control protein activity has many benefits such as high temporal and spatial control as well as being minimally invasive. This approach could also be expanded by using other photoreactive Uaa's that can undergo different reactions and hence cause different changes in activity. For example *p*-benzoyl-phenylalanine (bzPhe; Table 1.1) is a photocrosslinker (189) and has been incorporated into sfGFP with high orthogonality. The advantage of bzPhe is the reactive species is capable of relaxing back to the ground state if a crosslink is not formed which would increase yields of switching. Preliminary data showed that sfGFP^{BzF} variants did change fluorescent properties on irradiation, however the large size of bzPhe resulted in lower yields of functional protein produced. Other options include the photocrosslinking carbene precursor trifluoromethyl-diazirin-phenylalanine (190), photocaged amino acids such as nitrobenzyl-cysteine (Table 1.1) that release a natural amino acid on irradiation (13) and isomerizable amino acids like *p*-azo-benzene-phenylalanine that undergo a conformational change on irradiation (191).

This approach allies with the emerging discipline of optogenetics. The importance of autofluorescent proteins to cell imaging is clear, so our approach could be applied beyond sfGFP especially given the conserved nature of tyrosine in the chromophore or, as discussed, other proteins. Reprogrammed genetic code systems can now be implemented in eukaryotic (9) cells and whole animals (15, 16) making both the sfGFP molecular highlighters developed herein and the aryl azide-based photocontrol of proteins in general of broad applicability.

5. Elucidating the photochemical pathways of photocontrollable sfGFP variants

5.1 Introduction

In Chapter 4 azido-phenylalanine (azPhe) was incorporated into different residues throughout sfGFP. Some of these variants had fluorescent properties that could be modulated using light. This chapter aims to investigate these variants in more detail so as to understand the mechanism of action. Phenyl azides such as azPhe have well known and studied photoreactivity that can take multiple pathways and result in different chemical outcomes depending on local environment (Section 1.6.1; Figure 1.6). Outcomes include reduction, addition to double bonds, insertion in C-H and N-H bonds and crosslinking via attack by a nucleophile (Figure 1.6). It is important to understand the underlying molecular mechanism responsible for the photocontrollable properties of the sfGFP variants generated herein. This will aid in the placement of azPhe in future rational engineering attempts with different target proteins and could add to our understanding of phenyl azide chemistry in general. sfGFP^{Y66AzF} (activation), sfGFP^{F145AzF} (deactivation) and sfGFP^{H148AzF} (excitation switching) were deemed the most useful photoswitching properties (nature and stability of switch) and is thus the focus of this chapter.

5.2 Results

5.2.1 Investigating the molecular mechanism of sfGFP^{AzF} variants via incorporation of amino-phenylalanine

Incorporation of the Uaa *p*-amino-phenylalanine (amPhe, amF; Figure 3.5) allows one of the possible reaction endpoints to be probed so as to elucidate the underlying photochemical mechanisms responsible for the different photoreactive behaviours of sfGFP^{AzF} variants. As discussed above one potential pathway involves reduction of the reactive singlet nitrene to an aniline that can be mimicked by amPhe (Figure 5.8). By incorporating amPhe into sfGFP at the residues of interest (sfGFP^{AmF}) and comparing the spectral properties to the endpoints of sfGFP^{AzF} irradiation (and wild-type sfGFP) it was possible to confirm or rule-out the reduction pathway. The fluorescence spectra of sfGFP^{Y66AmF} were essentially identical to the irradiated form of sfGFP^{Y66AzF} (Figure 5.1A,B). sfGFP^{Y66AmF} had broad excitation and emission spectra peaking at 448 and 498 nm, respectively. Incorporation of amPhe at residue Phe145 (sfGFP^{F145AmF}) resulted in excitation and emission peaks at 488 nm and 510 nm, respectively, which were nearly identical to sfGFP (λ_{ex} 485 nm, λ_{em} 511 nm; Figure 5.1C,D). Incorporation

of amPhe at residue His148 did not generate the same fluorescence spectra observed for irradiated sfGFP^{H148AzF} (Figure 5.1E). Although the wavelength maxima of excitation and emission peaks were approximately the same for sfGFP^{H148AmF} and irradiated sfGFP^{H148AzF} the ratio of the two excitation peaks were very different. The major excitation peak in sfGFP^{H148AmF} was the neutral, 400 nm form which was the reverse of irradiated sfGFP^{H148AzF}. In fact, sfGFP^{H148AmF} was more similar to the non-irradiated form of sfGFP^{H148AzF} with a ($\lambda_{485}/\lambda_{400}$) ratio of 0.59 in comparison to 0.55 (Chapter 4, Figure 4.4).

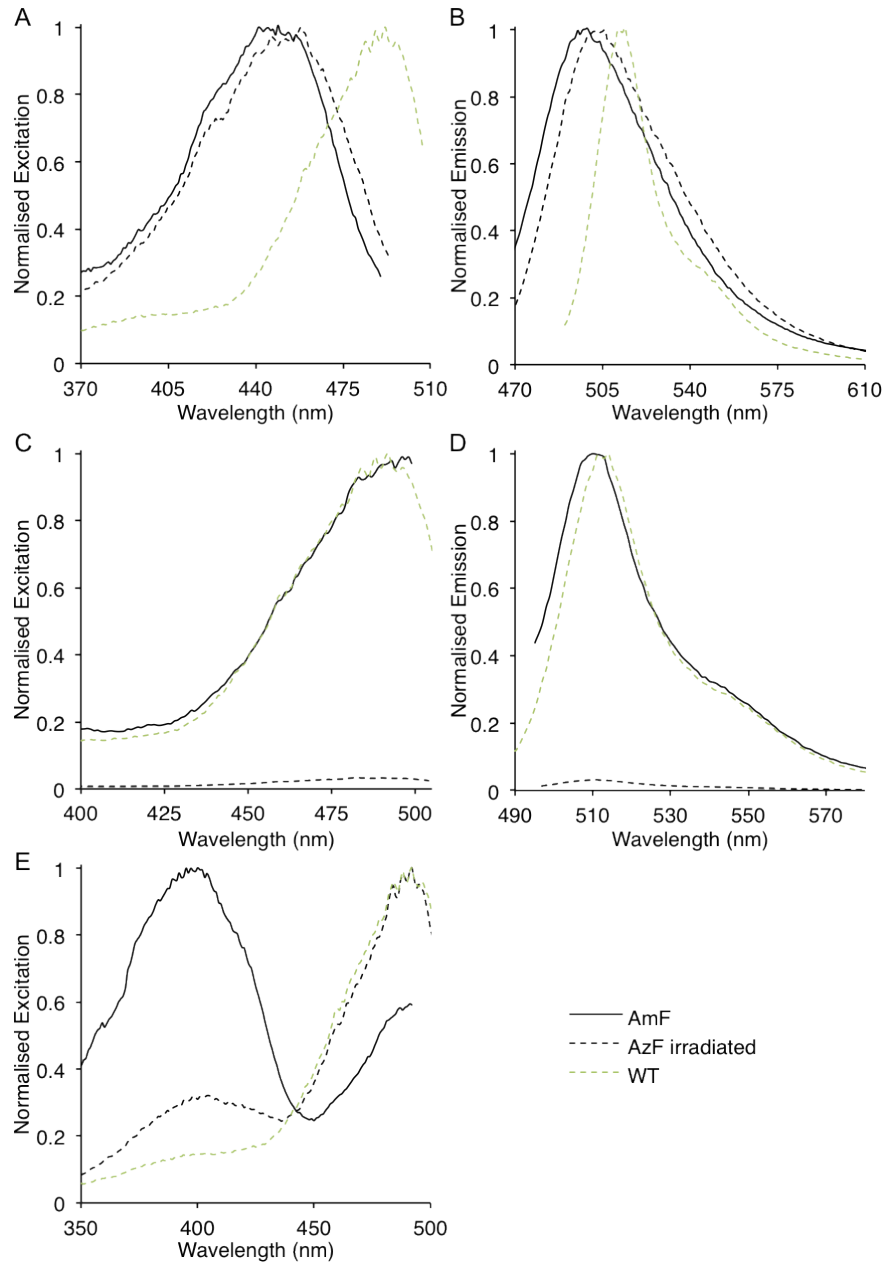


Figure 5.1. Spectral characterization of sfGFP variants containing *p*-amino-L-phenylalanine (amF). AmF was incorporated at residues Y66 (A, B), F145 (C, D) and H148 (E). Fluorescence excitation spectra of (A) sfGFP^{Y66AmF} (C) sfGFP^{F145AmF} and (E) sfGFP^{H148AmF}. Fluorescence emission spectra of (B) sfGFP^{Y66AmF} and (D) sfGFP^{F145AmF}. Spectra were recorded with 1 μ M protein in PBS. The spectra of wild-type sfGFP (green dashed) and the photoconverted (irradiated) forms of equivalent sfGFP^{AzF} variants (black dashed) are shown as a comparison to the sfGFP^{AmF} variants (solid black).

5.2.2 X-ray crystallography

5.2.2.1 Protein crystallization

The crystal structures of the dark and irradiated states of sfGFP^{Y66AzF} and sfGFP^{F145AzF} were solved so as to elucidate the photochemical pathway responsible for photoactivation and deactivation. Firstly, proteins were produced in high yield from large culture volumes and purified as in Section 2.5.3 and Section 3.2.3. The dark states of pure sfGFP^{Y66AzF} and sfGFP^{Y145AzF} were concentrated in 50 mM Tris-HCl, 150 mM NaCl, pH 8 (at 25 °C), to 28 and 19 mg/mL, respectively, and used to set up vapour diffusion crystal trays. Initially, small spiny crystals of sfGFP^{Y66AzF} grew in K&J screen condition A4 (100 mM Tris pH 8.3, 3 M (NH₄)₂SO₄) after approximately 2 months. Screens were then set up around this condition by varying the concentration of (NH₄)₂SO₄ (2 – 3.5 M) and pH (7.9 – 8.9). Protein and buffer were mixed in equal volume in all screens but with varying volumes (0.2, 0.5 or 1 μL) and equilibrated against 60, 85 or 150 μL of reservoir buffer at 4 or 18 °C. Also, further protein was produced, concentrated to 20 and 10 mg/mL and used to set up new crystal trays. The conditions that produced the best diffracting crystals for each variant can be seen in Table 5.1. At all stages of protein production, purification and crystallization, proteins were kept in the dark. After harvesting crystals of the dark state, crystal trays were irradiated as described in Table 5.1. Images of the crystals along with the X-ray diffraction detected can be seen in Figure 5.2.

Table 5.1. Crystallisation conditions of sfGFP^{AzF} variants that gave the best data sets. All of the crystals below were from plates incubated at 4 °C.

Variant	State	Buffer	pH	Precipitant	Protein concentration (mg/mL)	Volumes (μL; drop/reservoir)
sfGFP ^{Y66AzF}	Dark	100 mM Tris-HCl	8.3	2.8 M (NH ₄) ₂ SO ₄	10	0.2 / 60
sfGFP ^{Y66AzF}	Irradiated ^a	100 mM Tris-HCl	8.5	2.4 M (NH ₄) ₂ SO ₄	20	1 / 85
sfGFP ^{F145AzF}	Irradiated ^b	100 mM Tris-HCl	8.3	2 M (NH ₄) ₂ SO ₄	10	0.2 / 60

^a 3 × 8 W Hitachi FL8BL-B (λ_{\max} 350 nm) lamps, 3 × 8 W LZC-UVB (λ_{\max} 310 nm) lamps in a Luzchem LZC4 photoreactor for 30 mins. ^b 6 W UVM-57 mid-range UV handheld lamp (UVP; 1 cm, top down) for 90 min.

It is important to discuss the screens that failed to produce crystals and the crystals that did not diffract or diffracted to a lower resolution. Initially protein screens were set up using sfGFP^{Y66AzF} that was irradiated in solution, however this approach failed to produce crystals under a range of conditions. JBScreen Membrane (Jena Bioscience, Germany), K&J and Additive Screen MD1-11 (Molecular Dimensions, Suffolk, UK) were all assessed for crystal growth prior to finding the first condition that generated crystals (K&J screen condition A4). These small spiny crystals of sfGFP^{Y66AzF} were tested for X-ray diffraction but did not diffract to a usable level so a seeding screen was set up using fragments of these crystals in an attempt to initiate the crystallization process. Numerous crystals were generated from the final 100 mM Tris, (NH₄)₂SO₄ screen but only the crystals that diffracted to the highest resolution is discussed in depth below. Of note was a set of sfGFP^{F145AzF} crystals that crystallised in a different space group (hexagonal P6₅22) than the majority of the high-resolution crystals (orthorhombic P2₁2₁2₁). These crystals grew in 100 mM Tris pH 8.9, 2.6 M (NH₄)₂SO₄ and had a higher water content than the corresponding orthorhombic crystals. Unfortunately, these crystals diffracted to lower resolution (~2.6 – 2.9 Å) and the quality of data was insufficient to gain useful information about the photochemical pathway sampled. Also, the reflection data from one dark-state sfGFP^{Y66AzF} crystal that grew in 100 mM Tris pH 8.1, 2.6 M (NH₄)₂SO₄ could not be indexed (relating the data to atoms in the crystal and hence determination of unit cell dimensions and space group) and so could not be used.

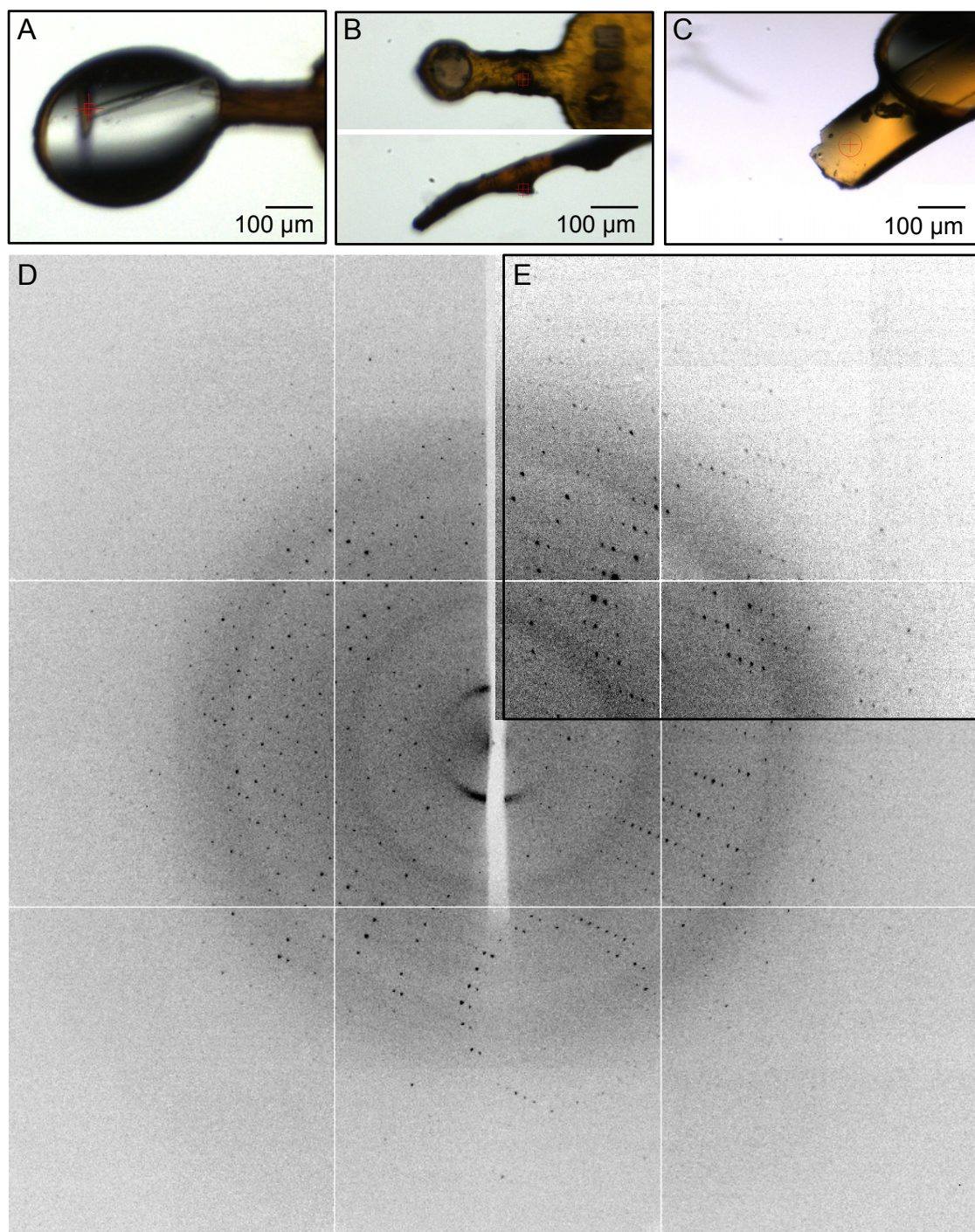


Figure 5.2. Protein crystals and X-ray diffraction. Images of crystals of (A) dark-state sfGFP^{Y66AzF}, (B) irradiated sfGFP^{Y66AzF} and (C) sfGFP^{F145AzF} on the X-ray beamline (red crosshair). Crystals are mounted in crystal loops (entering the image from the right) with irradiated sfGFP^{Y66AzF} (B) mounted on the stem of the crystal loop and shown from two angles. The crystals shown are the ones that gave the highest resolution data (Table 5.1 and Table 5.2). (D) Diffraction pattern generated from X-ray irradiation of sfGFP^{F145AzF} crystals (C; 1.26 Å). (E) Inset is a region of the diffraction pattern recorded in (D) with higher contrast to show the highest resolution diffraction.

5.2.2.2 Structural determination

Once crystals were grown, X-ray diffraction data was collected (Figure 5.2) at the Diamond Light Source (Oxfordshire, UK). The data was collected on the I03 and I04 beamlines (with Dr Pierre Rizkillah, School of Medicine, Cardiff University). Structures were solved by molecular replacement with wild-type sfGFP (PDB accession 2B3P) as a model using PHASER (127). Structures were adjusted manually using COOT (128) and refined by TLS restrained refinement (sfGFP^{Y66AzF} dark and irradiated) or mixed anisotropic/isotropic restrained refinement (sfGFP^{Y145AzF} irradiated) using RefMac. A more detailed description of structure determination is provided in Section 2.10. All structures were determined from proteins that crystallized in the orthorhombic space group P2₁2₁2₁ (full data collection and refinement statistics are summarized in Table 5.2). sfGFP^{Y66AzF} (dark and irradiated) and sfGFP^{F145AzF} (irradiated) contained two and one molecules in the asymmetric unit, respectively. The structures were determined to a resolution of 2.1 Å (sfGFP^{Y66AzF} dark), 2.1 Å (sfGFP^{Y66AzF} irradiated) and 1.26 Å (sfGFP^{F145AzF}), respectively.

Table 5.2. Crystal diffraction and refinement statistics.

Variant	Dark-state sfGFP ^{Y66AzF}	Irradiated-state sfGFP ^{Y66AzF}	Irradiated-state sfGFP ^{F145AzF}
PDB accession	4J88	4J89	4J8A
Data collection/reduction statistics			
Wavelength (Å)	0.9795	0.9795	0.9763
Beamline	I04	I04	I03
Space group	P2(1)2(1)2(1)	P2(1)2(1)2(1)	P2(1)2(1)2(1)
a (Å)	52.43	52.37	46.50
b (Å)	97.56	98.11	58.70
c (Å)	102.90	102.91	91.63
Resolution range (Å)	48.78 - 2.08	51.46 - 2.10	46.50 - 1.26
Total reflections measured	257300	245481	241496
Unique reflections	32504	31310	67853
Completeness (%) (last shell)	100.0 (100.0)	98.7 (100.0)	99.2 (98.8)
I/σ (last shell)	13.6 (3.1)	14.6 (3.3)	16.0 (2.2)
R(merge) ^a (%) (last shell)	13.0 (81.7)	9.2 (64.2)	3.2 (49.1)
B(iso) from Wilson (Å ²)	25.3	31.3	16.5
Refinement statistics			
Protein atoms excluding H	3662	3666	2028
Solvent molecules	351	326	323
R-factor ^b (%)	16.54	18.18	12.87
R-free ^c (%)	22.82	24.21	16.46
RMSD bond lengths (Å)	0.021	0.019	0.023
Ramachandran plot statistics			
Rmsd angles (°)	1.194	1.257	1.586
Core region (%)	97.39	96.49	97.73
Allowed region (%)	2.61	3.51	2.27
Additionally allowed region (%)	0.0	0.0	0.0
Disallowed region (%)	0.0	0.0	0.0

$$^a R_{merge} = \sum_h \sum_j (I_{hj} - \langle I_j \rangle) / \sum_h I_{hj}$$

$$^b R_{factor} = (\sum_h |F_{h,obs} - F_{h,calc}|) / (\sum_h F_{h,obs})$$

^c R_{free} is calculated from a set of 5% randomly selected reflections that were not included in the refinement.

5.2.2.3 General structural features of sfGFP^{AzF} variants

It is important to assess the effect of azPhe incorporation on protein structure in terms of gross structural differences as well as the more fine details around the mutated residue. This helps us to understand the effect of azPhe incorporation on protein structure and aid in the design of future mutations. The structures solved in this chapter were aligned with the parent scaffold, sfGFP (Table 5.3, Figure 5.3, Figure 5.4), to quantify differences in the backbones as well as between all individual residues (including side chains).

Table 5.3. RMSD measurements (in Å) between sfGFP^{AzF} variants and sfGFP

	Dark sfGFP ^{Y66AzF}		Irradiated sfGFP ^{Y66AzF}		Irradiated sfGFP ^{F145AzF}
	Chain 1 ^a	Chain 2	Chain 1	Chain 2	
C α only	0.322	0.447	0.433	0.488	0.624 (0.536) ^b
Backbone	0.331	0.444	0.432	0.491	0.613 (0.503)
All	0.412	0.499	0.506	0.567	0.703 (0.566)

^a dark and irradiated sfGFP^{Y66AzF} crystals had two molecules per asymmetric unit, which were aligned individually and shown as “Chain 1 or 2”.

^b the values in brackets are RMSD values calculated with only the major conformer for the Tyr143-His148 region of irradiated sfGFP^{F145AzF} (*vide infra*).

The dark structure of sfGFP^{Y66AzF} was very similar to wild-type sfGFP both in terms of C α , backbone and all residues with average RMSD values of the two molecules of 0.385, 0.388 and 0.456 Å, respectively (Table 5.3). The alignment can be seen in Figure 5.3A where the two proteins are effectively superimposable. A comparison of individual residue orientations showed a number of surface exposed residues (*eg.* Ser99, Tyr182 and Gln204) occupying different rotamers between sfGFP and sfGFP^{Y66AzF}, which is expected of dynamic residues in a crystal structure. A more detailed examination of the residues surrounding the chromophores also showed that the structures are very similar but indicated a few residues with different conformations (Figure 5.3B and C). His148 takes an alternate conformation in sfGFP^{Y66AzF} whereby the imidazole ring is flipped and the whole side chain is twisted away from the chromophore by ~0.4 Å. Thr203 has an alternate rotamer in sfGFP^{Y66AzF} where the side chain is rotated by 120° so the hydroxyl group no longer faces the *para* position of the chromophore as in sfGFP. The other significant change is twisting of Leu42 to face the former Thr65 moiety of the chromophore in sfGFP^{Y66AzF}.

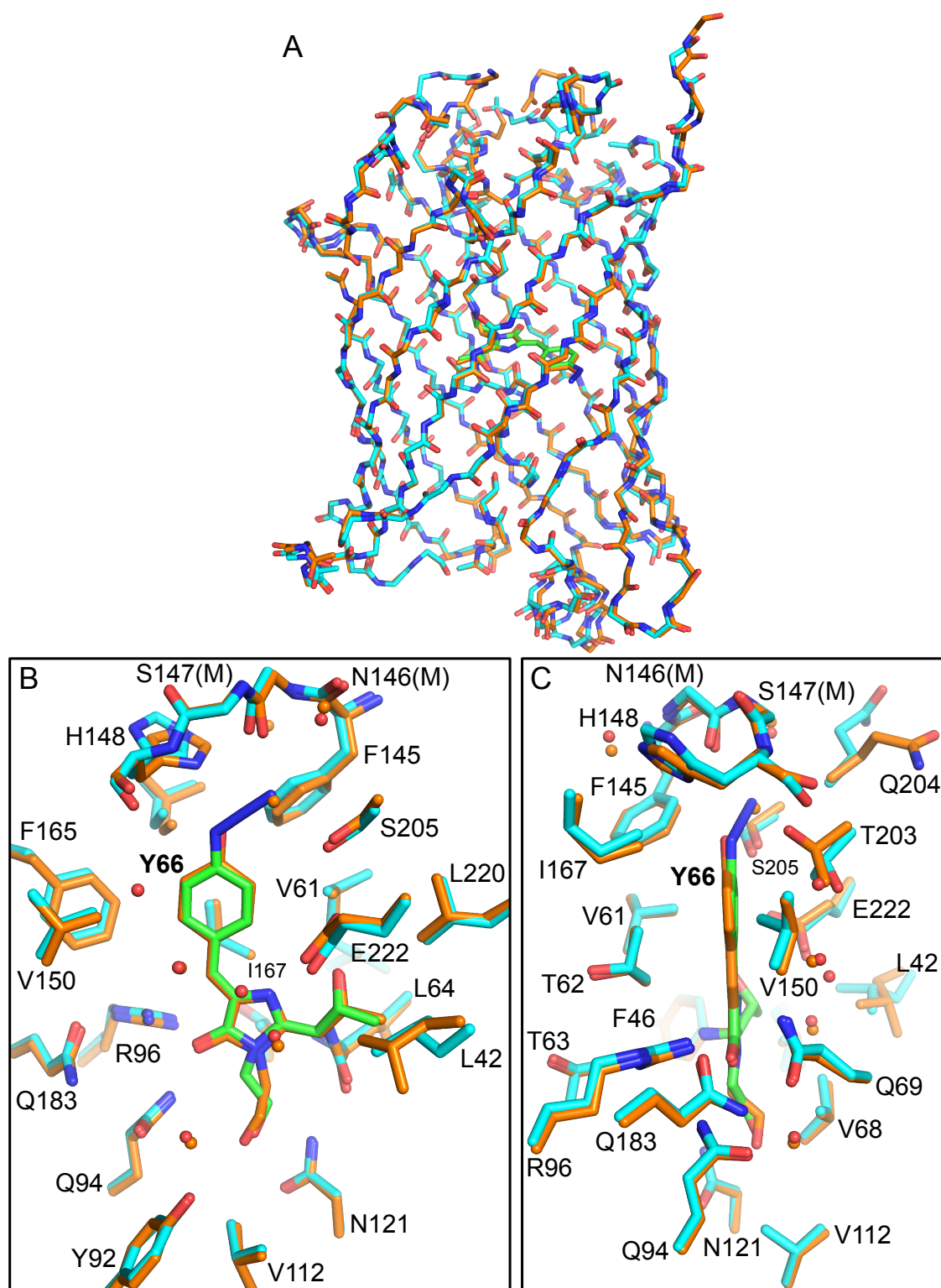


Figure 5.3. The effect of azPhe incorporation on sfGFP^{Y66AzF} structure. Structural alignments of wild-type sfGFP (orange) and sfGFP^{Y66AzF} (cyan) showing the (A) protein backbone only and (B and C) the environment of the chromophore. The chromophore in sfGFP^{Y66AzF} has carbons coloured green. The former Tyr66 residue (now part of the chromophore) is labelled in bold. Structural water molecules specific to wild-type and sfGFP^{Y66AzF} are shown as orange and red spheres, respectively. Only the main chain is shown for some residues as indicated by (M). PDB accessions 2B3P and 4J88.

Similarly to the dark form, the irradiated structure of sfGFP^{Y66AzF} aligned well with wild-type sfGFP giving very low RMSD values of 0.461, 0.462 and 0.537 Å, for C α 's, backbone traces and all atoms, respectively (Table 5.3, average of the two chains). Only small divergences were observed between backbone traces in flexible, surface regions of the proteins (eg. Phe100 – Thr105 and Gly189 – Leu195) with alternate rotamers for Arg73, Arg80, Lys113, Arg122, Glu124, Glu132, Glu172 and Gln204. Additionally, as with the other two structures solved here, the secondary structure of irradiated sfGFP^{Y66AzF} was unaffected by incorporation of azPhe.

The irradiated structure of sfGFP^{F145AzF} overlaid well with wild-type sfGFP (Figure 5.4A). Alignment revealed low RMSD values of 0.624 Å (C α), 0.613 Å (backbone) and 0.703 Å (all atoms) (Table 5.3). Some residues occupied different conformations in sfGFP^{F145AzF} and sfGFP, most of which were located in flexible loop regions (eg. Glu132 and Gln204). The only significant divergence between the backbone traces in the 'core' β -barrel was in the region surrounding residue 145. The backbone from Tyr143 to His148 in sfGFP^{F145AzF} diverged from sfGFP due to flipped peptide bonds between Tyr143 and Asn144, Asn144 and azPhe145, and Asn146 and Ser147 (Figure 5.4A). This alternate structure did not completely satisfy the electron density and the structure was eventually solved by including two different conformations of the Tyr143-His148 region (*vide infra*; Section 5.2.2.5). The second conformation was modelled as a minor (20 %) form and, like the major (80 %) form, had a different structure to wild-type sfGFP. When the alignment between sfGFP^{F145AzF} and sfGFP was repeated with just the major conformation the RMSD values were lower (C α , 0.536; backbone, 0.503; all atoms, 0.566). The double conformation of the Tyr143-His148 region was further investigated by calculating the thermal B-factors of sfGFP^{F145AzF} (Figure 5.4B). Structural (thermal) B-factors are an indication of thermal motion and hence dynamicity within a crystal structure. The B-factors were generally as expected with low values for structured β -strands and higher values in more flexible loop and turn regions. However, the Tyr143-His148 region showed unusually high B-factors (Figure 5.4B), which indicated a high degree of motion or a mixed population of conformers with highly different structures.

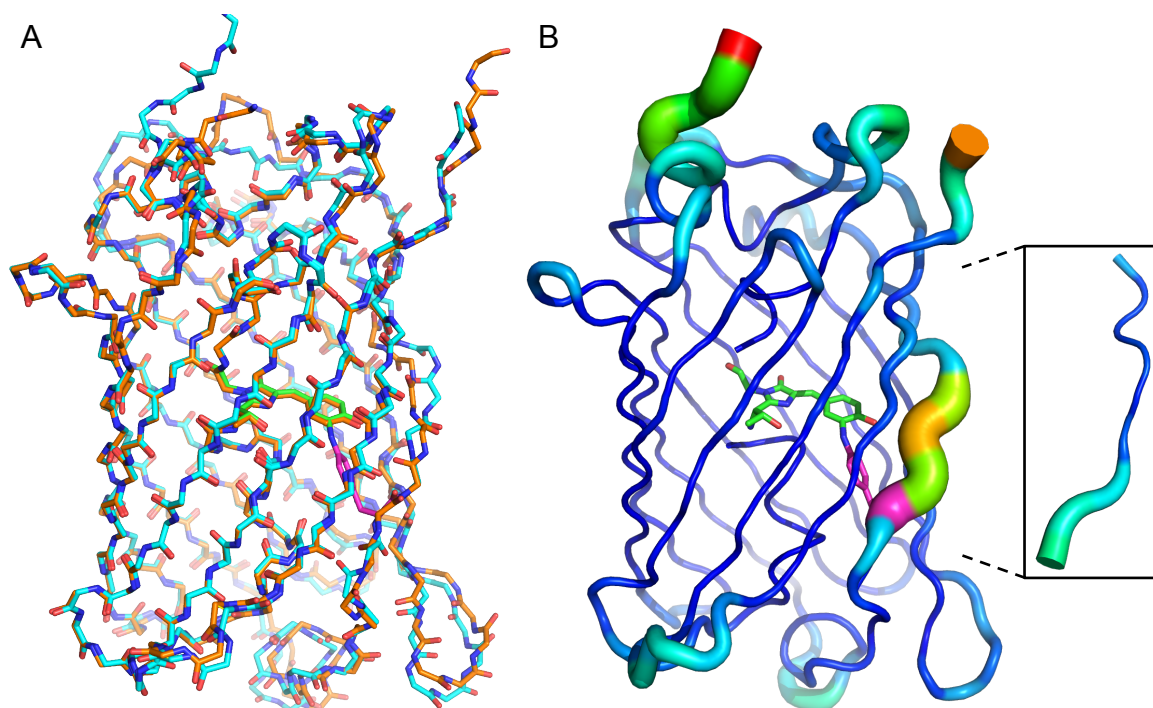


Figure 5.4. The effect of azPhe incorporation on sfGFP^{F145AzF} structure. (A) Structural alignments of the backbones of wild-type sfGFP (orange) and sfGFP^{F145AzF} (cyan). (B) Thermal B-factors of the irradiated structure of sfGFP^{F145AzF}. Thermal B-factors are shown on a relative scale by thickness (thin/low - thick/high) and colour spectrum progression (blue/low - red/high). The thermal B-factor representation of wild-type sfGFP for the backbone region harbouring residue 145 (Asn144 to Ile152) is shown inset (right). In both figures the chromophore and residue 145 in sfGFP^{F145AzF} have carbons coloured green and pink, respectively. PDB accessions 2B3P and 4J8A.

5.2.2.4 Structural comparison of the dark and irradiated states of sfGFP^{Y66AzF}

The structure of the dark state of sfGFP^{Y66AzF} confirmed the presence of the azido group at the *para* position of the tyrosyl moiety of the chromophore (Figure 5.5A and B). The azido group points away from the core of the protein into a cavity formed by Phe145, Ser205 and the main chains of Asn146 and Thr203. The terminal N of the azido group (N_γ) is within H-bonding distance of Asn146 (3.0 Å), Ser205 (3.0 Å) and Thr203 (3.2 Å). Generally the residues immediately surrounding the azido group (Phe145, Asn146 (main chain), Phe165, Ile167, Ser205 and Glu222) are shifted away by ~0.3 – 0.6 Å in comparison to wild-type sfGFP, probably as a result of the additional electron density of –N₃ compared to –OH (Figure 5.3B and C).

