From carbon nanotubes to zinc porphyrins: engineering proteins to interface with non-biological molecular systems

Benjamin James Bowen

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

School of Biosciences

Cardiff University

July 2020

Declaration

Acknowledgements

I would like to offer my thanks to my supervisor Dr. D. Dafydd Jones for all the guidance and support over the years, especially during the writing of this thesis. I also want to thank my wife Dawn James Bowen for her unending love, support and patience during this very long endeavour and my parents Wayne and Gwenda Bowen for their huge support since I left work to go back to school nine years ago. Apologies to all of the kids for having me stuck in the corner of the living room typing for so many months.

Many thanks to my review panel convenor Prof. Paola Borri and to Dr. Husam Al-Maslookhi, Dr. Harley Worthy, Rachel Johnson, Beckie Gwyther and Rochelle Ahmed from the DDJ lab at Cardiff School of Biosciences, for their help with all things protein related. Also, thanks to Prof. J. Emyr MacDonald, Dr. Martin Elliott, Dr. Suzanne Thomas and Dr. Adam Beachey from Cardiff School of Physics and Astronomy, for their collaboration with AFM and carbon nanotube work.

The β-lactamase sensing work would not have been possible without our collaborators at Queen Mary, University of London, Dr. Matteo Palma, Dr. Mark Freeley and Xinzhao Xu. Thank you, Dr. Chris Pudney at University of Bath for the analysis of cytochrome b_{562} REES data, Prof. Alexander Nemukhin at Moscow State University for molecular modelling and Prof. Timothy Palzkill at Baylor College of Medicine, Houston, for the BLIP-II and β-lactamase plasmids.

Finally, thank you to the BBSRC, for offering me this opportunity to study on your SWBio doctoral training program.

Abstract

A protein's function is dictated by its fold and surface chemistry, which are in turn dictated by its primary amino acid sequence. Protein structure and function can be engineered using protein structural data and *in silico* design methods to impart new functionality to natural proteins, by mutation of the native primary sequence using genetic manipulation techniques. The natural "toolbox" of 20 standard proteinogenic amino acids provides for enormous functional diversity in the context of complex and tightly regulated naturally-selected systems. However, the limited repertoire of amino acid chemistries can be limiting when engineering comparatively simple artificial protein systems. Genetic incorporation of non-natural amino acid (nnAAs) enables us to add new chemistries to engineered proteins, including fluorophores for spectroscopy, heavy elements to aid crystallography or bioorthogonal chemistries for selective reactions. More than 200 non-natural amino acids have been incorporated in proteins to date. This thesis describes the rational design of engineered proteins, which were made using site-directed mutagenesis with both canonical amino acids and the nnAA *p*-azido-L-phenylalanine (AzF).

In chapters 3 and 4 several variants of the β-lactamase binding protein BLIP-II were created, containing AzF at various defined locations. These were used to functionalise carbon nanotube (CNT) based field-effect transistor (NTFET) devices with populations of BLIP-II proteins all bound to the CNT surface in a defined and consistent orientation. These devices functioned as specific sensors for TEM β-lactamase analyte binding to surface-immobilised BLIP-II proteins. Devices functionalised with BLIP-II in different orientations were found to show distinct signal responses on binding of TEM, demonstrating recognition of analyte orientation. CNT-protein functionalisation using AzF was demonstrated via two chemical routes, using copper-free click chemistry (SPAAC) or UV-activated nitrene cycloaddition, demonstrating the versatility of AzF as a protein engineering tool. Surface functionalisation was analysed using atomic force microscopy (AFM).

In chapter 5, a variant of the small electron transport haem-protein cytochrome b_{562} was designed, produced and characterised, with its cofactor porphyrin binding affinity functionally switched from the native haem to the fluorogenic haem analogue zinc protoporphyrin IX (ZnPP). Cyt b_{562} haem binding was practically abolished while ZnPP affinity was increased 18-fold. On binding cyt b_{562}^{ZnPP}, the fluorescent emission of ZnPP increased ca. 70-fold, making cyt b_{562}^{ZnPP} potentially useful as a switchable, genetically encoded fluorescent label or as an optoelectronic component.

Abbreviations

2×YT	double yeast-tryptone media
ε	molar extinction coefficient
λ	wavelength
λ_{abs}	absorbance wavelength
λ_{em}	emission wavelength
λ_{ex}	excitation wavelength
λ_{max}	wavelength of highest absorbance
aaRS	amino-acyl tRNA-synthetase
AFM	atomic force microscopy
AMR	antimicrobial resistance
AzF	*p*-azido-L-phenylalanine
BCN	bicyclononyne
BLIP (-I, -II)	β-lactamase inhibitor protein
cyt b_{562}	cytochrome b_{562}
CNT	carbon nanotube
CuAAC	copper-catalysed azide-alkyne cycloaddition
DBCO	dibenzylcyclooctyne
DMF	dimethylformamide
DMSO	dimethyl sulphoxide
FPLC	fast protein liquid chromatography
GFP	green fluorescent protein
IMAC	immobilised metal affinity chromatography
K_D	dissociation constant
LB	lysogeny broth
nnAA	non-natural amino acid
NTFET	nanotube field-effect transistor
PCR	polymerase chain reaction
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SLIM	site-directed, ligase-independent mutagenesis
SPAAC	strain-promoted azide-alkyne cycloaddition
S-SWCNT	semiconducting, single walled carbon nanotube
TEMED	tetramethylethylenediamine
Tris	tris-hydroxymethyl-aminomethane
tRNA	transfer-RNA
v/v	volume per volume
w/v	weight per volume
WT	wild type
ZnPP	zinc (II) protoporphyrin IX

Contents

Declaration ... **ii**
Acknowledgements ... **iii**
Abstract ... **iv**
Abbreviations ... **v**
Contents ... **vi**
1 Introduction ... **1**
1.1 Engineering functional protein-nanomaterial hybrids ... 1
1.1.1 Protein engineering ... 1
1.1.2 Chemical diversity in natural proteins ... 5
1.1.3 Non-biological post-translational modifications ... 9
1.1.4 *p*-azido-L-phenylalanine: new bioorthogonal chemistry ... 13
1.2 Bionanotechnology ... 16
1.2.1 Nanotechnology for label-free biosensing ... 17
1.2.2 Carbon nanotube-protein NTFETs ... 18
1.3 Antimicrobial resistance: β-lactamases and BLIP-II ... 23
1.3.1 Class A β-lactamases ... 23
1.3.2 BLIP proteins: natural high-affinity β-lactamase inhibitors ... 26
1.4 Cytochrome b_{562}: a small metalloprotein as a nanoscale component ... 28
1.5 Aims and objectives ... 32
2 Materials & Methods ... **33**
2.1 Materials ... 33
2.1.1 Chemicals ... 33
2.1.2 Bacterial strains: *Escherichia coli* genotypes ... 33
2.1.3 Bacterial growth media ... 34
2.2 Methods ... 35
2.2.1 Generation of 3D structural models of proteins for illustrations ... 35
2.2.2 Molecular biology ... 36
2.2.3 Fast protein liquid chromatography (FPLC) ... 43
2.2.4 Recombinant cytochrome b_{562} production in *E. coli* ... 44
2.2.5 Recombinant BLIP-II production in *E. coli* ... 46
2.2.6 Recombinant CTX-M β-lactamase production in *E. coli* ... 47
2.2.7 Recombinant TEM β-lactamase production in *E. coli* ... 48

2.2.8 Protein analysis 48
2.2.9 Artificial post-translational modification of phenyl azide and cysteine protein variants 55
2.2.10 Construction and analysis of protein functionalised NTFET devices 57
2.2.11 Photo-functionalisation of CNTs with AzF-containing proteins 59

3 Integration of engineered proteins into a carbon nanotube based field-effect transistor sensing device 62

3.1 Introduction 62
3.2 Atomic force microscopy 67
3.3 Results and discussion 68

3.3.1 Production of TEM^{WT} 68
3.3.2 Characterisation of TEM^{WT} enzyme activity 70
3.3.3 BLIP-II variants: Design and production 71
3.3.4 BLIP-II variants: Modification and characterisation 85
3.3.5 Interfacing engineered β-lactamase inhibitor protein BLIP-II with carbon nanotubes via conjugated intermediates for binding event sensing 92

3.4 Conclusions 106

4 Site-specific photochemical covalent attachment of proteins to carbon nanotube side walls 110

4.1 Introduction 110
4.2 Results and discussion 115

4.2.1 Functionalisation of CNTs with $BLIP\text{-}II^{41AzF}$ via nitrene photoaddition for TEM β-lactamase binding and analysis using AFM 115
4.2.2 Functionalisation of CNTs with $BLIP\text{-}II^{213AzF}$ via nitrene photoaddition for CTX-M β-lactamase binding and analysis using TIRF microscopy 123
4.2.3 Functionalisation of CNTs with TEM^{87AzF} via nitrene photoaddition and analysis using AFM 132
4.2.4 Functionalisation of CNTs with cytochrome b_{562} variants via nitrene photoaddition and analysis using AFM 140

4.3 Conclusions 152

5 Switching protein metalloporphyrin binding specificity by design from iron to fluorogenic zinc 155

5.1 Introduction 155

5.1.1 Porphyrin cofactors 155

5.1.2 ZnPP fluorescence 156
5.2 Results and discussion 159
5.2.1 Computational design of zinc-optimised cytochrome b_{562} (performed with Adam McGarrity) 159
5.2.2 Design of cytochrome b_{562}^{ZnPP} construct 163
5.2.3 Cytochrome b_{562}^{ZnPP} production 163
5.2.4 Characterisation of cytochrome b_{562}^{ZnPP} 168
5.3 Conclusions 179
6 General discussion 183
6.1 Overview 183
6.2 Computational design of engineered proteins 184
6.3 Genetically-incorporated non-natural amino acids: expanding the chemical repertoire 185
6.4 Non-natural post-translational modification with AzF 186
6.5 Integration of proteins with carbon nanotubes for biosensing 188
6.6 Future work 189
7 Bibliography 192

1 Introduction

1.1 Engineering functional protein-nanomaterial hybrids

All known living organisms rely on protein-based systems that evolved over billions of years to fulfil a vast range of biological functions. Proteins have the greatest structural and functional diversity of all biological macromolecules, including: structural proteins such as the tubulins of the cellular cytoskeleton and collagens that underpin extracellular matrices; enzymes that catalyse metabolic pathways and synthesise other molecules; and proteins that facilitate signalling, movement, nerve transduction, immunity, growth and transport. The uniquely varied structural and functional properties of proteins make them an ideal platform for bionanotechnological uses.

Proteins are encoded by genes that use a well-understood DNA genetic code, allowing the final protein sequence to be programmed using established molecular biology methods such as site-directed mutagenesis or gene synthesis. A protein gene's DNA sequence is transcribed into mRNA, which is translated into its primary amino acid sequence, containing all information necessary for the protein to spontaneously fold into its thermodynamically favourable, functional native state[1]. Proteins can be considered examples of natural self-assembling nanomachines and can be adapted for use outside their native environment: in other organisms, in solution or interfaced with entirely non-biological systems. Many naturally evolved protein systems have potential for use outside a biological context for scientific or technological applications, e.g. enzyme catalysis, molecular recognition and sensing, electrochemistry, signal transduction or fluorescence. Protein-mediated processes can be superior to existing abiological approaches in many areas, e.g. catalytic rates at low temperature and pressure, chiral molecule selectivity or binding partner specificity. Direct functional integration of proteins with artificial nanomaterials and semiconductors, including carbon nanotubes, graphene, thin films or silicon nanowires in electronic or optoelectronic systems will enable the development of new technologies that can combine the enormous variety of protein function available in nature with the speed and simplicity of modern electronic and optical technologies.

1.1.1 Protein engineering

The wide range of physical and chemical functions of natural proteins have long made them attractive for research, biomedical and industrial applications. Natural enzymes have been used for biocatalysis since at least 1908[2], often showing superior specificity and performance to inorganic catalysts, especially in aqueous solution, at low

temperature and pressure[3]. Because proteins evolved in physiological settings, their function is often suboptimal when used outside their native environment; for example, natural enzymes are prone to inactivation by extremes of temperature, pH or oxidants used in industrial processes. Enzymes and binding proteins typically have some flexibility in substrate or binding partner specificity, and this can be engineered to switch specificity to other molecules. Improved understanding of protein structure, function and interactions gained during the last 100 years has allowed the modification of proteins to study proteins in their functional context, improve stability or function, customise specificity or add new functionality, using genetic manipulation techniques to revolutionise the human use of proteins.

The first approach to protein engineering was rational site-directed mutagenesis informed by crystallographic structural data about the subject protein, a method still widely used. An early success was modification of the *Bacillus* serine protease subtilisin by Estell *et al.* (1985)[4] to prevent oxidative inactivation of the enzyme when used with harsh detergent solutions containing hydrogen peroxide (H_2O_2). The active site adjacent residue methionine 222 was found to be vulnerable to oxidation so was mutated to all other amino acid permutations, resulting in an improved engineered subtilisin that tolerates H_2O_2 concentrations up to 1 M. Rational mutagenesis is difficult to plan without detailed 3D structural information, but another approach uses sequence alignment to compare the amino acid sequences of several homologous proteins from different species using online databases such as Uniprot (www.uniprot.org), to identify conserved functional residues and suggest residues suitable for mutagenesis[5].

Many other protein engineering approaches have been used since the mid-1980s, each having their own benefits and drawbacks. Directed evolution is an approach which creates a library of variants of the gene of interest containing random mutations, through methods such as error-prone PCR[6], or sequence saturation mutagenesis[7]. These genes are cloned into expression plasmids, transformed into a host organism and grown under artificial selective pressures (e.g. an antibiotic) to screen for beneficial mutations, with many repeated iterations to select for accumulated improvements in the gene[3]. This approach has proven successful in developing improved variants of enzymes like *para*-nitrobenzyl esterase, which was modified by directed evolution to improve its enzymatic function in a mixture of aqueous and organic solvents. A variant produced by random mutagenesis and selection was found to catalyse the target substrate in a 30% dimethylformamide (DMF) solution as well as the wild-type enzyme in aqueous buffer[8]. However, over many iterations of random mutation genes also tend to develop neutral and negative mutations. An improved variation on directed evolution used DNA shuffling, mimicking sexual recombination, in which homologous gene libraries are randomly

digested into oligonucleotides and reassembled by PCR, then cloned and transformed into a host for selection/screening. This process is repeated several times, including recombination with the original wild-type gene to minimise deleterious mutations and select for improved function[9]. This approach has been found to produce engineered proteins combining desirable properties from the different parental genes. Numerous variations of the shuffling approach were developed, including using libraries of degenerate synthesised oligonucleotides with optimised mutations for recombination[10].

Rationally designed mutagenesis has an advantage over directed evolution in directly targeting functionally relevant parts of the protein for modification, but is subject to failure because the full complexity of a protein's fold and function is often poorly understood. Mutations can result in unintended negative effects on protein function due to perturbation of the native conformation. The simplest approach to designing rational mutations is the manual inspection of 3D protein structures from x-ray crystallographic or NMR data, using molecular modelling software such as PyMOL[11] or CCP4MG[12], to inform the design of mutations. This is most useful for simple, single mutations such as introducing reactive residues distant from catalytic or binding sides, for labelling or conjugation purposes. Sites can be selected for mutation based on their location in more disordered regions of the protein, to minimise disruption of the native protein fold. Residues known to be important to protein function, e.g. from previous alanine scanning mutagenesis data, can also be avoided. Mutations can be assessed using *in silico* tools like Phyre2[13] (http://www.sbg.bio.ic.ac.uk/~phyre2) which attempts to predict the folded conformation of mutants based on comparison to a database of homologous protein structures, or ClusPro[14] (https://cluspro.bu.edu), which can simulate protein-substrate or protein-protein binding interactions (docking). Manually designed rational modification, with computational screening of designed variants, is still widely useful for simple mutations with atomic precision.

Manual selection of individual residues for mutagenesis becomes less feasible when trying to change fundamental protein functions such as binding affinity or substrate specificity. The complexity of protein binding behaviour makes estimating the effects of mutations on catalytic or binding sites difficult and unreliable. In these situations, computational tools can be used to predict beneficial mutations. Typically, a library of theoretical mutant protein structure permutations is produced, based on wild-type structural data, with binding site residues and surrounding amino acids mutated to every other possible amino acid. These can be optimised for energy minimisation in a simulated water solvation shell, to produce a library of mutant structures with minimised predicted steric clashes.