The *in crystallo* irradiated structure revealed no electron density past N_α of the former azido group indicating the loss of N₂. The density at the *para* position of the ring was therefore modelled as –NH₂ (Figure 5.5C and D). Generally, the final structures of dark and irradiated sfGFP^{Y66AzF} showed little positive (green) or negative (red) density (Figure 5.5). This implies fair refinement that is representative of the actual structure, which placed confidence in the

assignments. Density equivalent to N₂ was not detected, as the side chain of azPhe66 is near the exterior of the protein surrounded by dynamic residues on the β -barrel (eg. His148) allowing easy diffusion out of the protein. The structures of the dark and irradiated states are very similar with an RMSD for C α and backbone residues of 0.173 and 0.179 Å, respectively, and 0.227 Å for all atoms. Small shifts (0.7 – 1.2 Å) were observed in the water molecules that lie parallel to and in the plane of the chromophore between the dark and irradiated structures (Figure 5.5A versus C). Electron density associated with several residues suggested that their side-chains had double conformations. These included residues Ser147 in the dark-state and His148 in the irradiated-state that lie close to the chromophore (Figure 5.5). These double conformations are most likely due to the dynamic nature of these residues. After irradiation (Figure 5.5C and D), a water molecule takes the place of N γ of the azido group within hydrogen bonding distance (3 Å) of the putative amine. This water is conserved in most GFP-like proteins that contain tyrosyl chromophores and plays a role in defining the spectral properties of the protein (92).

The chromophore in the dark and irradiated structures has slightly different conformations with relation to the angle between the imidazolyl and tyrosyl moieties. The dark (*ie.* azido) chromophore is slightly less twisted than the irradiated (*ie.* amino) chromophore (135.4° versus 133.0°) that results in the *para* position of the chromophore shifting by ~0.4 Å. The residues surrounding the tyrosyl moiety of the chromophore reflect this small difference with similar ~0.4 Å shifts. The difference may be a result of the azido group fitting into the cavity (*vide supra*) with the minimal possible energy and hence the greatest distance from the surrounding residues.

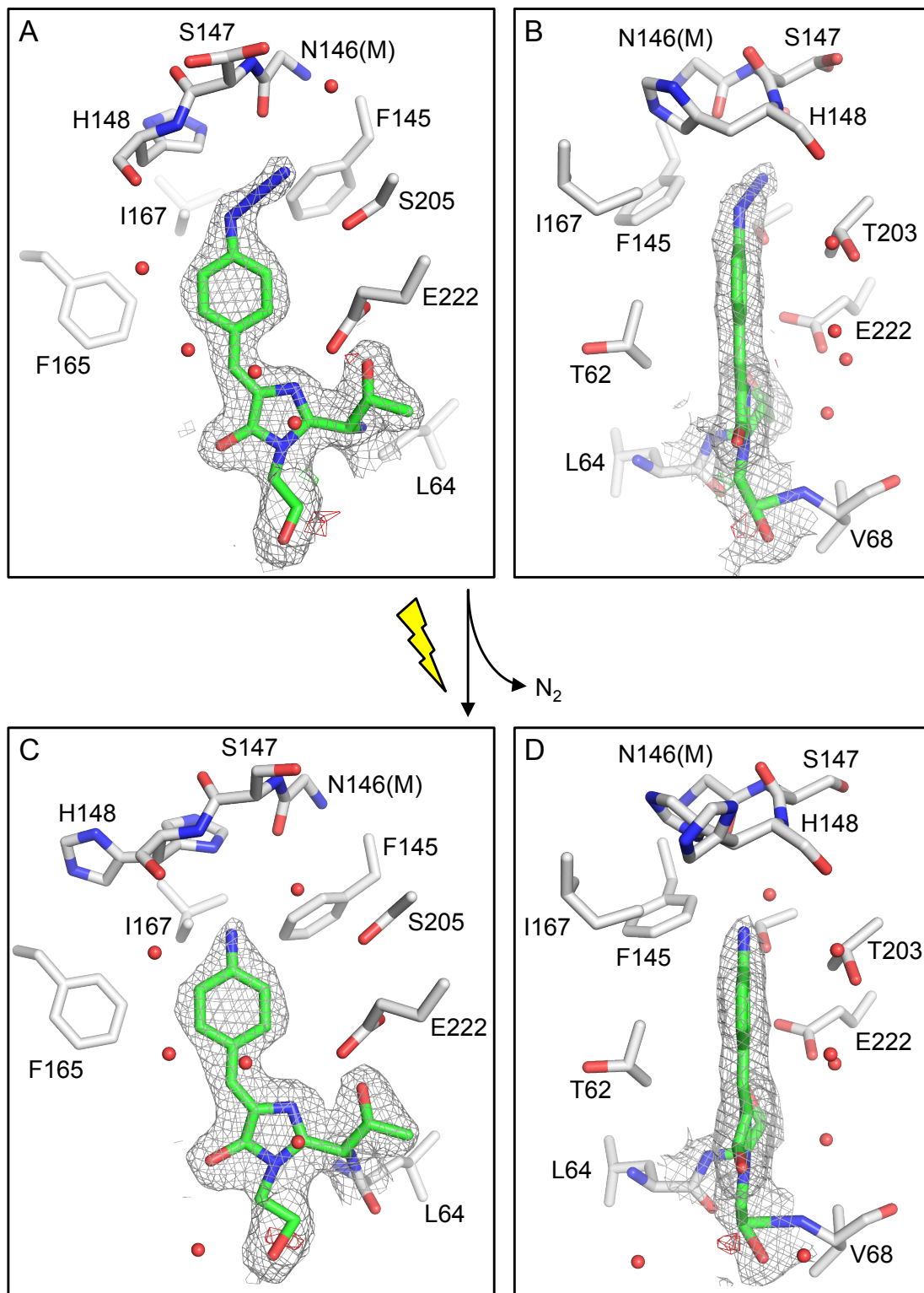


Figure 5.5. X-ray crystal structure study of sfGFP^{Y66AzF} photoreaction. 2Fo-Fc electron density maps of the (A and B) dark and (C and D) irradiated states of photoactivable sfGFP^{Y66AzF}. Electron density is displayed to 1 σ . Positive and negative electron density (if present) is coloured green and red, respectively, and displayed to $\pm 3 \sigma$. Atoms are coloured by element with chromophore carbons coloured green. Double conformers were observed for S147 in dark-state sfGFP^{Y66AzF} and H148 in irradiated sfGFP^{Y66AzF}. Only the main chain of N146 is shown and is indicated by (M). Water molecules are displayed as red spheres. Crystals of the sfGFP^{Y66AzF} variant were grown in the dark and irradiation was performed *in crystallo* as described in the text. PDB accession 4J88 and 4J89.

5.2.2.5 Structural analysis of the irradiated state of sfGFP^{F145AzF}

The crystal structure of the *in crystallo* irradiated sfGFP^{F145AzF} (1.26 Å resolution) revealed electron density for a crosslink from the nitrene N (of residue 145) to the *meta* carbon of the chromophore tyrosyl moiety (Figure 5.6A and B). When this new crosslinked conformation was modelled there was positive electron density along the Tyr143 to His148 region suggesting presence of a second conformer (Section 5.2.2.3). The final refined structure of irradiated sfGFP^{F145AzF} contains two conformations of the Tyr143 to His148 region, one where residue 145 is crosslinked to the chromophore that represents 80 % and one without the crosslink that has significantly different backbone and side chain structure that represents 20 % (Figure 5.6C and D). Although the second (non-crosslinked) conformation was modelled with a Phe at residue 145 (not azPhe) the presence of two conformations suggested incomplete photoconversion to the crosslinked form. The newly formed crosslink appears to have caused structural changes in the β-barrel region housing residue 145 (Tyr143 to His148) in comparison to the second non-crosslinked conformer and to wild-type sfGFP. The Tyr143-Asn144 peptide bond has flipped by ~150° in the crosslinked form allowing the backbone from Asn144 to Ser147 to take a different path with side chains occupying very different conformations in the non-crosslinked form (Figure 5.6C and D). In the crosslinked form of sfGFP^{F145AzF} the side chains of Asn144, Ser147 and His148 are also flipped in comparison to wild-type sfGFP (*vide infra*). However the backbone of the crosslinked form of sfGFP^{F145AzF} matches more closely with wild-type sfGFP than the non-crosslinked form. The result (or cause) of the structural rearrangements is the putative nitrene of residue 145 rotating and approaching the chromophore by ~2.3 – 2.5 Å to reach bonding distance.

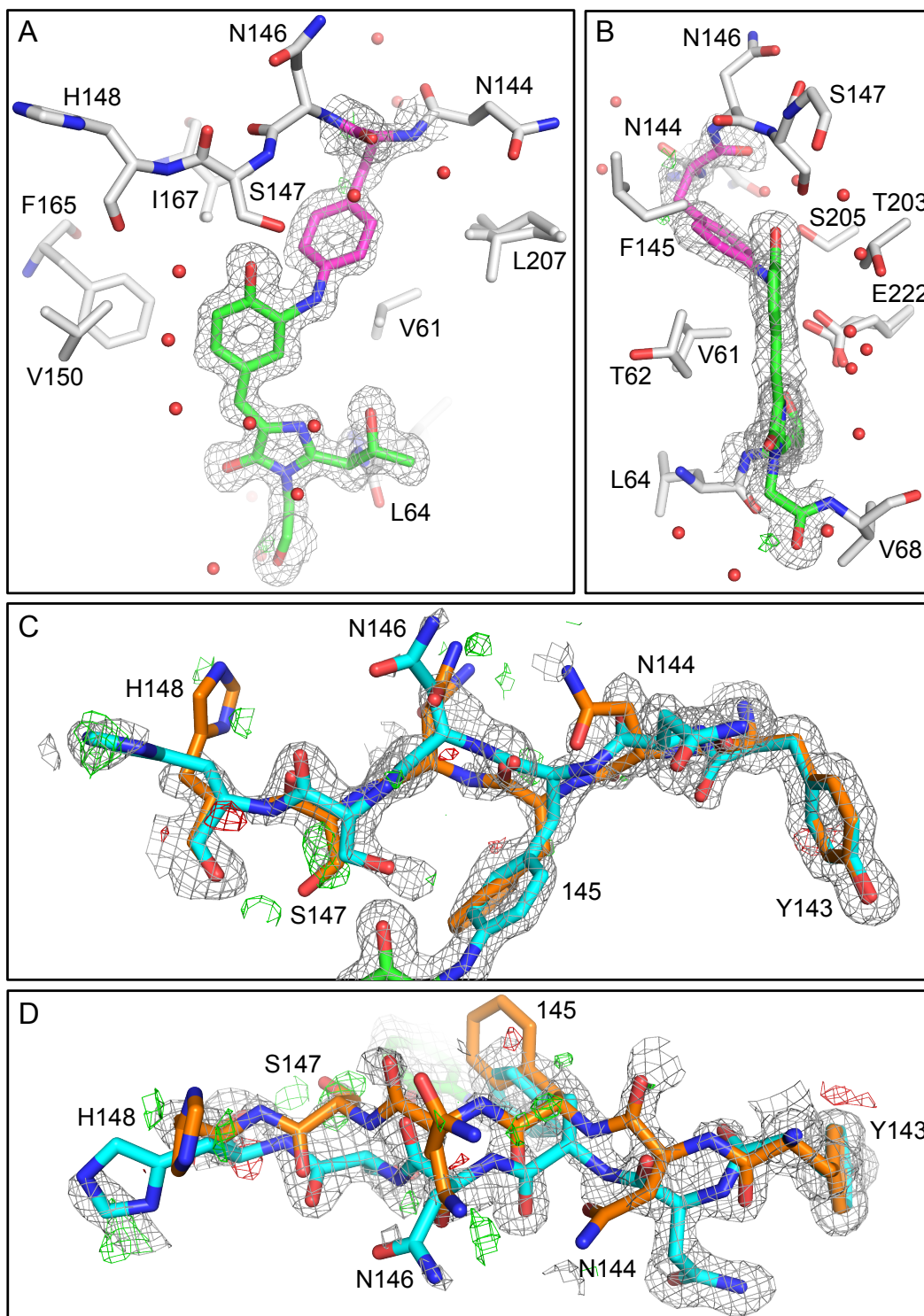


Figure 5.6. X-ray crystal structure study of sfGFP^{F145AzF} photoreaction. (A and B) 2Fo-Fc electron density map of the irradiated state of photodeactivable sfGFP^{F145AzF}. Electron density is displayed to 1σ . Positive and negative electron density is coloured green and red, respectively, and displayed to $\pm 3\sigma$. Atoms are coloured by element with chromophore carbons coloured green and residue 145 in pink. Water molecules are displayed as red spheres. Crystals of the sfGFP^{F145AzF} variant were grown in the dark and irradiation was performed *in crystallo* as described in the text. (C and D) 2Fo-Fc electron density map showing the two conformers of the Tyr143 to His148 region of irradiated sfGFP^{F145AzF}. The major (80 %) crosslinked form and the minor (20 %) non-crosslinked form have carbons coloured cyan and orange, respectively.

5.2.3 Investigating the molecular mechanism of photodeactivation in sfGFP^{F145AzF} by mass spectrometry

Trypsin-digest mass spectrometry (Section 2.9) was used to corroborate the crosslinked species observed in irradiated sfGFP^{F145AzF}. The fragment containing the chromophore (2415.3 Da for wild-type sfGFP including residues 54 to 74) could not be detected in any of the tested proteins, including wild-type sfGFP. The fragment housing residue 145 (residues 142 to 157) was observed. The mass of this fragment for wild-type sfGFP and dark state sfGFP^{F145AzF} were 1927.88 Da (predicted 1927.9 Da) and 1942.92 Da (predicted 1968.9 Da), respectively (Figure 5.7A). The mass difference was ~26 Da and corresponds to the loss of N₂ (28 Da) and addition of two hydrogen atoms (2 Da) to form an amine group. This mass matches the predicted mass of an amine species (predicted mass 1942.9 Da; Figure 5.7B) and was most likely due to the presence of thiol-based reducing agents in sample preparation resulting in chemical reduction of azPhe to an amPhe. On irradiation, the mass peak equivalent to the 145 fragment was absent for sfGFP^{F145AzF} (Figure 5.7C). Taken with the inability to detect the fragment housing the chromophore, the loss of the 145 fragment suggests a crosslink between fragment 66 and fragment 145. There was no evidence of a higher mass peak corresponding to a crosslinked fragment.

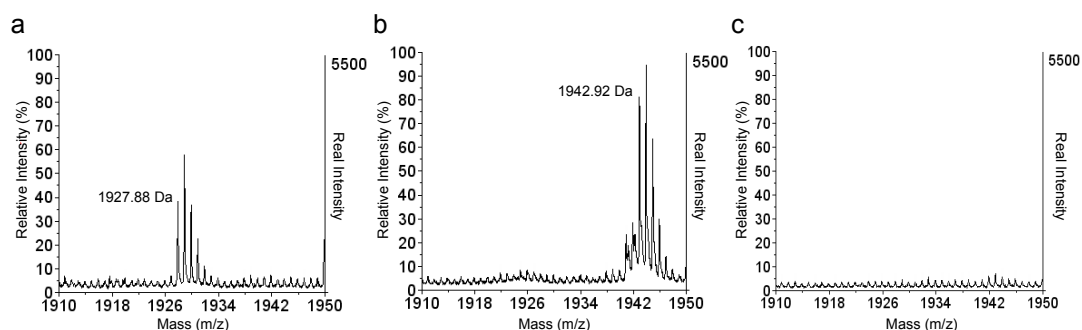


Figure 5.7. Investigating the molecular mechanism of photodeactivation in sfGFP^{F145AzF} by tandem mass spectrometry. Mass spectra of (A) wild-type sfGFP, (B) dark-state sfGFP^{F145AzF} and (C) irradiated sfGFP^{F145AzF}. The sample in (C) was irradiated in solution prior to sample preparation. The spectra are focused on the region with the expected digest fragment harbouring residue 145. The predicted masses of the fragment housing residue 145 are 1927.9 Da (F145 = wild-type), 1968.9 Da (AzF145) and 1942.9 Da (AmF145). The intact azido group (1968.9 Da) was not seen for (B) dark-state sfGFP^{F145AzF}. This is most likely due to the presence of thiol-based reducing agents in sample preparation or laser desorption resulting in reduction to an amine (1942.9 Da). Neither an azido group nor an amine was observed for (C) irradiated sfGFP^{F145AzF}.

5.3 Discussion

This chapter explored the underlying molecular basis for the photosensitive behaviours of sfGFP^{AzF} variants (Chapter 4). This was the first example of using azPhe to control protein activity using light so the photochemical pathways sampled by azPhe could not be well predicted. It is important to understand the photochemical basis in order to be able to make predictions about reactivity for other residues or even proteins. Phenyl azide photoreactivity is in itself an interesting subject that has been studied and exploited for over 90 years. Studies like these (and Chapter 6) add to our knowledge of the photoreactions in different, unique environments (*eg.* presence of hydrogen bond donors/acceptors, location with respect to chromophores). Fluorescence spectroscopy was used to compare sfGFP variants containing azPhe or amPhe to assess one of the potential endpoints of the azPhe photoreaction (reduction to phenyl amine). Moreover, X-ray crystallography, supported by mass spectrometry, provided strong evidence of the photochemical endpoints reached on photolysis.

5.3.1 Molecular explanation for fluorescence photoactivation in sfGFP^{Y66AzF}

sfGFP^{Y66AzF} is perhaps the most straight forward of the sfGFP^{AzF} variants studied to understand the underlying mechanism of action. The azido group is most likely responsible for the lack of fluorescence in sfGFP^{Y66AzF} as the electron rich N α is known to act as a quencher (176). Irradiation of sfGFP^{Y66AzF} results in a change in the electronic properties of the azido group and therefore removal of the quencher and gain in fluorescence. The low level of fluorescence observed in the dark state (Chapter 4; Figure 4.3) may be due to a small amount of photoconversion during protein purification or more likely by incomplete quenching by the azido group.

The X-ray crystal structures of the dark and irradiated states of sfGFP^{Y66AzF} (Figure 5.5) revealed the start and end points of the photochemical pathway, respectively. Electron density was seen for the azido group at the *para* position of the chromophore tyrosyl moiety in the dark structure. After irradiation no density was observed past N α indicating loss of molecular nitrogen (N₂). No other electron density was observed to suggest ring expansion or crosslinking so the irradiated state was modelled as an amine. Incorporation of amPhe at residue Tyr66 (sfGFP^{Y66AmF}) confirmed the photoreduction of azPhe to a phenyl amine (Figure 5.8; red). Spectral characterization clearly showed similarity of sfGFP^{Y66AmF} with the endpoint of sfGFP^{Y66AzF} irradiation (Figure 5.1A and B) (192). Additionally, these spectra were similar to those observed for amPhe incorporation at the same residue in wt-GFP (192) placing confidence in the assignment. To generate an amine the nitrene must gain two electrons and two protons that could come from any combination of individual electrons, protons or hydrogen atoms, from a range of sources. Singlet nitrene is a powerful base and can abstract

protons from a variety of sources (45, 193); more so in acidic environments or when the phenyl ring is substituted with electron donating groups. The imidazoline ring of the chromophore is a weak electron donor to the putative nitrene (Chapter 6; (194)) and the phenyl moiety of the chromophore is known to become more acidic on photo-excitation ((95), hence excited-state proton transfer; ESPT). The nitrene is surrounded by potential sources of H^+/H in the form of structured water molecules and basic side chains such as His148 (Figure 5.5 and Figure 1.9). Additionally, GFP has been shown to shuttle protons from the solvent to the chromophore along “wires” consisting mainly of carboxylate groups (101). The abstraction of protons would also require an electron source to balance the charge on the molecule, which could be due to the unique nature of GFP via an intramolecular transfer from the chromophore (195-197) or a sacrificial oxidant from the buffer or a neighbouring residue such as His148, Thr203 or Glu222. Alternatively, the nitrene generated on irradiation could abstract a hydrogen atom. Both singlet and triplet nitrenes are in theory capable of H-abstraction from a variety of sources as discussed in depth in Chapter 6 (Section 6.3.2; (198-200)). However, H-abstraction is generally thought to arise from the triplet spin state, which would only be a minor photoproduct at 277 K (Section 1.6.1 and Figure 5.8).

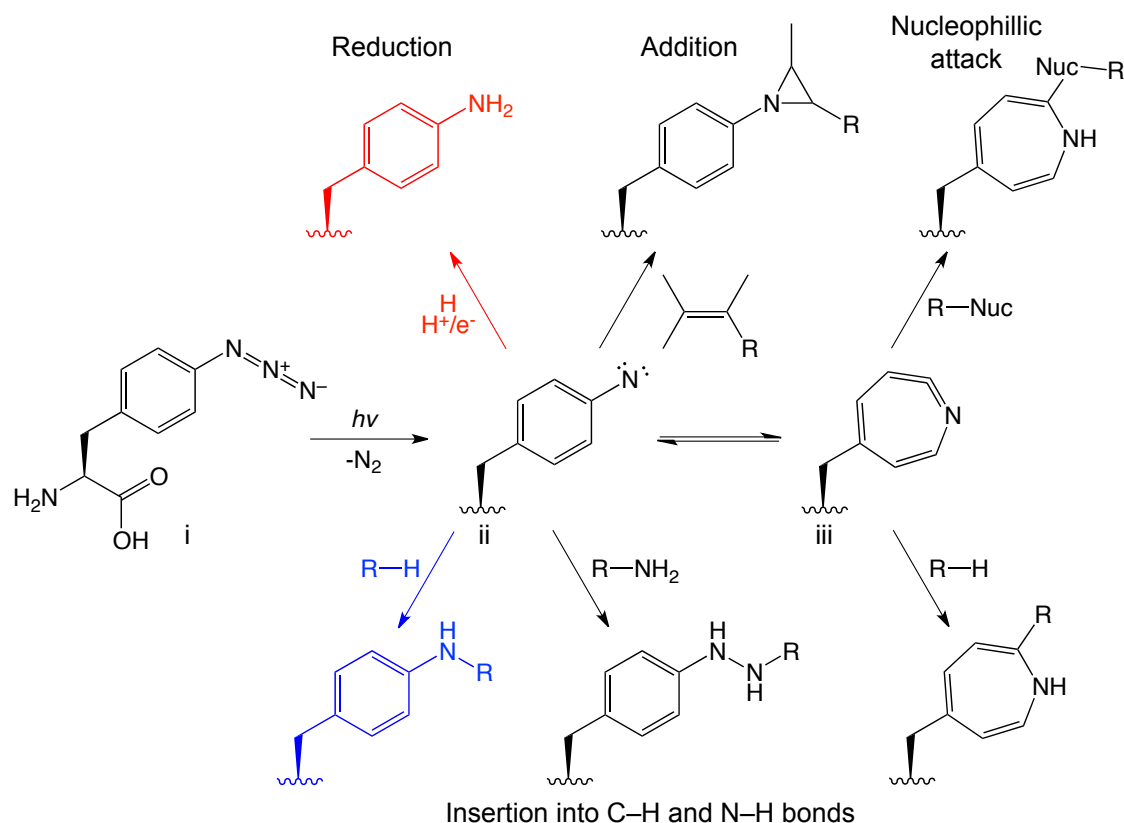


Figure 5.8. Photoreaction of *p*-azido-*L*-phenylalanine (azPhe; **i).** On irradiation, **i** loses molecular nitrogen to form a singlet nitrene (**ii**) which can undergo ring expansion to form a ketenimine (**iii**). **ii** and **iii** can undergo multiple reactions as indicated on the diagram. Above 165 K the singlet nitrene (**ii**) can abstract protons and hydrogen atoms from a variety of sources such as alcohols, amines and acidic C–H bonds. Nitrenes can also undergo cycloaddition reactions with alkenes or undergo ring expansion to form a 7-membered ketenimine (**iii**). **iii** can react with proximal nucleophiles. Additionally, both **ii** and **iii** can insert into active C–H and N–H bonds. The photochemical pathways proposed in this chapter are coloured; the reduction of sfGFP^{Y66AzF} to an amine is shown in red and photo-crosslinking of sfGFP^{F145AzF} in blue. **ii** is shown in the closed-shell electron configuration however an open-shelled configuration is also possible.

5.3.2 Molecular explanation for fluorescence photodeactivation in sfGFP^{F145AzF}

The photochemical pathway causing deactivation of sfGFP^{F145AzF} differs from that observed for sfGFP^{Y66AzF}. Incorporation of amPhe at residue Phe145 (sfGFP^{F145AmF}) results in spectral properties nearly identical to wild-type sfGFP (Figure 5.1C and D). This suggests that the endpoint of sfGFP^{F145AzF} irradiation is not an amine as with sfGFP^{Y66AzF}. The crystal structure of irradiated sfGFP^{F145AzF} revealed that the singlet nitrene is likely to attack the *meta* carbon of the chromophore tyrosyl moiety to form an N-phenyl crosslink (Figure 5.6A). Presence of the crosslink was indirectly confirmed by trypsin-digest mass spectrometry. The fragment containing the chromophore could not be detected in any of the tested proteins, which resulted in loss of the fragment housing residue 145 in the spectrum of irradiated sfGFP^{F145AzF} (Figure 5.7). The mechanism involves the phenolate (anionic) form of the

chromophore, as indicated by the absorbance spectrum for the dark form (Chapter 4; Figure 4.3; Section 1.7, (92)). The electrophilic singlet nitrene likely attacks the delocalised, electron-rich chromophore resulting in C–H bond insertion (Figure 5.8; blue). Preference for the *meta* position of the chromophore tyrosyl moiety is most likely dictated by proximity of residue 145. Alternatively (or in combination with a proximity effect), the strength of the electro- and nucleophilicity may play a role. Resonance of the singlet nitrene electrons with the phenyl ring reduce its electrophilicity making it a “soft” electrophile, which is also indicated by its lack of charge. Soft electrophiles are more likely to react with the soft nucleophiles, such as the *meta* position of the chromophore in comparison to the strong nucleophilic nature of the anionic hydroxyl group ($-O^-$). Loss of fluorescence is likely due to the introduction of a local quencher. Insertion of the out-of-plane phenyl nitrene (Figure 5.6B) and the associated nitrogen lone pair gives the possibility of photoinduced electron transfer as a non-radiative quenching pathway for the excited state. When the sfGFP chromophore is excited, an electron is raised to an excited state leaving a vacancy in the lower energy state. Normally for fluorescence, this electron relaxes back to the lower energy state and the difference in energy levels is emitted as visible light. In the crosslinked form, it is possible for the adjacent former nitrene N to act as either an electron donor filling the vacant lower energy level or as an acceptor of the excited, higher energy state electron (201, 202). Both scenarios prevent electronic relaxation and therefore quench fluorescence emission. This is supported by the non-radiative absorbance peaks at 386 and 500 nm after irradiation (Chapter 4; Figure 4.3), possibly corresponding to the phenyl nitrene moiety and the chromophore, respectively.

5.3.3 Photoswitching in sfGFP^{H148AzF}

In the previous chapter, photoswitching in terms of peak excitation wavelength was observed for sfGFP^{H148AzF}. The underlying mechanism was postulated to be due to modulation of the ground-state charge on the chromophore. Thus, it differs from that observed for sfGFP^{Y66AzF} and most likely from sfGFP^{F145AzF}. The result of photolysis is control over the protonation state of the chromophore in the ground state in which His148 plays a role (Section 1.7, (92)). As discussed in Section 1.7, GFP has two excitation peaks at 400 and 488 nm representing the protonated (neutral) and deprotonated (anionic) chromophore, respectively (92). The photochemical endpoint of sfGFP^{H148AzF} promotes the anionic form of the chromophore (λ_{exc} 485 nm; Chapter 4; Figure 4.4). It is unlikely that azPhe is reducing to the phenyl amine form observed for sfGFP^{Y66AzF}. Incorporation of amPhe at residue 148 does not generate the same fluorescence spectral properties observed for the irradiated sfGFP^{H148AzF} (Figure 5.1E). In fact, sfGFP^{H148AmF} more closely resembles the dark state of sfGFP^{H148AzF} suggesting that the presence of a phenyl amine at residue 148 does not promote chromophore

deprotonation. The protein is still fluorescent after irradiation suggesting that a crosslink with the chromophore is also not responsible for photoswitching in this variant. In sfGFP His148 promotes ionisation of the chromophore via H-bonding with the tyrosyl hydroxyl group (92). However amPhe148 promotes a neutral chromophore despite having the ability to hydrogen bond whereas irradiated azPhe148 once again promotes the anionic form (Figure 5.1E). One possibility is that azPhe148 cannot occupy a conformation that can hydrogen bond with the chromophore but on irradiation converts to a form with different dimensions that can, such as a derivative of the seven-membered ketenimine (Figure 5.8; iii). An X-ray crystal structure of the dark and particularly the irradiated forms of sfGFP^{His148AzF} should reveal the photochemical pathway sampled. This is currently under way in the Jones lab.

5.3.4 Effects of azPhe incorporation on the structure of sfGFP

As well as rationalising residues that were not tolerant to azPhe incorporation it is useful for future studies to understand how the structure as a whole accommodates Uaa's within the core of a protein. The tolerance of a particular amino acid to mutation, as with natural amino acid substitutions, can be a limiting factor to an approach. A major benefit of using azPhe to instil photocontrol is its size and similarity to natural amino acids, which should make it highly tolerable. Generally, azPhe incorporation was well tolerated as shown by the high yields of soluble sfGFP^{AzF} proteins produced (Chapter 3).

Tyr66 was tolerant to azPhe incorporation somewhat surprisingly considering its location in the protein core (compared to Glu222; Figure 1.9, Chapter 4). By comparison to the wild-type protein, incorporation of azPhe at residue 66 caused only minor structural changes (Figure 5.3). A number of surface exposed residues occupy different rotamers between sfGFP and sfGFP^{Y66AzF}, which is expected of dynamic residues in a crystal structure. Closer inspection of the residues surrounding the chromophore (Figure 5.3B and C) again shows only slight changes. The proximal residues are shifted away from the azido group by ~0.3 – 0.6 Å probably as a result of the additional electron density of –N₃. His148 takes an alternate conformation in sfGFP^{Y66AzF} with a flipped side chain that is twisted away from the chromophore. This shift could be the result of the dynamic nature of His148 sampling a different conformation in the crystal state or might reflect the additional space occupied by the azido group (compared to –OH in Tyr66). Alternatively, although the N α of the azido group is in theory capable of H-bonding, it may not be as accessible to His148 hence the flip that results in N₁ being further from the chromophore. Thr203 has an alternate rotamer in sfGFP^{Y66AzF} where the hydroxyl group no longer faces the *para* position of the chromophore and hence cannot H-bond. This probably reflects the hydroxyl → azido substitution knocking out the H-bond between Thr203 and the chromophore. In the irradiated (photoactivated)

sfGFP^{Y66AzF} structure Thr203 remains in a state where it cannot bond with the chromophore and hence is not essential for fluorescence. However, removal of this important H-bond that is found in many GFP structures (92, 104) may contribute to the altered spectral properties of photoactivated sfGFP^{Y66AzF} versus wild-type sfGFP (Chapter 4). The other significant change is twisting of Leu42 to face the former Thr65 moiety of the chromophore in sfGFP^{Y66AzF} (Figure 5.3B and C). If this is an accurate observation and not an artefact of crystallisation, then it could represent a change in packing at the core of the protein around the chromophore as a result of azPhe incorporation. On the whole, the structural waters around the chromophore matched up between sfGFP and sfGFP^{Y66AzF} with the absence or presence of some waters expected for the detection limits of the resolution. Interestingly, the azido group of sfGFP^{Y66AzF} displaces a conserved water molecule and points into a ~cavity created by surrounding residues (Figure 5.5), which may explain the tolerance of residue 66 to mutation. In general, azPhe incorporation was better tolerated in aromatic residues (Tyr66, Phe145) than smaller residues (Thr62, Thr203; Chapter 3, Table 3.1 and Chapter 4) as shown by higher levels of soluble protein produced. It is logical that a residue with a similar size and structure to azPhe would be more tolerant to substitution as the protein structure and folding would be less perturbed.

A dark-state structure of sfGFP^{F145AzF} is currently unavailable however comparison of the irradiated structure with wild-type sfGFP revealed, again, on the whole only minor structural differences (Section 5.2.2.3). The backbone traces (Figure 5.4A) are very similar with a few differences in flexible loop regions. A major difference to the 'core' β -barrel was seen from Tyr143 to His148. Here, the photo-induced crosslinking has resulted in the protein backbone diverging significantly from wild-type sfGFP with some peptide bonds and side chains flipped including Asn144, Asn146 and Ser147. The unusually high B-factors (Figure 5.4B) for the β -barrel around this region revealed that the poor density is somewhat expected. This put confidence in the assignment of two conformations because if two forms of the protein with highly different structures (Figure 5.6C and D) existed in the crystal, and hence the diffraction pattern, this would give the appearance of motion and manifest as high B-factors. Additionally, this highlights the degree of movement and structural rearrangement involved in forming the crosslink to the chromophore from the native structure.

As well as having high B-factors, the Tyr143-His148 region is different between wild-type sfGFP and the non-crosslinked conformer of irradiated sfGFP^{F145AzF}. This indicates that the non-crosslinked form is not simply equivalent to the native, wild-type structure but takes on a third form possibly as a result of photolysis. Alternatively these differences could have resulted from crystal packing effects as this region is on the surface of the protein. This is rational as sfGFP (2B3P) is space group P3₁21 and sfGFP^{F145AzF} (4J8A) is P2₁2₁2₁ and,

moreover, the crystal packing results in different protein-protein contacts in the crystal. By analysing the side chain orientation of Phe145 in wild-type sfGFP it is reasonable to predict that the azido group in the dark state of sfGFP^{F145AzF} sits in a space between the chromophore, Pro58 (main chain), Val61, Thr62, Ile167 and His169. This space also contains a structural water molecule near the *para* position of Phe145 that would probably be displaced to accept -N₃. There appears to be space for incorporation of the additional azido group (by displacement of a structural water molecule) hence the high yield of soluble protein produced. However, introduction of an additional azido group may result in the Tyr143-His148 region that houses 145 shifting away from the chromophore or rearranging which might indicate that the non-crosslinked conformer in sfGFP^{F145AzF} (Figure 5.6C and D) is in fact the dark-state of sfGFP^{F145AzF}.

5.3.5 Summary

This chapter has built on work in Chapter 4 by demonstrating the underlying photochemical routes responsible for photoactivation in sfGFP^{Y66AzF} (reduction to phenyl amine) and deactivation in sfGFP^{F145AzF} (insertion into the chromophore). The mechanism for sfGFP^{H148AzF} has also been discussed and although not conclusive, has ruled out some endpoints as a basis for future work. Such knowledge is important as it builds on our understanding of phenyl azide chemistry and helps design future experiments in different proteins and protein microenvironments.

The position of azPhe in sfGFP proved very important with the immediate protein environment having a profound influence on the photochemical pathway sampled. Each sfGFP^{AzF} variant follows a distinct reaction pathway on nitrene formation, which has different effects on sfGFP function. Photolysis of phenyl azides is normally achieved with high intensity UV light between 250-350 nm (46). However, in the current context low intensity blue light >400 nm (and even as high as 485 nm for sfGFP^{F145AzF}) is sufficient suggesting that the protein environment is also lowering the activation barrier to nitrene formation. Recent work has shown how FPs can photoreduce a range of chemicals such as FAD, NAD⁺ and even cytochrome *c* ((196), Chapter 6). It is possible that photoexcitation of the chromophore is lowering the barrier to azPhe activation and/or even directly photoreducing azPhe (sfGFP^{Y66AzF}). The decrease in activation energy is likely to be a FP-specific effect, however the approach of using azPhe to generate light-sensitive proteins should be transferable to any protein.

6. Caging reactive chemical intermediates using proteins

6.1 Introduction

The chemical conversions associated with phenyl azide photolysis (Figure 1.6) were first observed nearly 90 years ago (40). Phenyl azide photochemistry has since been exploited in many fields including photo-affinity labelling (PAL) of biomolecules, lithography and polymer chemistry (41, 203, 204). Critical to any application is to understand a process at the fundamental level to aid design and implementation. This is true of phenyl azide photochemistry for predicting its use for specific goals, such as PAL (205). As powerfully demonstrated in Chapter 4, one such use is the control of protein function through incorporation of azPhe as defined residues in sfGFP.

Elucidating the photochemical pathways and mechanisms of phenyl azides proved difficult due to the complexity of the reaction. Early studies showed the various pathways available to nitrenes such as various products of the ring-expanded ketenimine and anilines (206, 207). Deciphering the underlying mechanism has however been hampered by conflicting and misleading information generated from different experimental techniques (46, 47). For example, the IR and UV-vis spectra of the intermediate produced on photolysis of simple phenyl azide $C_6H_5N_3$ was disputed when studied by laser flash photolysis. Different groups used alternate solvents (inert such as 3-methylpentane to reactive such as nucleophilic diethylamine) and argued over which of a variety of intermediates including triplet nitrene, benzazirine and ketenimine (Figure 1.6) were responsible for the observed spectra (47, 208, 209). Furthermore, the major product following photolysis of phenyl azides in solution was often polymeric tar (199), which implies high levels of heterogeneous crosslinking. Despite setbacks, detailed spectroscopic and mechanistic analysis has provided the reaction pathways of many phenyl azides under a variety of conditions (46, 47, 210-212).

The majority of our understanding has come from matrix isolation studies, time-resolved spectroscopy, laser flash photolysis (LFP), *ab initio* calculations and the synergy between them (46, 213, 214). One ingenious development was the use of synthetic host molecules or “cages” that afford greater control over a reaction by defining the environment. Synthetic host molecules can be tailored to provide a defined chemical environment ranging from inert to one with specific chemistry. Nitrene reactivity has been modulated and studied within a variety of hosts including: calixarenes (56), hemicarcerands (58, 59), cavitands (57) and cryptands (55). One particular benefit of using synthetic host molecules for phenyl azide chemistry was the

elimination of the bimolecular reactions (*ie.* dimerization of triplet nitrene via an azo linkage (Section 1.6.1) that plagued many earlier phenyl nitrene studies. For example, Warmuth and Makowiec used a hemicarcerand host to cage the ring-expanded ketenimine (13) (**3**; Figure 1.6). Using different solvents, they were able to measure equilibrium kinetics and promote H₂O insertion into the ring-expanded ketenimine. In another case, Bucher *et al.* were able to use the protective cavity of a cryptand to trap a singlet nitrene following irradiation of phenyl azide (55). Such experiments gave important insights into the reaction including conversion yields under specific conditions and energetic barriers.

This chapter aims to explore the ability of proteins to act as environmentally defined cages for reactive organic intermediates. Nitrene chemistry is a good test-bed, not only because of the wealth of information available, but also because a nitrene precursor, azPhe, can be incorporated at defined positions and thus environments in a protein. Proteins offer themselves as ideal targets for caging studies because of the vast amount of structural knowledge available allowing for photochemistry to be probed in many diverse chemical environments. Proteins are also self-folding and/or assembling and easy to mutate. The combination of these features means proteins could effectively be thought of as a library of tuneable host molecules, where nature defines the caging matrix.

To test the applicability of using proteins to cage reactive intermediates, in this chapter azPhe was incorporated into two positions of sfGFP; a solvent-exposed surface residue and a residue buried in the core of the protein surrounded by a highly defined matrix (Section 1.7). sfGFP is a good target protein because a great deal is known about the central environment surrounding the active site of the protein; the chromophore. Additionally, high-resolution X-ray crystal structures are available allowing for the guided placement of azPhe and analysis of the reaction. The photolysis of azPhe in these two environments is compared in addition to a third environment; a hydrophobic cavity of T4 lysozyme (J. Morris; (121)). Photolysis was performed at 77 K and analysed by electron paramagnetic resonance (EPR) spectroscopy.

6.2 Results

6.2.1 Position of azPhe incorporation

To test the hypothesis that a protein could be used to cage a reactive chemical intermediate the nitrene precursor, azPhe was incorporated into two positions in sfGFP. High-resolution X-ray crystal structures were used to direct the incorporation of azPhe into sfGFP. Published structures (98, 104) were used in combination with structures of sfGFP^{Y66AzF} generated herein (Chapter 4) for analysis of photolysis. Figure 6.1 shows the chosen residues Tyr66 and Glu34 in the context of the protein structure as a whole (Figure 6.1A) and local environment of the residues (Figure 6.1B,C). Figure 1.9 shows the environment of Tyr66 in the context of the chromophore buried within the core of the protein. The tyrosine moiety resides in the central α -helix and makes contact with many residues including His148, Thr203 and Ser205 as well as structural water molecules (92, 104). The structure of sfGFP^{Y66AzF} (determined as outlined in Chapter 5) was used to compare the orientation of residue 66 and the side chain contacts and shows that azPhe66 occupies a very similar space to Tyr66 (Figure 6.1C). Glu34 is part of the β -barrel of sfGFP with an external-facing side chain that is surrounded by mainly polar and charged residues (Figure 6.1B).

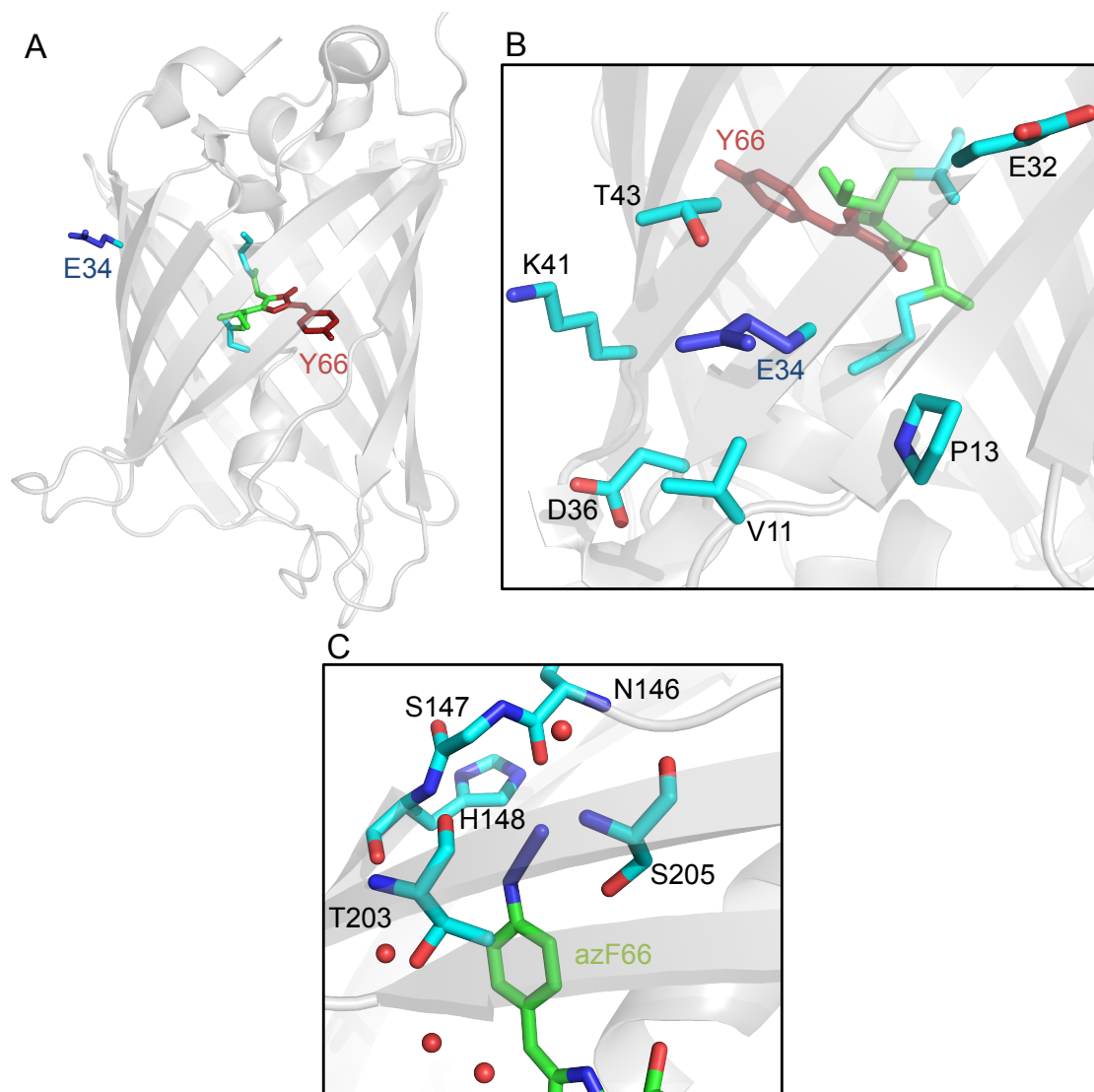


Figure 6.1. Structural environment of residues chosen for phenyl azide caging studies. (A) X-ray crystal structure of sfGFP (PDB accession 2B3P) showing the mutated residues Glu34 (blue) and Tyr66 (red) as sticks. The rest of the chromophore is coloured green. (B) Surface-exposed environment of Glu34. Glu34 and neighbouring residues are shown as sticks and coloured by element with Glu34 carbons coloured blue. (C) Environment of azPhe66 (carbons coloured green) within the core of sfGFP. Nearby residues are shown as sticks and labelled. PDB accession 4J88.

6.2.2 Photolysis of sfGFP^{AzF} variants and EPR measurements

sfGFP^{E34AzF} and sfGFP^{Y66AzF} were produced and purified as described in Chapter 3, Section 3.2.3.1. Electron paramagnetic resonance (EPR) spectroscopy was used to determine the species formed on photolysis as it is capable of detecting any species with unpaired electrons. EPR is theoretically analogous to NMR but exploits the magnetic moment of electrons associated with their spin, rather than nuclei. When placed in an external magnetic field the spin of an unpaired electron can align itself either parallel or anti-parallel to the magnetic field. Each spin state has an associated energy, which an unpaired electron can

transition between by absorbing or emitting a photon of energy. As the energy levels and therefore gap depend on the nature of the radical, the energy required to switch spin states can be used to identify specific chemical species. EPR was used to probe the reactions because, at 77 K, the photolysis of phenyl azides can favour the formation of triplet nitrene (52). It should then be possible to determine the species formed by comparison of the spectra with previous spectra of phenyl nitrene and its associated species, and with computational simulations.

Pure samples of the proteins were mixed with glycerol in EPR tubes in different ratios to find a glycerol concentration that would give a polycrystalline “glass” when snap-frozen at 77 K. A polycrystalline glass is important as this implies there is no localised order (in contrast to 100 % water) and therefore the molecules are randomly orientated. In this situation all protein orientations are sampled and the reading is an average of them, which removes angle specific effects and makes the data interpretable. A range of glycerol concentrations were tested (10, 25, 50, 75 % (v/v)) with 50 % (v/v) glycerol giving the best polycrystalline solution. Snap-frozen protein solutions were irradiated for 2 mins using a Hg/Xe arc lamp with a water filter, as described in Section 2.11. EPR tubes were rapidly transferred to an EPR spectrometer and CW-EPR (continuous wave) spectra were recorded at 10 K.

Analysis of sfGFP^{E34AzF} following irradiation revealed a featureless EPR spectrum with no peaks, and therefore the lack of any radical species. Any species containing unpaired electrons would give a peak in the EPR spectrum that corresponds to the energy gap between the two spin states of the electron (*vide supra*). The sfGFP^{E34AzF} variant was included as a comparison to the internal residue 66 as it is solvent-exposed and as such is not caged by the protein matrix.

The spectrum obtained after irradiation of sfGFP^{Y66AzF} (Figure 6.2) did not indicate any evidence of triplet nitrene formation (predicted ~7000 G field strength). However, an EPR signal was observed at 3400 G that was unusually wide (*ca.* 330 G) for an $S = \frac{1}{2}$ organic spin system.

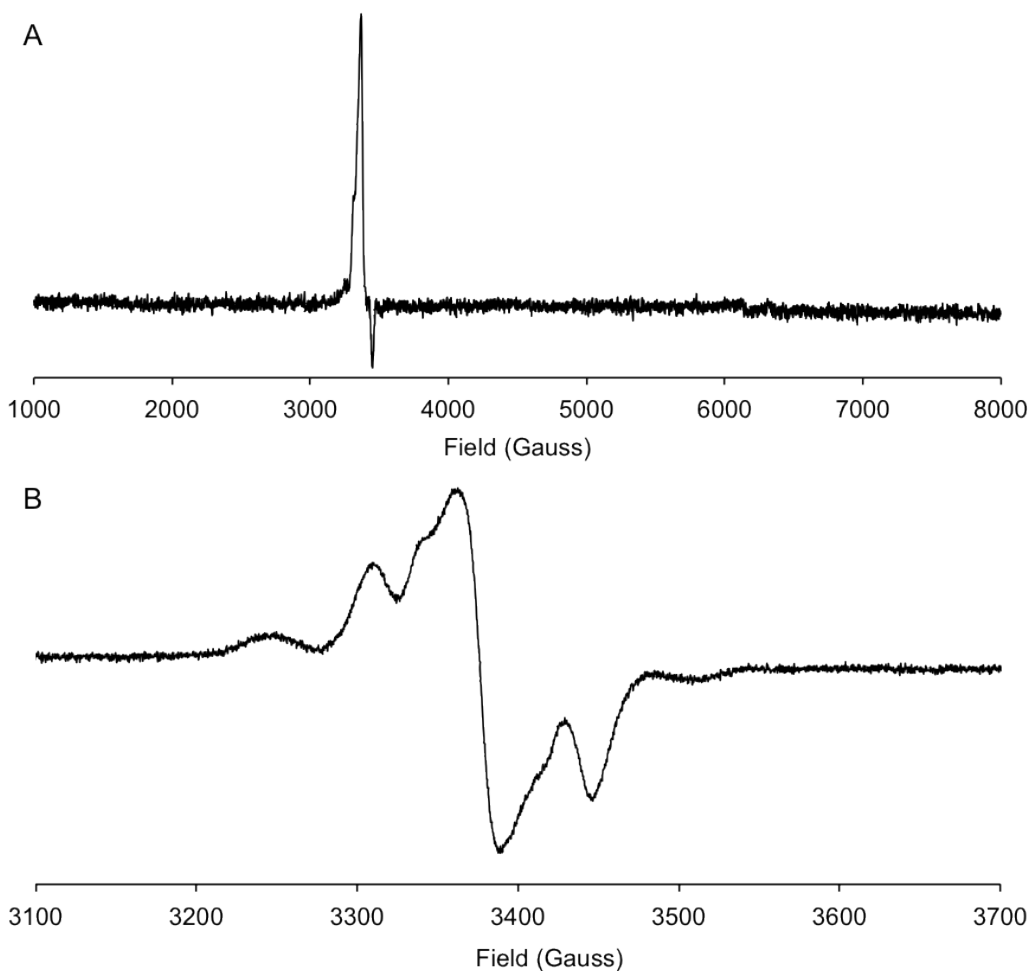


Figure 6.2. CW X-band EPR spectra (10K) of sfGFP^{Y66AzF} following photolysis at 77 K. (A) Spectrum of photolysed sfGFP^{Y66AzF} from 1000 - 8000 G field strength. (B) Spectrum of photolysed sfGFP^{Y66AzF} focusing on the feature at ~3400 G.

6.2.3 Simulation of the observed EPR spectrum

EPR simulations were performed (by D. Murphy) in order to help assign the spectrum observed for irradiated sfGFP^{Y66AzF}. Here, hyperfine coupling values are input to generate a simulated spectrum that is then compared to the observed spectrum and adjusted until the best match is found. Hyperfine couplings (hfc's) are an indication of the interaction between the paramagnetic electron and nuclei with a magnetic moment. Simulations (Figure 6.3) revealed large -NH (¹H, ¹⁴N) and phenyl (¹H_{ortho}) hyperfine couplings of ^H*a*_{iso} = 78.5 G (220 MHz), ^N*a*_{iso} = 31.5 G (88 MHz) and ^{Hortho}*a*_{iso} = 10 G (28 MHz), respectively (Figure 6.3A). The spectrum was ultimately assigned to an anilino radical (-NH[•]). A smaller background contribution (marked with the arrows in Figure 6.3A) is also observed assigned to a second unknown radical formed by generation of the anilino radical. The background radical was not added to

the simulation as the hyperfine coupling's were not sufficiently well resolved, hence the discrepancy in the fit between the experimental and simulated spectra.

Interestingly, the hyperfine coupling values observed for the chromophore-bound anilino radical are 3-4 times larger than those reported for the simpler $C_6H_5NH^\bullet$ anilino radical ($^H a_{iso} = 16.85G$, $^N a_{iso} = 8.85G$ and $^{H\alpha,p} a_{iso} = 6.47G$) in a solid crystal (215). This was rationalized by performing density functional theory (DFT) calculations (by J. Platts) on the parent $C_6H_5NH^\bullet$ anilino radical and the chromophore bound anilino radical. DFT calculations are used to investigate the electronic properties of a molecule and can predict the electrostatic potential and spin density over a molecule. These calculations revealed spin delocalisation over the whole chromophore as indicated by alternating up and down electron spins (Figure 6.3B). Delocalisation of the unpaired (anilino) electron over the entire π -system represents a lower energy state, and hence more stable state, compared to localisation of the electron on the anilino N. Calculations also showed a slight electron donating influence of the GFP chromophore, which increased the spin density on the anilino N atom (most negative energy (red); Figure 6.3C). These calculations suggest that the larger $-NH$ hyperfine couplings observed are expected for a system in conjugation with the chromophore.

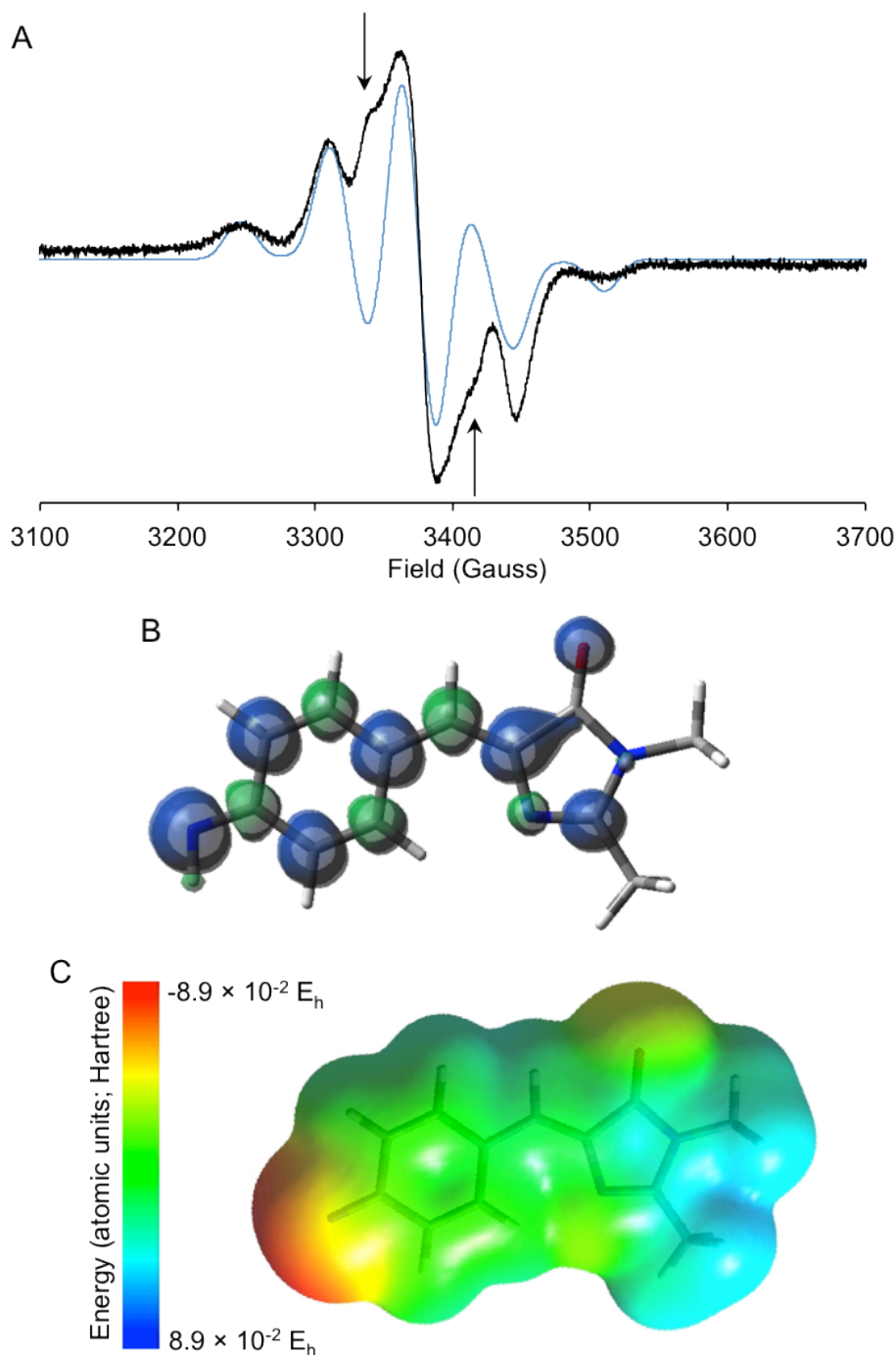


Figure 6.3. Simulation of putative anilino ($-\text{NH}^\bullet$) radical formed by the photolysis of sfGFP^{Y66A2F}. (A) Simulated EPR spectrum of the anilino radical (blue trace) overlaid with recorded spectrum (black trace; **Figure 6.2B**). Arrows indicate the contribution from the unknown and un-modelled secondary radical. (B) Spin density on the putative chromophore-bound anilino radical. Up and down electron spins are coloured green and blue, respectively. (C) Electrostatic potential mapped on the Van der Waal's surface of the putative chromophore-bound anilino radical. Colour represents the energy measured in atomic units (Hartree) as indicated in the scale bar shown on the left.

6.3 Discussion

In this chapter the photolabile nitrene precursor, azPhe, was site-specifically incorporated into proteins to probe the ability of a protein matrix to cage a reactive nitrene-associated radical. Proteins are well suited for caging studies because they are self-assembling, easy to mutate and offer a range of diverse chemical environments for investigation, which are supported by structural information. EPR was used to monitor radical species generated on protein photolysis, which was the first time EPR had been used to monitor caging in a protein containing a genetically encoded Uaa. By combining the experimental observations from EPR spectra with simulated spectra for different species and DFT calculations, a model can be built regarding the phenyl azide photolysis pathways within the context of a protein. This will in turn provide a fundamental insight concerning phenyl azide photolysis pathways.

6.3.1 Characterisation of irradiated sfGFP^{E34AzF}

For a species to be detected by EPR it must have unpaired electrons. The formation of species with unpaired electrons is accompanied by defined EPR signals, which depend on the nature of the radical form. For nitrene chemistry (Figure 1.6) this is limited to a number of species including triplet nitrene (^3N), nitrene radical anion ($\text{N}^{\cdot-}$) and anilino radical (NH^{\cdot}) (Figure 6.4). Also, formation of a nitrene radical anion or anilino radical may result in the generation of secondary radicals, such as the hydrogen atom donor on formation of an anilino radical via hydrogen abstraction.

No EPR signals were observed in the spectrum of sfGFP^{E34AzF} indicating the absence of radical species (Section 6.2.2). This can be explained by the exposure of residue 34 to the solvent in sfGFP (Figure 6.1). Nitrene radical species formed following photolysis are not caged by a protein matrix and will most likely react rapidly with aqueous solvent. Water is a weak nucleophile and can react with the ring-expanded ketenimine to form a stable product (Figure 1.6), which would not be detected by EPR. Intramolecular ring expansion can arise from both singlet and triplet nitrene (199, 205, 216) although the major route is from singlet nitrene, which would only be a minority form at 77K due to intersystem crossing (ISC) forming the triplet nitrene. The potential ring-expansion from either singlet or triplet nitrene would be rapid (picosecond timescale) and hence the nitrene intermediate is unlikely to be detected (217, 218). The surface-exposed residue 34 is also accessible for dimerization with a second sfGFP^{E34AzF} via formation of an azo linkage. Dimerization is a product of the triplet nitrene spin state (Figure 6.4) and takes place readily at 77 K (46), however should be minimal in the frozen matrix.

6.3.2 Characterisation of irradiated sfGFP^{Y66AzF}

The analysis of irradiated sfGFP^{Y66AzF} is more complex. As mentioned, the assignment of an anilino radical ($-\text{NH}^\bullet$) best satisfies experimental evidence and simulations, however is not conclusive. Confidence comes from the similarity between the simulated EPR spectrum and the observed spectrum (Figure 6.3A). Moreover, a closely-related EPR spectrum centred near 3250 G was reported on photolysis of 2,6-difluoro-phenyl-nitrene that was assigned to a phenyl anilino radical and the radical formed from its hydrogen abstraction (199). The spectrum shown in Figure 6.2 is unlikely to be a triplet nitrene by comparison to published spectra of triplet nitrenes (52). Triplet nitrene EPR spectra contain strong peaks between ~6300 – 7000 G (199, 205), as observed for the lysozyme^{F153AzF} mutant (*vide infra*; Figure 6.6C; Section 6.3.3), which was not seen with sfGFP^{Y66AzF}. Additionally, the feature is unlikely to represent a triplet radical owing to the absence of any $\Delta m_s = \pm 2$ transition (predicted ~1500 G).

Similarly, the nitrene radical anion ($\text{N}^{\bullet-}$; Figure 6.4) was ruled out because of the peak profile observed (Figure 6.2). As discussed above, the EPR spectrum recorded for sfGFP^{Y66AzF} shows line splitting indicative of an electron interacting with three nuclei, as would be the case for the proposed anilino radical ($-\text{NH}$, $-\text{NH}$, $-\text{CH}_{\text{ortho}}$; Section 6.3.3). The nitrene radical anion lacks a proton ($-\text{NH}$) compared to an anilino radical that is responsible for one of the line splittings and as such is unlikely to be the radical responsible for the observed spectrum.

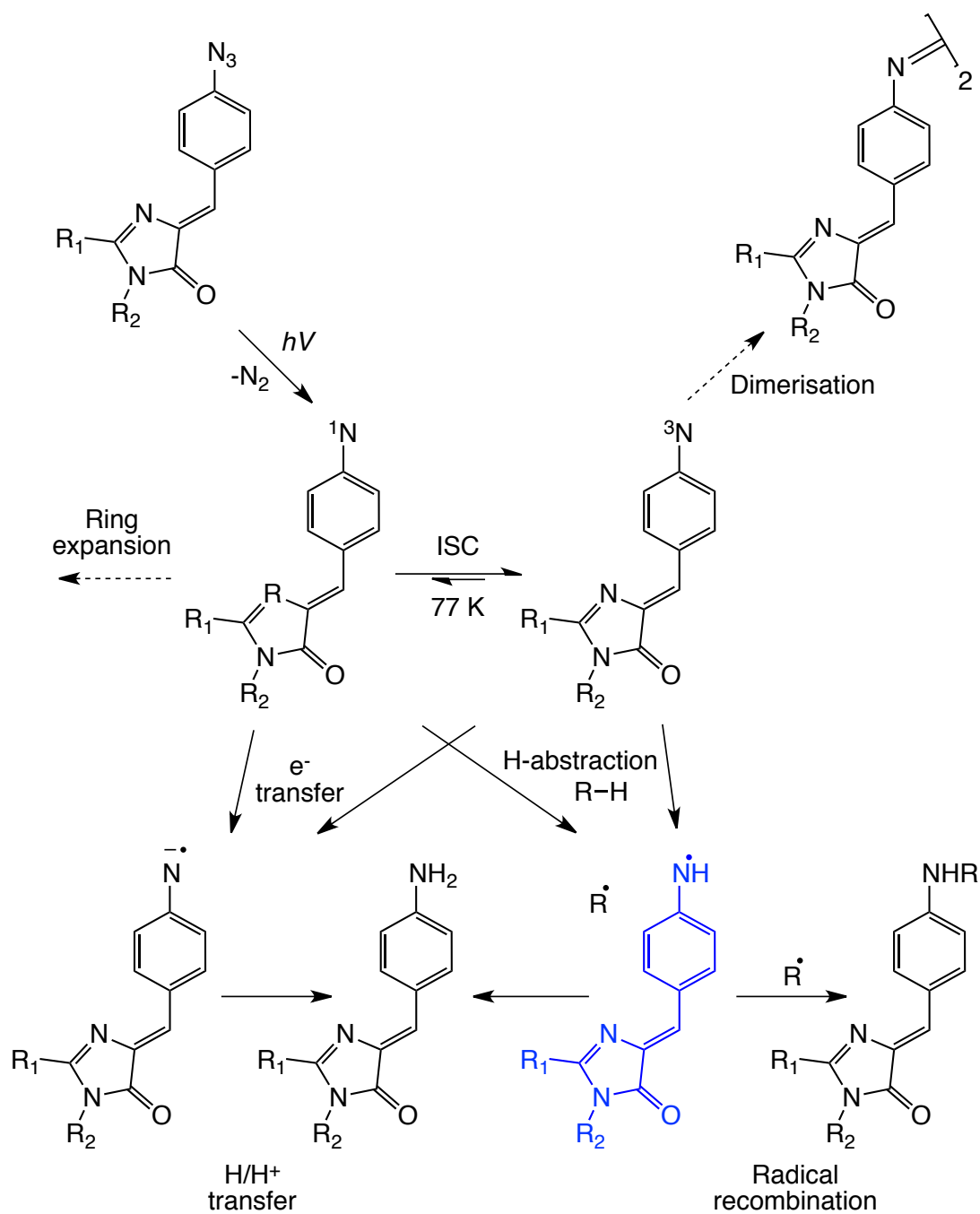


Figure 6.4. Summary of known and/or speculated mechanisms of phenyl azides. The potential photochemical pathways discussed herein are shown superimposed on the chromophore of sfGFP as in sfGFP^{Y66AzF}. The proposed anilino radical is coloured blue. Dimerization of triplet nitrene via an azo linkage is shown by a dashed arrow as this chemistry is excluded due to the protein cage. The intramolecular ring expansion chemistry is excluded from this figure (indicated by a dashed arrow) as this results from singlet nitrene but can be seen in **Figure 1.6**.

The photochemical mechanism for formation of the anilino radical is undetermined with the source and nature (H-atom *versus* H^+/e^-) of hydrogen debatable. Several groups have invoked a nitrene \rightarrow anilino radical \rightarrow aniline sequence in their mechanistic analysis of various nitrene reactions (195, 198, 200, 217, 219), including one with a protein (205). Within the literature, a variety of differing mechanisms (Figure 6.4) have been argued for, including: inter and intra-molecular H-abstraction, electron transfer, direct photo-reduction of the phenyl azide, and tunnelling, which are discussed below.

As mentioned above (199), an EPR spectrum with similar features to that observed here with irradiated sfGFP^{Y66AzF} (broad spectrum centered at ~ 3250 G) was recorded following irradiation of a simple phenyl nitrene. This spectrum was assigned to a radical pair from the phenyl nitrene abstracting hydrogen within frozen ethanol and 2-methyltetrahydrofuran matrices. Interestingly the spectrum was not observed in perfluoro-2-*n*-butyl-tetrahydrofuran, which has no hydrogens and hence H-abstraction is impossible (199). There are a variety of similar results in the literature that have postulated H-abstraction from a variety of nearby chemical groups (*eg.* CH₃, OH, NH₂) from sources including solvents such as hydrocarbons, EtOH and glycerol (199, 200, 205). The lowest electronic configuration of both the singlet and triplet states are open shelled therefore both spin states are, in theory, capable of H-abstraction (Figure 6.4). However, the prevailing view is that H-abstraction reactions from ¹N are generally not favoured because the intramolecular rearrangement (ring expansion) of the nitrene is so fast that the intermolecular reaction cannot compete (214). Also at the temperature of the reactions presented here (77 K) the rate of intersystem crossing (ISC) to the triplet is preferred to singlet intramolecular rearrangement. Further, Albini *et al.* (198) suggest the triplet state is responsible for H-abstraction, in their case from a nearby -CH₃ group or the solvent. Additionally, they report that the energy of the triplet state is increased by electron-donating ring substituents. The higher energy results in more efficient intermolecular H-abstraction in comparison to other pathways such as dimerization. As mentioned, in our case dimerization reactions are eliminated due to the protein cage. DFT calculations on the chromophore-bound anilino radical suggested a slight electron donating influence from the chromophore to the anilino (Figure 6.3C), which may therefore promote H-abstraction.

It is possible that the unique matrix surrounding the nitrene may affect the photochemistry and provide a hydrogen source. The anilino radical formed may be a consequence of proximal hydrogen-bond partners around residue 66 (Figure 6.1C and Figure 6.5). The highly ordered arrangement offers the nitrene multiple hydrogen sources including amino acid side chains and structural water molecules. As mentioned above, H-abstraction has been observed from a variety of donors including groups found in amino acid side and main chains (199, 200, 205). However, the energetics of such abstractions are not well understood and may not be favourable, particularly from water molecules. Interestingly, a similar EPR

spectrum was not observed when azPhe was placed in a protein environment lacking nearby hydrogen bonding groups (*vide infra*; Section 6.3.3). The potential for H-bond partners as sources of hydrogen atoms could be further investigated by mutating the interacting residues in sfGFP to an amino acid incapable of H-bonding (eg. His148Ala).