Local software such as AutoDock[15] or a supercomputer-based web service like ClusPro can be used to simulate protein-substrate or protein-protein docking with libraries of theoretical structure permutations, scoring each for free binding energy. This information can be used to narrow down a pool of likely candidates for protein mutagenesis and expression. Whole-protein molecular dynamics simulations require considerable computing power and even powerful personal computers (e.g. desktop PC or Mac Pro with Intel Xeon 10+ core CPU, dedicated graphics card (GPU) and 16+ GB RAM) can take days to run a single protein simulation in the nanosecond (ns) duration range. More complex or longer duration (μs) simulations involving multiple binding proteins are impractical using local PCs and must be performed using supercomputing clusters such as the Hawk cluster at Advanced Research Computing at Cardiff (ARCCA). Hawk allows the user to pre-book a number of CPU cores as required from its total of 19,416 cores[16]. Another option for supercomputing access is by using publicly-available web apps hosted by major research institutions like Boston University's ClusPro. In recent years, *de novo* protein design has become more feasible, using domains and motifs inspired by nature to build completely synthetic proteins. Currently, some catalytic and binding proteins have been designed[17] although native-type protein dynamics are difficult to emulate. This approach is becoming more prevalent as advances in methods and computing power continue.

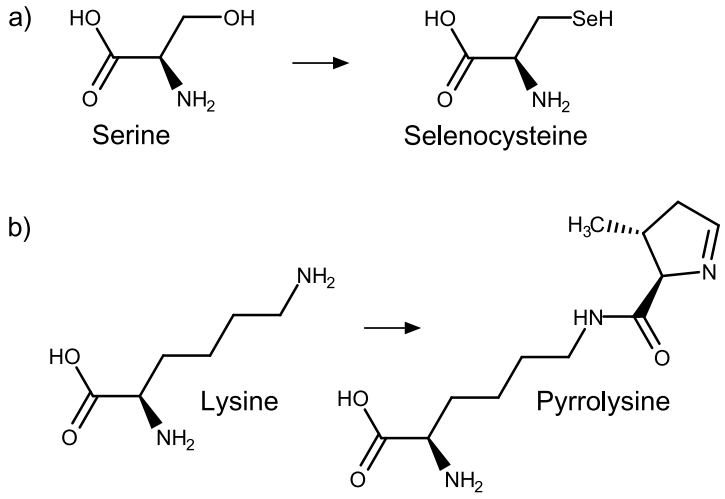
Historically, engineered proteins have mostly been designed to function within biological systems or to interact with natural biomolecules. Examples include engineered enzymes (proteases, DNA polymerases[18]), diagnostic assays (recombinant antibodies[19]) affinity fusion proteins (SNAP-tag[20], HaloTag[21]) and biomedical therapeutics (insulin, antibody-drug conjugates[22]). However, established protein engineering techniques also allow the useful properties of proteins to be integrated by design with non-biological materials, expanding their potential beyond purely biological systems towards new hybrid biotechnologies with metals, carbon nanomaterials, semiconductors or optics. Integrated biophysical systems can combine unique properties of proteins such as self-assembly, highly specific molecular recognition, enzyme catalysis, absorbance, fluorescence or electron transport, with the ability to directly couple protein function with electronic or optoelectronic technologies and record quantitative output in real time. Hybrid bioelectronic devices such as these have great potential in practical applications such as sensitive and specific molecular sensing for biomedical, environmental or industrial purposes.

This thesis explores the integration of several engineered proteins with non-biological materials, conferring new non-natural properties.

1.1.2 Chemical diversity in natural proteins

All organisms use a conserved 64-codon genetic code and a standard set of 20 amino acids to translate proteins, limiting the diversity of reactive groups available for covalently conjugating engineered proteins with non-biological materials to phenol (tyrosine), carboxyl (aspartate, glutamate), terminal amine (lysine), and thiol (cysteine) groups[23]. Two comparatively rare natural additions to the proteinogenic amino acid repertoire exist. The selenium-containing cysteine analogue selenocysteine (Sec) is the 21st amino acid in the proteomes of 48% of eukaryotes, 26% of bacteria and 13% of archaea[24], with 25 known selenoproteins in the human genome. The structure and chemical properties of selenocysteine (Se^{-}/SeH group) are similar to cysteine (S^{-}/SH group), with both able to act as coordinating ligands to metal ions and both able to form inter-and intra-chain covalent disulphide/diselenide bonds, but with some important differences. The selenocysteine side chain has a p*K*a of 5.2 compared to the cysteine side chain p*K*a of 8.3, so in the physiological pH range the selenocysteine is likely to be in the deprotonated Se^{-} form, while cysteine is likely to be protonated (SH). The selenocysteine side chain redox potential (-388 mV) is also much lower than cysteine (-220 mV), meaning that diselenide bonds are harder to break by reduction than disulphides[25]. Selenoproteins are thought to have prominent roles as antioxidants in response to oxidative stress and as redox signalling agents[26].

Selenocysteine is unusual as it is the only proteinogenic amino acid that is partly synthesised (from serine) while conjugated to its tRNA, and is incorporated into the peptide in response to a contextual UGA opal stop codon, only when a specific SECIS motif is present in the coding mRNA[27]. The only other known proteinogenic amino acid in nature is pyrrolysine (Pyl), a lysine analogue with a pyrroline group added to the side chain, originally discovered in the methanogenic archaean *Methanosarcina barkeri*[28]. Pyrrolysine was thought to be specific to methanogens, but pyrrolysine synthesis (from lysine) and incorporation genes were subsequently found in six separate archaean lineages[29]; however it remains the rarest proteinogenic amino acid. During translation pyrrolysine is incorporated by a UAG amber stop codon, with the pyrrolysine transfer RNA ($tRNA^{Pyl}$) acting as an amber suppressor by competing with release factor 1 (RF1) to prevent peptide termination. This $tRNA^{Pyl}$ and pyrrolysyl-tRNA synthetase ($aaRS^{Pyl}$) system has been engineered and widely used in synthetic biology as a method to incorporate a range of non-natural amino acids (nnAAs) into recombinant proteins, as described in section 1.1.3. Structures of selenocysteine and pyrrolysine are shown in Scheme 1.1.



Scheme 1.1 Structures of the 21st and 22nd natural amino acids. a) Selenocysteine, shown in comparison to serine, from which it is synthesised while conjugated to its tRNA. b) Pyrrolysine, shown in comparison to lysine, from which it is synthesised.

Despite the two exceptions of selenocysteine and pyrrolysine, most proteins in nature are limited to the standard 20 amino acids which provide a relatively limited repertoire of chemical reactivity. Current opinions place the number of potentially protein-coding genes in the human genome at ca. 20,000[30,31]. Many of these genes encode multiple protein isoforms using alternative promoters, alternative splicing and mRNA editing; over 250,000 different protein-coding mRNA transcripts have been identified in the human transcriptome to date[30]. The diversity of expressed proteins is further expanded by another order of magnitude by post-translational modifications (PTMs): the chemical modification of proteins as, or after, they are translated at the ribosome. The total number of proteoforms[32] in humans is not known but recent estimates are in the millions[33,34], with post-translational modification adding much of this variety (Figure 1.1).

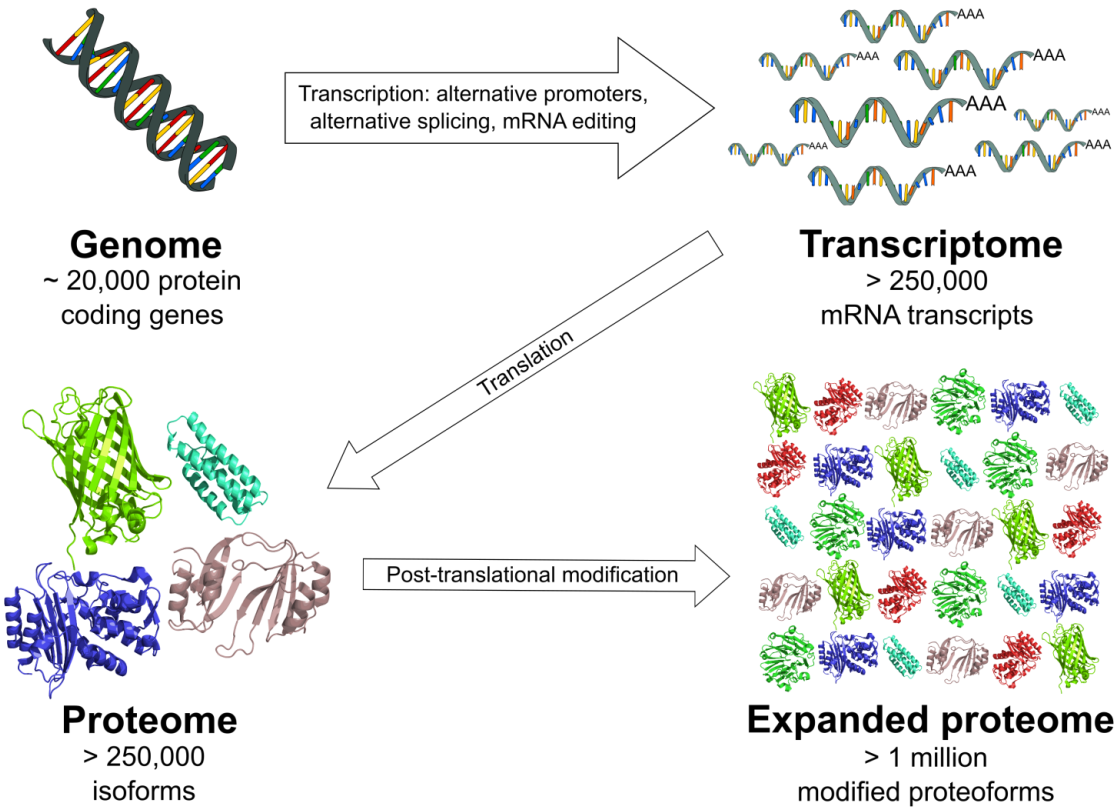


Figure 1.1 Two orders of magnitude increase in complexity from the same genomic information due to differential expression during transcription, and post-translational modification of the final protein. PDB: 2B3P[35], 1JTD[36], 1JTG[36], 1YLT[37] and 256B[38]. DNA and RNA adapted from "File:Difference DNA RNA-DE.svg" (Wikimedia Commons, licensed under CC BY-SA 3.0).

Most PTMs involve the site-specific covalent addition of a functional group to an amino acid side chain or N/C-termini of a protein by enzymes. Conjugating or removing modifications can quickly switch on or off protein function in response to stimuli, allowing a much faster and more energetically-efficient response than up- or downregulation of protein expression[39]. One of the most common modifications is phosphorylation, the addition of phosphate (PO_4^{2-}) to specific serine, threonine or tyrosine hydroxyl groups by protein kinases (PKs). PKs are a diverse family of over 500 enzymes involved in most cell signalling pathways, constituting 1.7% of the human proteome[40], and 48% of human proteins are subject to phosphorylation[41]. Phosphorylation often activates target enzymes by inducing a conformational shift, causing chain reactions of PKs phosphorylating other PKs in cascades involving thousands of proteins, for fast cellular signal transduction. Phosphate groups can be removed to deactivate these enzymes by phosphatases. Other modifications used in signalling include methylation or acetylation of specific amine groups on chromatin histone proteins, enabling rapid epigenetic regulation of gene transcription by histone methyltransferase and acetyltransferase enzymes. Covalent additions can be small, like methylation (14 Da), to full-sized proteins such as the immunoprotein ISG15 (15 kDa)[42] or multimeric chains of the 8.6 kDa protein ubiquitin, and can be reversible or irreversible. Ubiquitination is a coded label that marks

the target protein for degradation, trafficking to a different cellular localisation, activation or inactivation, depending on the number and location of conjugated ubiquitin subunits[43,44]. Other common PTMs include glycosylation, the conjugation of glycans to nitrogens (*N*-linked) or hydroxyls (*O*-linked) of amino acid side chains, involved in folding and cell surface recognition; and acylation, the attachment of acyl groups, commonly fatty acids, to thiol (*S*-linked) or hydroxyl (*O*-linked) groups, for membrane anchoring or trafficking. Structures of some common covalently added moieties are shown in Figure 1.2. Other PTMs include the deamidation of asparagine or glutamine, formation of disulphide bonds between cysteine residues, and cleavage of lengths of peptide by proteolytic enzymes[45].

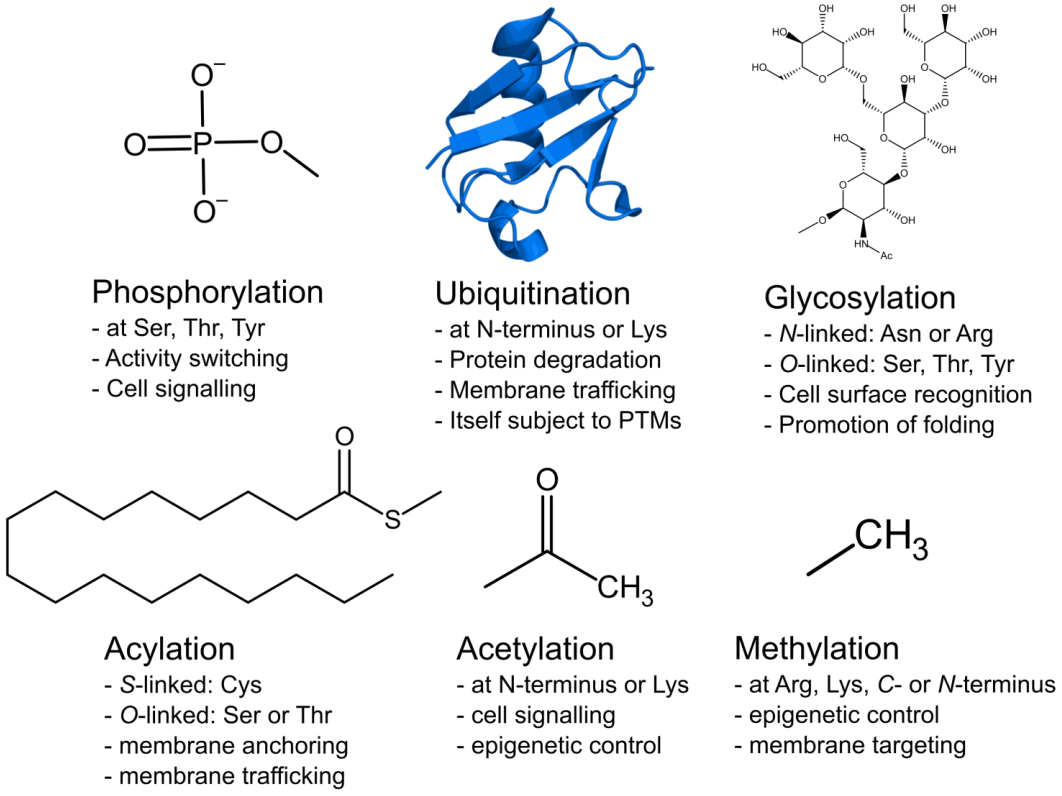


Figure 1.2 Structures of some more common covalent addition post-translational modifications, with their typical protein attachment locations and common functions. PDB: 4NQK[46].

While these natural methods of site-specific protein modification are extremely specific, fast and reliable for activity switching, regulation and interfacing with other biomolecules *in vivo*, they depend on a complex control network involving many other enzymes, which is only partly understood and currently impossible to replicate. This makes natural PTMs of limited use for engineering artificial interfaces between proteins and non-biological materials because we lack the capability to specifically modify one residue of many without off-target effects. Most proteins contain multiples of each amino acid, meaning that single-residue modification specificity is difficult to achieve using the canonical 20 amino acids.