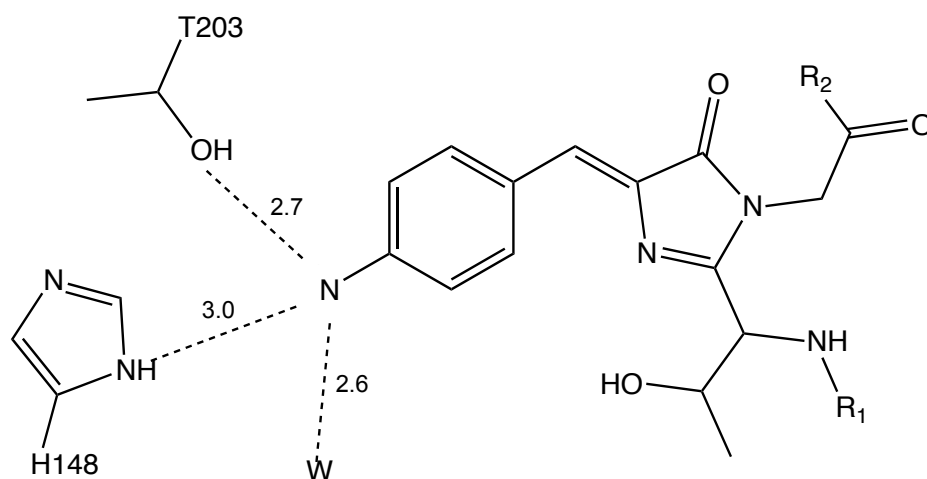


Figure 6.5. A model of the nitrene radical formed on irradiation of sfGFP^{Y66AzF}. On photolysis, chromophore-bound azPhe loses molecular nitrogen (N₂) producing a singlet nitrene. At 77 K the major conversion is from singlet nitrene to triplet nitrene via ISC. Distances (in Å) between the nitrene N of the three nearest hydrogen sources were determined from crystallographic data of wild-type sfGFP (PDB accession 2B3P).

GFP is a unique example because it is postulated to shuttle protons from the solvent to the chromophore and back to the solvent via “wires” (Section 1.7, (101)). The source of hydrogen could thus be H⁺/e⁻ recombination where the proton is “delivered” to the chromophore-bound nitrene on a proton wire. The ability of GFP to directly photoreduce various molecules has been recently reported (196). Thus, it may be conceivable that photoexcitation of the GFP chromophore itself contributes to the formation of the anilino radical via an intramolecular electron transfer from the imidazolyl moiety of chromophore (*vide infra*). Albini *et al.* (195) also suggest an intramolecular electron transfer from delocalized phenazines to the attached nitrene followed by protonation, in this case from the solvent. Similarly, CdS and CdSe nanoparticles have been used to directly photoreduce a variety of phenyl azides (220). Here, the proposed mechanism starts with single electron transfer followed by H-atom transfer (→ anilino radical) and subsequent proton transfer (→ aniline) (Figure 6.4; (47, 200, 220-222)). The element lacking in these explanations for the presented generation of an anilino radical is the electron source. Warriar and colleagues (220) used a sacrificial oxidant such as

sodium formate or methanol and postulated that many other chemicals may work. Conversely, in Bogdanov *et al.* (196), no such sacrificial oxidant is added as the chromophore itself is thought to be capable of two photoinduced oxidations. Subsequent QM/MM simulation showed the two electron process to form a radical followed by a stable cation on the imidazol moiety of the chromophore (197). However, these transfers were accompanied by significant shifts in the absorbance spectra; a blue-shift for the single electron transfer and a red-shift for the double (197). Absorbance spectra of sfGFP^{Y66AzF} were not recorded at 77 K but photolysis at room temperature resulted in a red-shift in absorbance from 390 to 446 nm (Chapter 4). Tunneling and solvent-assisted transfer have also been proposed as methods of proton and hydrogen transfer from distal sites in non-protic environments (200, 223, 224).

The simulated EPR spectrum of an anilino radical did not completely match the observed spectrum (Figure 6.3). The discrepancy is likely to be the result of a contribution from a second radical probably generated as a result of formation of the anilino radical. The source of the H-atom (H-abstraction) or electron (H⁺/e⁻ recombination) required to generate the anilino radical (*vide supra*) would itself become a radical and hence have an EPR signal. As discussed, this could be a neighbouring amino acid (*eg.* His148, Thr203), structural water molecule or even the imidazol moiety of the chromophore (following a single oxidation).

6.3.3 AzPhe photolysis in a hydrophobic cavity of a protein

In a separate study, azPhe photolysis was investigated in a disparate environment in a different protein; residue 153 in T4 lysozyme (lysozyme^{F153AzF}) produced and purified by J. Morris (121). Residue 153 is located in a well known and understood hydrophobic cavity of T4 lysozyme (36). The nearest six residues are all non-polar with approximate distances between the putative nitrene N and proximal C–H bonds ranging from ~ 3.3 – 5.5 Å (Figure 6.6A and B). When the lysozyme^{F153AzF} mutant was briefly irradiated at 77 K, an EPR spectrum characteristic of a triplet phenyl nitrene (S=1) was recorded (Figure 6.6C). It is important to note that the spectrum recorded contained an anomalous feature in the free spin region (~3500 G) caused by resonance from a cavity artefact in the EPR spectrophotometer. This feature is easy to distinguish from genuine signals as it is very narrow and has a saturation characteristic that is very different to an organic radical species. Confidence for the assignment of a triplet nitrene comes from the high degree of similarity in the spectrum compared with simulations (Figure 6.6D) and published EPR spectra attributed to triplet phenyl nitrenes (52, 55). The lack of hydrogen bond partners, and therefore proton/hydrogen sources, here supports the idea that the protein matrix is having an effect on the photochemical pathway sampled (*vide infra*; Section 6.3.3). As no hydrogen sources are proximal no hydrogen abstraction is observed whereas with sfGFP^{Y66AzF} neighbouring hydrogen-bonding

groups may give rise to the anilino radical. The generation and observation of two different radical species from the photolysis of phenyl azide is an exciting preliminary result for the approach of using defined protein matrices to cage and study reactive intermediates.

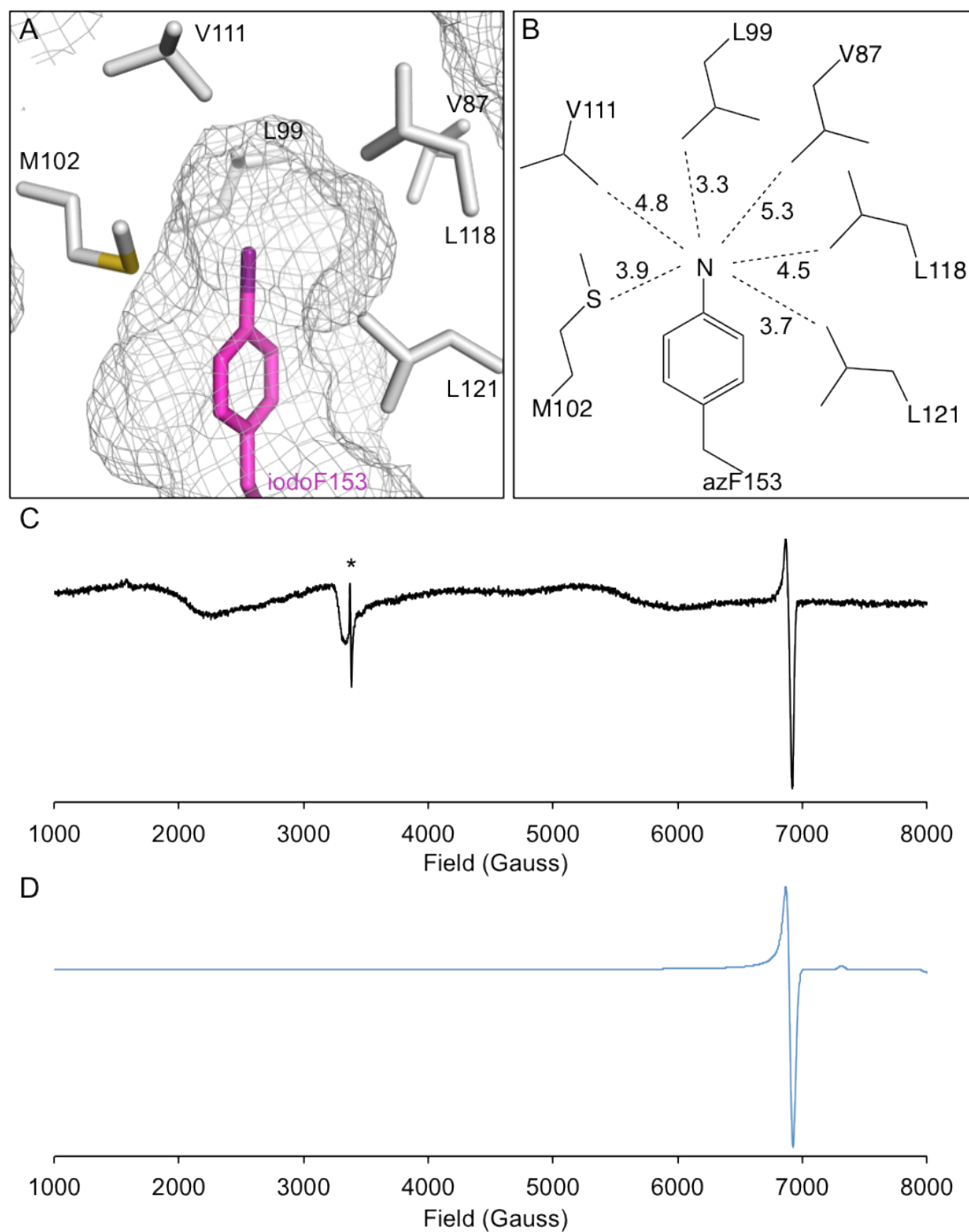


Figure 6.6. Modelling, photolysis and EPR of lysozyme F^{153AzF} . (A) 3D structure and (B) schematic representation of T4 lysozyme cavity (PDB accession 1T6H). In this structure residue 153 is a different UAA, *p*-iodo-phenylalanine (iodoF; pink) that has a single iodine atom at the *para* position of the Phe (36). The six nearest residues are labelled and displayed as sticks and coloured by atom. (B) Distances between the putative nitrene N and the six closest residues are displayed in Å. Distances were estimated from the iodine of iodoF. (C) CW-EPR spectrum of lysozyme F^{153AzF} following brief irradiation at 77 K. The spectrum shows a cavity artefact in the free spin region indicated by *. (D) Simulated EPR spectrum of a triplet nitrene generated as in Section 6.2.3.

6.3.4 Proteins as cages to study reactive organic intermediates

In summary, azPhe was genetically encoded into proteins to probe the ability of a protein to cage a highly reactive intermediate formed on photolysis of the Uaa. Evidence suggests an anilino and a triplet phenyl nitrene radical have been caged. The intermediate that is observed can be rationalized from the composition of the caging matrix provided by the protein. For example whether the radicals generated are proximal to sources of hydrogen atoms. Residue 66 of sfGFP is surrounded by hydrogen bond donors/acceptors and has the unique capability of proton shuttling. Also residue 66 is part of the delocalised sfGFP chromophore that donates electron density to azPhe and maybe capable of direct photoreduction. Photolysis of azPhe here generated an anilino radical most likely from triplet nitrene by either H-abstraction from a nearby amino acid or H^+/e^- recombination from the internal H-bonding network and an unknown source, respectively (Section 6.3.2). Conversely, the side chain of residue 153 of T4 lysozyme points into a cavity and is much further from a source of hydrogen and photolysis of azPhe here resulted in caging of the triplet nitrene intermediate (Section 6.3.3). Therefore, it was shown that proteins are capable of caging reactive organic intermediates and that the intermediate caged can be tuned by the choice of residue, and hence environment, in the protein. More work needs to be done, including photolysis of azPhe in a greater number of environments, in order to be able to predict the photochemical pathways (and therefore radicals) sampled from different protein sites. Although preliminary, this is an exciting alternative to synthetic caging approaches because of the huge diversity of structures seen in nature. The ever-increasing wealth of structural knowledge available will allow for precise placement of Uaa's in ever more proteins. The benefit of using proteins is the spontaneous folding into a complex 3-dimensional structure naturally defined by the amino acid sequence, alleviating the need for complex synthesis. Conversely, the complexity of protein structure could also be undesirable due to the lack of control over something as highly dynamic as a protein. Also background signal from the protein would make it difficult to perform other spectroscopy approaches (such as UV-vis absorption) to analyse the radical-protein cage. In some cases this could be alleviated by isotopically labelling the radical and/or residues thought to interact. As well as using different proteins, this technique could be expanded by encoding different Uaa precursors to reactive intermediates such as the carbene precursor trifluoromethyl-diazirin-phenylalanine (190).

7. Bio-orthogonal modification of proteins via genetically encoded azide chemistry

7.1 Modification of biomolecules

Biomolecule modification or, bioconjugation, offers a valuable way of studying, using and altering the properties of biomolecules such as proteins. Bioconjugation has created proteins with a range of novel properties including improved therapeutic properties (PEGylation) (225), the ability to bind nanoparticles (85, 86) and control over sub-cellular localisation (lipidation) (226, 227). Bio-orthogonal modification using Click chemistry is a powerful method for site selectively labelling proteins (as well as other biomolecules) that overcomes the limitations of modifying natural amino acids (228, 229). Traditional strategies for the covalent modification of proteins rely upon the chemistry of the 20 natural amino acids, principally amino groups (lysines and N-termini) and sulfhydryl groups (cysteines). A given protein often contains multiple of these target residues, which results in poor control over conjugation as the site or number of modifications cannot be defined. Additionally, more complex samples, such as cell lysates, may contain these functional groups as part of other classes of molecules such as DNA and lipids. This lack of specificity leads to a mixed population of modified proteins which is detrimental when a highly defined conjugate is required, such as for a bio-therapeutic, and can lead to loss of function of the target biomolecule (230). The site-specific incorporation of novel, bio-orthogonal functional groups, in the form of Uaa's, overcomes the inherent limitation of traditional labelling strategies with natural amino acids. Bio-orthogonal refers to the mutually exclusive reactivity of the new pair of chemical groups with any group found naturally in a biological system (*ie.* they do not cross-react). The result is control over target and site of bioconjugation even in complex biological mixtures (231). An additional benefit of using a bio-orthogonal reaction to label proteins is it can be used in conjunction with “natural” chemistries (*eg.* cysteine/maleimide). This allows a protein to be modified via two different, mutually exclusive reactions in a highly defined manor.

One such functional group is the azide, which reacts specifically with alkynes, neither of which are found naturally in biological systems. As discussed in Section 1.6.2, the model Click reaction (copper-catalysed azide-alkyne cycloaddition; CuAAC) was improved in terms of biocompatibility with the development of a copper-free variant; strain-promoted azide-alkyne cycloaddition (SPAAC). Both variations have been used to selectively label numerous biomolecules, including: RNA, lipids, glycans and proteins (67, 68, 229, 232). The biological

shortcoming of CuAAC is the requirement for Cu(I), which has been observed to be cytotoxic, cause peptide bond hydrolysis and/or to detrimentally affect downstream applications (66, 71, 233). SPAAC essentially overcomes such limitations but, as with Click chemistry as a whole, little is known about the influence of the local environment on the reaction efficiency or what affect the introduction of the Click handle itself may have on any given protein at a particular site. The evaluation and optimisation of the many facets concerning protein-related Click reactions is essential for future applications (79, 80, 234, 235) and this chapter aims to address this limited understanding.

7.1.1 Interfacing proteins with non-biological materials

The attachment of biomolecules to non-biological materials produces novel bioconjugates with applications including drug and contrast agent delivery (86, 236). Of particular interest to nanotechnology and material science is the defined and controllable attachment of proteins to surfaces (237). Surface labelling has many potential applications including the generation of biosensors (238, 239) and bioelectronic components such as protein transistors (119, 120), and ligand discovery and high-throughput screening via protein microarrays (230, 240). Additionally, surface labelling can be used to isolate and study proteins at the single molecule level, which is useful for gaining precise information about mechanism and dynamics of protein activity alleviating the averaging effects of bulk study. For example, time scales associated with different dynamic motions of lysozyme activity were measured by converting changes in protein motion into electrical signals (85). Valuable non-biological materials include gold, silicon and carbon allotropes such as carbon nanotubes (CNTs) and graphene (86). Graphene is billed as the material of this century because it has extraordinary electronic, mechanical and thermal properties (241). To exploit the properties of graphene and advance surface labelling in general, new approaches for defined labelling of materials is desirable. The importance of defined labelling was shown by Della Pia *et al* (119), where binding of cyt *b*₅₆₂ to gold via engineered thiol groups (in cysteine residues) generated bioconjugates with very different conductance properties depending on the orientation of binding. Uaa's such as azPhe provide new opportunities to generate protein-material interfaces due to the nature of azide chemistry and the ability to introduce the chemical handle at defined positions. Additionally, the bio-orthogonal reactivity removes issues associated with naturally occurring chemical handles or could be used in conjunction with natural amino acid modification to provide a second or third handle (Section 1.6.2). For example, pyrene (Figure 7.7A) is a planar group that has been shown to bind strongly to graphene via non-covalent π - π stacking (242, 243).

In this chapter the use of SPAAC to attach pyrene groups to sfGFP and cyt *b*₅₆₂ containing a surface exposed azido group was investigated. Pyrene-modified proteins were tested for the ability to bind to non-biological materials such as graphene, graphite and carbon nanotubes. The azide handle, in the form of the Uaa azPhe, was incorporated via a reprogrammed genetic code allowing the site of pyrene attachment and therefore orientation of binding to be highly defined.

7.2 Results

7.2.1 Effect of azPhe position on efficiency of sfGFP modification via Click chemistry

To test the effect of local protein environment on labelling efficiency by SPAAC, four surface exposed sfGFP residues were selected (Figure 7.1A and B). Each residue samples different features of sfGFP structure including surface accessibility, secondary structural element, local environment, native residue charge and location with respect to the chromophore. Lys26 is basic and has the highest surface exposure of the four residues at 169 Å². Lys26 resides at the boundary between a β-strand and a loop distal from the GFP chromophore (23 Å) with its side chain protruding fully into the solvent. Glu34 and Glu132 are acidic residues that reside in a β-strand and loop, respectively. Glu132 is the second most surface exposed residue (128 Å²) and the furthest from the chromophore (27 Å). Glu34 is less surface exposed at 83 Å², closer to the chromophore (14 Å) and is surrounded by charged, polar and hydrophobic residues. Gln204 is the only uncharged residue sampled and is located in a β-strand proximal to the chromophore (8 Å). Gln204 is the least surface exposed of the selected residues (73 Å²), and occupies an uncharged but polar local environment with a hydrophobic patch consisting of Phe223 and Val206.

The incorporation of azPhe at each of the four positions had little effect on the function of sfGFP. Absorption spectra of the sfGFP^{AzF} variants were similar to that of wild-type sfGFP (Figure 7.1C). It is important to note that the term wild-type is used here to represent the unaltered, parent sfGFP rather than a naturally occurring protein as sfGFP was generated via engineering of folding reporter GFP (104). Additionally, the fluorescent properties of the sfGFP^{AzF} variants were unchanged. No shifts in absorbance, excitation or emission wavelengths were observed indicating that the environment of the chromophore within sfGFP was unchanged. The only notable difference in the spectra was the extinction coefficient of sfGFP^{Q204AzF}, which was calculated as 55,000 M⁻¹.cm⁻¹ at 485 nm, higher than that calculated for the wild-type protein (49,000 M⁻¹.cm⁻¹; Figure 7.1C). The likely explanation is that incorporation of azPhe at position 204 promotes the deprotonated state (*ie.* phenolate; λ₄₈₅) of the sfGFP chromophore over the protonated form (*ie.* phenol; λ₄₀₀). This is evident by the reduction of the 400 nm shoulder region of the main absorbance peak. sfGFP^{AzF} variants remained predominately monomeric (Figure 7.1D) eluting from the column within 0.2 mL of each other with an average molecular mass of 28.6 kDa (pred. 27.8 kDa). Together these results indicated that incorporation of azPhe at the targeted residues did not significantly perturb structure or function of the protein.

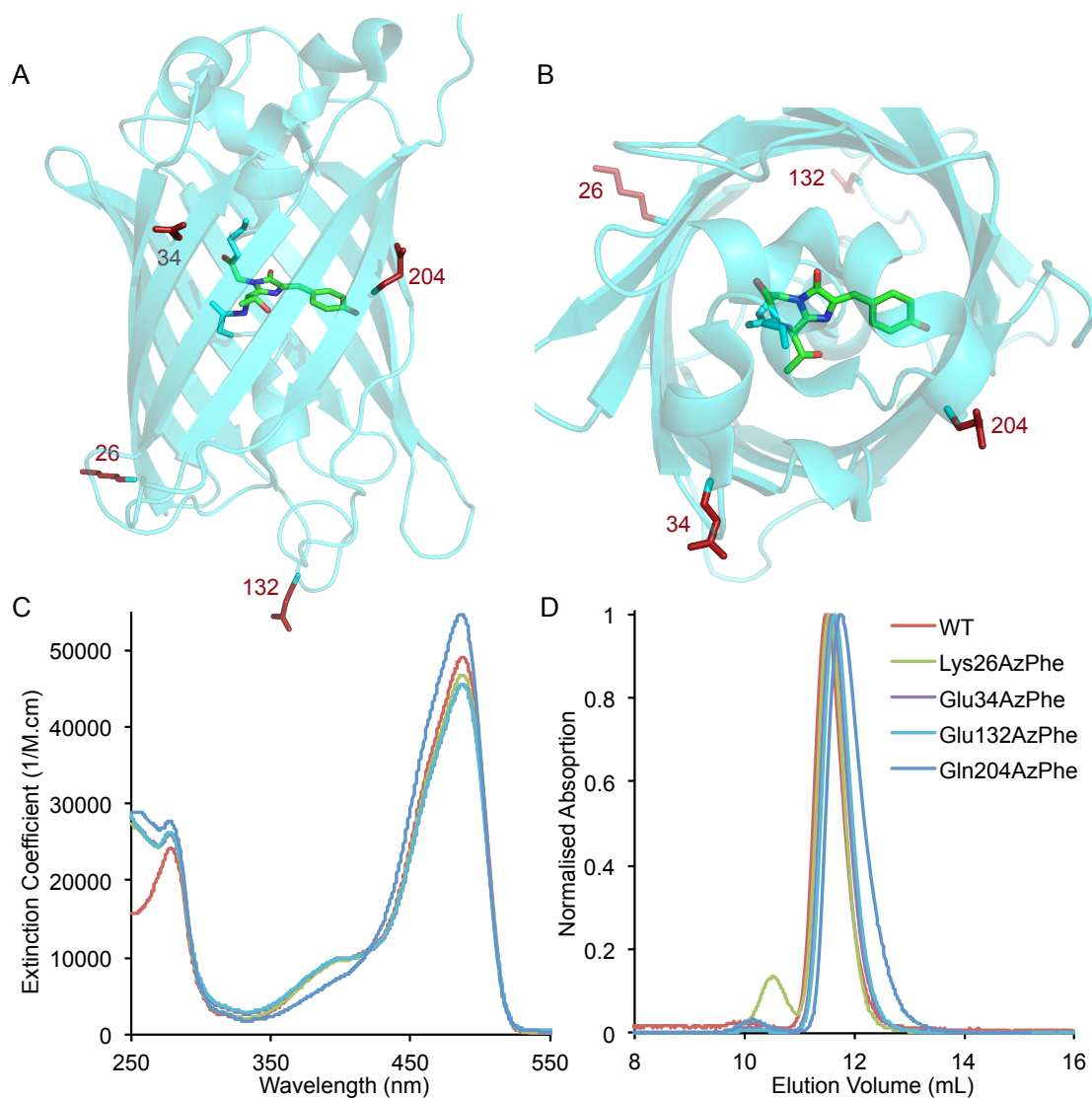


Figure 7.1. Residues of sfGFP chosen for modification via Click chemistry. Cartoon representation of sfGFP structure (PDB 2B3P) shown from two views; (A) side-on and (B) top-down. The 4 surface-exposed residues mutated to azPhe are shown as red sticks and labelled. (C) UV-visible absorption spectra of the sfGFP^{AzF} variants. Spectra were recorded using 12.5 μ M protein and converted to molar extinction coefficients according to Section 2.6.4. (D) Analytical gel filtration of sfGFP^{AzF} variants (10 μ M) as described in the Section 2.12.1. The average molecular weight of protein variants was calculated from the elution volume as 28.6 kDa (predicted 27.8 kDa). The small side peak of sfGFP^{K26AzF} ($V_e = 10.5$ mL) was calculated as 47.5 kDa (dimer predicted 55.6 kDa).

The sfGFP^{AzF} variants were subsequently modified with a dibenzylcyclooctyne functionalised red fluorescent probe based on sulforhodamine 101 (DBCO-585) by SPAAC (Figure 7.2A). Modifying a protein with a fluorescent tag is a useful test-bed for a reaction because of easy visualization and quantification. As well as being a probe for monitoring modification, DBCO-585 was chosen because its excitation spectrum overlaps with the emission spectrum of sfGFP potentially allowing for FRET from sfGFP to the attached DBCO-585 (*vide infra*).

Reactions were analysed by UV-visible spectrometry, fluorimetry and fluorescence imaging of SDS-PAGE gels (Section 2.12). Conjugation was performed using 10 μ M of each purified protein variant with 10, 20, 50 and 100 μ M of DBCO-585 (Section 2.12). DBCO-585 concentration did not affect the overall pattern of reaction yields from the different variants, however 100 μ M DBCO-585 gave lower yields for all of the variants (*vide infra*). As the reaction is second order and dependant on the concentration of both reactants, higher DBCO-585 concentrations gave higher reaction rates and so concentrations of either 20 or 50 μ M were subsequently used. The results in Figure 7.2 are from reactions with 50 μ M DBCO-585, incubated for 24 hours at 25 °C. Following the reactions, unreacted DBCO-585 had to be removed in order to spectroscopically analyse the reactions. The sfGFP^{AzF} variants, and therefore anything bound to them, were separated from unreacted DBCO-585 by Ni-affinity chromatography that specifically binds sfGFP. UV-visible absorbance of sfGFP (λ_{485}) and DBCO-585 (λ_{595}) (as in Figure 7.2D) were converted to concentrations using their extinction coefficients, as described in Section 2.12.3. The reaction yield was therefore calculated by the ratio of DBCO-585 bound to sfGFP.

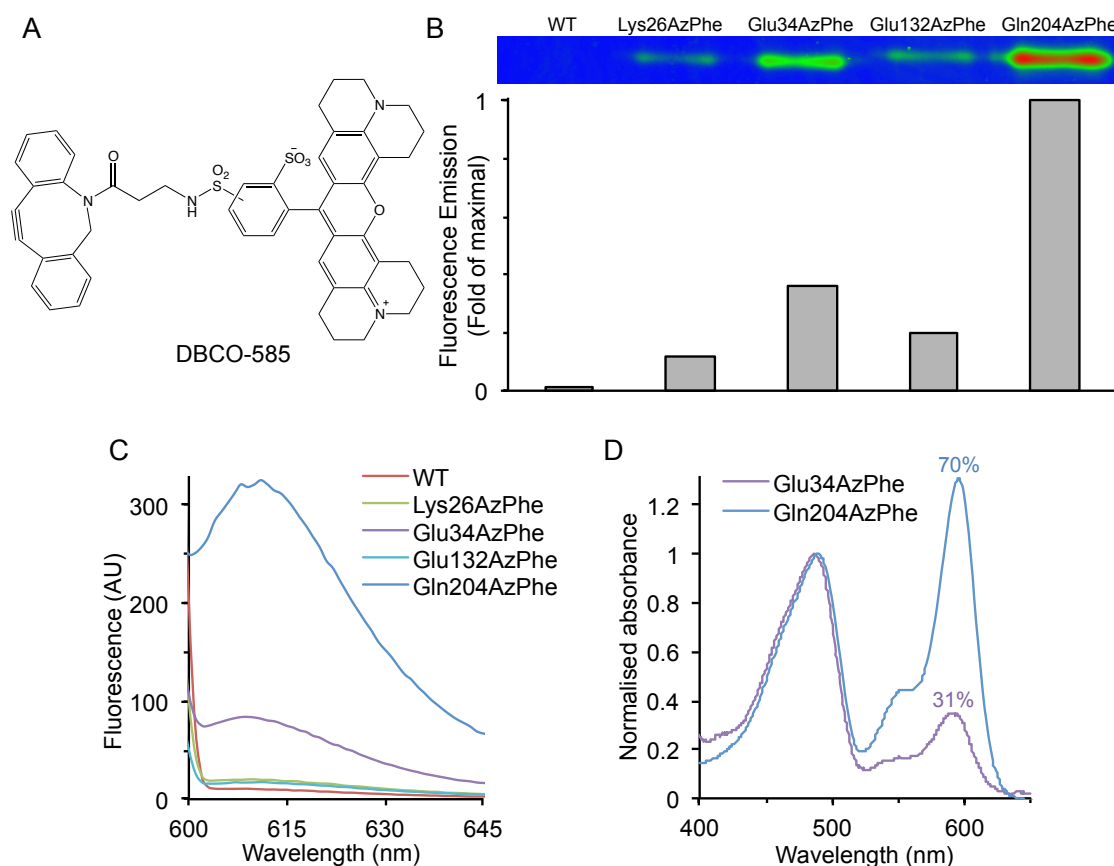


Figure 7.2. Modification of sfGFP^{AzF} variants with DBCO-585. (A) Chemical structure of DBCO-585. Reactions between 10 μ M sfGFP^{AzF} variants and 50 μ M DBCO-585 were incubated for 24 hours. Products of the reaction were analysed by (B) fluorescent SDS-PAGE analysis, (C) fluorescence emission spectroscopy ($\lambda_{exc} = 595$ nm). Fluorescent SDS-PAGE images were recorded on a Typhoon Imager with a 532 nm excitation laser and 610 nm emission filter. The image was false-coloured by a spectral progression from blue (0 %) to red (100%). Fluorescent bands were quantified using ImageJ software and plotted below. (D) UV-visible absorption analysis of the reactions of sfGFP^{E34AzF} and sfGFP^{Q204AzF}. Reaction yields are shown in percentage of sfGFP^{AzF} modified and were calculated using the extinction coefficients of the sfGFP^{AzF} variant and DBCO-585 as in Section 2.12.3.

A significant difference in the labelling efficiency of each sfGFP^{AzF} variant was observed (Figure 7.2). sfGFP^{Q204AzF} was modified to the greatest extent of the variants with between 68 and 76 % (from 6 reactions) of the protein modified with DBCO-585 after 24 hours. The next highest yield of modification was observed with sfGFP^{E34AzF} (26 to 31 %), which was nearly 3-fold lower than the sfGFP^{Q204AzF}. sfGFP^{K26AzF} and sfGFP^{E132AzF} were the least well modified variants with ~8 % and 15 % protein modified, respectively, equating to 9 and 5 fold lower yields compared to sfGFP^{Q204AzF}. No labelling was observed for wild-type sfGFP (Figure 7.2B and C) confirming that DBCO-585 is reacting selectively with azPhe and that Ni-affinity chromatography removed any unreacted DBCO-585.

Given that the nature of the SPAAC reaction, it should be compatible with cell-based systems. To test this, as *E. coli* are not permeable to DBCO-585, conjugation was performed using cell lysates to mimic intracellular conditions more closely. Cell lysates containing one of the sfGFP^{AzF} variants were standardised to cell number using absorbance at λ_{600} (of 1.0) and incubated with 100 μ M DBCO-585 for 20 hours. Each of the sfGFP^{AzF} variants was specifically modified with DBCO-585, with the modification yield following the same trend as with the purified protein (Figure 7.3). The discrepancy in protein mobility through the SDS-PAGE gel in Figure 7.3A between crude (C) and Ni-affinity purified (Ni) reactions was a result of the buffer. Following Ni-affinity purification, the sample buffer contains 500 mM imidazole that affects the mobility.

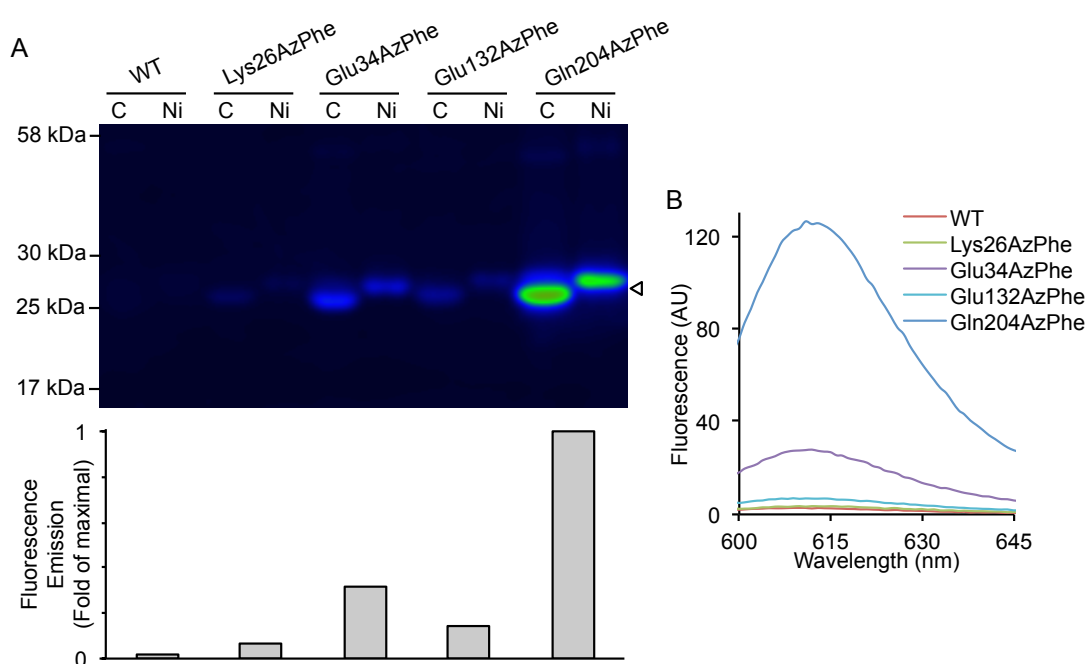


Figure 7.3. SPAAC modification of sfGFP^{AzF} variants in cell lysates. Fluorescent analysis of reactions between cell lysates (standardised to A_{600} of 1.0) and DBCO-585 (100 μ M). (A) Fluorescent SDS-PAGE images were recorded on a Typhoon Imager with a 532 nm excitation laser and 610 nm emission filter. For each sfGFP^{AzF} variant there are two lanes that represent the crude reaction (C) run directly on the gel and after Ni-purification (Ni) to remove unreacted DBCO-585. Molecular weight standards are shown on the left and a white arrow indicated the position of sfGFP. As in **Figure 7.2**, the image was false-coloured by a spectral progression from blue (0 %) to red (100%). Fluorescent bands were quantified using ImageJ software and plotted below. (B) Fluorescent emission spectra were recorded after excitation of DBCO-585 (595 nm) to measure the amount of DBCO-585 conjugated to each sfGFP variant. Unreacted DBCO-585 was removed by Ni-purification prior to fluorimetry. Reactions were incubated for 20 hours at 25 °C.