Many proteins have evolved to expand the range of available chemical or physical properties by binding to cofactors; these are non-protein molecules that bind to inactive apoproteins to form a functional holoprotein. Cofactor associations can be permanent, such as those bound by covalent or strong coordination chemistries, often called prosthetic groups, or transient and bound by weaker hydrophobic and polar interactions. Prosthetic groups can be structurally integrated and required for correct folding of the holoprotein. Common cofactors include vitamins such as ascorbic acid (electron donor) or biotin (CO_2 donor for carboxylases); adenosine triphosphate (ATP, phosphate donor); metal ions such as iron, zinc or magnesium, typically involved in catalysing redox reactions; and metal porphyrins like haem or chlorophyll, also redox-active components. Metal ions or metal porphyrins interact with proteins via coordination chemistry, with coordinating ligands provided by: backbone carbonyl[47] or amide[48] groups; the polar or negative side chains of histidine (imidazole), tyrosine (phenol), cysteine (thiol), methionine (thioether), aspartate or glutamate (carboxylate)[49]; or structural water molecules[50]. Metal coordination spheres can adopt a range of ligand geometries, most commonly with coordination numbers between 4 and 6[50,51]. In metal porphyrins, the tetrapyrrole ring nitrogens form four coordinating ligands to the metal centre, with the remaining one or two axial ligands provided by the protein, usually in a 5-coordinate tetragonal pyramidal or 6-coordinate octahedral geometry[52]. The haem-containing metalloprotein cytochrome b_{562} is engineered in chapters 4 and 5 of this thesis (see section 1.4).

1.1.3 Non-biological post-translational modifications

Though natural PTMs are difficult to mimic, the PTM concept is attractive from a protein engineering perspective because they could be exploited to create new protein-material interfaces or alter the function of proteins after translation. Artificial PTMs have been designed, using canonical chemistries, to modify proteins with fluorophores or affinity tags[53]; alter enzyme specificity[54]; improve solubility and reduce immunogenicity[55]; mimic natural PTMs[56]; or interface proteins with abiological nanomaterials[57]. Traditionally, these modifications have been made by chemically modifying, or conjugating to, the side chains of the standard 20 amino acids. An inherent limitation of this approach is that there are often multiples of the target amino acid type in the protein of interest, making the labelling of specific residues difficult. The main targets for covalent conjugation are primary amines (lysine), carboxylic acids (aspartic acid, glutamic acid) and thiols (cysteine). Cysteine is preferred as its occurrence in proteins is on average 4-5 times less frequent than lysine, aspartic acid or glutamic acid[58], with many proteins containing none. In these cases, a cysteine can be introduced by site-directed mutagenesis to

provide a specific reactive handle for labelling of conjugation, usually by efficient thiol-maleimide chemistry[59], a method used for specific fluorescent labelling of a protein in chapter 4 of this thesis.

To achieve the highest modification specificity in engineered PTMs, bioorthogonal chemistries can be used to eliminate off-target labelling or conjugation, allowing modification reactions to be targeted to the defined residue only. This allows the design of proteins that can self-assemble into specifically oriented structures via non-natural reactions, which is useful when interfacing protein populations with non-biological materials. Genetic code expansions have been developed to incorporate non-natural amino acids (nnAAs) bearing side chains with new physical and chemical properties at specific residues of an engineered protein. Genetic incorporation of nnAAs into proteins was first achieved using a cell-free transcription/translation system by the Peter Schultz lab at University of California, Berkeley, in 1989[60]. This first system used synthetic nnAAs chemically ligated to a yeast phenylalanyl transfer RNA ($tRNA^{Phe}$) with its anticodon mutated to CUA. The aminoacylated tRNA was used to incorporate nnAAs into recombinant proteins, in response to a TAG amber stop codon mutation in the coding gene. The engineered $tRNA_{CUA}$ competes with RF1 to incorporate an nnAA, suppressing the stop codon[61]. This system, known as amber suppression, was inspired by the naturally expanded genetic code discovered in pyrrolysine-containing archaea, which also uses a reassigned TAG stop codon. Radiolabelled phenylalanine and four different nnAAs were genetically incorporated into proteins using the artificially aminoacylated engineered $tRNA^{Phe}$, with expression yields up to 20% that of wild type protein.

The amber suppression approach for genetic code expansion and nnAA incorporation has been further evolved over the past 30 years, with more than 200 different synthetic amino acids genetically incorporated into recombinant proteins[62], offering a wide range of non-natural chemical groups. These have many purposes including: labelling proteins with heavy atoms such as iodine[63] or to aid determination of x-ray crystallographic structures; targeted bioorthogonal reactions with non-natural chemistries such as tetrazines[64] or strained cyclooctynes[65]; or photochemical reactivity with a benzophenone[66]. Non-natural side chains have been widely used to study and engineer protein structure and function, including studying the effects of residue fluorination on the folding and stability of hydrophobic protein regions by incorporation of fluorine-linked leucine analogues[67]; for residue-specific, defined conjugation of proteins to artificial surfaces and/or molecules such as biotin[68]; for defined self-assembly of artificial fluorescent protein homo- and heterodimers[69] via bioorthogonal chemistry; and for detecting transient expression of proteins in response to environmental pressures[70]. New applications are constantly being developed for nnAA incorporation, now an established

strategy in the research toolboxes of synthetic biology, proteomics, pharmacology and structural biology.

The most widespread current systems use live cell hosts with engineered tRNAs to incorporate the nnAA co-translationally at the ribosome, paired with an engineered amino acid-RNA synthetase (aaRS) to charge the tRNA with the nnAA[71] (wild type tRNA/aaRS shown in Figure 1.3). These incorporation components are taken from a different domain of life to the expression organism, to avoid interference with the native cellular machinery and prevent mistranslation. The aaRS are adapted to work with new amino acids by rounds of directed evolution using positive and negative selection[66].

For expression of recombinant proteins containing nnAAs in *Escherichia coli*, tRNA/aaRS pairs from methanogenic archaea are normally used, co-transformed along with the gene of interest on a separate plasmid. Pyrrolysyl-tRNA synthetases and $tRNA^{Pyl}$ from the methanogenic archaea *M. barkeri* or *M. mazei* are used as the basis for systems incorporating nnAAs with large linear side chains (similar to pyrrolysine), while a tyrosyl-tRNA synthetase and $tRNA^{Tyr}$ from the archaean *Methanocaldococcus jannaschii* are typically used as the basis for incorporating nnAAs with shorter, aromatic side chains (more similar to tyrosine). The $tRNA^{nnAA}_{CUA}$ anticodons are reprogrammed to incorporate the nnAA in response to a UAG amber stop codon in the template mRNA (TAG in gene) instead of terminating elongation[72,73] (Figure 1.4). To a lesser extent, an archaean leucyl-tRNA synthetase from *Methanobacterium thermoautotrophicum* and a $tRNA^{Leu}$ from *Halobacterium* have been engineered and used to incorporate nnAAs in response to either a TGA opal stop codon or the four-base codon AGGA[74].

The TAG amber codon is the least-common stop codon in the *E. coli* genome, terminating ca. 320 genes, around 8%, in *E. coli* K-12, so amber-suppressed cells remain viable[71,75]. Some *E. coli* nnAA expression systems have knocked-out expression of RF1 in the host genome, to further improve nnAA incorporation efficiency[76], and some have been modified to remove all native TAG stop codons[77]. Amber suppression allows relatively simple expression of recombinant proteins containing nnAAs at designed locations in *E. coli*, using standard site-directed mutagenesis to insert a TAG codon at the desired residue. The engineered $tRNA^{nnAA}_{CUA}$, and the aaRS that charges the tRNA with the nnAA, are encoded on a separate plasmid that is co-transformed into *E. coli* alongside the expression plasmid containing the target gene with TAG codon mutation. The growth media must be supplemented with the nnAA, or the TAG codon reverts to functioning as a standard stop codon resulting in a truncated protein; proteins are then expressed and purified as usual.

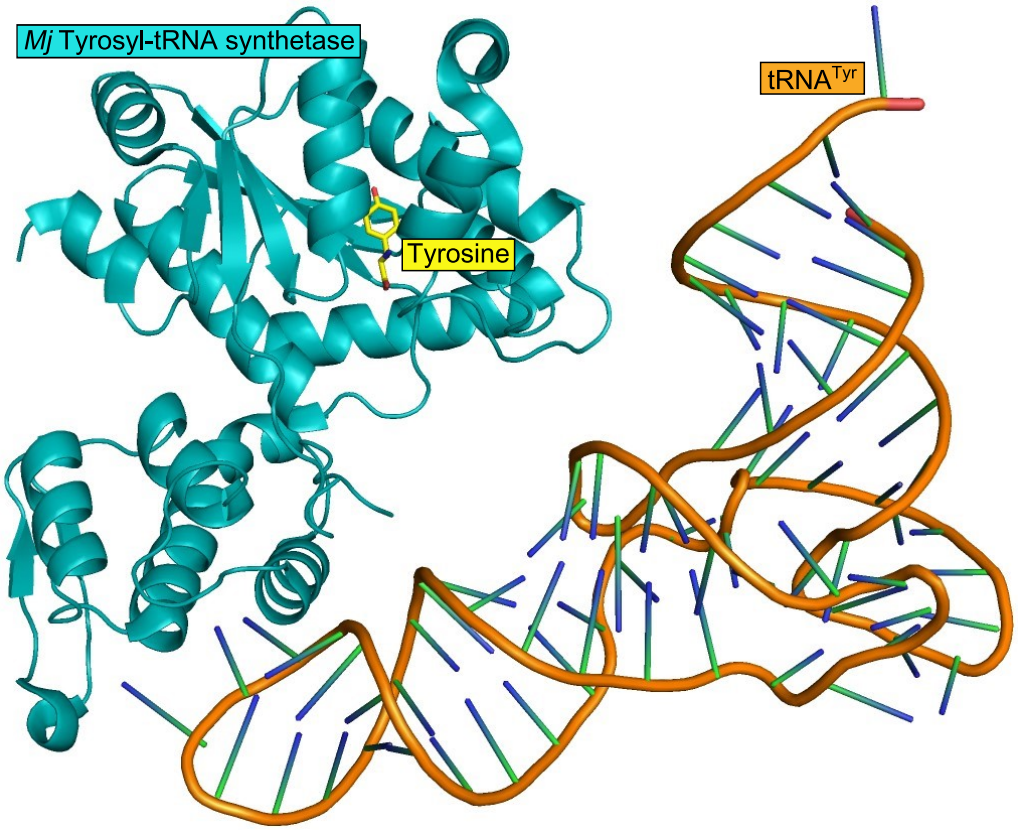


Figure 1.3 *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase and tRNA^{Tyr} pair. These components have been engineered to incorporate a range of different nnAAs in response to a TAG/UAG codon. PDB 1J1U[78].

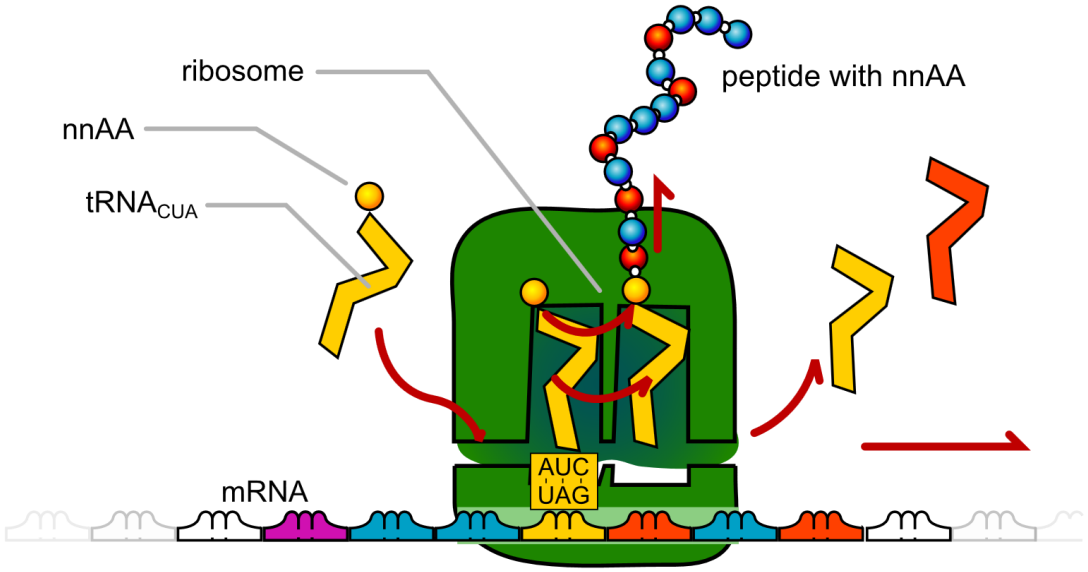
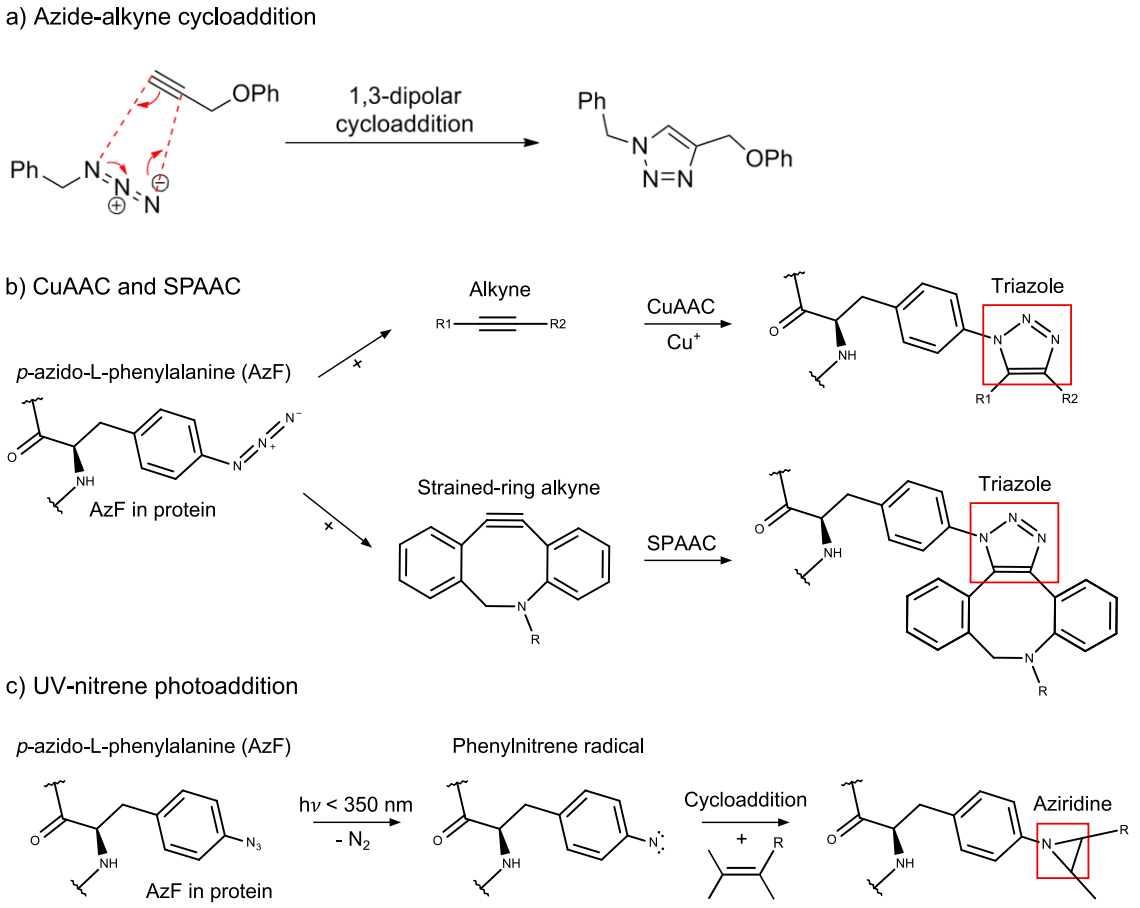


Figure 1.4 Schematic of nnAA incorporation during translation. The engineered tRNA$_{CUA}$ incorporates nnAA into the nascent peptide in response to a UAG amber stop codon in the mRNA transcript, suppressing normal termination of elongation by release factor 1 (RF1). Adapted from "File:TRNA ribosomes diagram en.svg" (Wikimedia Commons, public domain).