The rate of protein modification was then examined, as in biological applications and *in vivo* reactions this is often a critical parameter for assessing the temporal resolution achievable for real-time labelling and monitoring. Initially, the optimal molar ratio of

sfGFP^{Q204AzF} to DBCO-585 was investigated using a constant value of 10 μM protein with 1, 2, 5 or 10 molar equivalents of DBCO-585. The greatest extent of modification was observed with 2 or more equivalents (Figure 7.4A). However, at 10 molar equivalents (100 μM) of DBCO-585 there was a slight decrease in total modification yield possibly resulting from a decrease in protein solubility from increased DMSO (used to solubilise DBCO-585) in the reaction. As discussed, reactions were therefore performed with either 20 μM or 50 μM DBCO-585.

The reaction rate was elucidated for the two variants that modified to the highest yields, sfGFP^{Q204AzF} and sfGFP^{E34AzF} (10 μM), with 20 μM DBCO-585. The progress of the reaction was followed by FRET (*vide infra*) with sfGFP as the donor and DBCO-585 as the acceptor (Figure 7.4B). Reactions were performed in a fluorimetry cuvette and monitored by measuring DBCO-585 emission (λ_{611}) following excitation of sfGFP (λ_{485}). As FRET is heavily distance dependant, DBCO-585 emission would only be observed when it is attached to sfGFP. The end-point plateaus of the kinetics reactions matched reaction yields observed previously (Figure 7.2D). The second order rate constants calculated were very different with sfGFP^{Q204AzF} ($17.8 \text{ M}^{-1} \cdot \text{s}^{-1}$) modifying with a 10 fold higher rate than sfGFP^{E34AzF} ($1.6 \text{ M}^{-1} \cdot \text{s}^{-1}$).

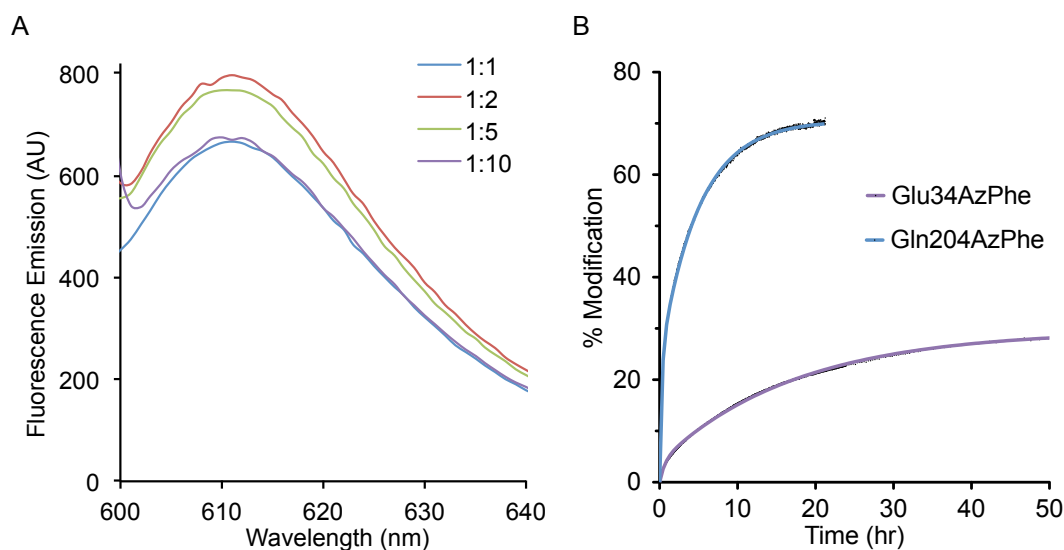


Figure 7.4. Reaction kinetics of sfGFP^{E34AzF} and sfGFP^{Q204AzF} modification with DBCO-585. (A) Optimisation of protein:probe molar ratio for the reaction between sfGFP^{Q204AzF} and DBCO-585. Reactions were assessed by fluorescence measuring amount of DBCO-585 ($\lambda_{\text{exc}} = 595 \text{ nm}$). sfGFP^{Q204AzF} concentration was kept constant at 10 μM and DBCO-585 concentration was varied from 10 μM (1:1, blue line), 20 μM (1:2, red line), 50 μM (1:5, green line) to 100 μM (1:10, purple line). Reactions were allowed to proceed for 20 hours. (B) Reaction progress for sfGFP^{E34AzF} and sfGFP^{Q204AzF} (10 μM) with DBCO-585 (20 μM) as measured by FRET and quantified by UV-visible absorption. Data points are the average from 3 repeats.

An important role of bioconjugation is to modify a protein's properties. Effectiveness is dependent on the site of modification and the nature of the incoming group. To demonstrate the effect of post-translational modification DBCO-585 was used to act as a FRET partner with sfGFP (Figure 7.5A) to extend sfGFP emission into the red region. sfGFP^{Q204AzF} was used to demonstrate FRET, as residue 204 is close to the chromophore (7–11 Å; Figure 7.1), allowing for efficient energy transfer. sfGFP^{Q204AzF} modified with DBCO-585 exhibited red fluorescence (λ_{em} 611 nm) when excited at the optimal excitation wavelength of sfGFP (λ_{exc} 485 nm). This single modification has increased the Stokes shift of the protein from 26 nm to 126 nm by red-shifting emission by 100 nm (Figure 7.5B).

Förster resonance energy transfer (FRET) events are extremely important in both nature (*eg.* light harvesting in photosynthetic complexes) and for research (*eg.* studying protein structure and dynamics) (244). The energy transfer efficiency is a crucial factor for whether a transfer is useful and can be used to calculate the distance between chromophores in FRET experiments. The FRET efficiency between the GFP chromophore (donor) and DBCO-585 (acceptor) was determined experimentally from the fluorescence spectra as 90 % (Section 2.12.5; Equations 2.4-2.6). Theoretical calculations were then used to both support the observed FRET and determine energy transfer efficiency. The inter-chromophore distance was estimated to be 30 Å by modelling the attachment of DBCO-585 to sfGFP^{Q204AzF}. The Förster distance or R_0 (distance that gives 50 % FRET efficiency) was calculated as 43 Å (Figure 7.5C) based on the determined spectral overlap of the donor–acceptor pair (J ; 6.3×10^{14} nm), the quantum yield of the donor (Q_D ; 0.65), the refractive index of the solvent (1.4 for aqueous solvent) and the assumed orientation factor (k^2 ; 2/3 for an undefined orientation) (Section 2.12.5; (245)). Using the Förster distance and the observed energy transfer efficiency (90 %; Figure 7.5C), it was then possible to back calculate the distance between chromophores (r) as ~30 Å. This calculated value coincides with the distance predicted from modelling (30 Å, *vide supra*) placing confidence in the FRET efficiency observations and calculations.

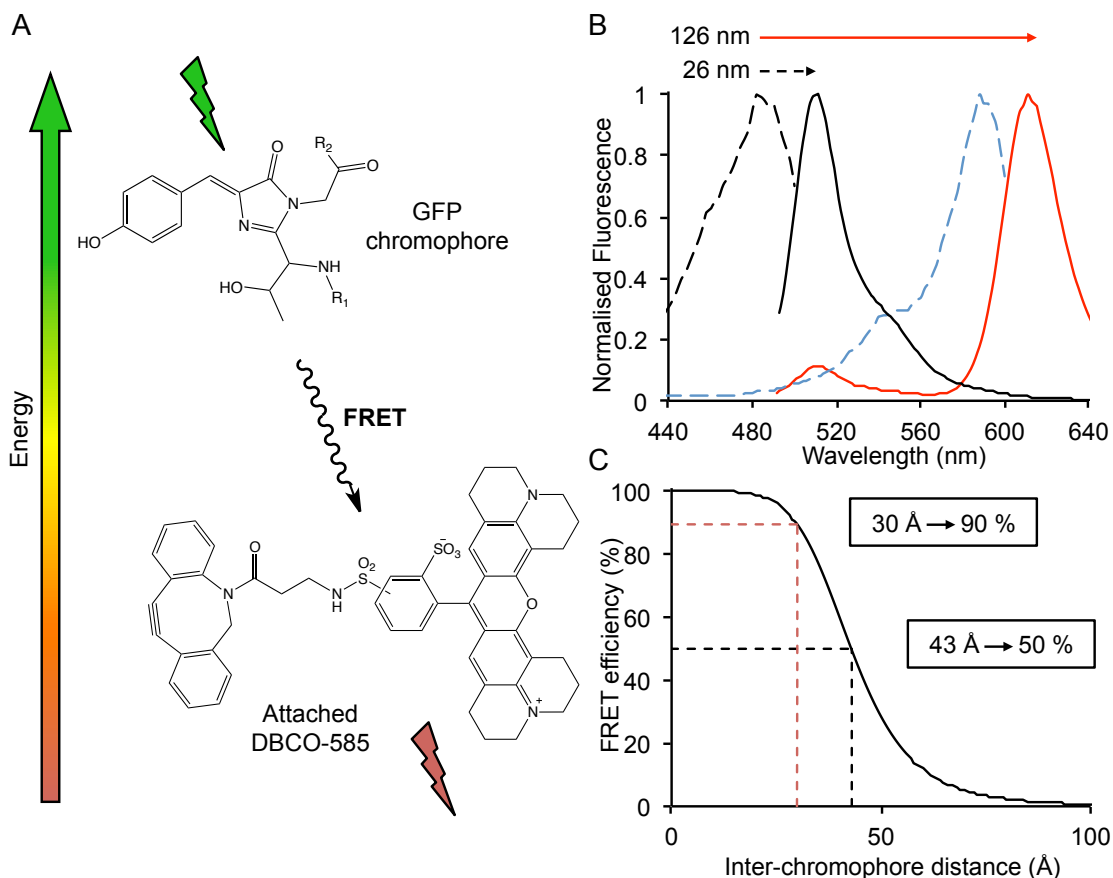


Figure 7.5. Förster resonance energy transfer of modified sfGFP^{Q204AzF}. (A) Schematic representation of FRET from the chromophore of sfGFP to the attached DBCO-585 dye. (B) Fluorescent properties of modified sfGFP^{Q204AzF}. Normalised emission spectra are shown for sfGFP^{Q204AzF} (black line) and sfGFP^{Q204AzF} modified with DBCO-585 (red line) both after excitation at 485 nm). The excitation spectra (λ_{em} monitored at 511 nm) of unmodified sfGFP^{Q204AzF} (dashed black line) and unmodified DBCO-585 (dashed blue line) are also shown. The arrows above represent the Stokes shift of unmodified (black dashes) and modified (red) sfGFP^{Q204AzF}. (C) Theoretical FRET efficiency dependence on inter-chromophore distance. The Förster distance (R_0) was calculated as 43 Å and the inter-chromophore distance (30 Å) estimated from modelling is indicated by red dashed lines.

7.2.2 Effect of azPhe position on the efficiency of cyt b_{562} modification via Click chemistry

To follow up the observation that azPhe position within a protein (sfGFP, Section 7.2.1) can result in vastly different SPAAC reaction yields the experiment was repeated with cyt b_{562} . AzPhe was incorporated into cyt b_{562} at four different surface exposed positions, which were previously selected for their ability to bind gold electrodes (119). As discussed above (Section 7.1.1 and Chapter 3), the cyt b_{562}^{AzF} variants generated herein were engineered with partnering cysteine mutations at the opposite position to azPhe. This created four variants that fall into two pairs depending on the orientation of the mutations called short axis (SA; Asp5 and Lys104) and long axis (LA; Asp21 and Asp50) as seen in Figure 7.6A and B. Asp5 and Lys104 are located in well defined α -helices (helix 1 and 4, respectively) in the holo-protein but these regions become dynamic in absence of haem. The residues are approximately 12 Å from the central haem. Asp21 and Asp50 are located further from the haem (20 and 32 Å) in a turn between helices 1 and 2 and a long flexible loop that links, helices 2 and 3, respectively. The cyt b_{562}^{AzF} variants generated were analysed by PAGE and UV-vis absorption to assess whether the incorporation of azPhe impaired the structure or function of the proteins. Both denaturing and native PAGE, as with gel filtration chromatography during purification (Chapter 3), indicated that the cyt b_{562}^{AzF} variants had a greater propensity to form dimers than wild-type cyt b_{562} . This difference was most likely due to the complementary cysteine mutations introduced at surface exposed residues on the opposite side of the protein to the azPhe mutations (Section 1.8). The high concentration of cyt b_{562}^{AzF} variants during production and purification increases the probability of forming disulphide bridges with other cyt b_{562}^{AzF} molecules under oxidising conditions. To prevent dimer formation, these variants were purified and stored in buffers containing the reducing agent ascorbic acid.

Mutations did not have a major effect on the wavelengths of absorbance (Figure 7.6C and D) however the SA mutations (Asp5 and Lys104) resulted in lower extinction coefficients for the haem-associated absorbance peaks ($\lambda_{418/427}$) under both oxidizing and reducing conditions (Figure 7.6D). This decreased absorbance indicates a change (and possible reduction) in the haem binding ability of the protein, which is rationalized by the proximity of residues 5 and 104 to the haem binding site including the coordinating residues Met7 and His102 (Figure 7.6A).

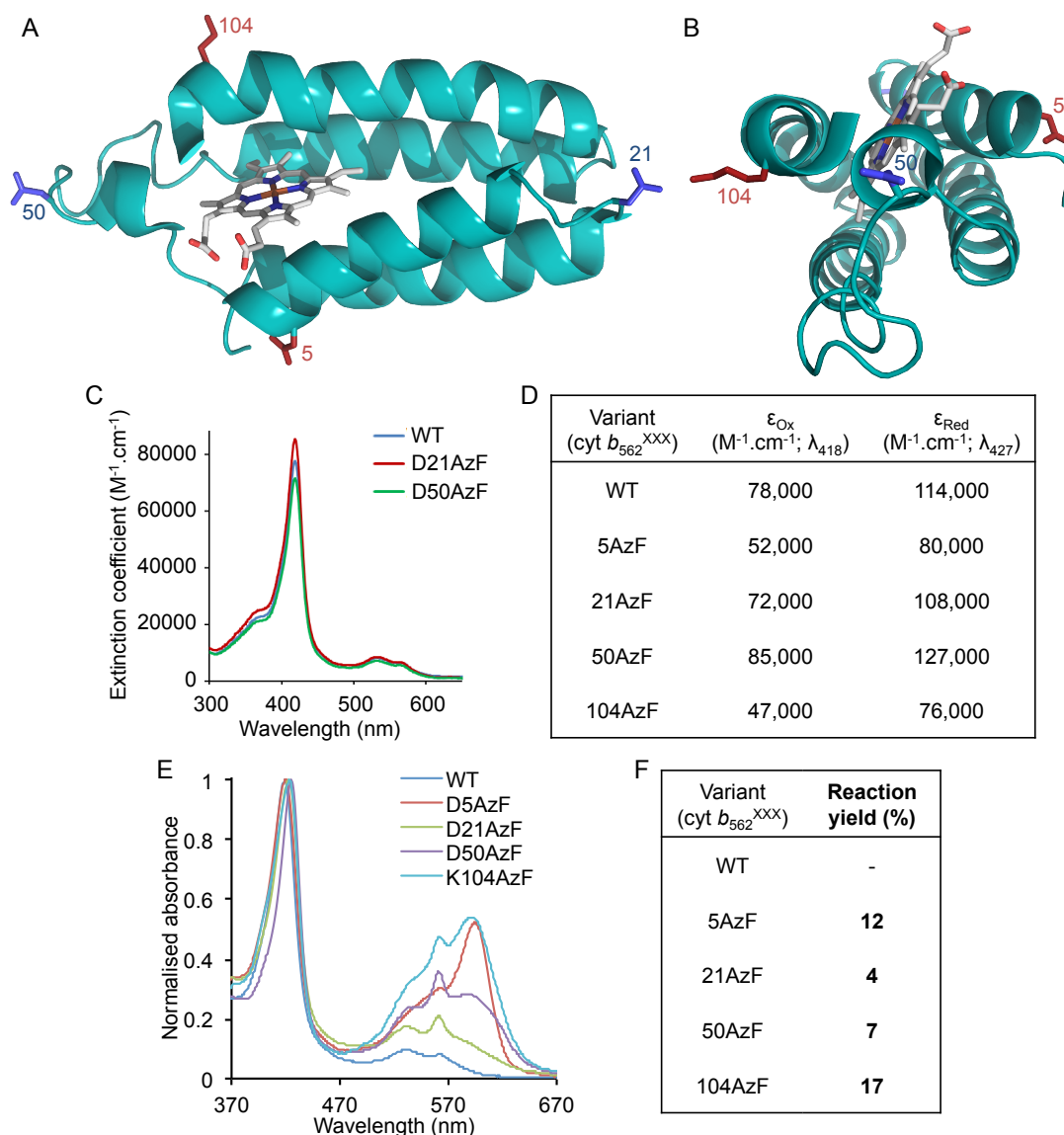


Figure 7.6. Engineering and modification of cyt b_{562}^{AzF} variants via Click chemistry. 3-dimensional representation of cyt b_{562} (PDB accession 1QPU) from two views; (A) side-on and (B) top-down. The 4 surface-exposed residues mutated to azPhe are shown as red and blue sticks. Cyt b_{562}^{AzF} variants were engineered with partnering cysteine mutations at the opposite residue to the azPhe. The two pairs are shown as red (short axis) and blue (long axis) sticks. (C) Comparison of the UV-vis absorption wavelength of wild-type cyt b_{562} and cyt b_{562}^{D21AzF} and cyt b_{562}^{D50AzF} under oxidizing (ϵ_{Ox}) conditions. (D) Extinction coefficients (ϵ) of the cyt b_{562}^{AzF} variants under oxidizing (ϵ_{Ox}) and reducing (ϵ_{Red}) conditions in comparison to wild-type cyt b_{562} . (E) UV-visible absorption analysis of reactions between 10 μ M holo cyt b_{562}^{AzF} variants and 20 μ M DBCO-585 after overnight incubation. Unreacted DBCO-585 was removed by gel filtration. (F) ϵ were used to calculate the reaction yields from (E) (DBCO-585 $\epsilon = 102,000$).

Purified cyt b_{562}^{AzF} variants (10 μM) were modified with DBCO-585 (20 μM) and reactions were analysed by UV-vis absorption (Figure 7.6C). Again, the position of azPhe incorporation had a significant effect on reaction yield (Figure 7.6D) with endpoints ranging from 4 to 17 % of the total protein modified (Lys104 > Asp5 > Asp50 > Asp21). Generally, reaction yields were much lower than for sfGFP with average yields of 10 % of total cyt b_{562}^{AzF} modified *versus* ~33 % for sfGFP^{AzF}. Also, the maximum observed yield for a cyt b_{562}^{AzF} variant (cyt b_{562}^{K104AzF}) was lower at 17 % protein modified *versus* 76 % for sfGFP^{Q204AzF} (Section 7.2.1).

7.2.3 Modifying sfGFP and cyt b_{562} with DBCO-pyrene

One specific goal of developing the SPAAC reaction was to generate proteins that could bind to graphene in a defined, precise and stable manner (Section 7.1.1). This can be achieved by modifying a protein at defined residue positions with chemical adducts that promote π stacking with graphene. One such compound is pyrene (Figure 7.7A). While cysteine-based chemistry is available for attaching pyrene, this becomes problematic when a protein contains native cysteines. Thus the SPAAC-based approach allows for defined placement of pyrene in the context of the protein. This required the synthesis of an azide-reactive pyrene probe for modifying proteins via SPAAC (DBCO-pyrene; Figure 7.7A). DBCO-pyrene was synthesised by a nucleophilic substitution ($\text{S}_{\text{N}}2$) reaction of two commercially available reagents, one containing the DBCO group (DBCO-amine) and one with the pyrene (pyrene butanoic acid, succinmidyl ester). DBCO-amine (18 mg, 65 μmol) and pyrene butanoic acid, succinmidyl ester (25 mg, 65 μmol) were mixed in 4 mL anhydrous THF under an argon atmosphere. TEA (9 μL , 71.5 μmol) was added to the solution to prevent protonation of DBCO-amine and hence promote the reaction. The resulting yellow solution was stirred at room temperature until the starting materials had been consumed. The reaction progress was monitored by thin layer chromatography (TLC), which allowed visualisation of DBCO-amine (by ninhydrin) and pyrene butanoic acid, succinmidyl ester (by UV) and indicated complete conversion in ~ 12 hours. The resulting compound was purified as described in Section 2.13.1 and analysed by TLC (silica gel, EtOAc, $R_f = 0.42$), $^1\text{H-NMR}$, $^{13}\text{C-NMR}$ and mass spectrometry. The predicted mass of 547.24 Da was observed and NMR confirmed the designed product, DBCO-pyrene. 30 mg of DBCO-pyrene was purified giving a yield of 84 % for the synthesis and purification.

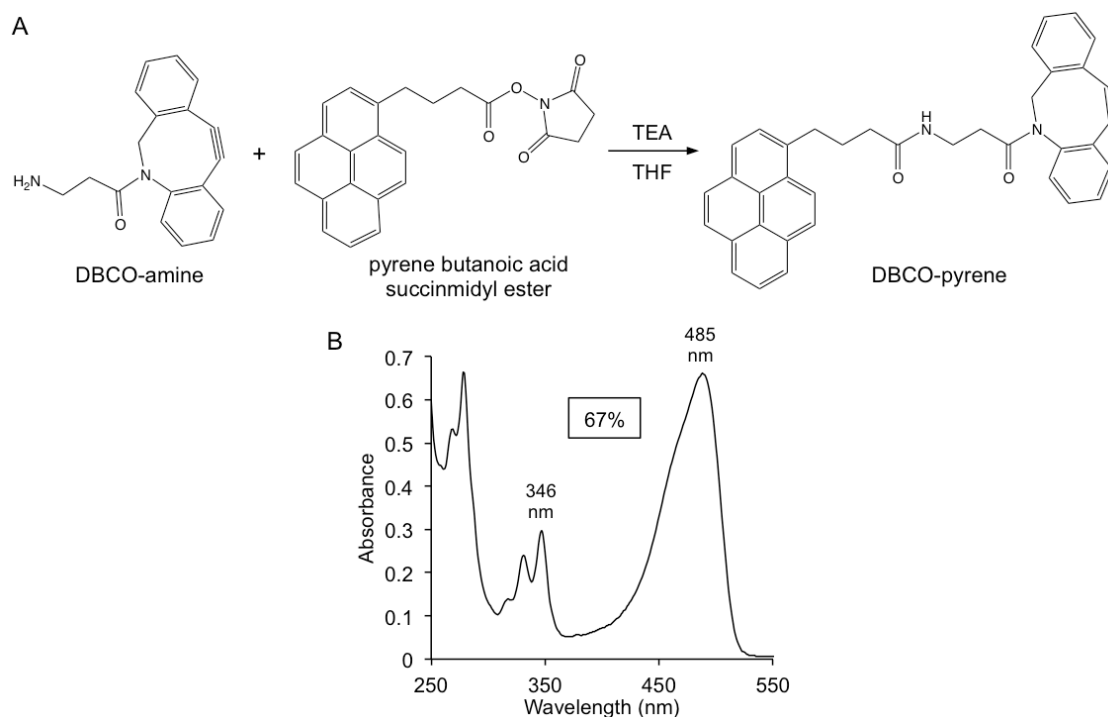


Figure 7.7. Synthesising DBCO-pyrene. (A) Schematic of the reaction between DBCO-amine and pyrene butanoic acid, succinimidyl ester forming DBCO-pyrene. (B) UV-visible absorbance analysis of the reaction of sfGFP^{Q204AzF} (20 μ M) and DBCO-pyrene (100 μ M). The reaction yield, as shown in the box, was 67 % of sfGFP^{Q204AzF} modified. The yield was calculated using the extinction coefficients of sfGFP^{Q204AzF} (55,000 $M^{-1}cm^{-1}$) and DBCO-pyrene (37,000 $M^{-1}cm^{-1}$) at the indicated wavelengths, as in Section 2.12.3.

DBCO-pyrene was used to modify sfGFP and cyt b_{562} variants containing surface accessible azPhe residues, as described above (Figure 7.2 and Figure 7.6). Based on experiments outlined in Sections 7.2.1 (Figure 7.2) and 7.2.2 (Figure 7.6), variants displaying the best modification efficiency with DBCO-585, namely sfGFP^{Q204AzF} and cyt $b_{562}^{K104AzF}$, were selected for modification with pyrene. Conjugation was performed using 20 μ M purified protein with a 5 molar excess of DBCO-pyrene, as in Section 2.12.2 and 7.2.1. Reactions were stopped by removing unreacted DBCO-pyrene and analysed by UV-vis absorption based on the absorbance of pyrene (λ_{346}) and either sfGFP (λ_{485}) or cyt b_{562} (λ_{418}) (Figure 7.7B). Modification yields with sfGFP^{Q204AzF} were similar to those with the fluorescent probe DBCO-585 with 67 % of the protein labelled (Figure 7.7B). Conversely, the yield of the cyt b_{562}^{D5AzF} SPAAC reaction with DBCO-pyrene (40 %; with A. McGarrity) was significantly higher than initial attempts with DBCO-585 (12 %; Figure 7.6).

After removal of unreacted DBCO-pyrene, modified sfGFP^{Q204AzF} was separated from unmodified sfGFP^{Q204AzF} by anion exchange chromatography using a MonoQ column (Section 2.5.3.4). Elution from the column was monitored for absorbance at both at 346 nm and 485 nm indicative of pyrene and sfGFP, respectively. Separation of the sfGFP^{Q204AzF} reaction mixture

gave two distinct elution peaks, one that absorbed at just 485 nm (110 mM NaCl) and one that absorbed at both 346 and 485 nm (145 mM NaCl). The elution that absorbed at 346 and 485 nm was therefore thought to be modified protein (sfGFP^{Q204Pyr}). Unfortunately, a full UV-vis absorption spectrum of this elution peak indicated that the protein was not fully modified (by comparison of the relative intensities of the pyrene to sfGFP absorption peaks with their known extinction coefficients; *vide supra*).

7.2.4 Attaching proteins to carbon allotropes

As stated above, the main reason for attaching pyrene to a protein is to promote defined and stable interactions with carbon allotropes, such as graphene. To investigate this, sfGFP^{Q204AzF} modified with pyrene (sfGFP^{Q204Pyr}) was tested for the ability to bind to small flakes of highly ordered pyrolytic graphite (HOPG) (Figure 7.8). HOPG is a good substitute for graphene for preliminary experiments as it has similar character being essentially multiple stacked layers of graphene. sfGFP^{Q204Pyr} was incubated with HOPG with a range of different conditions including protein concentration (0.05 to 5 μ M), incubation time (5 min – 2 hours) and solvent additives (>50 % (v/v) acetonitrile). The surfaces were imaged using atomic force microscopy (AFM) (with A. Moskalenko, A. Zaki and A. McGarrity). AFM is a type of scanning microscopy that uses an oscillating probe to examine the profile of a surface with high-resolution. AFM gives precise information about the dimensions (particularly the height) of features on a sample surface and so is ideal for assessing HOPG for protein binding.

Initially a drop of protein in phosphate buffer (typically 10 μ L) was spotted onto a HOPG sample, incubated and then washed with dH₂O to remove weakly bound proteins. Using this approach no protein was detected bound to the surface (by AFM) so an alternative deposition technique was used whereby the HOPG sample was immersed in ~25 μ L protein sample. The immersion approach was successful and resulted in surface decorated proteins where the extent of protein attachment (monolayer to individual proteins) depended on protein concentration (Figure 7.8A,C,E). The surface features that were assigned as proteins here had lateral dimensions of ~10-15 nm with heights of 4 – 5 nm. The lateral resolution of AFM is lower than height resulting in the observed dimensions appearing to large (predicted ~5.5 x 4 nm). The height dimensions matched closely with those predicted for sfGFP bound with the plane of the β -barrel parallel to the surface (~4 nm), providing confidence in their assignment as proteins. The incubation time (5 min) and washing procedure (stirred in ~100 mL dH₂O) was kept constant and protein concentration was varied from 0.05 to 5 μ M. At 5 μ M proteins formed a monolayer over the HOPG surface with small circular gaps containing no protein. When the concentration was decreased to 0.5 μ M, proteins formed monolayer “islands”, almost the reciprocal of 5 μ M (Figure 7.8A). Importantly, at the same concentration, wild-type

sfGFP did not decorate the surface to the same extent or with the same regular pattern (Figure 7.8B). At 0.25 μM , sfGFP^{Q204Pyr} covered the HOPG surface but single protein molecules were observed (Figure 7.8C), whereas wild-type sfGFP bound to a far lesser extent (Figure 7.8D). Although the wild-type protein was shown to have some affinity for HOPG at these concentrations, this protein was easily moved and brushed off of the surface by the AFM probe during imaging indicating a weak, non-specific interaction. The best conditions for protein attachment were observed with 0.05 μM sfGFP^{Q204Pyr} (in PB; 5 min incubation) as they resulted in individual, well-separated proteins (Figure 7.8E).

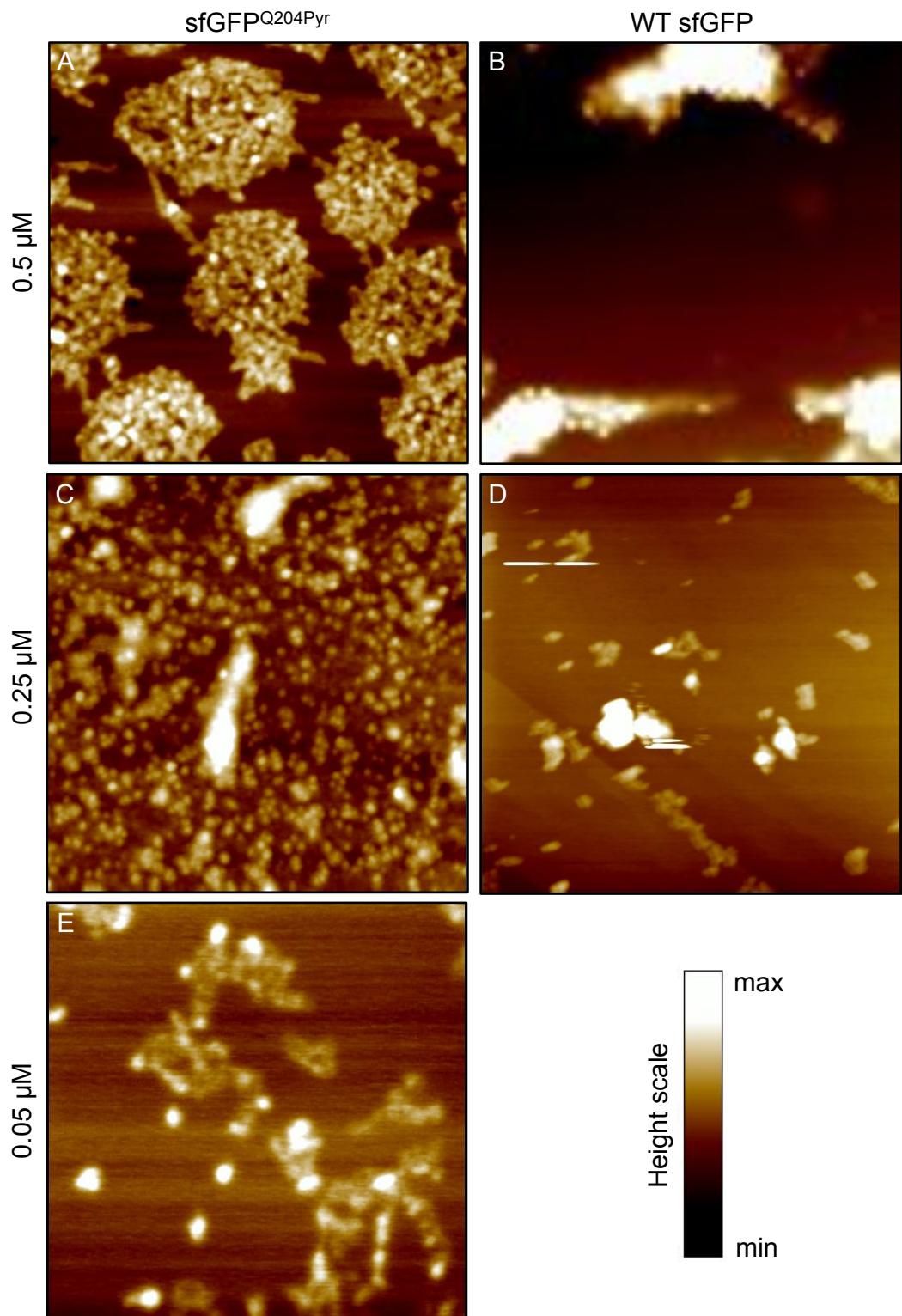


Figure 7.8. Labelling graphene with sfGFP^{Q204Pyr}. AFM images showing HOPG after incubation with sfGFP^{Q204Pyr} or wild-type (WT) sfGFP for 5 min. Images are coloured by height profile from black-through gold-to white (bottom right). (A) 0.5 μM sfGFP^{Q204Pyr}, height scale = 7 nm. (B) 0.5 μM wild-type sfGFP, height scale = 30 nm. (C) 0.25 μM sfGFP^{Q204Pyr}, height scale = 15 nm. (D) 0.25 μM wild-type sfGFP, height scale = 20 nm. (E) 0.05 μM sfGFP^{Q204Pyr}, height scale = 5 nm. All images are 1 \times 1 μm apart from (E) which is 500 \times 500 nm. The predicted dimensions of sfGFP^{Q204Pyr} are 5.5 \times 4 nm, 4 nm height.

7.3 Discussion

Bioconjugation is a useful way of studying, exploiting and altering the properties of proteins (228, 229). Protein conjugates allow tracking (77), easy purification (232) and the creation of novel protein-based devices for purposes from energy harvesting to medical diagnostics (86). Traditionally, protein bioconjugation exploits the chemistry of natural amino acids, such as lysine, cysteine and the amino group of the N-terminal residue. These residues are common and the reactive chemical groups (amino, sulfhydryl) are found in other classes of molecules (*eg.* DNA and lipids). The result is poor control over the site and number of modifications, and incompatibility of the reaction with complex mixtures or *in vivo* reactions. Uaa's are valuable additions to the bioconjugation toolbox as they allow for highly specific modification that does not cross-react with any chemical groups found naturally in a cell (*ie.* bio-orthogonal). One such Uaa is azPhe, which can undergo cycloaddition reactions with alkyne groups to form stable covalent linkages. The selectivity of the reactions between azides and alkynes and their absence from nature, in combination with an expanded genetic code approach for incorporation of azPhe, allows for site-specific modification of target proteins in complex mixtures.