1.1.4 *p*-azido-L-phenylalanine: new bioorthogonal chemistry

The nnAA genetically incorporated into proteins in this thesis was *p*-azido-L-phenylalanine (AzF), a versatile tyrosine-derived synthetic nnAA that carries a bioorthogonal phenyl azide (Phe-N_3) side chain. The AzF incorporation system was engineered by the Schultz lab[79] by directed evolution, using the *M. jannaschii* aaRS/$tRNA_{CUA}$ pair (see 1.1.3) as a basis. The phenyl azide group of AzF has a range of unusual properties, including its distinctive infrared absorbance at 2100 cm^{-1} making it useful as an IR spectroscopy probe. The activation of the G-protein coupled receptor (GPCR) rhodopsin was first tracked by monitoring the IR signatures of AzF residues in a mutant rhodopsin as the protein's conformation shifted in response to light stimuli[80]. AzF also offers two different chemical routes to artificial PTM via covalent addition: the first route is via azide-alkyne cycloaddition (AAC), often described as "click chemistry" due to its high specificity and ease of joining individual chemical modules[81]. In this reaction the azide group of AzF reacts specifically with an alkyne group of a target molecule to form a five-membered triazole ring joining the two. A copper (I) catalyst is required for the reaction to occur at room temperature (CuAAC), however Cu^+ ions damage proteins by catalysing production of reactive oxygen species such as hydroxyl radicals[82]. To avoid protein degradation by Cu^+, the strained-ring alkyne reactive group dibenzylcyclooctyne (DBCO) can be used instead, for conjugation to protein AzF residues via their phenyl azide side chain. This reaction, strain-promoted AAC (SPAAC), occurs at room temperature in aqueous buffers without catalyst to form a covalent triazole ring linkage (Scheme 1.2a-b). The bioorthogonal specificity of azide-alkyne click chemistry allows targeted conjugation to a single AzF residue of a single protein, even in complex protein mixtures such as cell lysate.

A second route to covalent modification of AzF uses ultraviolet (UV) light to activate the phenyl azide group, via the extrusion of molecular nitrogen (N_2), to a highly reactive phenylnitrene radical that can undergo a wide range of reactions. These include reduction to amine, ring expansion to a seven-membered ketenimine ring, insertion into N-H or C-H σ-bonds or [2+1] nitrene cycloaddition to π-bonds[83]. For this work, genetically-incorporated AzF residues were covalently modified by nitrene cycloaddition, with the phenyl azide activated to nitrene by UV irradiation, followed by formation of a triangular aziridine covalent linkage with the electron-rich π-electron system of alkene/phenyl carbons (Scheme 1.2c). See chapter 4.1 for further information on the photochemical reactions of AzF.



Scheme 1.2 AzF click reactions used in in this thesis. a) Mechanism of azide-alkyne cycloaddition (AAC) by 1,3-dipolar cycloaddition. b) Copper-catalysed (CuAAC) and strain-promoted (SPAAC) approaches; azide binds alkyne with a copper (I) catalyst, or a strained-ring alkyne without catalyst, forming triazole ring. c) UV-nitrene photoaddition: phenyl azide is photoactivated to phenylnitrene radical by UV irradiation, then adds to electron rich bond (e.g. alkene) to form aziridine ring. Image in a) from "File: Example of 1,3-dipolar cycloaddition.tif" (Wikimedia Commons, public domain).

AzF has an extensive 18-year history of use in non-natural protein engineering and is arguably one of the most popular nnAAs, due to its range of useful non-natural chemical properties. Over the last decade, the Dr. D. Dafydd Jones lab here at Cardiff University has created many engineered proteins genetically incorporating nnAAs, with 13 papers published using AzF in many different contexts. The specific bioorthogonal SPAAC reactivity of AzF as used in chapter 3 of this thesis has been used extensively, including for labelling an engineered variant of *Aequorea victoria* hydrozoan green fluorescent protein (GFP) or TEM β-lactamase (see 1.3.1) with exactly one synthetic DNA oligonucleotide per protein at a specified AzF residue, for incorporation into DNA origami arrays. Proteins were immobilised in defined orientations by bottom-up assembly on a DNA origami tile, resulting in increased TEM catalytic rates compared to TEM in solution[84]. This demonstrated the utility of rationally-designed AzF mutations for designing oriented self-assembling nanostructures. In another paper, further nanoscale structures were assembled, interfacing proteins with a wholly non-biological material: the

ends of DBCO-functionalised carbon nanotubes (CNTs) were conjugated to individual GFP molecules using SPAAC, with AzF mutations in the GFP variants placed to orient the proteins in two distinct orientations relative to the CNT. GFP fluorescence was found to be unaffected when conjugated to individual CNTs in one orientation, but quenched by 80% when attached in a 90° rotated orientation. This implied direct communication between the GFP chromophore and the CNT in the orientation where the distance between them was 1.6 nm, but not where the CNT-chromophore distance was 3 nm. GFP mutants containing two AzF residues were also conjugated to CNT ends, resulting in evidence of CNT-GFP-CNT molecular junctions, confirmed by atomic force microscopy[85]. In a 2019 paper, GFP was again used as a platform to construct self-assembling nanostructures by SPAAC, for investigating communication between adjacent chromophores. Two GFP variants were produced containing AzF residues at position 148 or 204, and two variants were produced containing the SPAAC-active nnAA cyclooctyne-lysine (SCO) at position 148 or 204. These were used to construct a range of artificial dimers by inter-protein SPAAC, which were purified and spectrally characterised, finding that dimer spectral properties varied based on the inter-chromophore geometry and range. Synergistic emission effects suggested direct energy transfer between some chromophore pairs, similar to natural light harvesting systems[69].

The photochemical reactive properties of AzF have also been widely used by the Jones lab to probe the structure-function relationships of proteins and construct artificial nanostructures. The earliest paper genetically incorporated AzF into T4 lysozyme and GFP to analyse the short-lived radical intermediates formed when AzF phenyl azide is irradiated with UV-B light. AzF was inserted into a hydrophobic pocket of lysozyme and into the chromophore of GFP, surrounding the phenyl azide with residues to cage the reactive intermediate. On UV irradiation and analysis with electron paramagnetic resonance (EPR) spectroscopy, signatures for anilino radical (lysozyme) and a triplet aryl nitrene (GFP) were detected[86]. Further experiments with AzF incorporated into GFP developed three different photo-switchable GFP variants with the AzF placed in the chromophore at Y66AzF or nearby at F145AzF or H148AzF; these variants showed structural and fluorescence changes after irradiation with UV-B light. GFP Y66AzF was a largely inactive variant with ca. 5% of WT emission, that was activated after reduction of the phenyl azide by a 30 second UV-B irradiation, switching to WT-equivalent emission, effectively switching on. F145AzF was found to crosslink to the chromophore tyrosine after a 30 second irradiation, reducing emission from WT-equivalence by ca. 85%, effectively switching off. The third variant, H148AzF, swapped the ratios of its two excitation maxima after a 4 minute irradiation, effectively switching its major excitation wavelength (λ_{ex}) from 490 nm to 511 nm[87]. These demonstrated the utility of AzF to effect

designed post-translational structural modifications in engineered proteins by light irradiation alone, without chemical modification. Similar experiments were performed using mCherry, an engineered version of the red fluorescent protein DsRed from *Discosoma* corallimorphs, a distant evolutionary relative of GFP with high structural homology. Variants of mCherry were produced with photoswitching properties similar to those induced in the GFP variants described above: the inactive variant I197AzF was found to be activated by 30 seconds of UV-B irradiation, while active variant W143AzF was deactivated by the same irradiation[88]. This demonstrated the wider applicability of designed structural and functional modulation of AzF-containing proteins by light exposure. Fluorescent protein photoswitching mutations were also randomly induced using directed evolution rather than rational design, identifying novel positions for mutations that affect the chromophore not previously considered[89]. AzF photochemistry has also been used by the Jones lab to interface proteins directly with graphitic carbon nanomaterials such as graphene or CNT surfaces for functional integration of proteins into nanoelectronic systems, as in chapter 4 of this thesis. UV-activated [2+1] nitrene cycloaddition is a useful method of covalently functionalising nanomaterials with populations of engineered proteins, with all proteins conjugated in predefined spatial orientations. Graphene surfaces were functionalised with populations of engineered TEM, GFP or cytochrome b_{562} in mild biologic aqueous conditions, with the surface shown to be decorated with individual proteins via atomic force microscopy[90]. This demonstrated the practicality of interfacing proteins with nanomaterials using genetically incorporated AzF chemistry.

1.2 Bionanotechnology

Nanotechnology is the application of the physical sciences at the molecular level, generally considered to apply to structures, systems and phenomena at a scale below 100 nm[91], and is still a relatively new field. Even more recent is the incorporation of biological materials into hybrid nanoscale systems, sometimes called *bionanotechnology*. This includes the interfacing of biological molecules with artificial nanosystems, like exploiting the self-assembling nature of many proteins to create artificial biomaterials with superior mechanical properties[92] or using the unique base-pairing properties of DNA as a self-assembling structural nanocomponent, rather than as a store of genetic information[93]. Proteins can themselves be considered a naturally evolved example of nanomachinery and have been incorporated into artificial nanosystems for solubilising carbon nanotubes (CNTs) in aqueous solution[94] and for rendering CNTs biocompatible as the structural basis of artificial extracellular matrix scaffolds[95]. Interfacing engineered proteins with synthetic nanomaterials also shows

great potential for the construction of advanced biosensors that improve on traditional methods of analyte detection like western blotting or ELISA, in terms of sensitivity and speed. Bionanotechnological devices combine the high specificity of biomolecular interactions with the sensitivity and ease of integration with electronic or optical systems offered by nanomaterials such as CNTs, thin films and semiconductors.

1.2.1 Nanotechnology for label-free biosensing

Real-time, label-free sensing of biological molecules is of great technological interest for a wide range of biomedical, commercial and environmental applications, and many different nanoscale approaches are under development. Most current biosensor designs detect biological analytes using surface-immobilised recognition elements such as antibodies, binding proteins, nucleic acid aptamers etc., with detection via either optical or electrical signals. The most widely-used optical method is surface plasmon resonance (SPR) spectroscopy, a method where a metal thin-film substrate is functionalised with a population of recognition molecules such as antibodies, receptors or aptamers, then subsequent analyte binding is quantified by the absorbance of polarised light reflected from the substrate. This is a powerful technique able to characterise kinetic parameters (k_{on}, k_{off}, K_D) for the binding interaction in real time and has high reproducibility. Recognition/binding molecules are immobilised on the gold substrates in defined orientations via engineered cysteine residues or linker molecules carrying thiol residues. Linker molecules can be conjugated via engineered nnAAs using click chemistry for more precise control over the orientation of recognition elements[96]. Another thin-film optical biosensing technology is reflectometric interference spectroscopy (RIfS), which uses antibody-functionalised tantalum oxide (Ta_2O_5) and silicon dioxide (SiO_2) film layers or more recently a nanoporous aluminium substrate, in which analyte binding to the functionalised substrate is measured by its refractive effect on interference in a beam of light. SiO_2 substrates can be functionalised by coating with silane-amine-streptavidin complexes, which allows immobilisation of biotin-linked proteins[97], while aluminium substrates are gold-coated by deposition so proteins can be linked via thiols[98] (as with SPR). RIfS also allows for oriented immobilisation of recognition proteins using nnAAs to define conjugation points. Though these optical methods have been miniaturised using lab-on-a-chip microfluidic biosensing technologies[98,99], a major downside is the requirement for large, specialised analytical equipment costing tens to hundreds of thousands of dollars, necessitating laboratory analysis of samples and ruling out point-of-care or field testing of analytes.

Electronic biosensors are typically based on transduction by field-effect transistors (BioFETs), comprising a semiconducting silicon transistor channel between source and

drain electrodes. The most common type of BioFET is the ion-sensitive FET (ISFET), in which the channel current is gated by an electric field modulated by interactions between electrolytes, immobilised recognition elements and analytes on the other side of an SiO_2 insulating membrane[100]. Recognition biomolecules are typically immobilised on the SiO_2 membrane similarly to RIfS as mentioned above: a silane monolayer coating is adsorbed to the surface, providing amine groups for intimate chemical attachment via a range of reactive linker molecules. Still greater sensitivity has been achieved with silicon nanowire field-effect transistors (SiNW-FETs), in which the channel is a silicon nanowire of ca. 100 nm diameter, due to the higher surface area to volume ratio of the channel material. SiNW channels are usually functionalised by the same methods is ISFETs[101].

The most sensitive bioelectronic sensors so far created are based on transistor channels composed of single-walled semiconducting carbon nanotubes (S-SWCNTs). CNT-based biosensing devices offer advantages over both thin-film optical methods and silicon semiconducting FETs, such as the inherently high sensitivity of CNTs, their low cost and ease of integration with standard electronic components. CNTs are readily functionalised with biomolecules such as DNA aptamers or proteins using acid oxidation to provide reactive defects for covalent conjugation, conjugation to DNA-wrapped CNTs, adsorption via non-covalent linker molecules such as pyrene, or direct UV-mediated nitrene cycloaddition (as mentioned in section 0). Antibody-functionalised CNT-based field effect transistors (NTFETs) have shown extreme sensitivity for specific sensing of C-reactive protein down to 4 pM (0.1 ng/mL) with a wide detection range up to 4 µM (100 µg/mL)[102]. The label-free protein sensing lower limit of SPR sensitivity is at least two orders of magnitude higher at ca. 100-200 pM[103,104], and the lower limit for sensing of the small molecule diclofenac using RIfS is ca. 400 pM. The equipment requirement for use of an electronic NTFET sensing device can be minimal making the approach well-suited to point-of-care applications, and the component itself can even be disposable[102], in contrast to the high equipment requirement for optical methods.

1.2.2 Carbon nanotube-protein NTFETs

Carbon nanotubes (CNTs) are graphitic nanomaterials composed of graphene rolled into tubes with diameter: length aspect ratios typically greater than 1:100. They are considered almost ideal one-dimensional electrical conductors and have uniquely high surface area to volume ratios. CNT conduction properties vary depending on the geometry of their graphitic network. Metallic-type CNTs have the “armchair” plane of the graphitic lattice aligned around their circumference and exhibit metal-like conduction properties, while the circumference of semiconducting CNTs is aligned along the 30°-rotated “zigzag” plane as shown in Figure 1.5. Armchair-oriented CNTs have no bandgap

between their valence and conduction bands, hence their metallic conduction properties; zigzag CNTs exhibit a bandgap and semiconducting properties.

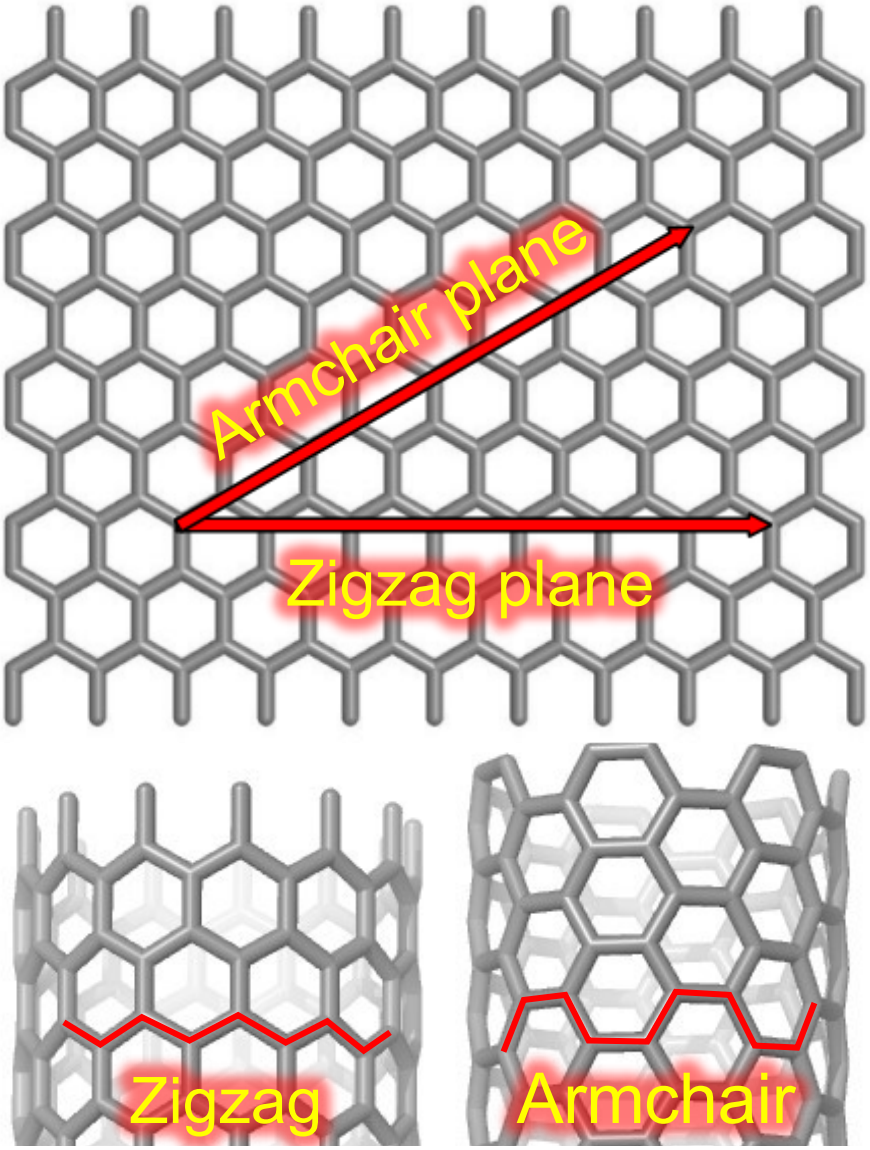


Figure 1.5 Effect of graphitic network geometry on CNT electrical properties. CNTs with the "zigzag" plane around their circumference show semiconducting properties while CNTs with the "armchair" plane around their circumference have metallic properties.

Semiconducting CNTs are of particular interest as the basis of molecular sensing devices due to their high sensitivity to the local chemical environment[105]. NTFETs work on the same principles as SiNW-FETs, but with greater sensitivity because CNTs have low charge carrier density compared to other materials, with all conduction occurring at their surface, so semiconducting single-walled carbon nanotube (S-SWCNT) conduction is readily modulated by weak electric fields. It has been shown that charged molecules adsorbed or bound within one Debye length (λ_D) of the surface of S-SWCNTs can modulate the conductive properties of the material, through local depletion or concentration of charge carriers (electrons in n-type CNTs, holes in p-type CNTs), a

phenomenon called electrostatic gating[106]. A Debye length is ca. 0.7 nm at a typical physiological ionic strength (ca. 160 mM) because of the screening effect from counter-charged ions[107,108], this range is sufficient for charged surfaces of adsorbed biomolecules to interact with the CNT. Sensors have been made that exploit these effects for bulk chemical sensing of adsorbed small molecules through detection of changes in resistivity of S-SWCNTs[109–111] forming the transistor channel of an NTFET[105]. An interesting goal is the study of more complex biomolecular interactions by an adsorbed molecule's electrostatic effect on a current signal through an NTFET.

NTFET sensors have been constructed to detect biomolecular binding events or protein conformational changes at the CNT surface, using various recognition/binding elements and conjugation strategies. In Xu *et al.* (2018)[112] NTFETs were functionalised with synthesised DNA aptamers as binding elements for three biomarkers of human stress and neurotrauma. Aptamers are synthetic 20-80 base DNA or RNA molecules that bind target molecules through their folded 3D structure and many sequences are available from aptamer databases such as Aptagen (www.aptagen.com/apta-index). DNA aptamers are more commonly used due to the vulnerability of RNA to degradation. CNTs were solubilised for aqueous solution by wrapping with DNA oligonucleotides carrying a bicyclononyne (BCN) strained-ring alkyne group; these were functionalised with azide-linked DNA aptamers using SPAAC click chemistry. The aptamer-functionalised devices specifically detected the biomarkers cortisol, dehydroepiandrosterone sulphate (DHEAS), and neuropeptide Y (NPY) in real time as changes in current signal through the CNT channel. While successful in this case, and having the advantage of stability and cheap, easy synthesis, DNA aptamers have serious shortcomings that limit their usefulness in applications such as NTFETs. New aptamers are designed by a time-consuming iterative process of *in vitro* evolution to select aptamer variants with binding affinity for the target molecule called Systematic Evolution of Ligands by Exponential enrichment (SELEX). A relatively limited range of aptamers are available and typical analyte binding affinity tends to be inferior to that of antibodies or specific binding proteins[113]. Affinity is often further reduced by addition of a reactive group and conjugation to a substrate such as a CNT[114].

Antibodies have been used to functionalise NTFETs with success, albeit with relatively poor sensitivity due to the large size of antibodies binding their analytes outside the Debye length range from the CNT, even in dilute buffers. This has been improved by using the smaller antigen-binding Fab domain fragments of antibodies rather than whole antibodies, increasing analyte detection sensitivity a million-fold from 1 µg/mL to 1 pg/mL (7 fM)[115]. Other small binding proteins that have been used as for localisation and recognition include nanobodies: small single chain antibodies that originated in some

shark or camelid species; and affimers: engineered small binding proteins that are modelled on the antigen binding sites of antibodies[116]. While these antibody-based recognition elements are promising, high affinity and specificity non-antibody binding proteins or enzymes already exist for many analytes, are much smaller than antibodies and often undergo conformational shifts on analyte binding that can be detected by the NTFET.

Choi *et al.* (2013)[117] functionalised single CNTs with individual lysozyme molecules via a pyrene linker molecule conjugated to a single cysteine residue placed at ten different locations on the enzyme surface in ten variants. These devices showed varying electric current responses on addition of peptidoglycan substrate to the device, with lysozyme variants in different orientations showing distinct and consistent signal changes, related to conformational shifts of the charged protein surfaces close to the CNT surface. Detailed catalytic, turnover and conformational information about individual lysozyme molecules was recorded in real-time in this paper and its 2015 follow-up[118].

Another protein-NTFET device, described by Hatada *et al.* (2018)[119], was functionalised with a population of the protein SocA, a high-specificity binding protein for fructosyl valine (FV), a biomarker for long-term raised blood glucose concentration. SocA was conjugated to the CNT channel via random amine groups, rather than in a defined orientation. Conformational changes induced in SocA at the CNT surface on FV binding were detected as distinct current changes through the device, resulting in a sensor 14 times more sensitive than a previous fluorescence-based method. This demonstrated the utility of a population of small binding proteins as NTFET recognition elements, but for analysis or detection of specific protein effects at the single-molecule level, it is preferable that all recognition/binding proteins are bound to a CNT surface in the same orientation, as shown in Figure 1.6. A mixed population of proteins in random orientations gives an ensemble signal, with subtle characteristics lost to averaging across a range of binding orientations.

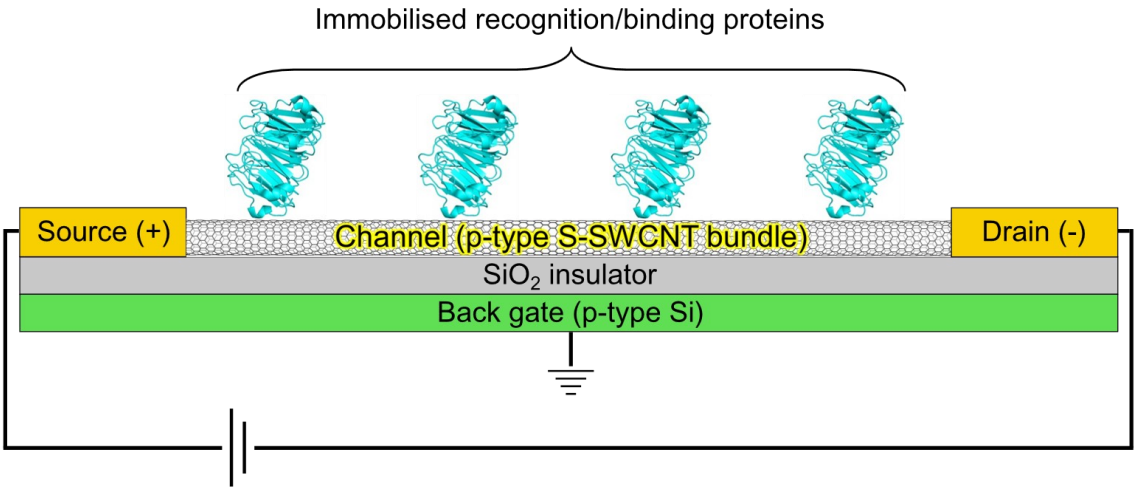


Figure 1.6 Schematic of an NTFET functionalised with specifically oriented binding proteins. Changes in the electrostatic environment within 1 Debye length (λ_D) of the CNT channel will modulate channel conductivity on target protein binding. A fixed source-drain potential results in measurable changes in current. PDB: 1JTD[36].

The development of an expanded repertoire of amino acid chemistry available for protein engineering offers interesting new opportunities for bionanotechnology. Genetically encoding non-natural amino acids at specific positions within a protein enables new approaches to interfacing proteins with nanomaterials with high specificity, while defining the orientation of attached proteins. The reactive side chain of AzF[79], phenyl azide, is especially well suited for functionalising carbon nanomaterials like graphene, CNTs and fullerenes. On UV-B irradiation, the phenyl azide forms a singlet phenylnitrene radical[120] that is able to insert into the π-bond system by electrophilic [2+1] cycloaddition to form a stable covalent aziridine attachment, while introducing few sidewall defects[121–123]. Phenyl azide is also able to undergo azide-alkyne cycloaddition (AAC) as described in section 0, to form a covalent triazole linkage to practically any attached moiety. This method is useful to form interfaces between proteins and carbon nanomaterials via DNA[124] or pyrene[125] intermediates that can bind carbon surfaces like CNT sidewalls non-covalently by π-π-stacking[126]. Both these approaches are explored in this thesis.

A major aim of this thesis is to make and examine interfaces between S-SWCNTs and engineered proteins attached in defined orientations, to create hybrid nanomaterial-biomaterial constructs. We aim to functionalise CNT devices with small (< 30 kDa) proteins that show high affinity for a target analyte, while minimising size to bind analytes within the Debye range of the CNT surface. These will form the basis of a CNT-based sensor that can detect specific protein-protein interactions by the effects on CNT conduction and is readily integrated into electronic devices. This could provide a new type of diagnostic assay that is faster and more sensitive than current antibody-based protein assays like Western blotting or ELISA. This CNT-based work was in partnership

with the Prof Emyr Macdonald and Dr. Martin Elliott labs at Cardiff University School of Physics and the Dr. Matteo Palma lab at Queen Mary, University of London School of Biological and Chemical Sciences.

1.3 Antimicrobial resistance: β-lactamases and BLIP-II

Modern healthcare is heavily dependent on antibiotic drugs to treat common bacterial infections that were often fatal throughout most of human history, infected wounds or transmitted diseases were commonly incurable before the discovery and exploitation of antibiotic compounds. The first of these was arsphenamine, first marketed in 1910 for treating syphilis, followed by the less toxic and more broadly applicable penicillin, in use from 1942[127]. Antibiotics are widely considered one of the most important human discoveries, having saved countless lives since the mid-20th century. However, since the 1960s, the efficacy of many antibiotics has been threatened by increasing incidence of antimicrobial resistance (AMR) in bacterial populations: bacterial ability to survive previously effective antibiotics.

AMR is not a new phenomenon, having evolved in bacteria over millions of years[128]. The resistance genes that encode antibiotic-inactivating enzymes are often plasmid-borne and have spread widely between bacterial species and populations through horizontal gene transfer[129,130]. Most classes of antibiotics in current use derive from soil fungi and bacteria that were thought to have co-evolved in an “arms race” of competing selective pressures, evolving antibiotics, resistance enzymes, and inhibitors to these enzymes. However, the natural functions of antibiotics are unknown and possibly unrelated to their antibacterial properties[129]. Widespread human and livestock overuse of antibiotics since the 20th century has created new selective pressures, speeding up bacterial evolution accordingly[131]. This has greatly increased the incidence of AMR in populations of human pathogenic bacteria, and the number of multi-drug resistant bacterial infections that are difficult or impossible to treat[132]; AMR is one of the biggest challenges facing health systems worldwide.

1.3.1 Class A β-lactamases

Penicillin was the first of the β-lactam antibiotics, a major class of antibiotics that inactivate the DD-transpeptidase enzymes that catalyse crosslinking of peptidoglycans during the synthesis of bacterial cell walls[131]. Transpeptidase inactivation by a β-lactam is lethal to bacteria as a weakened cell wall causes cell lysis. All β-lactam antibiotics share a central reactive 2-azetidinone ring[133] (Scheme 1.3a-b) that acts as a “suicide substrate”, with the carbonyl carbon of this ring covalently binding to the transpeptidase active site serine and preventing catalysis[134].

The most common class of antibiotic resistance genes in the environment are those that encode β-lactamase enzymes that inactivate β-lactam antibiotics by cleaving the reactive 2-azetidinone ring at their centre (Scheme 1.3c). The β-lactamase family of enzymes is closely related to, and evolved from, the DD-transpeptidases, acquiring the capability to hydrolyse and disable β-lactams[135]. The β-lactamases are classified by protein sequence similarity into four classes: A, C and D having a catalytic serine hydroxyl group as the nucleophile that attacks the β-lactam carbonyl carbon, and class B (metallo-β-lactamases) having a zinc(II)-activated water molecule as the nucleophile[136]. Of these, class A are the most prevalent, especially the extended-spectrum β-lactamases TEM (named for the original patient, Temoneira: Figure 1.7), CTX-M (named for its cefotaxime hydrolysis, isolated in Munich) and SHV (named for "sulfhydryl reagent variable")[137,138].

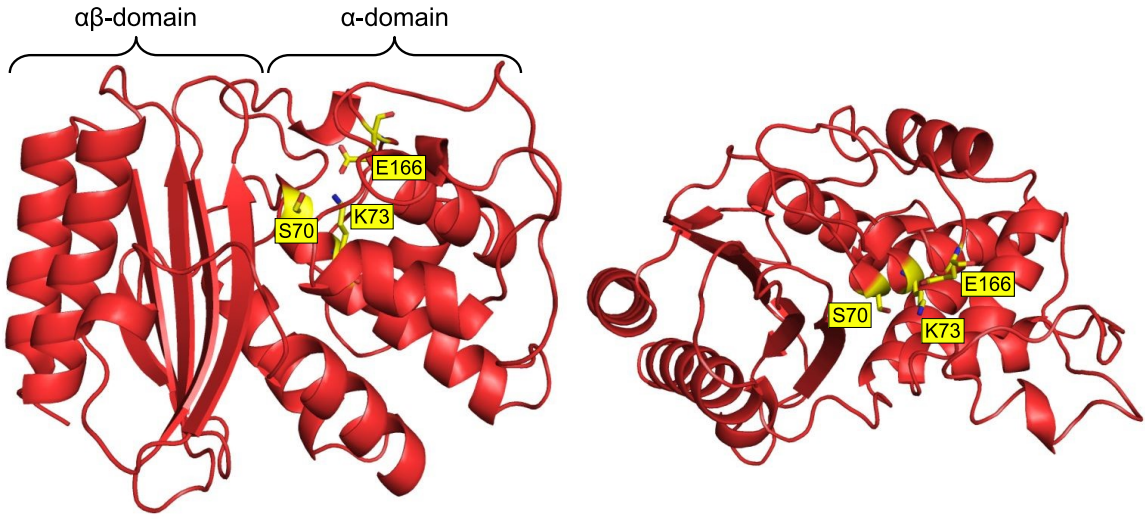
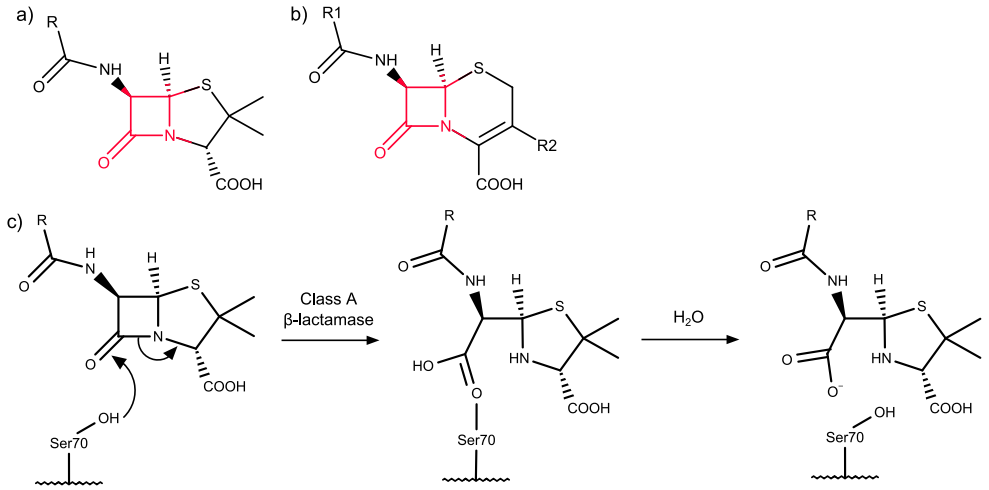


Figure 1.7 Structure of TEM β-lactamase. All class A serine β-lactamases share the same fold as TEM, having two major domains, one comprising five α-helices and the other a mixture of α-helices and β-sheets. The β-lactam binding site is between the two domains and the catalytic residues are S70, K73 and E166.