7.3.1 Precision of SPAAC based protein modification

As mentioned, the introduction of novel chemical handles such as azido groups into proteins opens up the possibility of new post-translational modification approaches by biocompatible reactions such as SPAAC. In the current context, the precision or bio-orthogonality of the reaction appears to be exquisite both on isolated proteins and in complex mixtures of biological molecules. Purified samples of wild-type sfGFP or cyt *b*₅₆₂ lacking azPhe were not modified (Figure 7.2 and Figure 7.6). In contrast, variants engineered to contain azPhe at defined positions could be selectively modified with a number of DBCO-containing probes (Figure 7.2, Figure 7.6, Figure 7.7). Furthermore, conjugation in cell lysates to mimic intracellular conditions more closely also showed highly precise labelling of the target protein (Figure 7.3). Cell lysates are complex, highly heterogeneous mixtures so are also a good test of bio-orthogonality. Only cell lysates containing the engineered sfGFP^{AzF} variants displayed any significant labelling capacity. No significant wild-type target protein or native protein appeared to be modified with DBCO-585 (Figure 7.3). Together these results confirm that the SPAAC reaction is bio-orthogonal and compatible with cellular environments (70).

7.3.2 Effect of azPhe position on SPAAC efficiency

Proteins are highly complex and dynamic molecules with each residue sampling a unique local environment. This makes the prediction of the optimal site for modification difficult because not only does the influence on protein function need to be considered but so do other factors such as accessibility, local charge and ultimately residue reactivity. These factors vary not only from protein to protein but also from site to site within the same protein. An expanded genetic code approach to incorporate azPhe into target proteins at defined positions allows the effect of azPhe position on SPAAC efficiency to be assessed.

Each of the variants generated allows azPhe to sample different structural and chemical environments and thus allows their influence on the SPAAC reaction to be evaluated. Firstly, the mutation of the endogenous residue to azPhe did not significantly affect the structure, function or oligomeric state of the protein (Figure 7.1), which is an important prerequisite for use of an Uaa for modification. Of the four sfGFP^{AzF} variants studied here sfGFP^{Q204AzF} predominantly and sfGFP^{E34AzF} to a lesser degree were modified to the highest extents. The only apparent connection or similarity between these variants is that they have azPhe incorporated within β -strands of the highly structured β -barrel of sfGFP and they are the least solvent exposed of the four residues (Glu34, 83 Å²; Gln204, 73 Å²; Section 7.2.1). The low accessibility and position within β -strands decreases the conformational freedom compared to more exposed and structurally heterogeneous residues on flexible loops. The reduced flexibility may act to present the azido group to an incoming alkyne whereas a flexible residue may spend a proportion of time occupying an inaccessible conformation. sfGFP^{K26AzF} and sfGFP^{E132AzF} were modified to far lower extents and have azPhe incorporated at the end of a β -strand and in a flexible loop region, respectively. Lys26 and Glu132 have the highest surface exposure of the selected positions (169 Å² and 128 Å²) and are located in more flexible regions. These residues have greater conformational freedom than Glu34 and Gln204 with fewer neighbouring residues (Figure 7.9A and C). Also, Lys26 and Glu132 have multiple charged and polar residues nearby (Figure 7.9A and C).

The prevailing view of protein modification is that the more surface accessible and flexible a residue is, the better it will be labelled with many reactive handles placed in loop regions. Results presented here contradict this with surface accessibility inversely proportional to reaction efficiency and the two residues in defined secondary structures (sfGFP^{Q204AzF} and sfGFP^{E34AzF}) modifying to the highest extent. It seems counterintuitive that higher surface exposure and thus accessibility hinders efficiency and rate. Therefore, the local surface microenvironment of a protein may be the critical factor in defining the SPAAC reaction, as has been suggested recently for antibody labelling via cysteine–maleimide linkage (246) as well as other reactions (247). Gln204 occupies an amphipathic, but uncharged environment

(Figure 7.9D). The hydrophobic patch created by Val206 and Phe223 next to Gln204 may facilitate docking of the hydrophobic DBCO group, which could promote the reaction as SPAAC is proximity dependent. Similarly, Glu34 occupies an amphipathic environment with neighbouring hydrophobic patch (Val11, Pro13, Gly33), however Glu34 is also close to charged residues (Figure 7.9B).

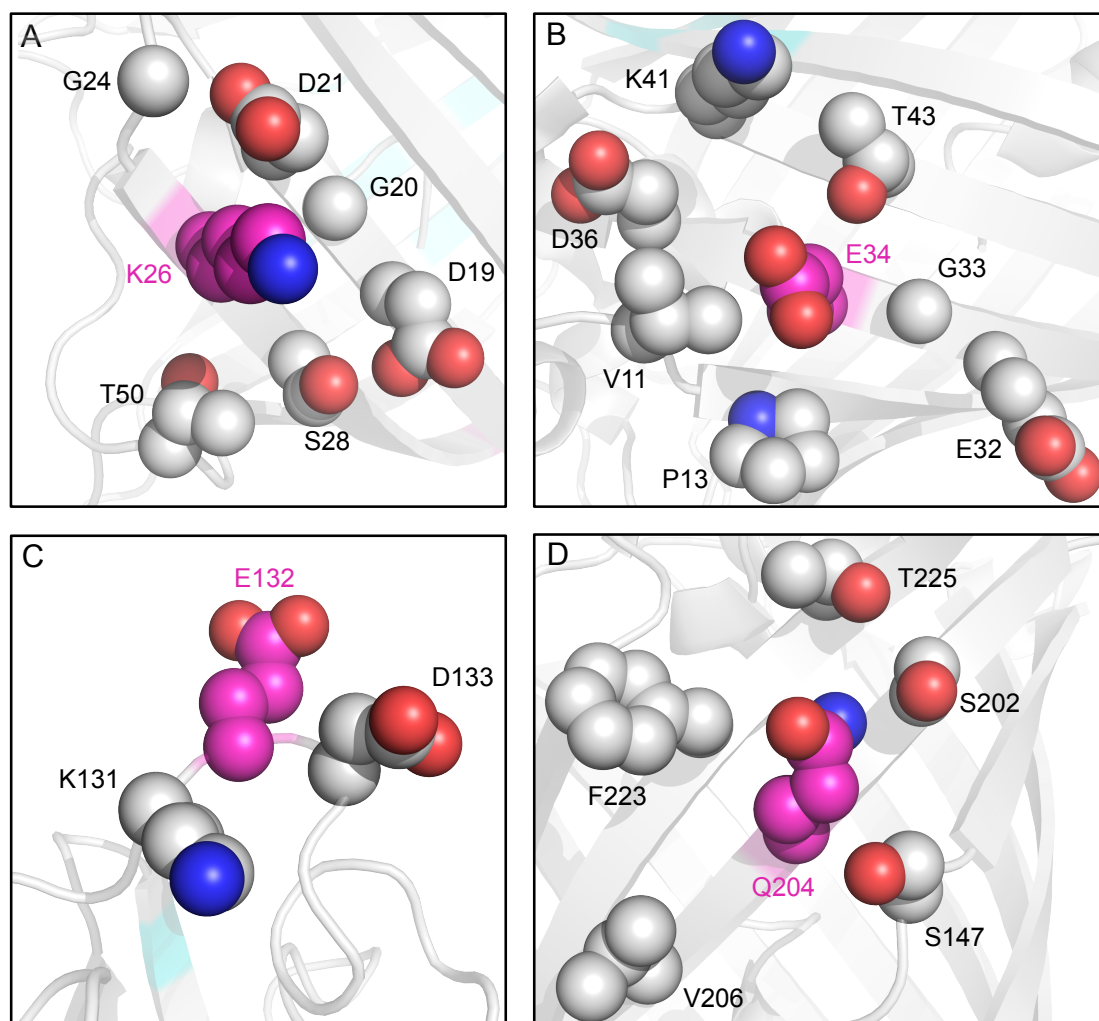


Figure 7.9. Local microenvironments of residues mutated to azPhe in sfGFP variants. Structure of sfGFP (PDB accession 2B3P) showing the environment of (A) Lys26, (B) Glu34, (C) Glu132 and (D) Gln204. Neighbouring residues are shown as spheres and coloured by element. The target residues are coloured by element with carbons coloured pink.

This examination of residue choice was repeated with cyt b_{562} to try and understand and draw a set of rules for SPAAC modification of proteins (Figure 7.6). As with sfGFP, reduced surface exposure and therefore residue accessibility correlated with improved reaction yields. The Asp5 and Lys104 mutations comprising the cyt b_{562} SA variant (Figure 7.6A) have considerably lower surface accessibilities (77 \AA^2 and 90 \AA^2 , respectively) compared to Asp21

and Asp50 residues alter in the *cyt b₅₆₂* LA mutants (119 Å² and 148 Å², respectively) but exhibited the highest reaction yields. Again, this suggests that the accessibility of a residue is not the main contributor to the extent of modification.

The turn and loop regions of *cyt b₅₆₂* that house Asp21 and Asp50 (Figure 7.10B and C), respectively, are more dynamic than the α -helices that house Asp5 and Lys104 (Figure 7.10A and D) as reflected by average main chain thermal B-factors of ~23 compared to 14, respectively. This suggests that less dynamic regions are modified more efficiently possibly because residues are held in a more defined orientation for conjugation with the incoming probe. This is supported by comparison to the sfGFP^{AzF} reactions as residues in β -strands (less dynamic) modified to a greater extent than residues in more flexible regions (Figure 7.2). However, *cyt b₅₆₂*^{AzF} variants in defined secondary structures were modified to lower extents than the equivalent in sfGFP^{AzF}. This may be explained by comparison of the surface microenvironments of the residues and neighbouring residues (Figure 7.10). The most efficiently modified variant, Lys104 (Figure 7.10D), does have a hydrophobic patch created by Ala100, Tyr101 and Tyr105 that seems to be important for modification (Figure 7.9B and D). However, Lys104 is also fully solvent exposed on one side (left in Figure 7.10D). Asp5 (Figure 7.10A) has the benefit of a defined secondary structure but does not have a hydrophobic patch (apart from the nearby, flexible Ala1) and is surrounded by mainly charged and some polar residues. The microenvironments of Asp21 (Figure 7.10B) and Asp50 (Figure 7.10C) show the more exposed, dynamic nature of these residues and with neighbouring charged and polar residues as well as the lack of any hydrophobic patches.

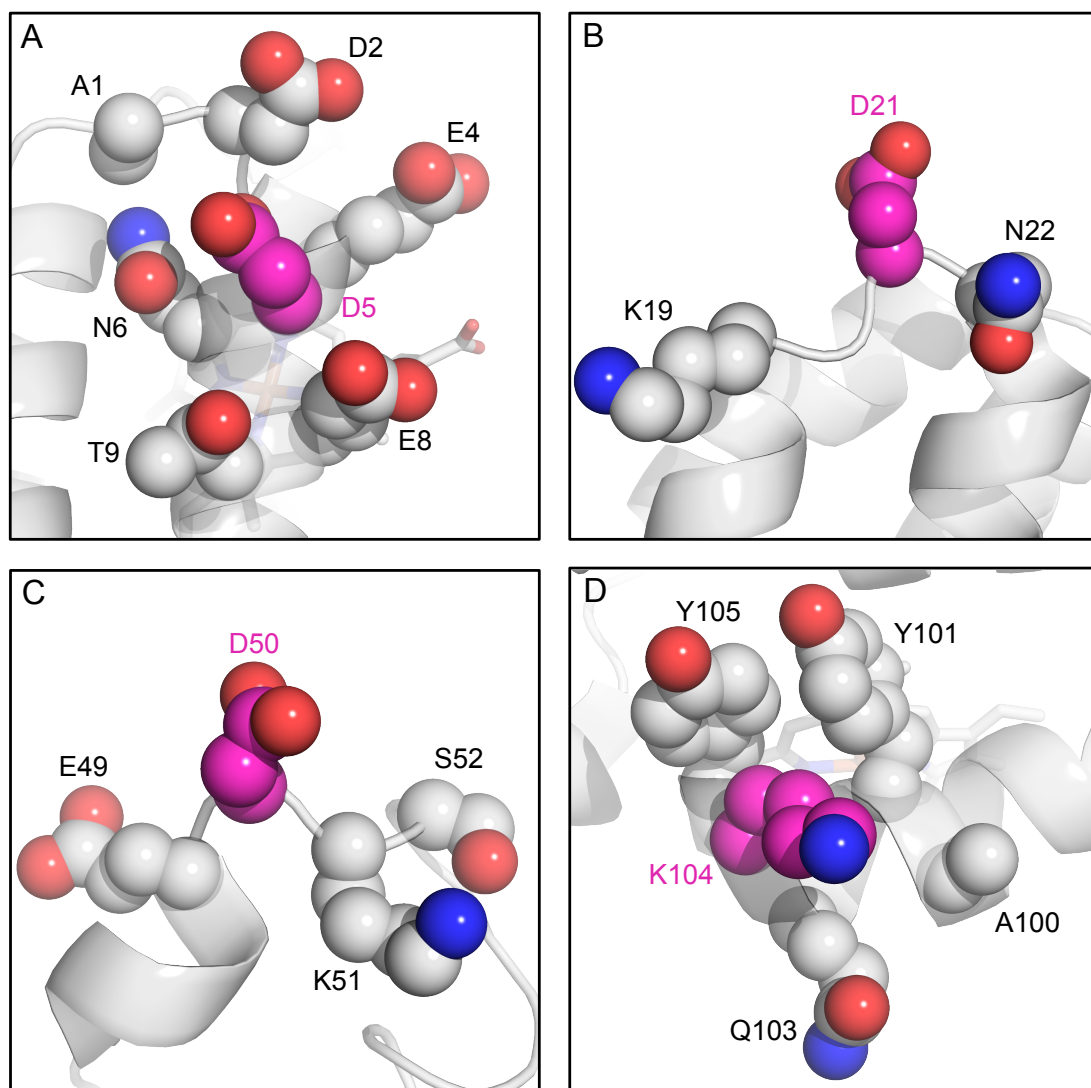


Figure 7.10. Local microenvironments of residues mutated to azPhe in *cyt b₅₆₂^{AzF}* variants. Structure of *cyt b₅₆₂* (PDB accession 1QPU) showing the environment of (A) Asp5, (B) Asp21, (C) Asp50 and (D) Lys104. Neighbouring residues are shown as spheres and coloured by element. The target residues are coloured by element with carbons coloured pink.

It is clear that the choice of residue is important in terms of efficiency of modification and the influence on function, which has implications for protein bioconjugation in general (248). The relationship between position and modification is complex with all of the facets discussed (accessibility, secondary structure, flexibility, surface charge) likely contributing to the overall efficiency. However, the counterintuitive results presented here for sfGFP and *cyt b₅₆₂* suggest that the local surface microenvironment of a protein and the structural element housing the residue may prove to be the most important considerations. The former would require a different criterion for residue choice for each new reaction studied, as the character of reactive handles would differ from azide/DBCO. The surface charge profile would need to

complement the incoming reactive group just as the hydrophobic patches surrounding Glu34 and Gln204 may attract DBCO (in SPAAC). For example, polar residues may attract polar reactants such as aminoxy groups for oxime ligations (249).

Given the lack of definitive rules for residue choice, optimising the placement of the azide handle using recently developed scanning mutagenesis approaches as in Chapter 4 (TriNEx; (171)) should be advantageous. This would facilitate discovery of the residues that modify to the greatest extent as well as the subsequently modified proteins with the most useful properties. In this vein, a random approach to residue choice would allow proteins to be optimised for various different processes through defined bioconjugation events such as PEGylation for improved therapeutic proteins. Whether by scanning mutagenesis approaches or rational selection, studying more residues and therefore environments help provide a set of general rules for the placement of azPhe.

7.3.3 Effect of azPhe position on the influence of sfGFP SPAAC modification

As discussed, protein modification is useful for changing the properties of proteins for studying or using the proteins in biotechnological applications. sfGFP was the main target of modification in this study and the property of sfGFP that was manipulated was its fluorescence emission wavelength. Modulating the emission of fluorescent proteins is useful for creating new variants that can be used as FRET partners with other colour variants and for different imaging applications (*eg.* red-shifted variants for in tissue imaging). By selecting an azide-reactive fluorescent probe with complementary spectral properties to sfGFP (DBCO-585) it was in theory possible to get energy transfer from sfGFP to the attached probe. One variant (sfGFP^{Q204AzF}) when modified with DBCO-585 exhibited a highly efficient energy transfer of 90 % (Figure 7.5). A transfer efficiency of 90 % is high compared to existing fluorescent protein based FRET systems (92, 250, 251), which was not possible with the other azPhe positions studied, highlighting the power of being able to select and define the site of modification. The fluorescent emission of the resulting conjugate was red-shifted by 100 nm (to 611 nm) making it effectively an orange-red fluorescent protein. This magnitude of shift in wavelength is not easily achieved through traditional protein engineering approaches such as mutation of residues of the GFP chromophore, which again shows the power of a technique for site-specific modification. Furthermore, the Stoke's shift of the modified protein (126 nm) is unprecedented (252) and was previously only possible with quantum dots or tandem dyes (253). One specific benefit of red-shifted fluorescent protein variants over standard GFPs is better tissue penetration of longer wavelengths of light for use with whole tissue imaging.

7.3.4 SPAAC reaction kinetics

Although not always critical for *in vitro* modification, reaction rate is important when modification needs to be on the biological time frame such as for monitoring the production of newly synthesised proteins (81, 82). The rates of reaction observed in this chapter are 10 to 50 fold higher than that observed previously for SPAAC reactions involving small molecules in organic solvents (80, 235). This is surprising considering the additional complexity involved in protein modification. A macromolecular crowding effect may be responsible where the physical presence of a large molecule can enhance a contact-dependant reaction (254). Such an effect is encouraged in DNA ligation reactions with the addition of PEG and may be the result of an apparent increase in the concentration of reactants (254). Alternatively, the rate enhancement observed might be an effect of the protein surface microenvironment presenting the azido group or promoting approach of DBCO-585. These rates are higher than those observed for norbornene cycloadditions (*vide infra*) and equivalent to CuAAC's (229), and should be sufficiently fast to study dynamics of live systems such as changes in glycan populations (69).

7.3.5 Alternative bio-orthogonal chemistries

Other bio-orthogonal chemistries exist that utilise Uaa's that have various advantages and disadvantages. One of the first Uaa's explored was *p*-acetyl-phenylalanine (acPhe) that contains a terminal carbonyl group (255). Carbonyl groups react selectively with aminoxy and hydrazine groups via oxime and hydrazone ligation, respectively (249, 256). acPhe is incorporated into protein with high yields however the ubiquitous nature of carbonyl groups and the harsh reaction conditions (pH 4) make oxime and hydrazone ligation unsuitable for biological uses. Recently, norbornene was incorporated into proteins in *E. coli* and mammalian cells in the form of N^ε-5-norbornene-2-ylloxycarbonyl-lysine (norLys) (257). Tetrazines selectively react with strained double bonds, such as in norbornene, with high rates of between 1 and 9 M⁻¹sec⁻¹ (257). However, due to the bulky nature of norLys proteins may be less tolerant to incorporation resulting in lower yields of soluble protein (< 4 mg/L culture). Alternatively to SPAAC and CuAAC, azides also have reactivity with phosphines via Staudinger ligation (73, 258-260). The reaction rate of Staudinger ligation is lower than CuAAC and phosphines also suffer from oxidation in air.

Overall azPhe conjugation via SPAAC is currently the best approach for protein modification. The superior reaction kinetics (Section 7.3.4) in combination with mild reaction conditions and no requirement for a catalyst make it suitable for *in vivo* modification on the biological timescale. SPAAC could be combined with other reactions (and Uaa's) to allow

defined protein modification at two or more sites or even proteins. An approach like this could be useful for studying proteins with multiple, different post-translational modifications (261) or even creating complex protein constructs.

7.3.6 Labelling graphene with pyrene-functionalised proteins

In this study a novel approach to label graphene (and potentially other carbon allotropes such as CNT's) with any target protein was demonstrated. The approach uses a genetically encoded azPhe residue that is covalently modified with a DBCO-pyrene linker to give highly defined surface labelling (Sections 7.2.3 and 7.2.4). Preliminary experiments demonstrated that HOPG could be decorated with single, separated protein molecules (Figure 7.8E). Further optimisation will fine-tune the protein concentrations required to give different protein coverage's for different applications. It should be possible to reduce the amount of protein required by increasing the incubation time or improving the protein-surface interaction with less-polar solvent mixtures (*eg.* 25 % DMSO in dH₂O). This will be useful for poorly expressed proteins and will reduce the quantity of DBCO-pyrene required, making the approach more efficient. As well as optimising the surface labelling of the pyrene-modified proteins, the purification of modified from unmodified protein following the SPAAC reaction (Section 7.2.3) will need to be investigated. The initial failed attempts at isolating modified protein could have been the result of unresolved elution peaks from ion exchange chromatography. If the modified and unmodified protein elutions overlapped the result would be a mixed population that resembles the original mix. To address this, ion exchange chromatography could be reattempted with a shallower elution gradient (of increasing NaCl concentration) or hydrophobic interaction chromatography could be used to exploit the attachment of the hydrophobic pyrene. Alternatively, modification of sfGFP^{Q204AzF} with DBCO-pyrene could have changed the extinction coefficient of either the sfGFP chromophore or the pyrene. As these values were used to calculate the extent modification an altered value would give an inaccurate interpretation. Ideally, when the surface labelling is optimised this step can be eliminated, as the affinity of the pyrene for graphene should act as the purification step.

Recently a number of approaches have been proposed such as non-specific labelling via aromatic residues (262), using pre-functionalized materials (*eg.* carboxylated-graphene via diazonium salts) (263) and even using pyrene via maleimide (85) and succinimidyl ester (264) linkages. Techniques for site-specific modification, such as with pyrene via SPAAC, are more desirable as they afford greater control over the orientation of protein binding to the surface. Biotechnological applications and single molecule analyses require control over orientation as

a heterogeneous attachment gives a mixture of protein orientations resulting in an averaging of any readings or effects. The disadvantage of using pre-functionalized materials is the decreased connection between the protein and material caused by long and sometimes undefined linkages. Also, excessive modification could affect the desirable properties of the material by, for example, disrupting the π -conjugated nature of graphene. By modifying proteins at defined positions with short, variable adapters (such as DBCO-pyrene) the protein-material link is intimate and can be tailored.

An alternative approach to covalently link proteins to graphene is currently being investigated using the photoreactivity of azPhe (Section 1.6.1 and Chapter 4) to crosslink via a cycloaddition reaction to graphene. Direct crosslinking would give an even more intimate linkage between the properties of the protein and graphene. Graphene (265) and CNTs (266) have been successfully photolabeled with simple phenyl azides but not directly with azPhe-containing proteins. It is yet to be seen whether breaking the delocalized, sp^2 -hybridized character of the graphene (via the addition reaction with singlet nitrene (Figure 1.6) would have a positive or negative effect (267). The delocalized electronic character is responsible for graphene's impressive electron mobility and conductance (268). Although initial thoughts might be that disrupting this character would be detrimental, attaching a protein via such a linkage might provide better communication with the protein.

"Natural" chemistries for protein functionalization such as a maleimide-cysteine linkage have the advantage of not requiring the Uaa incorporation machinery or genetic engineering. Conversely, proteins often contain multiple surface-exposed cysteines (or lysines) therefore removing uniformity and definition of orientation. If a protein contains multiple or no suitable cysteines then the genes must be engineered anyway. The additional benefit of using an Uaa with bio-orthogonal reactivity is that it can be used in combination with a natural (or different unnatural) chemistry. One specific use of this is analysis of conductance of a material-bound protein by Scanning Tunneling Microscopy (STM). For STM, contact must be made with the protein from both the surface (*eg.* graphene) and the STM probe (*eg.* gold tip). For this, azPhe can be used to attach the protein to the material (*eg.* via pyrene) leaving an engineered cysteine at the opposite side of the protein to contact a gold STM electrode via the strong noble metal-thiol interaction (as in Section 7.2.2 and Chapter 3; (119)).

The myriad uses of defined labelling of graphene make it an extremely promising and fast-moving area of research (86). One particular goal is studying intricate details of function and dynamics of a protein by measuring the conductance of the graphene surface, as for CNTs in Choi *et al.* (85). Here, enzyme activity (and therefore protein motion) resulted in changes in

current that were used to elucidate and calculate constants for seven different timescales that make-up lysozymes activity. As an expanded genetic code approach combined with bio-orthogonal Click chemistry is applicable to any protein this opens up massive possibilities for studying proteins.

8. Discussion

Uaa's provide a powerful approach to expand the functionality of the proteome as they confer new physical and chemical properties to proteins. In this thesis, Uaa's have been site-specifically incorporated into sfGFP and *cyt b₅₆₂* in high yields using an expanded genetic code approach (Chapter 3). Uaa's were well structurally and functionally tolerated and, for the most, introduced with high orthogonality (Chapter 3-7). Uaa analogs of tyrosine (cyPhe, tfmPhe, azPhe and napAla) with varying polarity and size were used to alter the fluorescent properties of sfGFP and could be used as a general way of altering protein function (Chapter 4). The majority of this work focused on azPhe as it has a number of properties that would be desirable in the context of proteins such as photoreactivity and selective reactivity with alkynes. AzPhe proved to be a useful tool for controlling protein activity using light and was used to create sfGFP variants capable of photo-activation, -deactivation and -switching (Chapter 4 and 5). The low temperature (77 K) photochemistry of phenyl nitrene was also studied in defined protein environments by genetically encoding azPhe, which provides a useful alternative to synthetic cages. Two radicals were caged and detected (anilino and triplet phenyl nitrene), with the radical observed dependant on the immediate protein environment (Chapter 6). Finally, the bio-orthogonal reactivity of azPhe was exploited to create proteins capable of site-specific modification via SPAAC (Chapter 7). The position of azPhe was shown to have a significant effect (<10 fold difference) on reaction yield and rate. Modification was used to install proteins with useful novel properties such as red-shifted fluorescence emission (sfGFP with DBCO-585 via FRET) and the ability to bind to non-biological materials like graphene (with attached pyrene via π - π stacking).

8.1 The advantages and disadvantages of genetically encoding Uaa's as a protein engineering technique

Numerous methods have been developed to incorporate Uaa's into proteins as discussed in Chapter 1 (Section 1.1). These include solid-phase peptide synthesis, auxotrophic doping with structural analogues, post-translational modification and the use of chemically aminoacylated tRNAs, all of which are useful but have drawbacks associated with them. By far the most useful approach to incorporate Uaa's is via an expanded genetic code where a codon is reprogrammed to code for the Uaa. This provides greater control over protein production by allowing placement of the Uaa within a target protein to be genetically defined. Additionally, there is a useful level of control of protein production involved with providing (or not) the required incorporation machinery and Uaa to the cells. This co-translational

approach also has the benefits of recombinant protein expression whereby bacteria (or other organisms) autonomously produce the Uaa-containing proteins in high yields alleviating the need for complex synthesis or external interference. Genetic coding means that both rational design and newly emerging directed evolution (172, 269, 270) protein engineering approaches can be applied (as demonstrated in Chapter 4). Additionally, because the proteins are produced *in situ* it allows their study and manipulation within the cellular context.

There are currently disadvantages to genetically encoding Uaa's, many of which are methodological aspects as a result of it being a relatively new discipline. Noticeable is the poor fidelity of some of the aaRS's used to bind and load tRNA's with Uaa's. Particularly apparent in this work was amPheRS (to incorporate *p*-amino-phenylalanine) that clearly incorporated a natural amino acid (most likely tyrosine) in the absence of amPhe (Chapter 4). This is a problem because the result is mixed population of a target protein containing the Uaa or a natural amino acid, and therefore loss of control over incorporation. Additionally some of the aaRS's used are poor enzymes with low K_m and k_{cat} parameters and k_{cat}/K_m ratios 10^3 times lower compared to the wild-type aaRS (147). The lower incorporation efficiency of Uaa's compared to natural amino acids creates the requirement for large excesses of the Uaa (and sometimes aaRS). Both of these aaRS issues could be remedied through a more stringent protein evolution strategy. The current approach involves randomisation of 6 active site residues that make direct contact with the amino acid R-group. Directed evolution of the entire protein would allow the aaRS library to sample the important effects of second (or third) shell mutations or more dramatic changes in protein architecture that could result in changes to the amino acid binding region. For example, change in secondary structure caused by a distal mutation might open or shift the amino acid binding site and/or even change the side chains that interact with the Uaa. This strategy would require a large library with multiple rounds of evolution (and could benefit from recombination of variants), which would be a long and arduous process but should produce more efficient aaRS^{Uaa}'s. Additionally, the design of new and improved aaRS^{Uaa}'s could benefit from recent advances in computational design (271-273). The potential for using computational design has been shown by demonstrating the fidelity of a range of aaRS's (274-276). Computational design of aaRS's mainly focuses on the amino acid binding site and the optimisation of interactions between the protein and the substrate, by a number of different methods. In one study Zhang and colleagues (277) used a clash opportunity progressive (COP) method that focuses on a few key residues and first minimises clashes with the substrate then finds mutations that improve binding. These mutations were combined and used to find an aaRS for *o*-methyl-tyrosine that was very similar to the one evolved (7, 26).

Amber stop codon suppression has been used to produce high levels of proteins containing Uaa's (Chapter 3), however, as with recoding any codon (including rare, degenerate amino

acid-coding codons) there are negative effects associated with ‘hijacking’ the function of that codon. An alternative is to use a unique codon such as a quadruplet codon however this requires additional machinery and currently gives lower protein production yields. One exciting alternative is Multiplex Automated Genome Engineering (MAGE), which is capable of creating a blank codon (to code for an Uaa for example) by replacing every one of a chosen codon in a genome with another.

8.2 AzPhe as an addition to the protein engineering toolbox

AzPhe offers novel ways to control protein function post-translationally. The two approaches used here exploit the photoreactivity of azPhe (Chapter 4) and the selective reactivity of azPhe (Chapter 7) to modify protein function with light and by attachment of new chemical moieties, respectively.

The photocontrollable sfGFP variants (sfGFP^{L44AzF}, sfGFP^{Y66AzF}, sfGFP^{F145AzF}, sfGFP^{H148AzF} and sfGFP^{T203AzF}) generated in Chapter 4 open up the possibility to use azPhe as a small photo-trigger for use in optogenetics. AzPhe offers several advantages over current optogenetic techniques such as alleviating the need for whole protein domains to elicit a photo-response. Only a limited number of these photosensitive proteins exist, which could limit the range of responses sampled by a target protein and additionally some rely on cofactors so limiting general applicability. As discussed, a reprogrammed genetic code means that azPhe could be incorporated into potentially any protein making it generally applicable. Uaa optogenetics is also expandable with the use of other photoreactive Uaa’s such as benzoyl-phenylalanine (photocrosslinker), trifluoromethyl-diazirin-phenylalanine (carbene precursor), nitrobenzyl-cysteine (photocaged natural amino acid) and azo-benzene-phenylalanine (photo-isomerizable).

Conversely, the energy requirement for azPhe photolysis (classically <310 nm) is not ideal for use with biological samples. The increase in wavelength required (405 – 485 nm) for the sfGFP^{AzF} variants generated here may be a GFP-specific effect as a result of the unique photochemistry of the chromophore. Although lower wavelengths (and hence higher energies) of light are more damaging this should not be a limiting factor. Although highly reactive, the nitrene produced on photolysis of azPhe may cause a relatively small change in terms of a large protein. Nevertheless, put in the right place such as in or near an enzyme active site (eg. sfGFP^{F145AzF}), on an important dynamic region associated with protein function, at a residue distal from the active site (eg. sfGFP^{L44AzF}) or at a protein-protein interface even a relatively ‘small’ modification, or moreover a crosslinking event, can have a significant effect (278). Also, simply the loss of molecular nitrogen from azPhe to form the reactive nitrene could be enough to elicit a change in function if in a key area. Although it is a specific example, the

active site Ser221Cys mutation in subtilisin (effectively only an O → S substitution) results in a massive 10^5 drop in protease activity (279, 280). The importance of positioning is again shown by the Val112Ala and Leu250Ser mutations in metallo- β -lactamase; both residues are distal from the active site but result in increased activity (281). An example from this thesis comes from comparison of sfGFP^{F145AzF} and sfGFP^{H148AzF}. Residue 145 and 148 are ~adjacent in sfGFP and both face the chromophore tyrosyl moiety at a distance of ~2.7 Å however irradiation of these variants (*ie.* photolysis of azPhe) gives two very different effects on activity (deactivation *versus* excitation switching) from two different photochemical pathways (Chapter 4 and 5). Therefore, to maximise the chances of finding useful residues it is important to not just focus on active site residues, although this is often a good place to start considering the generally more extreme effects (278). Directed evolution approaches (such as TriNEx as in Chapter 3 and 4) allow less obvious, but equally important, residues to be sampled. Also, more will be learned about the photochemical mechanism (and preferences) of azPhe when more residues, and therefore protein environments, have been targeted for azPhe optogenetics. This will aid in the design of future rational mutations.

Using light to alter protein activity (photocontrol) is powerful because it allows precise control in both space and time and is minimally invasive. Photocontrol of proteins could be useful for studying important cellular processes such as enzyme cascades, cell signalling networks or as an alternative to conditional knockout/knock-in mutations. For example by creating a photo-activatable or –deactivatable receptor a signalling pathway could be selectively switched on by irradiation resulting in a downstream effect such as cell differentiation (187, 188). An approach like this could therefore afford total user control over cellular function. Taken together, azPhe has massive potential and is of broad applicability as an optogenic probe, especially as Uaa's can now be readily incorporated into mammalian cells and whole animals.

The second use of azPhe is the ability to post-translationally attach new chemical groups to proteins with Click chemistry (such as SPAAC; Chapter 7). One sfGFP variant generated here (sfGFP^{Q204AzF}) exhibited a large 100 nm red-shift in fluorescence emission after attachment of the red-fluorescent dye, DBCO-585 (Figure 7.5). This shift was protein context-dependant as it was a result of FRET from sfGFP to the attached dye. This post-translational Click modification approach holds more general promise beyond this specific example, with many diverse applications available. It offers a way of tagging any protein with any desired chemical group whether it is a fluorescent probe, as above, for tracking, or a molecule such as biotin for easy and specific purification. A large hydrophobic group could be attached to change the structure and stability or even membrane-permeability of a protein, or molecules could be added that block enzyme active sites or trap dynamic protein regions in specific

conformations. One biotechnological application is the attachment of long-chain PEG to therapeutic proteins to improve their pharmacokinetic properties (65, 225). It would even be possible to combine the two approaches for post-translationally controlling protein function via azPhe (*vide supra*). This would involve attaching a photo- (or otherwise)-controllable chemical moiety to a target protein by Click chemistry. The initial attachment could alter protein activity (as above) then the attached moiety could be switched to cause a secondary response and therefore extra level of control.