Scheme 1.3 β-lactam antibiotics. The common 2-azetidinone ring highlighted in red: a) Penicillins. b) Cephalosporins. c) Hydrolysis of the β-lactam 2-azetidinone amide by class A serine β-lactamases.

Currently, clinical testing for AMR requires samples from an infected patient to be sent to a laboratory for culture on a range of antibiotics to determine an organism's susceptibility or resistance. The result can take several days and cannot definitively identify the resistance protein being expressed. Testing strips based on chromogenic β-lactamase substrates such as nitrocefin are also available but are also non-specific and are especially prone to false negatives and positives[139]. A device that could specifically detect β-lactamase proteins in a patient sample, or even detect *which* β-lactamase is present, would be a useful tool in the clinical environment, informing the choice of antibiotic and β-lactamase inhibitor prescribed.

β-lactamases have been extensively engineered to characterise their structure, function, dynamics and mutation potential, due to their biomedical importance. Information for many engineered β-lactamases are stored in the Lactamase Engineering Database[140] (www.laced.uni-stuttgart.de). Because they have been so thoroughly characterised, β-lactamases have also served as model systems for elucidating general aspects of protein function and evolution, using protein engineering. As an example, the natural evolution of β-lactamase function in ancestral DD-transpeptidases was investigated by engineering a current PBP2X DD-transpeptidase to acquire β-lactamase function, using directed evolution and random mutagenesis. β-lactamase function was acquired, albeit with slow catalytic turnover compared to natural β-lactamases, providing evolutionary insights into development of β-lactam resistance in bacteria[141]. The Jones lab at Cardiff University have previously created engineered TEM β-lactamase variants containing AzF residues. Hartley *et al.* (2015) used the SPAAC reactivity of the AzF phenyl azide to post-translationally modulate enzyme activity with targeted click reactions in two different TEM mutants. A DBCO-linked amine moiety was conjugated at two AzF mutation sites and was found to 1) inhibit enzyme activity by 75% when conjugated at residue Y105AzF or 2) restore wild-type level enzyme activity when conjugated to the functionally-attenuated P174AzF variant[142]. This demonstrated the feasibility of designed bioorthogonal PTMs as molecular switches able to modulate enzyme structure and activity. TEM variants with surface AzF mutations were also covalently conjugated to DNA oligonucleotides via SPAAC and interfaced directly with DNA origami arrays to create spontaneous self-assembling nanostructures[84], which enhanced TEM activity as mentioned in section 0 above. Utilising the photochemical attachment properties of AzF, TEM variants were created to functionalise graphene sheets with populations of TEM proteins in specifically defined orientations[90]. This work proved the concept and feasibility of oriented protein conjugation to graphitic nanomaterials via UV-nitrene cycloaddition and led directly to the UV-nitrene functionalisation of CNTs with three unrelated proteins in chapter 4 of this thesis. Chapters 3 and 4 describe the bottom-up assembly of

nanoscale CNT-based devices functionalised with proteins via SPAAC and UV-nitrene photoaddition, respectively. A major aim of this work was to use β-lactamase inhibitory/binding proteins as recognition elements to electronically detect the closely related β-lactamases TEM and CTX-M as target analytes. These two β-lactamases have high structural homology but only 38% sequence identity (via EMBOSS[143]), and distinct surface electrostatic properties; creation of a bioelectronic device able to distinguish between β-lactamases is the ultimate aim.

1.3.2 BLIP proteins: natural high-affinity β-lactamase inhibitors

Several natural proteins exist in soil bacteria that are extremely specific and potent inhibitors of a wide spectrum of β-lactamase enzymes, with binding affinities many orders of magnitude stronger than the commonly used inhibitors clavulanate, sulbactam and tazobactam[144]. BLIP and BLIP-I are 17.5 kDa inhibitory proteins isolated from *Streptomyces exfoliatus*, they are close structural homologues with 38% sequence identity and very high binding affinity for many β-lactamases: TEM dissociation constant (K_D): BLIP = 110 pM[145]; BLIP-I = 47 pM[146]. BLIP and BLIP-I bind to the catalytic residues of β-lactamases, disabling them. BLIP-II is an evolutionarily-unrelated, 28 kDa protein from *S. clavuligerus* with a seven bladed β-propeller structure and even higher affinity for β-lactamases: TEM and CTX-M K_D = 0.5 pM[147]; KPC-2 K_D = 80 fM[147] - some of the strongest known protein-protein interactions (Figure 1.8). BLIP-II does not bind to β-lactamase catalytic sites, but to a nearby loop-helix domain, sterically blocking the catalytic site. The main β-lactamase binding residues of BLIP-II form a circle of aromatic residues in which many residues contribute to binding. This concerted binding is thought to contribute to the broad spectrum specificity of BLIP-II for the whole class A β-lactamase family[148].

The high affinity β-lactamase interactions common to BLIP proteins have made them a subject of interest as model systems for protein engineering, with much of the significant characterisation and engineering carried out by the Prof Timothy Palzkill lab at Baylor College of Medicine, Houston, USA. The TEM β-lactamase – BLIP interaction was used as a basis for engineering an artificial allosteric modulation site into an end loop of GFP, by inserting TEM into GFP then subjecting the fusion protein to random mutagenesis and directed evolution, creating a fusion protein with fluorescence activated by BLIP binding[149]. This enzyme-inhibitor pair were also used for a similar sensing/switching purpose, integrating a fluorophore into TEM that activates on binding to BLIP[150]. BLIP has also been fused to TEM via a cleavable peptide linker, to construct an autoinhibitory enzyme that can be activated by proteolytic cleavage for use as a reporter in protease activity assays[151]. BLIP has had its binding site residues engineered to reduce its broad

spectrum of β-lactamase enzyme specificity to effectively bind and inhibit only KPC β-lactamase. This BLIP variant could prove useful as the recognition component of a KPC-specific sensing device or assay[152]. BLIP residue Y50 was found to be central to β-lactamase binding specificity[153]. By comparison, BLIP-I and BLIP-II have been subject to little engineering, other than the characterisation of BLIP-II binding interactions with β-lactamases and other bacterial proteins by alanine scanning mutagenesis[147,154–156]. This is possibly because they were discovered later than BLIP (Doran *et al.* 1990[157]), with BLIP-II discovered in 1998[158] and BLIP-I in 2000[146].

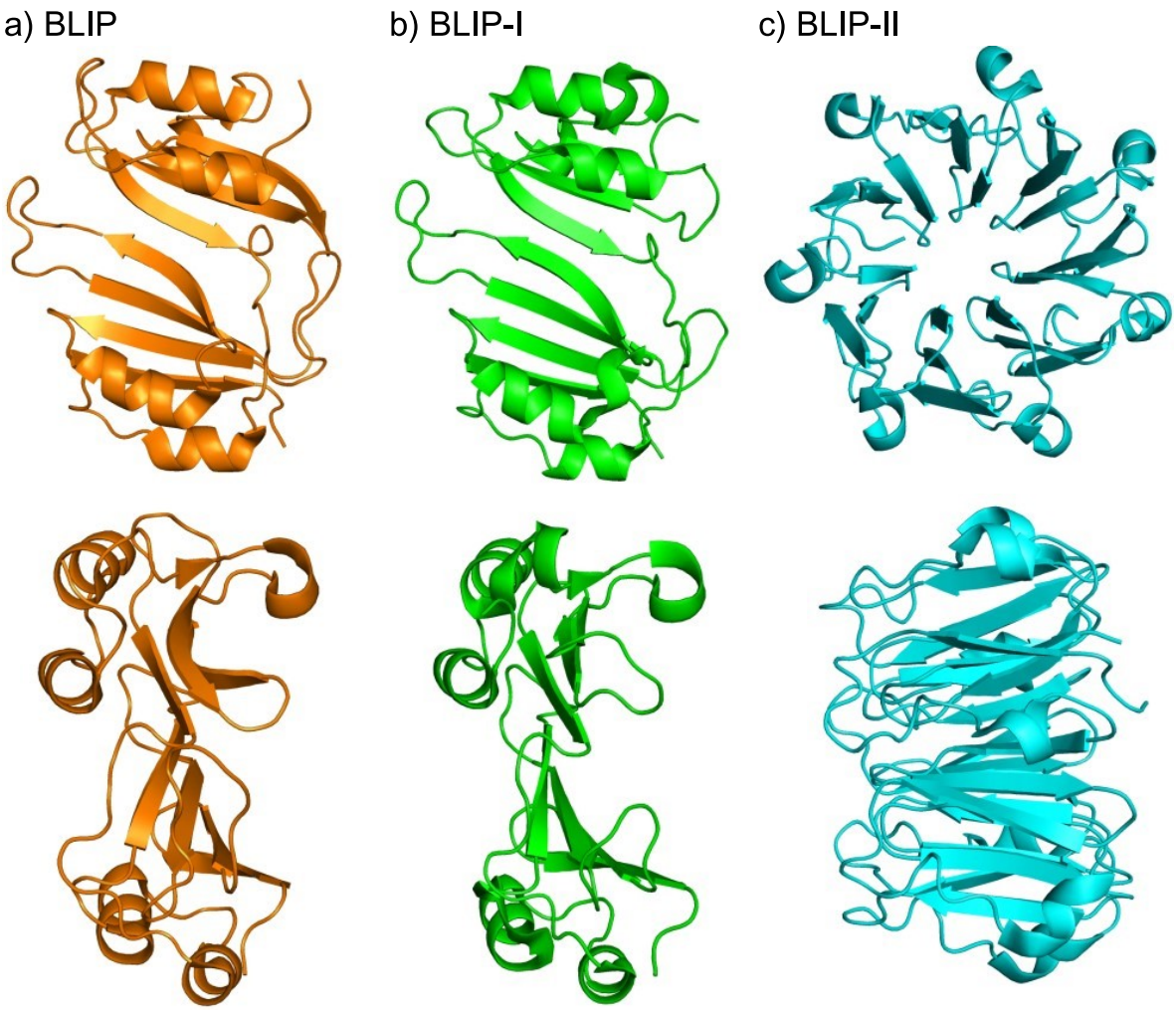


Figure 1.8 β-lactamase inhibitor proteins. a) BLIP and b) BLIP-I share close structural homology. c) BLIP-II is unrelated and has a seven-bladed β-propeller fold. PDB: 1JTD/1JTG[36], 3GMW[159].

The inherent tight binding affinity and broad range of specificity makes BLIP-II attractive as a “universal antibody” recognition element for a class A β-lactamase sensing device. BLIP-II proteins are also considerably smaller than traditional antibodies with a maximum height of ca. 4 nm compared to a typical IgG (> 8.5 nm), potentially making BLIP-II superior for functionalising nanodevices that require intimate contact between analyte and CNT. The circular β-propeller structure of BLIP-II also offers seven repeating

domains which could be used to easily present the protein in a range of orientations at a CNT surface by introducing AzF mutations on these domains for CNT conjugation. This could allow the CNT to sample different electrostatic faces of β-lactamase analytes in different orientations. This idea could be extended to discerning between different β-lactamases by their different electrostatic effects on the signal though an NTFET.

Chapters 3 and 4 of this thesis address the functionalisation of NTFET devices with BLIP-II, with the aim of specifically sensing β-lactamase enzymes. Engineered variants of BLIP-II containing genetically-incorporated AzF residues were designed and produced for immobilisation on CNT surfaces by two distinct chemical routes (see 0). To examine the broader applicability of the photochemical method, an engineered AzF containing variant of TEM β-lactamase was also used to functionalise CNTs. The eventual aim of this research is to develop a biomedical point-of-care test that can rapidly identify β-lactamase types present in a patient sample, to assist treatment decisions.

1.4 Cytochrome b_{562}: a small metalloprotein as a nanoscale component

Some of the most important proteins in living organisms require a metal porphyrin cofactor to function, with haem the most common of these. Haemproteins exist in all domains of life and are integral to many highly conserved systems such as the mitochondrial electron transport complexes of oxidative phosphorylation in eukaryotes; the cytochrome P450 enzyme family that catalyse a wide range of synthetic and catabolic functions; and oxygen transport and storage via haemoglobin and myoglobin in vertebrates.

Cytochrome b_{562} is a small (12 kDa) soluble electron transport haemprotein found in the periplasm of *E. coli*, consisting of a bundle of four α-helices surrounding a single haem cofactor. The haem iron is hexacoordinated by the porphyrin tetrapyrrole, with the side chains of two cyt b_{562} residues providing axial ligands: methionine 7 and histidine 102[160]. The exact natural function of soluble bacterial cyt b_{562} is not known, but its structural and functional similarity to cytochrome *c* [161] implies a related role in catalysing redox reactions and electron transport for oxidative phosphorylation. The protein has many attributes that make it a popular basis for protein engineering. Cyt b_{562} is a robust protein that remains stable and soluble up to extremes of pH, temperature, salt/chaotropic concentration[162] and ultraviolet radiation[163] and is also tolerant to mutagenesis[160]. Denatured cyt b_{562} assumes a random coil state and in the absence of haem the apo-protein remains largely unstructured, but spontaneously refolds in the presence of haem (or other metalloporphyrins, see chapter 5) into the native structured holo-protein. Helix 4 does

not fully form until cyt b_{562} binds to haem (Figure 1.9). The simple, self-assembling four-helix structure of cyt b_{562} has inspired the design of fully synthetic metalloprotein structures (maquettes) that help understand the principles underlying protein structure and function[164].

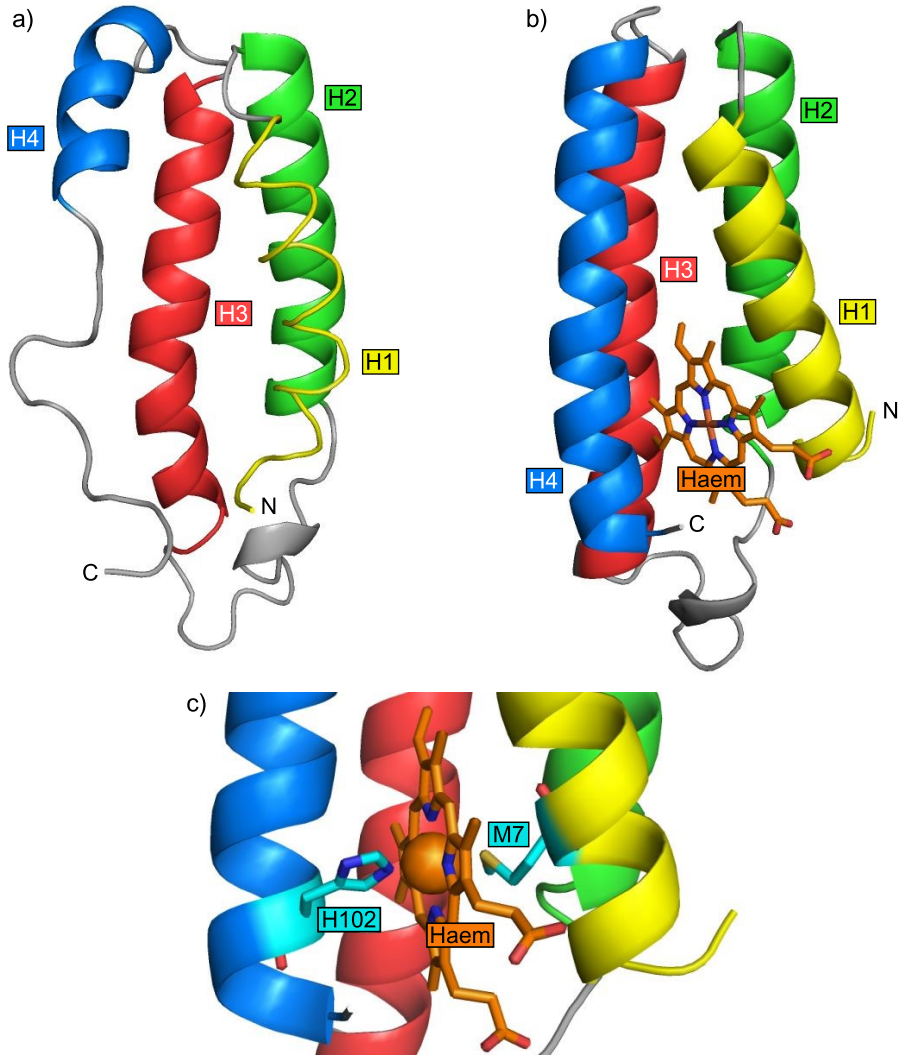


Figure 1.9 Cytochrome b_{562}. a) Cyt b_{562} apo-protein exists in a disordered, semi-folded state with helix 4 mostly unstructured. b) On haem cofactor binding, cyt b_{562} folds around haem to form the fully structured holo-protein, with helix 4 adopting its native structure. c) Detail of haem iron coordinating axial ligands M7 and H102 (cyan). PDB: 1APC[165], 256B[38].