8.3 Using SPAAC-based Click chemistry for protein modification

Many reactions fall under the philosophy of Click chemistry each with benefits and drawbacks. One of the most recently developed reactions was strain-promoted azide-alkyne cycloaddition (SPAAC) as used in Chapter 7. SPAAC offers many advantages over other Click reactions such as fast kinetics, no dependence on toxic catalysts (*versus* CuAAC) and no issues with air oxidation (*versus* Staudinger ligation). One negative of SPAAC is the large and somewhat hydrophobic nature of the reactive group, DBCO. Although no problems were encountered herein, attaching a hydrophobic group could cause proteins to destabilise and denature or precipitate out of solution. Moreover the large DBCO group may not be able to enter (or be transported into) cells limiting the use of SPAAC in living cells without external manipulation.

The use of Uaa's in general for selective post-translational modification (PTM) of proteins is valuable. The major benefit is specificity because a chemical group that is not normally found in nature is incorporated into the protein (as an Uaa) that will only react with its partner group (azide with alkyne here) that is also not found naturally. This means the reaction is target specific and, in combination with a reprogrammed genetic code, position specific. The alternative and current paradigm is to modify a natural amino acid, often cysteine, lysine or the N-terminus of a protein. Proteins often contain multiple of these residues and, moreover, complex samples such as cells or their lysates will contain the modifiable groups (-NH₃, -SH) across the whole proteome and as part of other classes of molecules such as DNA and lipids. The result is lack of control over the target and number of PTM's, which leads to a mixed population of modified proteins that is impossible to resolve. Despite this, in cases where PTM is performed on pure protein and when position is not important, modification of natural amino acids is the best choice as there is no need for extra engineering (introduction of an amber stop codon, incorporation machinery, *etc*). Another drawback of SPAAC (and CuAAC to a lesser extent) is the smaller number and range of commercially available reagents compared with reagents for natural amino acids (*eg*.

maleimide, NHS ester). Although there are currently less available, the number is rapidly increasing along with the popularity of SPAAC.

8.4 Interfacing proteins with non-biological materials

The potential uses of PTM of proteins are extensive with examples from bio-therapeutics to detection of newly synthesised proteins (228), as discussed above in Section 8.2 and in Chapter 7. One goal that is particularly interesting as explored in Chapter 7 is the attachment of proteins to non-biological materials like graphene, CNTs and gold. Non-biological materials can be used to study proteins at the single molecule level, which provides a greater level of detail compared to a study in bulk solution where the observations are averaged and hence intricate molecular events are lost. There is also potential to modify the properties of the material by attaching proteins with specific functions thereby creating useful hybrid materials (*vide infra*). There are many potential non-biological materials that have useful properties, with graphene being perhaps the most exciting recent advance. Graphene has extraordinary electronic and thermal conductivity, high mechanical strength and is lightweight. These properties make graphene research a highly active area with the initial experiments towards isolating and characterising graphene earning the Nobel prize in Physics in 2010 (241). Graphene would be useful for myriad physical and also some biological applications from biomolecular sensors (282-284) to tissue engineering (285, 286), that require defined functionalization with biological molecules (264). The strategy here to attach proteins to graphene was to use the strong, non-covalent π -stacking interaction between pyrene and graphene. SPAAC is an excellent and novel way to attach pyrene groups to proteins because of the positional control afforded by combining a selectively reactive Uaa (azPhe) with a reprogrammed genetic code approach (as discussed in Section 8.3). The result is a protein of interest with an attached pyrene group at a genetically defined position via a variable, user-controlled linker, that can be used bind to graphene. The high conductivity of graphene means that structural and/or functional changes in an attached protein are linked through to changes in conductance of the graphene surface. This approach can be used to study function and dynamics of a protein at an extremely precise, single molecule level, as for CNTs in Choi *et al.* (85). Here, enzyme activity (and therefore protein motion) resulted in changes in current that were used to elucidate and calculate constants for seven different timescales that make-up lysozymes activity.

Although a (very) long-term goal, proteins could be used as biological components in computers or other devices due to unique functions like electron transport (cytochromes). Cyt b_{562} has been attached to gold and shown to exhibit very high and tuneable conductance through the protein (118-120), which could be expanded for use with graphene and other proteins. Another future application of graphene is the construction of complex cellular

networks for tissue engineering (264). Graphene is ideal because it allows electrical as well as chemical signaling in combination with micro-patterning techniques to form the networks (via pyrene labeled cell-surface proteins) (285, 286).

Proteins are ideal for use in devices as they are nanoscale and work at the single molecule level, they encode all the information within their sequence to self-assemble into complex functional structures. Also, proteins exist with a wide variety of functions that have exquisite molecular recognition properties and are “programmable” through protein engineering. For example, cyt *b*₅₆₂ measures approximately 2.5×4.8 nm (*versus* ~ 45 nm for a modern transistor), binds haem in a defined and specific manner and can be programmed to act as a voltage-gated conducting bridge between metallic electrodes. The drawback of using proteins in non-biological environments is they are naturally evolved to work in (often) complex solutions whereas traditional devices usually work in air. There is a level of fragility associated with proteins with the possibility of denaturation at non-optimal conditions and also particular working conditions (temperature, pH, *etc*) that could limit their use. Additionally, the shorter life span of an organic molecule like a protein, due to denaturation or proteolysis, compared to a silicon device could be a shortcoming. Despite the drawbacks, the huge potential of combining the useful properties of proteins with those of novel materials via Uaa approaches could form the template for the next generation of materials.

8.5 Future work

The work demonstrating photocontrol of sfGFP (Chapter 4) can potentially be expanded to proteins in general and is an exciting future development. Firstly, finishing some of the work started here by elucidating the photochemical endpoint and therefore pathway sampled by sfGFP^{H148AzF} would help build on our knowledge of azPhe photoreactions for future engineering. X-ray crystal structures of the dark and irradiated states of sfGFP^{H148AzF} are currently being pursued, with preliminary crystal structures being solved as this thesis is submitted. It would be interesting to expand the photocontrollable FPs generated in Chapter 4 by azPhe incorporation to other existing FP scaffolds. For example TagRFP has the conserved tyrosine in the chromophore and equivalents to the neighbouring Phe145 and His148 (287). TagRFP has the benefit of being red-shifted in relation to sfGFP, which is a useful property of FPs for whole animal imaging and for use as FRET partners. Additionally it will be important to expand the technique to non-FPs. A genetically encoded azPhe has many benefits as an optogenetic probe and huge potential (Section 8.2). It would be interesting to try controlling an enzyme with light and work has been started with β -lactamase and protein kinase A. Also, (potentially) more difficult and rewarding target examples including membrane channels and

protein complexes. As with examples in Section 8.2, having complete spatial and temporal control over important proteins like these could shape the future of optogenetics.

The results towards photo-caging reactive chemical intermediates of azPhe in Chapter 6 are preliminary but prove that proteins have potential as a new type of cage. It would be interesting and would help understand the photochemistry to try a more diverse range of micro-environments (*ie.* residues) in sfGFP and T4 lysozyme as well as other proteins. Particularly interesting is the combination of azPhe incorporation with traditional amino acid substitutions to tailor the protein matrix surrounding the radical. For example, using the lysozyme^{F153AzF} variant the hydrophobic cavity could be manipulated to introduce different types of hydrogen/proton sources (*eg.* aprotic *versus* uncharged *versus* basic/acidic H-bonders) and compare the products generated. Additionally, other reactive Uaa's can be site-specifically incorporated into proteins thereby opening up this approach as a general way of studying and trapping reactive organic intermediates. Although limited to small peptides, solid-phase peptide synthesis could also widen the range of Uaa's that can be introduced and therefore studied.

The research group will continue to perfect the deposition of proteins onto graphene in terms of buffers used, protein concentrations and how the graphene is washed to remove non-specifically bound molecules. The protein concentration that is best for labelling may need to be adjusted for each new protein (with varying molecular weight, shape and hydrophobicity) or general rules may emerge. An alternative approach that is similar to the one used here is to first deposit the bi-functional DBCO-pyrene on graphene, then perform the SPAAC reaction on the graphene surface. Another option that is possible with azPhe is the direct, covalent attachment to graphene using photochemistry (Section 8.2). The nitrene generated on photolysis can react with double bonds such as in sp²-hybridised graphene and has been used to functionalise graphene with small phenyl azides such as perfluoro-phenylazide (265). This approach has not been reported with whole proteins but offers a way of creating an intimate linkage between the protein and graphene. It is yet to be seen whether this linkage may be too close and whether the disruption of conjugated bonding network (by the addition reaction of azPhe) would be a positive or negative in comparison to the non-perturbative π -stacking approach (*vide supra*).

Once the technique of attaching proteins to graphene in defined orientations is developed the work will focus less on sfGFP and more on cyt *b*₅₆₂ and other proteins that are useful for studying and developing devices, such as protein kinases. Longer-term goals are extensive (as discussed above in Section 8.4) and include creating sensing devices (*eg.* glucose sensor (264)), single molecule analysis (85) and even bio-molecular electronics (118, 120).

8.6 Summary and achievements

Incorporating Uaa's via an expanded genetic code is a relatively new approach to protein engineering but has rapidly become an established technique for modulating the properties of proteins. It offers a way of incorporating Uaa's into proteins *in vivo* at genetically defined positions, which is a useful way of extending the physical and chemical functionality of proteins beyond what is possible in nature. This thesis focussed mainly on the Uaa *p*-azido-phenylalanine (azPhe) because of the useful properties of the amino acid that include photoreactivity and selective reactivity by Click chemistry. Despite this Chapter 4 used a number of other Uaa's with unique polarity and structure to alter the spectral properties of sfGFP. This was more of a "classic" approach to protein engineering that allies with mutagenesis with natural amino acids. It resulted in mostly subtle, but some more extreme, changes in sfGFP properties and led to the generation of variants with red-shifted fluorescence and altered excitation spectra by sampling residues proximal to and distal from the chromophore (Chapter 4).

Also, a novel approach was established to photocontrol proteins that exploits the photoreactivity of a single azPhe incorporated into a key residue to modulate the properties of the whole protein. This azPhe optogenetic approach generated sfGFP variants capable of activation, deactivation and switching (with relation to fluorescence excitation) on application of low energy light, both *in vitro* and in cells (Chapter 4, Section 4.2.2). These, and similar variants should be useful as molecular highlighters for the emerging techniques of super-resolution microscopy (*eg.* PALM) and as photobleachers in FRAP-like experiments. The photochemical pathway sampled by azPhe that caused these changes in protein activity was elucidated by fluorescence, absorption and mass spectroscopy and X-ray crystallography (Chapter 5). Each variant underwent a different pathway on photolysis with the local protein environment playing an important role. This technique can be applied to any protein as well as in mammalian cells and whole animals so could be expanded to generate photocontrollable enzymes, enzyme cascades, signalling pathways, transcription factors, differentiation, cell division and even whole cells.

In a similar approach, the low temperature (77 K) photochemistry of azPhe was studied in unique protein environments with the intension of caging particular reactive intermediates and controlling the reaction pathway sampled. This study generated and detected a triplet nitrene and the somewhat elusive anilino radical, which were rationalised by analysing the unique protein matrix (Chapter 6, (194)). This was the first report of using genetically encoded photoreactive Uaa as a precursor to cage a reactive organic intermediate and opens up new possibilities for studying photochemistry.

Another property of azPhe is selective reactivity with alkyne groups. Although some natural amino acids have specific reactivities (cysteine, lysine, *etc*), it is useful to have

reactivity that is not found naturally in proteins, or even cells as a whole (*ie.* bio-orthogonal). sfGFP and cyt *b*₅₆₂ containing azPhe were modified with a red-fluorescent dye (DBCO-585) and a probe containing pyrene (DBCO-pyrene) by SPAAC, a type of Click chemistry. By analysing multiple surface-exposed residues it was possible to compare reactivities and interestingly, residue choice had a huge effect on efficiency of modification (Chapter 7, (139)). Surface accessibility had an inverse relationship with reaction efficiency, which contradicts the standard approach of selecting the most accessibility and flexible residue. It seemed that the protein micro-environment had a significant effect on reaction efficiency with residues in defined secondary structures and complementary surface chemistry to the incoming probe (*ie.* uncharged, hydrophobic) modifying to the greatest extent. The newfound knowledge of the best residues for modification was applied to attach DBCO-pyrene to sfGFP and cyt *b*₅₆₂. Pyrene-modified labelled proteins were then attached to the graphene substitute, HOPG, via π -stacking, a technique that can potentially be used for graphene, graphite, CNT's and fullerenes (Chapter 7). This work opens up many future avenues of investigation (Section 8.5) that hold great promise for studying proteins and creating new and useful variants and even graphene-based devices. Uaa's are an excellent addition to the proteome and offer new ways of modulating protein structure and function not normally possible in nature.

8.7 Publications associated with this thesis

Review

Genetically encoding phenyl azide chemistry: news uses and ideas for classical biochemistry.

Reddington SC, Watson P, Rizkallah P, Tippmann EM and Jones DD.

2013. *Biochem Soc T.* 41 (5), 1177-1182

Chapters 4 and 5

Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence.

Reddington SC, Rizkallah P, Watson P, Pearson R, Tippmann EM and Jones DD.

2013. *Angew Chem Int Edit.* 52 (23), 6090-6093

Chapter 6

Aryl azide photochemistry in defined protein environments.

Reddington SC, Morris JL, Murphy DM, Jones DD, Platts JA and Tippmann EM.

2013. *Org Lett.* 15 (4), 728-31

Chapter 7

Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry.

Reddington SC, Tippmann EM and Jones DD.

2012. *Chem Commun.* 48 (67), 8419-21

9. References

1. Berg JM, Tymoczko JL, Stryer L, Stryer L. Biochemistry. 5th edit. WH Freeman and Company New York; 2002.
2. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F. Selenocysteine: The 21st amino acid. *Mol Microbiol.* 1991. **5** (3) 515-20.
3. Srinivasan G, James C, Krzycki J. Pyrrolysine encoded by UAG in archaea: Charging of a uag-decoding specialized tRNA. *Science.* 2002. **296** (5572) 1459-62.
4. Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, Huber R. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur J Biochem.* 1995. **230** (2) 788-96.
5. Bain J, Diala E, Glabe C, Dix T, Chamberlin A. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *J Am Chem Soc.* 1989. **111** (20) 8013-4.
6. Nowak M, Kearney P, Sampson J, Saks M, Labarca C, Silverman S, Zhong W, Thorson J, Abelson J, Davidson N. Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science.* 1995. **268** (5209) 439-42.
7. Wang L, Brock A, Herberich B, Schultz P. Expanding the genetic code of *Escherichia coli*. *Science.* 2001. **292** (5516) 498-500.
8. Neumann H. Rewiring translation - genetic code expansion and its applications. *FEBS Lett.* 2012. **586** (15) 2057-64.
9. Liu CC, Schultz PG. Adding new chemistries to the genetic code. *Annu Rev Biochem.* 2010. **79** 413-44.
10. Wang L, Schultz P. A general approach for the generation of orthogonal tRNAs. *Chem Biol.* 2001. **8** (9) 883-90.
11. Chin J, Santoro S, Martin A, King D, Wang L, Schultz P. Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J Am Chem Soc.* 2002. **124** (31) 9026-7.
12. Chin J, Cropp T, Anderson J, Mukherji M, Zhang Z, Schultz P. An expanded eukaryotic genetic code. *Science.* 2003. **301** (5635) 964-7.
13. Wu N, Deiters A, Cropp T, King D, Schultz P. A genetically encoded photocaged amino acid. *J Am Chem Soc.* 2004. **126** (44) 14306-7.
14. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, Soma A, Kobayashi T, Kitabatake M, Takio K, Saito K. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* 2002. **30** (21) 4692-9.
15. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. Expanding the genetic code of *Drosophila melanogaster*. *Nat Chem Biol.* 2012. **8** (9) 748-50.
16. Greiss S, Chin JW. Expanding the genetic code of an animal. *J Am Chem Soc.* 2011. **133** (36) 14196-9.
17. Crick F. Central dogma of molecular biology. *Nature.* 1970. **227** (5258) 561-3.
18. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. The complete genome sequence of *Escherichia coli* K-12. *Science.* 1997. **277** (5331) 1453-62.
19. Benzer S, Champe S. A change from nonsense to sense in the genetic code. *P Natl Acad Sci USA.* 1962. **48** (7) 1114-21.
20. Zhang Z, Alfonta L, Tian F, Bursulaya B, Uryu S, King D, Schultz P. Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells. *P Natl Acad Sci USA.* 2004. **101** (24) 8882-7.
21. Hohsaka T, Sato K, Sisido M, Takai K, Yokoyama S. Site-specific incorporation of photofunctional nonnatural amino-acids into a polypeptide through *in-vitro* protein-biosynthesis. *FEBS Lett.* 1994. **344** (2-3) 171-4.
22. Anderson J, Magliery T, Schultz P. Exploring the limits of codon and anticodon size. *Chem Biol.* 2002. **9** (2) 237-44.

23. Hohsaka T, Ashizuka Y, Taira H, Murakami H, Sisido M. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an *Escherichia coli in vitro* translation system. *Biochemistry*. 2001. **40** (37) 11060-4.
24. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin J. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*. 2010. 441-4.
25. Mehl R, Anderson J, Santoro S, Wang L, Martin A, King D, Horn D, Schultz P. Generation of a bacterium with a 21 amino acid genetic code. *J Am Chem Soc*. 2003. **125** (4) 935-9.
26. Wang L, Magliery T, Liu D, Schultz P. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the *in vivo* incorporation of unnatural amino acids into proteins. *J Am Chem Soc*. 2000. **122** (20) 5010-1.
27. Fechter P, Rudinger-Thirion J, Tukalo M, Giegè R. Major tyrosine identity determinants in methanococcus jannaschii and saccharomyces cerevisiae trnatyr are conserved but expressed differently. *Eur J Biochem*. 2001. **268** (3) 761-7.
28. Steer B, Schimmel P. Major anticodon-binding region missing from an archaeobacterial tRNA synthetase. *J Biol Chem*. 1999. **274** (50) 35601-6.
29. Jakubowski H, Goldman E. Editing of errors in selection of amino acids for protein synthesis. *Microbiol Mol Biol R*. 1992. **56** (3) 412-29.
30. Kobayashi T, Nureki O, Ishitani R, Yaremchuk A, Tukalo M, Cusack S, Sakamoto K, Yokoyama S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat Struct Mol Biol*. 2003. **10** (6) 425-32.
31. Zhang Y, Wang L, Schultz P, Wilson I. Crystal structures of apo wild-type *M. jannaschii* tyrosyl-tRNA synthetase (tyrRS) and an engineered tyrRS specific for *o*-methyl-L-tyrosine. *Protein Sci*. 2005. **14** (5) 1340-9.
32. Xie JM, Schultz PG. Innovation: A chemical toolkit for proteins - an expanded genetic code. *Nat Rev Mol Cell Bio*. 2006. **7** (10) 775-82.
33. Zhang Z, Smith B, Wang L, Brock A, Cho C, Schultz P. A new strategy for the site-specific modification of proteins *in vivo*. *Biochemistry*. 2003. **42** (22) 6735-46.
34. Nguyen DP, Alai MMG, Kapadnis PB, Neumann H, Chin JW. Genetically encoding *n*-epsilon-methyl-L-lysine in recombinant histones. *J Am Chem Soc*. 2009. **131** (40) 14194-5.
35. Summerer D, Chen S, Wu N, Deiters A, Chin J, Schultz P. A genetically encoded fluorescent amino acid. *P Natl Acad Sci USA*. 2006. **103** (26) 9785-9.
36. Xie J, Wang L, Wu N, Brock A, Spraggon G, Schultz P. The site-specific incorporation of *p*-iodo-L-phenylalanine into proteins for structure determination. *Nat Biotechnol*. 2004. **22** (10) 1297-301.
37. Mori H, Ito K. Different modes of secy-seca interactions revealed by site-directed *in vivo* photo-cross-linking. *P Natl Acad Sci USA*. 2006. **103** (44) 16159-64.
38. Xie JM, Liu WS, Schultz PG. A genetically encoded bidentate, metal-binding amino acid. *Angew Chem Int Edit*. 2007. **46** (48) 9239-42.
39. Wuttke DS, Gray HB, Fisher SL, Imperiali B. Semisynthesis of bipyridyl alanine cytochrome-c mutants - novel proteins with enhanced electron-transfer properties. *J Am Chem Soc*. 1993. **115** (18) 8455-6.
40. Bertho A. The decay of phenyl azide in benzoyl and in *p*-xylo. *Ber Dtsch Chem Ges*. 1924. **57** 1138-42.
41. Fleet GWJ, Porter RR, Knowles JR. Affinity labelling of antibodies with aryl nitrene as reactive group. *Nature*. 1969. **224** (5218) 511-2.
42. Chen Y, Ebright Y, Ebright R. Identification of the target of a transcription activator protein by protein-protein photocrosslinking. *Science*. 1994. **265** (5168) 90-2.
43. Schuster G, Platz M. Photochemistry of phenyl azide. *Adv Photochem*. 1992. **17** 69-143.
44. Keana JFW, Cai SX. New reagents for photoaffinity-labeling - synthesis and photolysis of functionalized perfluorophenyl azides. *J Org Chem*. 1990. **55** (11) 3640-7.
45. Voskresenska V, Wilson RM, Panov M, Tarnovsky AN, Krause JA, Vyas S, Winter AH, Hadad CM. Photoaffinity labeling via nitrenium ion chemistry: Protonation of the nitrene

- derived from 4-amino-3-nitrophenyl azide to afford reactive nitrenium ion pairs. *J Am Chem Soc.* 2009. **131** (32) 11535-47.
46. Gritsan N, Platz M. Kinetics, spectroscopy, and computational chemistry of arylnitrenes. *Chem Rev.* 2006. **106** (9) 3844-67.
 47. Schrock AK, Schuster GB. Photochemistry of phenyl azide - chemical-properties of the transient intermediates. *J Am Chem Soc.* 1984. **106** (18) 5228-34.
 48. Mishra A, Rice SN, Lwowski W. Singlet and triplet nitrenes. 3. Addition of carbethoxynitrene to 1,3-dienes. *J Org Chem.* 1968. **33** (2) 481-6.
 49. Wang J, Burdzinski G, Zhu ZD, Platz MS, Carra C, Bally T. Ultrafast spectroscopic and matrix isolation studies of *p*-biphenyl, *o*-biphenyl, and 1-naphthyl nitrenium cations. *J Am Chem Soc.* 2007. **129** (26) 8380-8.
 50. Schnapp KA, Platz MS. A laser flash-photolysis study of difluorinated, trifluorinated and tetrafluorinated phenyl nitrenes - implications for photoaffinity-labeling. *Bioconjugate Chem.* 1993. **4** (2) 178-83.
 51. Rizk MS, Shi XF, Platz MS. Lifetimes and reactivities of some 1,2-didehydroazepines commonly used in photoaffinity labeling experiments in aqueous solutions. *Biochemistry.* 2006. **45** (2) 543-51.
 52. Smolinsky G, Wasserman E, Yager WA. EPR of ground state triplet nitrenes. *J Am Chem Soc.* 1962. **84** (16) 3220-1.
 53. Leyva E, Platz MS, Persy G, Wirz J. Photochemistry of phenyl azide - the role of singlet and triplet phenyl nitrene as transient intermediates. *J Am Chem Soc.* 1986. **108** (13) 3783-90.
 54. Shields CJ, Chrisope DR, Schuster GB, Dixon AJ, Poliakoff M, Turner JJ. Photochemistry of aryl azides - detection and characterization of a dehydroazepine by time-resolved infrared-spectroscopy and flash-photolysis at room-temperature. *J Am Chem Soc.* 1987. **109** (15) 4723-6.
 55. Bucher G, Tonshoff C, Nicolaidis A. Photochemistry of an azido-functionalized cryptand: Controlling the reactivity of an extremely long-lived singlet aryl nitrene by complexation to alkali cations. *J Am Chem Soc.* 2005. **127** (18) 6883-92.
 56. Tokitoh N, Saiki T, Okazaki R. Abnormal reaction of an aryl azide confined in a calix[6]arene skeleton. *J Chem Soc Chem Comm.* 1995. **18** 1899-900.
 57. Wagner G, Arion VB, Brecker L, Krantz C, Mieusset JL, Brinker UH. Controllable selective functionalization of a cavitand via solid state photolysis of an encapsulated phenyl azide. *Org Lett.* 2009. **11** (14) 3056-8.
 58. Warmuth R, Makowiec S. The phenyl nitrene rearrangement in the inner phase of a hemiacarcerand. *J Am Chem Soc.* 2005. **127** (4) 1084-5.
 59. Warmuth R, Makowiec S. Photochemical and thermal reactions of intermediates in the phenyl nitrene rearrangement inside a hemiacarcerand. *J Am Chem Soc.* 2007. **129** (5) 1233-41.
 60. Kolb H, Finn M, Sharpless K. Click chemistry: Diverse chemical function from a few good reactions. *Angew Chem Int Edit.* 2001. **40** (11) 2004-21.
 61. Huisgen R. Kinetics and mechanism of 1, 3-dipolar cycloadditions. *Angew Chem Int Edit.* 1963. **2** (11) 633-45.
 62. Tornøe C, Christensen C, Meldal M. Peptidotriazoles on solid phase:[1, 2, 3]-triazoles by regioselective copper (i)-catalyzed 1, 3-dipolar cycloadditions of terminal alkynes to azides. *J Org Chem.* 2002. **67** (9) 3057-64.
 63. Rostovtsev V, Green L, Fokin V, Sharpless K. A stepwise huisgen cycloaddition process: Copper (i)-catalyzed regioselective ligation of azides and terminal alkynes. *Angew Chem Int Edit.* 2002. **114** (14) 2708-11.
 64. Dirks A, Berkel S, Hatzakis N, Opsteen J, Delft F, Cornelissen J, Rowan A, Hest J, Rutjes F, Nolte R. Preparation of biohybrid amphiphiles via the copper catalysed huisgen [3+2] dipolar cycloaddition reaction. *Chem Commun.* 2005. **2005** (33) 4172-4.
 65. Deiters A, Cropp T, Summerer D, Mukherji M, Schultz P. Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorgan Med Chem.* 2004. **14** (23) 5743-5.
 66. Link A, Tirrell D. Cell surface labeling of *Escherichia coli* via copper (i)-catalyzed [3+2] cycloaddition. *J Am Chem Soc.* 2003. **125** (37) 11164-5.

67. Phelps KJ, Morris AA, Beal PA. Novel modifications in RNA. *ACS Chem Biol.* 2012. **7** (1) 100-9.
68. Neef AB, Schultz C. Selective fluorescence labeling of lipids in living cells. *Angew Chem Int Edit.* 2009. **48** (8) 1498-500.
69. Laughlin ST, Baskin JM, Amacher SL, Bertozzi CR. *In vivo* imaging of membrane-associated glycans in developing zebrafish. *Science.* 2008. **320** (5876) 664-7.
70. Agard N, Prescher J, Bertozzi C. A strain-promoted [3+ 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc.* 2004. **126** (46) 15046-7.
71. Kennedy DC, McKay CS, Legault MCB, Danielson DC, Blake JA, Pegoraro AF, Stolow A, Mester Z, Pezacki JP. Cellular consequences of copper complexes used to catalyze bioorthogonal click reactions. *J Am Chem Soc.* 2011. **133** (44) 17993-8001.
72. Prescher J, Dube D, Bertozzi C. Chemical remodelling of cell surfaces in living animals. *Nature.* 2004. **430** (7002) 873-7.
73. Staudinger H, Meyer J. On new organic phosphorus bonding iii phosphine methylene derivatives and phosphinimine. *Helv Chim Acta.* 1919. **2** 635-46.
74. Ess DH, Houk K. Distortion/interaction energy control of 1, 3-dipolar cycloaddition reactivity. *J Am Chem Soc.* 2007. **129** (35) 10646-7.
75. Ess DH, Jones GO, Houk K. Transition states of strain-promoted metal-free click chemistry: 1, 3-dipolar cycloadditions of phenyl azide and cyclooctynes. *Org Lett.* 2008. **10** (8) 1633-6.
76. Campbell - Verduyn LS, Mirfeizi L, Schoonen AK, Dierckx RA, Elsinga PH, Feringa BL. Strain - promoted copper - free "click" chemistry for ¹⁸F radiolabeling of bombesin. *Angew Chem Int Edit.* 2011. **50** (47) 11117-20.
77. Baskin J, Prescher J, Laughlin S, Agard N, Chang P, Miller I, Lo A, Codelli J, Bertozzi C. Copper-free click chemistry for dynamic *in vivo* imaging. *P Natl Acad Sci USA.* 2007. **104** (43) 16793-7.
78. Ning X, Guo J, Wolfert MA, Boons GJ. Visualizing metabolically labeled glycoconjugates of living cells by copper free and fast Huisgen cycloadditions. *Angew Chem Int Edit.* 2008. **120** (12) 2285-7.
79. Sletten EM, Bertozzi CR. A hydrophilic azacyclooctyne for Cu-free Click chemistry. *Org Lett.* 2008. **10** (14) 3097-9.
80. Debets MF, Van Berkel SS, Schoffelen S, Rutjes FPJT, van Hest JCM, Van Delft FL. Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition. *Chem Commun.* 2009. **46** (1) 97-9.
81. Dieterich DC, Link AJ, Graumann J, Tirrell DA, Schuman EM. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *P Natl Acad Sci USA.* 2006. **103** (25) 9482-7.
82. Dieterich DC, Lee JJ, Link AJ, Graumann J, Tirrell DA, Schuman EM. Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat Protoc.* 2007. **2** (3) 532-40.
83. Harris J, Martin N, Modi M. PEGylation: A novel process for modifying pharmacokinetics. *Clin Pharmacokinet.* 2001. **40** (7) 539-51.
84. Veronese F, Pasut G. PEGylation, successful approach to drug delivery. *Drug Discov Today.* 2005. **10** (21) 1451-8.
85. Choi YK, Moody IS, Sims PC, Hunt SR, Corso BL, Perez I, Weiss GA, Collins PG. Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science.* 2012. **335** (6066) 319-24.
86. Algar WR, Prasuhn DE, Stewart MH, Jennings TL, Blanco-Canosa JB, Dawson PE, Medintz IL. The controlled display of biomolecules on nanoparticles: A challenge suited to bioorthogonal chemistry. *Bioconjugate Chem.* 2011. **22** (5) 825-58.
87. Akerman ME, Chan WCW, Laakkonen P, Bhatia SN, Ruoslahti E. Nanocrystal targeting *in vivo*. *P Natl Acad Sci USA.* 2002. **99** (20) 12617-21.

88. Ye S, Huber T, Vogel R, Sakmar T. FTIR analysis of GPCR activation using azido probes. *Nat Chem Biol*. 2009. **5** (6) 397-9.
89. Shimomura O, Johnson FH, Saiga Y. Extraction, purification and properties of aequorin, a bioluminescent protein from luminous hydromedusan, aequorea. *J Cell Compar Physl*. 1962. **59** (3) 223-39.
90. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene*. 1992. **111** (2) 229-33.
91. Nobelprize.org. The nobel prize in chemistry 2008. [cited 2012 5 Nov]; Available from: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2008/.
92. Tsien R. The green fluorescent protein. *Annu Rev Biochem*. 1998. **67** (1) 509-44.
93. Sniegowski JA, Lappe JW, Patel HN, Huffman HA, Wachter RM. Base catalysis of chromophore formation in Arg(96) and Glu(222) variants of green fluorescent protein. *J Biol Chem*. 2005. **280** (28) 26248-55.
94. Johnson FH, Gershman LC, Waters JR, Reynolds GT, Saiga Y, Shimomura O. Quantum efficiency of *Cypridina* luminescence, with a note on that of *Aequorea*. *J Cell Compar Physl*. 1962. **60** (1) 85-103.
95. Chattoraj M, King BA, Bublitz GU, Boxer SG. Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *P Natl Acad Sci USA*. 1996. **93** (16) 8362-7.
96. Cubitt AB, Woollenweber LA, Heim R. Understanding structure-function relationships in the *Aequorea victoria* green fluorescent protein. *Method Cell Biol*. 1999. **58** 19-30.
97. Wachter R, Elsliger M, Kallio K, Hanson G, Remington S. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure*. 1998. **6** (10) 1267-77.
98. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science*. 1996. **273** (5280) 1392-5.
99. Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M, Remington SJ. Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *P Natl Acad Sci USA*. 1997. **94** (6) 2306-11.
100. Arpino JAJ, Rizkallah PJ, Jones DD. Crystal structure of enhanced green fluorescent protein to 1.35 angstrom resolution reveals alternative conformations for Glu222. *PLoS One*. 2012. **7** (10).
101. Shinobu A, Palm GJ, Schierbeek AJ, Agmon N. Visualizing proton antenna in a high-resolution green fluorescent protein structure. *J Am Chem Soc*. 2010. **132** (32) 11093-102.
102. Heim R, Cubitt A, Tsien R. Improved green fluorescence. *Nature*. 1995. **373** (6516) 663.
103. Cormack B, Valdivia R, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. 1996. **173** (1) 33-8.
104. Pedelacq J, Cabantous S, Tran T, Terwilliger T, Waldo G. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol*. 2005. **24** (1) 79-88.
105. Springs SL, Bass SE, McLendon GL. Cytochrome *b*₅₆₂ variants: A library for examining redox potential evolution. *Biochemistry*. 2000. **39** (20) 6075-82.
106. Fujita T, Sato R. Soluble cytochromes in *Escherichia coli*. *Biochim Biophys Acta*. 1963. **77** (4) 690.
107. Itagaki E, Palmer G, Hager LP. Studies on cytochrome *b*₅₆₂ of *Escherichia coli* .2. Reconstitution of cytochrome *b*₅₆₂ from apoprotein and hemin. *J Biol Chem*. 1967. **242** (9) 2272-7.
108. Arnesano F, Banci L, Bertini I, Faraone-Mennella J, Rosato A, Barker PD, Fersht AR. The solution structure of oxidized *Escherichia coli* cytochrome *b*(562). *Biochemistry*. 1999. **38** (27) 8657-70.
109. Hamada K, Bethge PH, Mathews FS. Refined structure of cytochrome *b*(562) from *Escherichia-coli* at 1.4 angstrom resolution. *J Mol Biol*. 1995. **247** (5) 947-62.
110. Feng YQ, Sligar SG, Wand AJ. Solution structure of apocytochrome *b*(562). *Nat Struct Biol*. 1994. **1** (1) 30-5.

111. Nikkila H, Gennis RB, Sligar SG. Cloning and expression of the gene encoding the soluble cytochrome b562 of escherichia coli. *Eur J Biochem.* 1991. **202** (2) 309-13.
112. Springs SL, Bass SE, McLendon GL. Cytochrome *b*(562) variants: A library for examining redox potential evolution. *Biochemistry.* 2000. **39** (20) 6075-82.
113. Della Pia EA, Chi QJ, Elliott M, Macdonald JE, Ulstrup J, Jones DD. Redox tuning of cytochrome *b*(562) through facile metal porphyrin substitution. *Chem Commun.* 2012. **48** (86) 10624-6.
114. Edwards WR, Busse K, Allemann RK, Jones DD. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Res.* 2008. **36** (13) 78-87.
115. Jones DD, Barker PD. Controlling self-assembly by linking protein folding, DNA binding, and the redox chemistry of heme. *Angew Chem Int Edit.* 2005. **44** (39) 6337-41.
116. Zuo P, Albrecht T, Barker PD, Murgida DH, Hildebrandt P. Interfacial redox processes of cytochrome *b*(562). *Phys Chem Chem Phys.* 2009. **11** (34) 7430-6.
117. Della Pia EA, Chi QJ, Jones DD, Macdonald JE, Ulstrup J, Elliott M. Single-molecule mapping of long-range electron transport for a cytochrome *b*(562) variant. *Nano Lett.* 2011. **11** (1) 176-82.
118. Della Pia EA, Chi Q, Macdonald JE, Ulstrup J, Jones DD, Elliott M. Fast electron transfer through a single molecule natively structured redox protein. *Nanoscale.* 2012. **4** (22) 7106-13.
119. Della Pia EA, Elliott M, Jones DD, Macdonald JE. Orientation-dependent electron transport in a single redox protein. *ACS Nano.* 2012. **6** (1) 355-61.
120. Della Pia EA, Macdonald JE, Elliott M, Jones DD. Direct binding of a redox protein for single-molecule electron transfer measurements. *Small.* 2012. **8** (15) 2341-4.
121. Morris JL, Reddington SC, Murphy DM, Jones DD, Platts JA, Tippmann EM. Aryl azide photochemistry in defined protein environments. *Org Lett.* 2013. **15** (4) 728-31.
122. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Express Purif.* 2005. **41** (1) 207-34.
123. Laemmli UK. Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature.* 1970. **227** (5259) 680-5.
124. Winter G. XIA2: An expert system for macromolecular crystallography data reduction. *J Appl Crystallogr.* 2010. **43** 186-90.
125. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D.* 2006. **62** 72-82.
126. Bailey S. The CCP4 suite - programs for protein crystallography. *Acta Crystallogr D.* 1994. **50** 760-3.
127. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007. **40** 658-74.
128. Emsley P, Cowtan K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr D.* 2004. **60** (Pt 12 Pt 1) 2126-32.
129. Stoll S, Schweiger A. Easyspin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson.* 2006. **178** (1) 42-55.
130. Becke AD. Density-functional thermochemistry. 3. The role of exact exchange. *J Chem Phys.* 1993. **98** (7) 5648-52.
131. Ditchfie.R, Hehre WJ, Pople JA. Self-consistent molecular-orbital methods. 9. Extended gaussian-type basis for molecular-orbital studies of organic molecules. *J Chem Phys.* 1971. **54** (2) 724-9.
132. Frisch MJ, Pople JA, Binkley JS. Self-consistent molecular-orbital methods. 25. Supplementary functions for gaussian-basis sets. *J Chem Phys.* 1984. **80** (7) 3265-9.
133. Barone V, Adamo C. Proton transfer in the ground and lowest excited states of malonaldehyde: A comparative density functional and post-hartree-fock study. *J Chem Phys.* 1996. **105** (24) 11007-19.
134. Neese F. The ORCA program system. *Wires Comput Mol Sci.* 2012. **2** (1) 73-8.
135. Anslyn EV, Dougherty DA. Modern physical organic chemistry: University Science Books; 2006.

136. Hammill JT, Miyake-Stoner S, Hazen JL, Jackson JC, Mehl RA. Preparation of site-specifically labeled fluorinated proteins for ^{19}F -NMR structural characterization. *Nat Protoc.* 2007. **2** (10) 2601-7.
137. Schleif R. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.* 2000. **16** (12) 559-65.
138. Miyake-Stoner SJ, Refakis CA, Hammill JT, Lusic H, Hazen JL, Deiters A, Mehl RA. Generating permissive site-specific unnatural aminoacyl-tRNA synthetases. *Biochemistry.* 2010. **49** (8) 1667-77.
139. Reddington SC, Tippmann EM, Jones DD. Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted click chemistry. *Chem Commun.* 2012. **48** (67) 8419-21.
140. Wang K, Neumann H, Peak-Chew SY, Chin JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol.* 2007. **25** (7) 770-7.
141. Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. An expanded genetic code with a functional quadruplet codon. *P Natl Acad Sci USA.* 2004. **101** (20) 7566-71.
142. Brown CM, Tate WP. Direct recognition of messenger-RNA stop signals by *Escherichia coli* polypeptide-chain release factor-2. *J Biol Chem.* 1994. **269** (52) 33164-70.
143. Johnson DBF, Wang C, Xu JF, Schultz MD, Schmitz RJ, Ecker JR, Wang L. Release factor one is nonessential in *Escherichia coli*. *ACS Chem Biol.* 2012. **7** (8) 1337-44.
144. Eriani G, Delarue M, Poch O, Gangloff J, Moras D. Partition of transfer-RNA synthetases into 2 classes based on mutually exclusive sets of sequence motifs. *Nature.* 1990. **347** (6289) 203-6.
145. Young TS, Ahmad I, Yin JA, Schultz PG. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J Mol Biol.* 2010. **395** (2) 361-74.
146. Loscha KV, Herlt AJ, Qi RH, Huber T, Ozawa K, Otting G. Multiple-site labeling of proteins with unnatural amino acids. *Angew Chem Int Edit.* 2012. **51** (9) 2243-6.
147. Kiga D, Sakamoto K, Kodama K, Kigawa T, Matsuda T, Yabuki T, Shirouzu M, Harada Y, Nakayama H, Takio K, Hasegawa Y, Endo Y, Hirao I, Yokoyama S. An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *P Natl Acad Sci USA.* 2002. **99** (15) 9715-20.
148. Hoesl MG, Budisa N. *In vivo* incorporation of multiple noncanonical amino acids into proteins. *Angew Chem Int Edit.* 2011. **50** (13) 2896-902.
149. Wan W, Huang Y, Wang Z, Russell WK, Pai PJ, Russell DH, Liu WR. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew Chem Int Edit.* 2010. **49** (18) 3211-4.
150. Rydén S, Isaksson L. A temperature-sensitive mutant of *Escherichia coli* that shows enhanced misreading of UAG/A and increased efficiency for tRNA nonsense suppressors. *Mol Gen Genet.* 1984. **193** (1) 38-45.
151. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, Kraal L, Tolonen AC, Gianoulis TA, Goodman DB, Reppas NB. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. *Science.* 2011. **333** (6040) 348-53.
152. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. Programming cells by multiplex genome engineering and accelerated evolution. *Nature.* 2009. **460** (7257) 894-8.
153. Taskent-Sezgin H, Marek P, Thomas R, Goldberg D, Chung J, Carrico I, Raleigh DP. Modulation of *p*-cyanophenylalanine fluorescence by amino acid side chains and rational design of fluorescence probes of alpha-helix formation. *Biochemistry.* 2010. **49** (29) 6290-5.
154. Tucker MJ, Oyola R, Gai F. A novel fluorescent probe for protein binding and folding studies: *p*-cyano-phenylalanine. *Biopolymers.* 2006. **83** (6) 571-6.
155. Rogers JMG, Lippert LG, Gai F. Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications. *Anal Biochem.* 2010. **399** (2) 182-9.

156. O'connor MJ, Boblak KN, Topinka MJ, Kindelin PJ, Briski JM, Zheng C, Klumpp DA. Superelectrophiles and the effects of trifluoromethyl substituents. *J Am Chem Soc.* 2010. **132** (10) 3266-7.
157. Yoder NC, Kumar K. Fluorinated amino acids in protein design and engineering. *Chemical Soc Rev.* 2002. **31** (6) 335-41.
158. Wang L, Brock A, Schultz PG. Adding l-3-(2-naphthyl)alanine to the genetic code of *E. coli*. *J Am Chem Soc.* 2002. **124** (9) 1836-7.
159. Deisseroth K. Optogenetics. *Nat Methods.* 2011. **8** (1) 26-9.
160. Toettcher JE, Voigt CA, Weiner OD, Lim WA. The promise of optogenetics in cell biology: Interrogating molecular circuits in space and time. *Nat Methods.* 2011. **8** (1) 35-8.
161. Baratta MV NS, Dobelis P, Pomrenze MB, Dolzani SD & Cooper DC. Optogenetic control of genetically-targeted pyramidal neuron activity in prefrontal cortex. *Nat Precedings.* 2012. (2 April).
162. Witten IB, Lin SC, Brodsky M, Prakash R, Diester I, Anikeeva P, Gradinaru V, Ramakrishnan C, Deisseroth K. Cholinergic interneurons control local circuit activity and cocaine conditioning. *Science.* 2010. **330** (6011) 1677-81.
163. Hegemann P, Moglich A. Channelrhodopsin engineering and exploration of new optogenetic tools. *Nat Methods.* 2011. **8** (1) 39-42.
164. Lippincott-Schwartz J, Patterson GH. Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging. *Trends Cell Biol.* 2009. **19** (11) 555-65.
165. Patterson GH, Lippincott-Schwartz J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science.* 2002. **297** (5588) 1873-7.
166. Chudakov DM, Verkhusha VV, Staroverov DB, Souslova EA, Lukyanov S, Lukyanov KA. Photoswitchable cyan fluorescent protein for protein tracking. *Nat Biotechnol.* 2004. **22** (11) 1435-9.
167. Arnold FH. Combinatorial and computational challenges for biocatalyst design. *Nature.* 2001. **409** (6817) 253-7.
168. Dalby PA. Optimising enzyme function by directed evolution. *Curr Opin Struct Biol.* 2003. **13** (4) 500-5.
169. Yuen CM, Liu DR. Dissecting protein structure and function using directed evolution. *Nat Methods.* 2007. **4** (12) 995-7.
170. Kim D, Rhee Y, Rhodes D, Sharma V, Sorenson O, Greener A, Smider V. Directed evolution and identification of control regions of colE1 plasmid replication origins using only nucleotide deletions. *J Mol Biol.* 2005. **351** (4) 763-75.
171. Baldwin A, Arpino J, Edwards W, Tippmann E, Jones D. Expanded chemical diversity sampling through whole protein evolution. *Mol Biosyst.* 2009. **5** (7) 764-6.
172. Baldwin A, Busse K, Simm A, Jones D. Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx). *Nucleic Acids Res.* 2008. **36** (13) 77-86.
173. Zacharias DA, Tsien RY. Molecular biology and mutation of green fluorescent protein. *Method Biochem Anal.* 2006. **47** 83-120.
174. Kadokura H, Katzen F, Beckwith J. Protein disulfide bond formation in prokaryotes. *Annu Rev Biochem.* 2003. **72** 111-35.
175. Ostergaard H, Henriksen A, Hansen FG, Winther JR. Shedding light on disulfide bond formation: Engineering a redox switch in green fluorescent protein. *EMBO J.* 2001. **20** (21) 5853-62.
176. Sivakumar K, Xie F, Cash BM, Long S, Barnhill HN, Wang Q. A fluorogenic 1,3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Org Lett.* 2004. **6** (24) 4603-6.
177. Hansch C, Leo A, Taft RW. A survey of hammett substituent constants and resonance and field parameters. *Chem Rev.* 1991. **91** (2) 165-95.
178. Thompson MA, Biteen JS, Lord SJ, Conley NR, Moerner WE. Molecules and methods for super-resolution imaging. *Methods in Enzymology, Vol 475: Single Molecule Tools, Pt B.* 2010. **475** 27-59.

179. Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *P Natl Acad Sci USA*. 2002. **99** (20) 12651-6.
180. Andresen M, Stiel AC, Trowitzsch S, Weber G, Eggeling C, Wahl MC, Hell SW, Jakobs S. Structural basis for reversible photoswitching in Dronpa. *P Natl Acad Sci USA*. 2007. **104** (32) 13005-9.
181. Habuchi S, Ando R, Dedecker P, Verheijen W, Mizuno H, Miyawaki A, Hofkens J. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *P Natl Acad Sci USA*. 2005. **102** (27) 9511-6.
182. Shroff H, Galbraith CG, Galbraith JA, White H, Gillette J, Olenych S, Davidson MW, Betzig E. Dual-color superresolution imaging of genetically expressed probes within individual adhesion complexes. *P Natl Acad Sci USA*. 2007. **104** (51) 20308-13.
183. Renz M, Daniels BR, Vamosi G, Arias IM, Lippincott-Schwartz J. Plasticity of the asialoglycoprotein receptor deciphered by ensemble fret imaging and single-molecule counting PALM imaging. *P Natl Acad Sci USA*. 2012. **109** (44) E2989-E97.
184. Subach FV, Patterson GH, Renz M, Lippincott-Schwartz J, Verkhusha VV. Bright monomeric photoactivatable red fluorescent protein for two-color super-resolution PALM of live cells. *J Am Chem Soc*. 2010. **132** (18) 6481-91.
185. Lippincott-Schwartz J, Altan-Bonnet N, Patterson GH. Photobleaching and photoactivation: Following protein dynamics in living cells. *Nat Cell Biol*. 2003. S7-S14.
186. Patterson G. Photobleaching and photoactivation of fluorescent proteins for studies in cell biology. *Microsc Microanal*. 2007. **13** (S02) 294-5.
187. Gautier A, Deiters A, Chin JW. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J Am Chem Soc*. 2011. **133** (7) 2124-7.
188. Hahn ME, Muir TW. Photocontrol of SMAD2, a multiphosphorylated cell-signaling protein, through caging of activating phosphoserines. *Angew Chem Int Edit*. 2004. **43** (43) 5800-3.
189. Umanah G, Huang LY, Schultz PG, Naider F, Becker JM. Incorporation of the unnatural amino acid *p*-benzoyl-l-phenylalanine (Bpa) into a G protein-coupled receptor in its native context. *Adv Exp Med Biol*. 2009. **611** 333-5.
190. Tippmann EM, Liu W, Summerer D, Mack AV, Schultz PG. A genetically encoded diazirine photocrosslinker in escherichia coli. *Chembiochem*. 2007. **8** (18) 2210-4.
191. Bose M, Groff D, Xie JM, Brustad E, Schultz PG. The incorporation of a photoisomerizable amino acid into proteins in *E. coli*. *J Am Chem Soc*. 2006. **128** (2) 388-9.
192. Wang L, Xie J, Deniz A, Schultz P. Unnatural amino acid mutagenesis of green fluorescent protein. *J Org Chem*. 2003. **68** (1) 174-6.
193. Wang J, Burdzinski G, Platz MS. Solvent effects on intermolecular proton transfer: The rates of nitrene protonation and their correlation with Swain acidity. *Org Lett*. 2007. **9** (25) 5211-4.
194. Reddington SC MJ, Murphy DM, Jones DD, Platts JA and Tippmann EM. Aryl azide photochemistry in defined protein environments. *Org Lett*. 2013. **15** (4) 728-31.
195. Albini A, Bettinetti G, Minoli G. Intermolecular and intramolecular reactions of nitrenes and their cyclic isomers in the photodecomposition of some substituted 2-azidophenazines. *J Org Chem*. 1987. **52** (7) 1245-51.
196. Bogdanov AM, Mishin AS, Yampolsky IV, Belousov VV, Chudakov DM, Subach FV, Verkhusha VV, Lukyanov S, Lukyanov KA. Green fluorescent proteins are light-induced electron donors. *Nat Chem Biol*. 2009. **5** (7) 459-61.
197. Epifanovsky E, Polyakov I, Grigorenko B, Nemukhin A, Krylov AI. The effect of oxidation on the electronic structure of the green fluorescent protein chromophore. *J Chem Phys*. 2010. **132** (11) 115104-13.
198. Albini A, Bettinetti G, Minoli G. Photodecomposition of some para-substituted 2-pyrazolylphenyl azides. Substituents affect the phenylnitrene s-t gap more than the barrier to ring expansion. *J Am Chem Soc*. 1999. **121** (13) 3104-13.
199. Leyva E, Young MJT, Platz MS. High yields of formal CH insertion products in the reactions of polyfluorinated aromatic nitrenes. *J Am Chem Soc*. 1986. **108** (26) 8307-9.

200. Murata S, Nakatsuji R, Tomioka H. Mechanistic studies of pyrene-sensitized decomposition of *p*-butylphenyl azide - generation of nitrene radical-anion through a sensitizer-mediated electron-transfer from amines to the azide. *J Chem Soc Perk T 2*. 1995. **2** (4) 793-9.
201. Torimura M, Kurata S, Yamada K, Yokomaku T, Kamagata Y, Kanagawa T, Kurane R. Fluorescence-quenching phenomenon by photoinduced electron transfer between a fluorescent dye and a nucleotide base. *Anal Sci*. 2001. **17** (1) 155-60.
202. Fischer AB, Bronsteinbonte I. Photoinduced electron-transfer quenching of rhodamine-b in polymer-films. *J Photochem*. 1985. **30** (4) 475-85.
203. Niino H, Koga Y, Yabe A. Surface reaction of organic materials by laser ablation of matrix-isolated photoreactive aromatic azido compound. *J Photoch Photobio A*. 1997. **106** (1-3) 9-13.
204. Gritsan NP, Tigelaar D, Platz MS. A laser flash photolysis study of some simple para-substituted derivatives of singlet phenyl nitrene. *J Phys Chem A*. 1999. **103** (23) 4465-9.
205. Kanakarajan K, Goodrich R, Young MJT, Soundararajan S, Platz MS. Electron-paramagnetic-res spectroscopy of triplet aryl nitrenes covalently bound to alpha-chymotrypsin - application of low-temperature methods to photoaffinity-labeling. *J Am Chem Soc*. 1988. **110** (19) 6536-41.
206. Doering WVE, Odum RA. Ring enlargement in photolysis of phenyl azide. *Tetrahedron*. 1966. **22** (1) 81-93.
207. Splitter JS, Calvin M. Irradiation of 3-substituted-2-phenyloxaziridines - direct evidence for phenylnitrene. *Tetrahedron Lett*. 1968. (12) 1445-8.
208. Waddell WH, Feilchenfeld NB. Photoinitiated chain decomposition of phenyl isocyanate via its reaction with phenylnitrene. *J Am Chem Soc*. 1983. **105** (16) 5499-500.
209. Sundberg RJ, Suter SR, Brenner M. Photolysis of ortho-substituted aryl azides in diethylamine - formation and autoxidation of 2-diethylamino-1h-azepine intermediates. *J Am Chem Soc*. 1972. **94** (2) 513-20.
210. Hayes JC, Sheridan RS. Infrared-spectrum of triplet phenylnitrene - on the origin of didehydroazepine in low-temperature matrices. *J Am Chem Soc*. 1990. **112** (15) 5879-81.
211. Li YZ, Kirby JP, George MW, Poliakoff M, Schuster GB. 1,2-didehydroazepines from the photolysis of substituted aryl azides - analysis of their chemical and physical-properties by time-resolved spectroscopic methods. *J Am Chem Soc*. 1988. **110** (24) 8092-8.
212. Gritsan NP, Zhu ZD, Hadad CM, Platz MS. Laser flash photolysis and computational study of singlet phenylnitrene. *J Am Chem Soc*. 1999. **121** (6) 1202-7.
213. Donnelly T, Dunkin IR, Norwood DSD, Prentice A, Shields CJ, Thomson PCP. Didehydroazepines from the photolysis of phenyl azide and 3-substituted and 4-substituted phenyl azides isolated in low-temperature matrices. *J Chem Soc Perk T 2*. 1985. **2** 307-10.
214. Borden WT, Gritsan NP, Hadad CM, Karney WL, Kemnitz CR, Platz MS. The interplay of theory and experiment in the study of phenylnitrene. *Accounts Chem Res*. 2000. **33** (11) 765-71.
215. Sayin U, Turkkan E, Dereli O, Yuksel H, Birey M. EPR study of gamma-irradiated single crystal 4-phenylsemicarbazide. *Radiat Phys Chem*. 2010. **79** (8) 863-9.
216. Chapman OL, Leroux JP. 1-aza-1,2,4,6-cycloheptatetraene. *J Am Chem Soc*. 1978. **100** (1) 282-5.
217. Albin A, Bettinetti G, Minoli G. Reactivity of singlet and triplet aryl nitrenes: Temperature-dependent photodecomposition of 1-(2-azidophenyl)-3,5-dimethylpyrazole. *J Am Chem Soc*. 1997. **119** (31) 7308-15.
218. Platz MS. Comparison of phenylcarbene and phenylnitrene. *Accounts Chemical Res*. 1995. **28** (12) 487-92.
219. Sankaranarayanan J, Rajam S, Hadad CM, Gudmundsdottir AD. The ability of triplet nitrenes to abstract hydrogen atoms. *J Phys Org Chem*. 2010. **23** (4) 370-5.
220. Warriar M, Lo MKF, Monbouquette H, Garcia-Garibay MA. Photocatalytic reduction of aromatic azides to amines using CdS and CdSe nanoparticles. *Photoch Photobio Sci*. 2004. **3** (9) 859-63.

221. Liang TY, Schuster GB. Photochemistry of para-nitrophenyl azide - single-electron-transfer reaction of the triplet nitrene. *J Am Chem Soc.* 1986. **108** (3) 546-8.
222. Herbranson DE, Hawley MD. The electrochemical reduction of para-nitrophenyl azide - evidence consistent with the formation of the para-nitrophenyl nitrene anion radical as a short-lived intermediate. *Abstr Pap Am Chem S.* 1989. **198** 4297-303.
223. Platz MS, Carrol G, Pierrat F, Zayas J, Auster S. Some observations of various 1,8 naphthoquinodimethane biradicals by electron-spin resonance spectroscopy. *Tetrahedron.* 1982. **38** (6) 777-85.
224. Polshakov D, Rai S, Wilson RM, Mack ET, Vogel M, Krause JA, Burdzinski G, Platz MS. Photoaffinity labeling with 8-azidoadenosine and its derivatives: Chemistry of closed and opened adenosine diazaquinodimethanes. *Biochemistry.* 2005. **44** (33) 11241-53.
225. Jevsevar S, Kunstelj M, Porekar VG. PEGylation of therapeutic proteins. *Biotechnol J.* 2010. **5** (1) 113-28.
226. Wittinghofer A, Waldmann H. Ras - a molecular switch involved in tumor formation. *Angew Chem Int Edit.* 2000. **39** (23) 4193-214.
227. Kho Y, Kim SC, Jiang C, Barma D, Kwon SW, Cheng J, Jaunbergs J, Weinbaum C, Tamanoi F, Falck J, Zhao Y. A tagging-via-substrate technology for detection and proteomics of farnesylated proteins. *P Natl Acad Sci USA.* 2004. **101** (34) 12479-84.
228. Best MD. Click chemistry and bioorthogonal reactions: Unprecedented selectivity in the labeling of biological molecules. *Biochemistry.* 2009. **48** (28) 6571-84.
229. Boyce M, Bertozzi CR. Bringing chemistry to life. *Nat Methods.* 2011. **8** (8) 638-42.
230. Kalia J, Raines RT. Advances in bioconjugation. *Curr Org Chem.* 2010. **14** (2) 138-47.
231. Prescher JA, Bertozzi CR. Chemistry in living systems. *Nat Chem Biol.* 2005. **1** (1) 13-21.
232. Szychowski J, Mahdavi A, Hodas J, Bagert JD, Ngo JT, Landgraf P, Dieterich DC, Schuman EM, Tirrell DA. Cleavable biotin probes for labeling of biomolecules via azide-alkyne cycloaddition. *J Am Chem Soc.* 2010. **132** (51) 18351-60.
233. Wang Q, Chan TR, Hilgraf R, Fokin VV, Sharpless KB, Finn M. Bioconjugation by copper (i)-catalyzed azide-alkyne [3+ 2] cycloaddition. *J Am Chem Soc.* 2003. **125** (11) 3192-3.
234. Agard NJ, Baskin JM, Prescher JA, Lo A, Bertozzi CR. A comparative study of bioorthogonal reactions with azides. *ACS Chem Biol.* 2006. **1** (10) 644-8.
235. Jewett JC, Sletten EM, Bertozzi CR. Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. *J Am Chem Soc.* 2010. **132** (11) 3688-90.
236. Farkas P, Korcova J, Kronek J, Bystricky S. Preparation of synthetic polyoxazoline based carrier and vibrio cholerae o-specific polysaccharide conjugate vaccine. *Eur J Med Chem.* 2010. **45** (2) 795-9.
237. Mendes PM. Stimuli-responsive surfaces for bio-applications. *Chem Soc Rev.* 2008. **37** (11) 2512-29.
238. Tan L, Zhou KG, Zhang YH, Wang HX, Wang XD, Guo YF, Zhang HL. Nanomolar detection of dopamine in the presence of ascorbic acid at beta-cyclodextrin/graphene nanocomposite platform. *Electrochem Commun.* 2010. **12** (4) 557-60.
239. Kuila T, Bose S, Khanra P, Mishra AK, Kim NH, Lee JH. Recent advances in graphene-based biosensors. *Biosens Bioelectron.* 2011. **26** (12) 4637-48.
240. Jonkheijm P, Weinrich D, Schroder H, Niemeyer CM, Waldmann H. Chemical strategies for generating protein biochips. *Angew Chem Int Edit.* 2008. **47** (50) 9618-47.
241. Allen MJ, Tung VC, Kaner RB. Honeycomb carbon: A review of graphene. *Chem Rev.* 2010. **110** (1) 132-45.
242. Jaegfeldt H, Kuwana T, Johansson G. Electrochemical stability of catechols with a pyrene side-chain strongly adsorbed on graphite-electrodes for catalytic-oxidation of dihydronicotinamide adenine-dinucleotide. *J Am Chem Soc.* 1983. **105** (7) 1805-14.
243. Chen RJ, Zhang YG, Wang DW, Dai HJ. Noncovalent sidewall functionalization of single-walled carbon nanotubes for protein immobilization. *J Am Chem Soc.* 2001. **123** (16) 3838-9.

244. Selvin PR. The renaissance of fluorescence resonance energy transfer. *Nat Struct Biol.* 2000. **7** (9) 730-4.
245. Lakowicz JR. Principles of fluorescence spectroscopy 2nd Ed. *Kluwer Academic/Plenum Press: Chapters.* **1** 14.
246. Shen BQ, Xu K, Liu L, Raab H, Bhakta S, Kenrick M, Parsons-Repointe KL, Tien J, Yu SF, Mai E. Conjugation site modulates the *in vivo* stability and therapeutic activity of antibody-drug conjugates. *Nat Biotechnol.* 2012. **30** (2) 184-9.
247. Smith JJ, Conrad DW, Cuneo MJ, Hellinga HW. Orthogonal site-specific protein modification by engineering reversible thiol protection mechanisms. *Protein Sci.* 2005. **14** (1) 64-73.
248. Spicer CD, Davis BG. Palladium-mediated site-selective Suzuki–Miyaura protein modification at genetically encoded aryl halides. *Chem Commun.* 2011. **47** 1698-700.
249. Dirksen A, Dawson PE. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjugate Chem.* 2008. **19** (12) 2543-8.
250. Piston DW, Kremers GJ. Fluorescent protein FRET: The good, the bad and the ugly. *Trends Biochem Sci.* 2007. **32** (9) 407-14.
251. VanEngelenburg SB, Palmer AE. Fluorescent biosensors of protein function. *Curr Opin Chem Biol.* 2008. **12** (1) 60-5.
252. Kuhn SM, Rubini M, Muller MA, Skerra A. Biosynthesis of a fluorescent protein with extreme pseudo-stokes shift by introducing a genetically encoded non-natural amino acid outside the fluorophore. *J Am Chem Soc.* 2011. **133** (11) 3708-11.
253. Rantanen T, Pakkila H, Jamsen L, Kuningas K, Ukonaho T, Lovgren T, Soukka T. Tandem dye acceptor used to enhance upconversion fluorescence resonance energy transfer in homogeneous assays. *Anal Chem.* 2007. **79** (16) 6312-8.
254. Ellis RJ. Macromolecular crowding: Obvious but underappreciated. *Trends Biochem Sci.* 2001. **26** (10) 597-604.
255. Wang L, Zhang ZW, Brock A, Schultz PG. Addition of the keto functional group to the genetic code of escherichia coli. *P Natl Acad Sci USA.* 2003. **100** (1) 56-61.
256. Sletten EM, Bertozzi CR. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew Chem Int Edit.* 2009. **48** (38) 6974-98.
257. Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem.* 2012. **4** 298-304.
258. Saxon E, Bertozzi CR. Cell surface engineering by a modified staudinger reaction. *Science.* 2000. **287** (5460) 2007-10.
259. Kohn M, Breinbauer R. The Staudinger ligation - a gift to chemical biology. *Angew Chem Int Edit.* 2004. **43** (24) 3106-16.
260. Saxon E, Armstrong JI, Bertozzi CR. A "traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. *Org Lett.* 2000. **2** (14) 2141-3.
261. van Kasteren SI, Kramer HB, Jensen HH, Campbell SJ, Kirkpatrick J, Oldham NJ, Anthony DC, Davis BG. Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature.* 2007. **446** (7139) 1105-9.
262. Zorbas V, Smith AL, Xie H, Ortiz-Acevedo A, Dalton AB, Dieckmann GR, Draper RK, Baughman RH, Musselman IH. Importance of aromatic content for peptide/single-walled carbon nanotube interactions. *J Am Chem Soc.* 2005. **127** (35) 12323-8.
263. Lu Y, Lerner MB, Qi ZJ, Mitala JJ, Lim JH, Discher BM, Johnson ATC. Graphene-protein bioelectronic devices with wavelength-dependent photoresponse. *Appl Phys Lett.* 2012. **100** (3) 33110-3.
264. Kodali VK, Scrimgeour J, Kim S, Hankinson JH, Carroll KM, de Heer WA, Berger C, Curtis JE. Nonperturbative chemical modification of graphene for protein micropatterning. *Langmuir.* 2011. **27** (3) 863-5.
265. Liu LH, Lerner MM, Yan MD. Derivatization of pristine graphene with well-defined chemical functionalities. *Nano Lett.* 2010. **10** (9) 3754-6.

266. Pastine SJ, Okawa D, Kessler B, Rolandi M, Llorente M, Zettl A, Frechet JMJ. A facile and patternable method for the surface modification of carbon nanotube forests using perfluoroarylazides. *J Am Chem Soc.* 2008. **130** (13) 4238-9.
267. Georgakilas V, Otyepka M, Bourlinos AB, Chandra V, Kim N, Kemp KC, Hobza P, Zboril R, Kim KS. Functionalization of graphene: Covalent and non-covalent approaches, derivatives and applications. *Chem Rev.* 2012. **112** (11) 6156-214.
268. Novoselov KS, Fal'ko VI, Colombo L, Gellert PR, Schwab MG, Kim K. A roadmap for graphene. *Nature.* 2012. **490** (7419) 192-200.
269. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A. How enzymes adapt: Lessons from directed evolution. *Trends Biochem Sci.* 2001. **26** (2) 100-6.
270. Williams GJ, Nelson AS, Berry A. Directed evolution of enzymes for biocatalysis and the life sciences. *Cell Mol Life Sci.* 2004. **61** (24) 3034-46.
271. Davis IW, Baker D. Rosettaligand docking with full ligand and receptor flexibility. *J Mol Biol.* 2009. **385** (2) 381-92.
272. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban YEA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P. Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. *Methods in Enzymology, Vol 487: Computer Methods, Pt C.* 2011. 545-74.
273. Looger LL, Dwyer MA, Smith JJ, Hellinga HW. Computational design of receptor and sensor proteins with novel functions. *Nature.* 2003. **423** (6936) 185-90.
274. Datta D, Vaidehi N, Zhang DQ, Goddard WA. Selectivity and specificity of substrate binding in methionyl-tRNA synthetase. *Protein Sci.* 2004. **13** (10) 2693-705.
275. Kekenus-Huskey PM, Vaidehi N, Floriano WB, Goddard WA. Fidelity of phenylalanyl-tRNA synthetase in binding the natural amino acids. *J Phys Chem B.* 2003. **107** (41) 11549-57.
276. McClendon CL, Vaidehi N, Kam VWT, Zhang DQ, Goddard WA. Fidelity of seryl-tRNA synthetase to binding of natural amino acids from hierdock first principles computations. *Protein Eng Des Sel.* 2006. **19** (5) 195-203.
277. Zhang DQ, Vaidehi N, Goddard WA, Danzer JF, Debe D. Structure-based design of mutant *Methanococcus jannaschii* tyrosyl-tRNA synthetase for incorporation of *o*-methyl-L-tyrosine. *P Natl Acad Sci USA.* 2002. **99** (10) 6579-84.
278. Toscano MD, Woycechowsky KJ, Hilvert D. Minimalist active-site redesign: Teaching old enzymes new tricks. *Angew Chem Int Edit.* 2007. **46** (18) 3212-36.
279. Polgar L, Bender ML. Reactivity of thiol-subtilisin an enzyme containing a synthetic functional group. *Biochemistry.* 1967. **6** (2) 610-20.
280. Neet KE, Nanci A, Koshland DE. Properties of thiol-subtilisin - consequences of converting active serine residue to cysteine in a serine protease. *J Biol Chem.* 1968. **243** (24) 6392-401.
281. Tomatis PE, Rasia RM, Segovia L, Vila AJ. Mimicking natural evolution in metallo-beta-lactamases through second-shell ligand mutations. *P Natl Acad Sci USA.* 2005. **102** (39) 13761-6.
282. Tu Y, Lin YH, Ren ZF. Nanoelectrode arrays based on low site density aligned carbon nanotubes. *Nano Lett.* 2003. **3** (1) 107-9.
283. Wang J, Liu GD, Jan MR. Ultrasensitive electrical biosensing of proteins and DNA: Carbon-nanotube derived amplification of the recognition and transduction events. *J Am Chem Soc.* 2004. **126** (10) 3010-1.
284. Besteman K, Lee JO, Wiertz FGM, Heering HA, Dekker C. Enzyme-coated carbon nanotubes as single-molecule biosensors. *Nano Lett.* 2003. **3** (6) 727-30.
285. Lovat V, Pantarotto D, Lagostena L, Cacciari B, Grandolfo M, Righi M, Spalluto G, Prato M, Ballerini L. Carbon nanotube substrates boost neuronal electrical signaling. *Nano Lett.* 2005. **5** (6) 1107-10.
286. Mattson MP, Haddon RC, Rao AM. Molecular functionalization of carbon nanotubes and use as substrates for neuronal growth. *J Mol Neurosci.* 2000. **14** (3) 175-82.

287. Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Gadella TWJ, Chudakov DM. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat Methods*. 2007. **4** (7) 555-7.