The high haem binding affinity (K_D ca. 9 nM) of cyt b_{562} was exploited by Oohora *et al.* (2012) to use the protein as a structural subunit in haem-mediated self-assembling nanostructures. Synthetic haem molecules were tethered to an engineered cysteine residue on a cyt b_{562} variant via a maleimide linker chain. On extraction of the native haem cofactor, the semi-folded cyt b_{562} proteins reconstituted around the tethered haem groups of other cyt b_{562} proteins, forming linear fibrils of proteins joined by the protein-haem binding interactions of subunits[166]. This arrangement is analogous to the bottom-up assembly of natural fibrous proteins such as actin, and similar systems could lead to

development of designed, spontaneously self-assembling scaffolds for nanotechnological uses.

Previous work from the Jones lab at Cardiff has demonstrated many interesting characteristics and applications of cyt b_{562}. A cyt b_{562}-TEM β-lactamase fusion protein was created by directed evolution-domain insertion in which TEM enzyme activity was near-disabled by binding of haem to cyt b_{562}[167]. This effectively repurposed cyt b_{562} as a small genetically-encoded integral haem receptor or biosensor, coupling the function of the two unrelated proteins. Functional coupling was again demonstrated when cyt b_{562} was fused with GFP by directed evolution-domain insertion, creating a system in which energy from fluorescent excitation at the GFP chromophore λ_{max} was transferred with ca. 100% efficiency to the haem group of the attached cyt b_{562}. The distance of the resonance energy transfer between chromophores in this artificial scaffold was ca. 1.7 nm, displaying efficiency and range comparable to energy transfer in nature such as plant light harvesting systems[168].

This remote communication between two natural chromophores inspired further work: engineering cyt b_{562} to interface with non-biological materials, including conjugation to gold (III) surfaces via engineered cysteine thiol groups. A cyt b_{562} variant adsorbed to gold was shown to undergo electron transfer with the gold over a distance of 2 nm, demonstrating significant molecular conductance[169]. Further, variants of cyt b_{562} were shown by AFM height measurement to assemble on a gold surface in two distinct orientations, lying down (5Cys, 104Cys: short axis variant) or standing up (21Cys, 50Cys: long axis variant), depending on placement of the cysteine conjugation residues on the protein surface[170] (Figure 1.10). This showed that the self-assembled orientation of cyt b_{562} on non-biological substrates can be defined by design. Analysis by scanning tunnelling microscopy (STM) revealed that direct electrical measurements could be taken across individual proteins adsorbed on gold by contacting the second cysteine at the opposite side of cyt b_{562} with a gold STM probe to form a protein "molecular bridge" between STM tip and surface[170].

All these findings make cyt b_{562} an interesting subject for further electrochemical investigation and construction of new nanoscale structures. Two cyt b_{562} variants were created to investigate conjugation of proteins to graphitic surfaces; these contained mutations to AzF and cysteine at positions 5AzF, 104Cys (short axis or cyt b_{562}^{SA}) and 21Cys, 50AzF (long axis or cyt b_{562}^{LA}). Graphene sheet was functionalised with cyt b_{562}^{LA} proteins by UV-activated nitrene photoaddition (see section 0), demonstrating the feasibility of this approach to oriented conjugation of proteins to carbon nanomaterials[90]. The variants cyt b_{562}^{SA} and cyt b_{562}^{LA} are used in chapter 4 of this thesis, for oriented

functionalisation of CNTs with cyt b_{562} via AzF photochemistry, and for cyt b_{562}-cyt b_{562} dimerisation via inter-chain disulphide bridges.

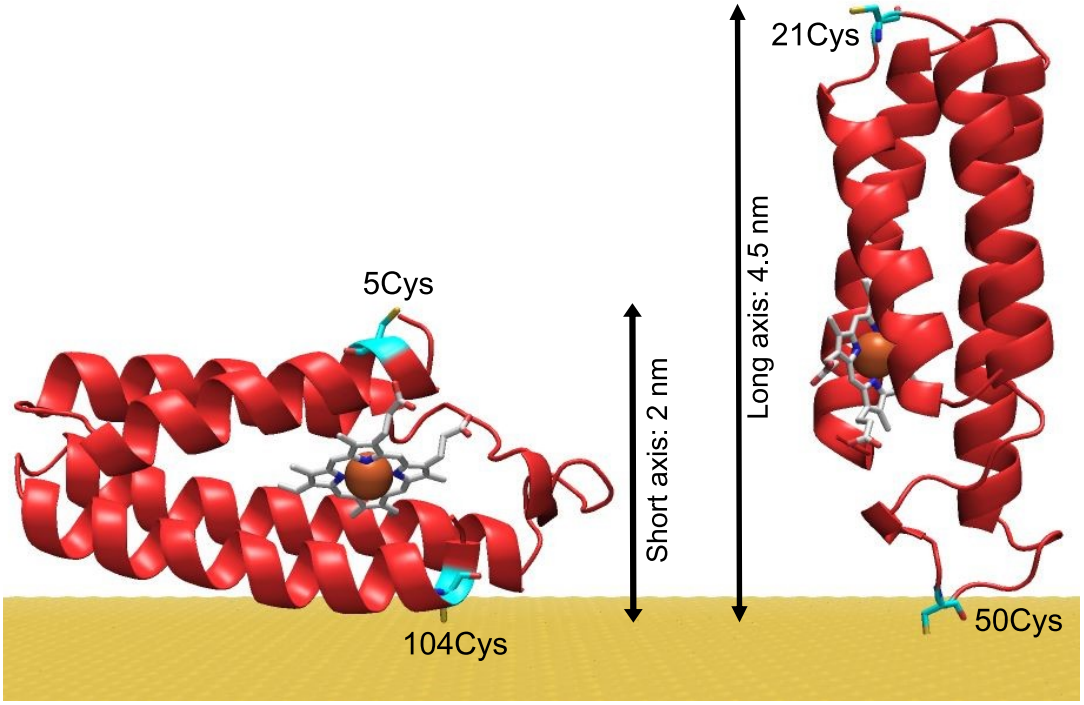


Figure 1.10 Two orientations of cytochrome b_{562} immobilised on gold surface (Della Pia et al. 2012[170]). Engineered cysteine mutations for binding to surface and STM probe shown in cyan. Short axis: cysteine at residues 5 and 104. Long axis: cysteine at residues 21 and 50.

While cyt b_{562} natively binds a haem cofactor, the apo-protein is also able to bind other metalloporphyrins, wild-type cyt b_{562} and engineered variants have been reconstituted with a range of natural or synthetic porphyrins, conferring altered electrochemical and photochemical properties. Della Pia *et al.* (2012)[171] reconstituted cyt b_{562} with protoporphyrin IX variants with copper (CuPP) or zinc (ZnPP) metal centres, albeit at lower affinity than haem. This showed the ability to easily tune the electron transport and redox properties of a protein by switching out the porphyrin cofactor, with potential catalytic applications. In Sommer *et al.* (2016)[172], cyt b_{562} variants were engineered by substituting the haem centre for cobalt protoporphyrin IX (CoPP), resulting in a cyt b_{562} variant that catalysed the photoinduced reduction of protons to molecular hydrogen (H_2). Several variants were made with mutations to the porphyrin axial ligand methionine 7, resulting in a variant with 2.6 times the catalytic activity of WT, demonstrating the potential of such engineered protein-porphyrin constructs in biocatalysis. Apo-cyt b_{562} itself has also shown catalytic function, promoting the metalation of an empty protoporphyrin IX with copper (Cu^{2+}) or zinc (Zn^{2+}) ions[173].

The established long-distance conduction properties of cyt b_{562} and its ability to integrate with conductive surfaces in defined orientations make cyt b_{562} an attractive platform as a

molecular electronics component. Cyt b_{562} binds many optically-active porphyrins with relatively high affinity such as ZnPP, an intrinsic fluorophore and photosensitiser, and could form part of porphyrin-based artificial light harvesting systems[174] or light-activated molecular sensors[175]. Single ZnPP molecules have semiconducting current-voltage (I-V) characteristics[176] potentially making cyt b_{562}-like proteins a useful self-assembling scaffold for building arrays of ZnPP-based transistors, especially for use in aqueous solution as porphyrins are poorly soluble. In chapter 5 of this thesis, *in silico* modelling was used to predict the mutations required to switch the cofactor binding specificity of cyt b_{562} from haem to ZnPP.

1.5 Aims and objectives

The aims of this thesis are to use protein engineering techniques to intimately interface natural proteins with non-natural materials or molecules. Chapters 3 and 4 will use rational design, informed by crystallographic models and *in silico* structure prediction tools, to insert the genetically encoded non-natural amino acid AzF into the proteins of interest. The non-natural phenyl azide side chain of AzF provides a bioorthogonal reactive handle for specific, targeted reactions, allowing proteins to be conjugated to CNTs in specific orientations, or labelled with other molecules at a precise surface residue. Chapter 3 will describe CNT-based NTFET sensing devices the functionalised with populations of the high-affinity β-lactamase binding protein BLIP-II, organised in defined orientations via their designed AzF mutation. These proteins will be interfaced with CNTs via SPAAC click chemistry conjugation with linker molecules adsorbed to CNT sidewalls via pyrene groups. This chapter will encompass the design, production and characterisation of the engineered proteins and the construction and testing of the NTFET β-lactamase sensing devices. Chapter 4 will cover the functionalisation of CNTs with populations of BLIP-II, TEM β-lactamase or cytochrome b_{562} proteins via direct UV-activated nitrene cycloaddition to CNT sidewalls via their designed AzF mutations. Protein immobilisation on CNT surfaces will be analysed using data from atomic force microscopy, as well as subsequent formation of BLIP-II – β-lactamase complexes at the CNT surface by protein-protein interactions.

Chapter 5 will report the optimisation of cytochrome b_{562} to bind a ZnPP cofactor instead of its native haem. The chapter will describe the *in silico* design and optimisation of the cyt b_{562}^{ZnPP} variant, made with Adam McGarrity, with its predicted switch in binding specificity from haem to ZnPP. The production and purification of cyt b_{562}^{ZnPP} will be described, and the new physical and optical properties of the variant will be characterised, including its changed binding affinity and new fluorescent properties.

2 Materials & Methods

2.1 Materials

2.1.1 Chemicals

Antibiotics were dissolved in ddH_2O to a 1000× stock concentration, except tetracycline which was dissolved in ethanol to a 200× stock concentration. Antibiotics were supplied by Melford and sterilised using 0.2 µm syringe filters from VWR, Lutterworth, UK. Final working concentrations were: carbenicillin 50 µg/mL, kanamycin 25 µg/mL, chloramphenicol 35 µg/mL, tetracycline 25 µg/mL.

The non-natural amino acid *p*-azido-L-phenylalanine (AzF, Bachem or Click Chemistry Tools) was prepared by dissolution in 1 M NaOH (ca. 4 µL per milligram, Melford) and added to growth media to a working concentration of 1 mM. Care was taken to keep AzF in dark conditions to prevent photolysis of its azide group.

The cephalosporin nitrocefin (Cambridge Biosciences), a chromogenic substrate for β-lactamases, was prepared as concentrated stock freshly each day by dissolution in a minimal volume of dimethyl sulphoxide (DMSO, Sigma-Aldrich) then diluted with ddH_2O as required. Concentration was measured by absorbance in water using a molar extinction coefficient at 390 nm (ε_{390}) of 17,700 $M^{-1}\cdot cm^{-1}$.

The fluorescent dyes used for site-specific labelling of proteins, DBCO-TAMRA and Cy3B-maleimide were dissolved in a minimal volume of DMSO then diluted with ddH_2O as required. Concentration was measured by absorbance: DBCO-TAMRA, ε_{548} = 92,000 $M^{-1}\cdot cm^{-1}$ in water; Cy3B-maleimide, ε_{559} = 130,000 $M^{-1}\cdot cm^{-1}$ in methanol (Thermo Fisher).

Metal porphyrins were prepared as concentrated stocks freshly each day and diluted as required. Haem in the form of haemin (Sigma-Aldrich) was dissolved in 1 M NaOH. Zinc protoporphyrin IX (ZnPP, Sigma-Aldrich) was dissolved in 1 M NaOH or DMSO. Concentration was measured by absorbance: Haem ε_{403} = 170,000 $M^{-1}\cdot cm^{-1}$ in 80% DMSO; ZnPP ε_{425} = 135,000 $M^{-1}\cdot cm^{-1}$ in pyridine (Sigma-Aldrich).

2.1.2 Bacterial strains: *Escherichia coli* genotypes

BL21 (DE3) electrocompetent: Used for expression of BLIP-II, TEM and CTX-M in pET plasmids. Supplied by New England Biolabs (NEB).

E. coli str. B F^- *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$(λ^S)

NEB 5-alpha (DH5α) electrocompetent: Used for archiving of plasmids. Supplied by NEB.

F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ^-

Invitrogen One Shot TOP10 electrocompetent: Used for expression of cytochrome b_{562} in pBAD plasmids. Supplied by Thermo Fisher.

F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G

2.1.3 Bacterial growth media

All growth media was prepared with ddH_2O and sterilised by autoclaving for 15 minutes at > 121 °C. Filter sterilisation used 0.2 µm syringe filters or 0.2 µm 47 mm ⌀ vacuum filters.

For cell recovery after transformation by electroporation, cells were incubated in super optimal broth with catabolite repression (SOC). This consisted of 31 g/L super optimal broth (SOB) powder sterilised by autoclave, supplemented with a final concentration of 20 mM filter-sterilised glucose. Lysogeny broth (LB) liquid growth media was made with 20 g/L LB broth powder LB agar solid growth media was made with 31 g/L LB agar powder. Double yeast extract-tryptone (2×YT) liquid growth media was made with 31 g/L 2×YT powder. All growth media powders were from Melford.

Media for the modified Studier ZYM-5052 arabinose auto-induction protocol[177] consisted of 31% w/v 2×YT powder, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM $MgSO_4$, 4 µM $CaCl_2$, 2 µM $MnCl_2$, 2 µM $ZnSO_4$, 0.4 µM $CoCl_2$, 0.4 µM $CuCl_2$, 0.4 µM $NiCl_2$, 0.4 µM Na_2MoO_4, 0.4µM H_3BO_3, 10µM $FeCl_3$, 0.5% v/v glycerol, 0.05% w/v glucose, 0.2% w/v lactose and 0.05% w/v arabinose. All chemicals from Melford.

2.2 Methods

2.2.1 Generation of 3D structural models of proteins for illustrations

Three dimensional structural models were made representing the engineered proteins in this thesis. These were produced by manually editing the crystal structures sourced from the Protein Data Bank (PDB, www.rcsb.org) using PyMOL molecular graphics software[11] with the SwissSidechain PyMOL Plugin[178] installed to provide a library of nnAA structures including AzF. BLIP-II and TEM β-lactamase variants were based on structure 1JTD[36], CTX-M β-lactamase variants on 1YLT[37] and cyt b_{562} variants on 256B[38]. AzF or cysteine mutations were introduced into structures using the PyMOL mutagenesis tool to replace individual residues at the desired position. The relevant residue was selected, then the command for mutation to AzF was “cmd.get_wizard().set_mode("4BF")” and for cysteine was “cmd.get_wizard().set_mode("CYS")”, then changes were applied. Sequences for protein variants were submitted to the Phyre2 web app[13] (www.sbg.bio.ic.ac.uk/~phyre2) to predict a probable conformation that the mutant residues would take. AzF was substituted by tyrosine for this process as Phyre2 cannot recognise nnAAs.

Protein structures containing optimised AzF mutations were manually docked and conjugated to a carbon nanotube model in PyMOL to make approximate models for illustrative purposes. A 1 nm diameter, 250 nm long CNT model was created in PDB format using Nanotube Modeler software[179] and imported into PyMOL alongside the protein. Protein conjugation to the CNT by UV-nitrene photoaddition was modelled by deleting the terminal two nitrogen atoms of the protein’s AzF phenyl azide (N_3) side chain and joining the remaining nitrogen to a carbon of the CNT graphitic network using the FUSE command to form one bond of the aziridine. The nitrogen was then bonded to a second carbon of the CNT using the BOND command, and the aziridine manually adjusted to approximately the correct bond lengths (ca. 1.5 Å) and angles (ca. 60°)[180]. Models representing protein-CNT conjugation via a DBCO-pyrene linker were constructed using a geometry-optimised model of the AzF-triazole-DBCO-pyrene linker construct provided by Prof Alexander Nemukhin, Moscow State University. The model represents the lowest free energy conformation of the linker. The linker (adsorbed to CNT) was conjugated to the protein by deleting the side chain of AzF then connecting the triazole of the linker to its α-carbon using the FUSE command in PyMOL. For both aziridine and triazole-linker conjugated proteins, position relative to the CNT was manually adjusted by rotating the AzF side chain around the α-β and β-γ carbon bonds until the protein-CNT or protein-linker-CNT complexes were in sterically reasonable positions. This was defined as a position where no steric clashes were apparent between the protein, linker and CNT structures when in van der Waals radius spherical

representation. These conformations were also manually adjusted to suggest a hypothetical rationale for surface electrostatic effects that affected NTFET conductance in chapter 3.3.5.2, and to provide illustrations.

Cytochrome b_{562} dimers were modelled by changing residue 104 (cyt b_{562}^{SA}) or 21 (cyt b_{562}^{LA}) of structure 256B to cysteine using the PyMOL mutagenesis tool. These were optimised using Phyre2 as above. Two copies of these structures were then uploaded to the ClusPro web server[14] (https://cluspro.bu.edu). This was to find likely homodimer conformations by docking simulation, with restriction parameters specified to force contact (< 3 Å) between cysteine residues. The docked protein structures with the largest cluster size and/or lowest free energy score output by ClusPro were used as a basis to make hypothetical dimer models. Cyt b_{562} monomer pairs were merged into a single object in PyMOL and their adjacent cysteine thiols were joined using the BOND command. AzF residues were inserted and conjugated to CNT models as previously. These structures were used to make representative illustrations of the cyt b_{562}^{SA} and cyt b_{562}^{LA} dimers for use in chapter 4, and to assess the possible orientations of cyt b_{562} dimers on a CNT surface.

2.2.2 Molecular biology

2.2.2.1 Non-natural amino acid incorporation plasmids

For expression of proteins containing AzF, cells must contain the cellular machinery needed to incorporate non-natural amino into proteins during protein synthesis by the ribosome in response to a TAG codon. A plasmid carrying the genes for expressing the engineered *Methanocaldococcus jannaschii* tyrosyl-aaRS and $tRNA_{CUA}$ pair must be co-transformed along with plasmid for expressing the target protein of interest. Two such plasmids, carrying the same $tRNA_{CUA}$/aaRS genes, were used for incorporation of AzF (Figure 2.1): 1. pDULE-AzF (also called pDule-pspPhe[181] or pDULE-pCNF/CCF[182], Addgene: 85494), created by the Ryan Mehl lab, which confers tetracycline resistance and constitutively expresses the tRNA and aaRS. 2. pEVOL-AzF (or pEVOL-pAzF[79], Addgene: 31186) from the Peter Schultz lab which confers chloramphenicol resistance, constitutively expresses the tRNA, with the aaRS induced by arabinose under an araBAD promoter. pDULE-AzF was used for expression of AzF-containing cytochrome b_{562} proteins and pEVOL-AzF was used to express AzF-containing BLIP-II or TEM proteins.

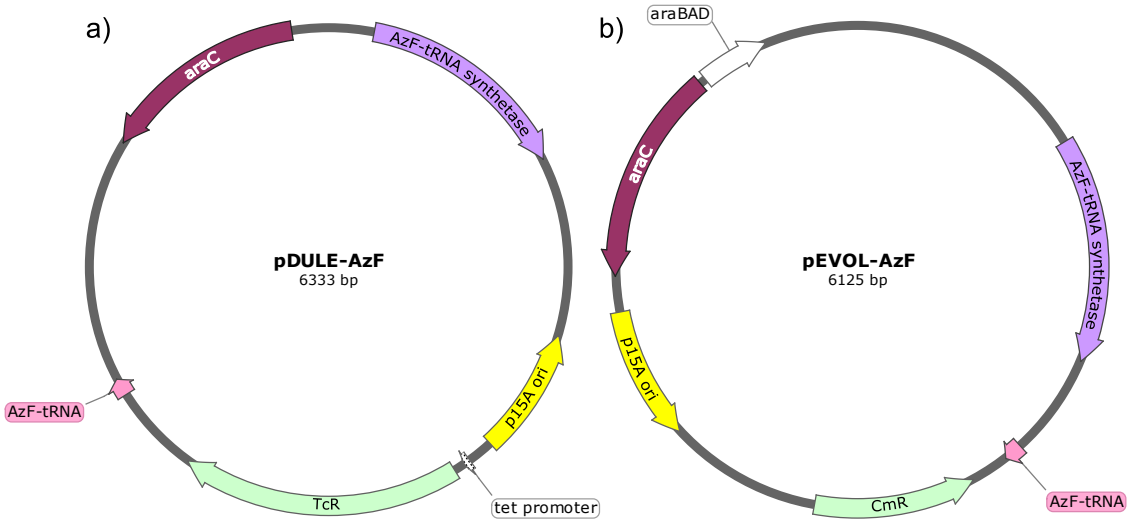


Figure 2.1 Graphic maps of the two plasmids used that encode the bioorthogonal $tRNA_{CUA}$/aaRS genes required for protein incorporation of AzF. Both are based on the p15A origin of replication (yellow). a) pDULE-AzF confers tetracycline resistance (green) and constitutively expresses the tRNA (pink) and aaRS (violet). b) pEVOL-AzF confers chloramphenicol resistance (green), constitutively expresses the tRNA (pink), with the aaRS (violet) induced by arabinose under an araBAD promoter (white). Made using Snapgene software[183].

2.2.2.2 Recombinant protein expression plasmids

Cytochrome b_{562} proteins were expressed using a pBAD plasmid containing variants of the cyt b_{562} gene (Figure 2.2a). The original plasmid was constructed by Dr. Samuel Reddington, who also made the genes for the cysteine and AzF-containing mutants cyt b_{562}^{SA} and cyt b_{562}^{LA}. The cyt b_{562}^{ZnPP} variant protein used in chapter 5 was also based on pBAD-cyt b_{562}. Expression of the cyt b_{562} gene in pBAD-cyt b_{562} is induced by arabinose and the plasmid confers ampicillin (or carbenicillin) resistance. BLIP-II and TEM proteins were expressed using pET24a vectors containing the relevant genes. CTX-M was expressed using a pET28 vector containing the CTX-M gene. The pET-BLIP-II[154] and pET-CTXM14[184] plasmids were made by the Timothy Palzkill lab, pET-TEM1[185] was made by the Marvin Makinen lab. Expression of these proteins is induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) and they confer kanamycin resistance. Expression of AzF-containing variants of these proteins requires co-transformation with either pDULE-AzF or pEVOL-AzF (2.2.2.1) to incorporate AzF in the protein. Plasmid maps are shown in Figure 2.2.

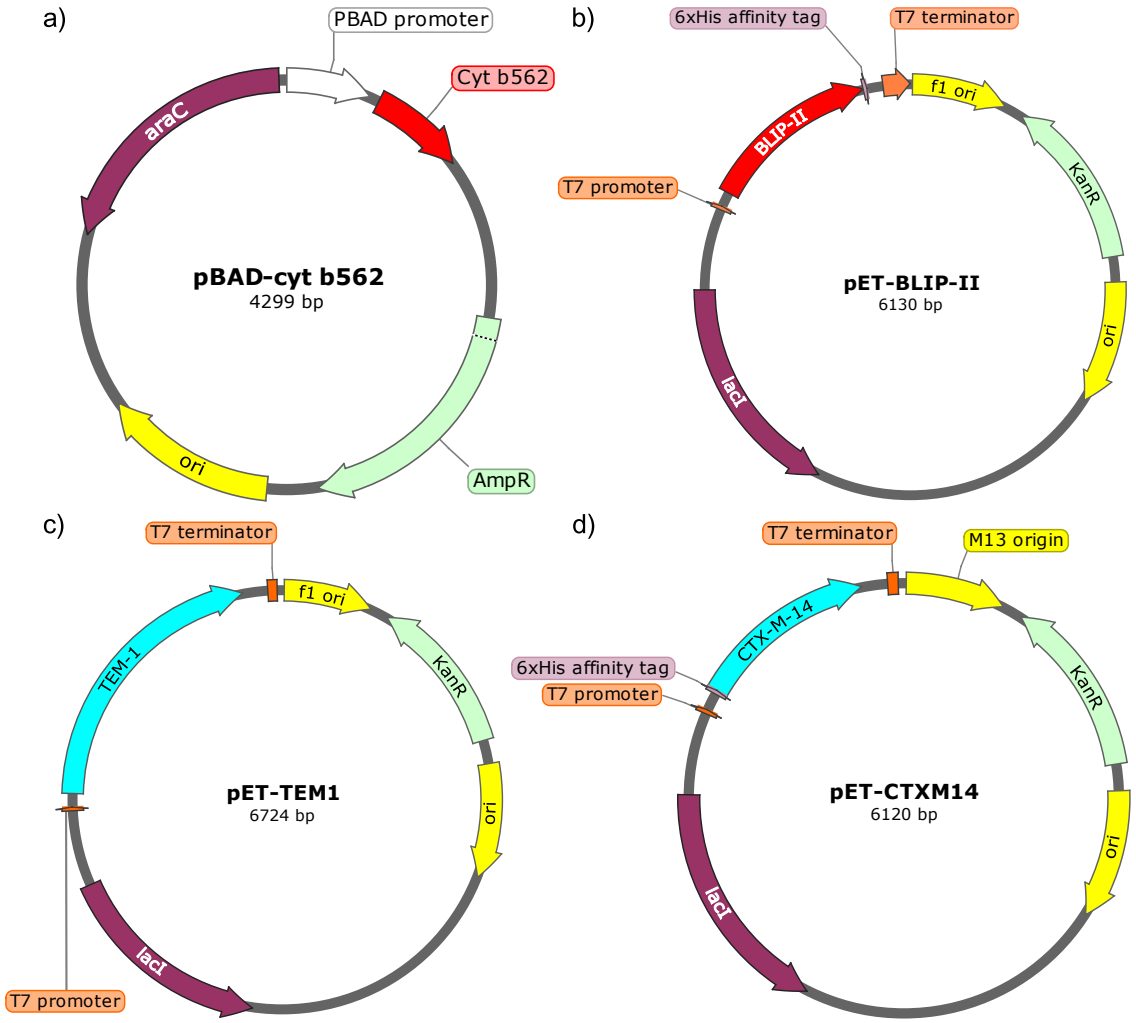


Figure 2.2 Graphic maps of the expression plasmids used in this thesis. a) pBAD-cyt b_{562} confers ampicillin resistance (green), with cyt b_{562} (red) expression induced by arabinose under a pBAD promoter (white). b) pET-BLIP-II confers kanamycin resistance (green), with BLIP-II (red) expression induced by IPTG under a T7 promoter (orange). BLIP-II has a C-terminal hexahistidine tag (His-tag, pink) for purification by immobilised metal affinity chromatography (IMAC). c) pET-TEM1 confers kanamycin resistance (green), with TEM (blue) expression induced by IPTG under a T7 promoter (orange). d) pET-CTXM14 confers kanamycin resistance (green), with CTX-M (blue) expression induced by IPTG under a T7 promoter (orange). CTX-M has an N-terminal His-tag (pink) for purification by IMAC. Made using Snapgene software[183].

2.2.2.3 DNA oligonucleotides

Oligonucleotide primers for site-directed mutagenesis were designed using Serial Cloner[186] software and optimised and assessed with Amplifx[187] software and are shown in Table 2.1. Primers were supplied by Integrated DNA Technologies, except T7/T7 terminator sequencing primers by Eurofins Genomics.

Table 2.1 Names and sequences of oligonucleotide primers used for sequencing, colony PCR or PCR mutagenesis. Mutations introduced are highlighted in red.

pET plasmids sequencing	**Name**	**Sequence (mutations in red)**
Forward	IDT: T7 Promoter	TAA TAC GAC TCA CTA TAG GG
Reverse	IDT: T7 Terminator	GCT AGT TAT TGC TCA GCG G
Forward	Eurofins: T7	TAA TAC GAC TCA CTA TAG GG
Reverse	Eurofins: T7term	CTA GTT ATT GCT CAG CGG T
pBAD plasmid sequencing		
Forward	SRpBADSqF*	ATG CCA TAG CAT TTT TAT CC
Reverse	SRpBADSqR*	GAT TTA ATC TGT ATC AGG
pET-BLIP-II		
A41TAG mutation		
Forward	B2iPalz_A41TAG_F	TAT ACA TAT GTA GAC CTC GGT GGT G
Reverse	B2iPalz_A41_S43_TAG_R	TCT CCT TCT TAA AGT TAA ACA AAA TTA TTT CTA
S43TAG mutation		
Forward	B2iPalz_S43TAG_F	TAT ACA TAT GGC AAC CTA GGT GG
Reverse	B2iPalz_A41_S43_TAG_R	TCT CCT TCT TAA AGT TAA ACA AAA TTA TTT CTA
G49TAG mutation		
Forward	B2iPalz_G49TAG_F	CCT GGG GCT AGA ACA ATG ACT G
Reverse	B2iPalz_G49_N50_TAG_R	CCA CCA CCG AGG TTG CCA TA
T213TAG mutation		
Forward	B2iPalz_T213TAG_F	TTG GCC AGA CCT AGG TGC
Reverse	B2iPalz_T213TAG_R	AAT AAT TAT CCC CCC ACG CAA TAA CAC
pET-TEM1		
G87TAG mutation		
Forward	HS&PalzTEM_G87TAG_F	CGT GTT GAC GCC TAG CAA GAG CAA CT
Reverse	HS&PalzTEM_G87TAG_R	GGA TAA TAC CGC GCC ACA TAG CAG AAC TT
pET-CTXM14		
T86C mutation		
Forward	CTX_T86mut_F	AGC TGC TTA ATC AGC CTG TCG AGA
Reverse	CTX_T86TGC_R	GCT TTT GGC ATT CAC TCT GCT TAA GCA C
pBAD-cyt b_{562}		
D5TAG mutation		
Forward	SRc5TAGF*	ACC CTC AAC GAC AAT TTA AAA GTG ATC G
Reverse	SRc5TAGR*	TTC CAT ATT CTA TTC AAG ATC TGC GG
K104C mutation		
Forward	SRc104CYSF*	CCT ATC ACC AGT GTT ATC GTT
Reverse	SRc104CYSR*	CGT TGC GGG TCG TTT TCA GTT G
D21C mutation		
Forward	SRc21CYSF*	GTG ATC GAA AAA GCG TGT AAC GCG GCG CAA GTC
Reverse	SRc21CYSR*	CTT TGG GAG TTG CTG TTA AAT TTT

D50TAG mutation		
Forward	SRc50TAGF*	ACA GCC CGG AAA TGA AAG ATT TC
Reverse	SRc50TAGR*	CCG GTG ATT TCT ATT CGA GCT TCG
ZnPP mutations		
Forward (short)	pbAD WT b562 > 6x vf**	CAA CGA CAA TTT AAA AGT GAT C
Reverse (tailed)	pbAD WT b562 > 6x ovr**	CGG CAT TCA CAA TTG TCG TCC TTA TCT GCC ATG GTT AAT TCC
Forward (tailed)	pbAD WT b562 > 6x ovf**	AAG GAC GAC AAT TGT GAA TGC CGC AAC GAC AAT TTA AAA GTG ATC
Reverse (short)	pbAD WT b562 > 6x vr**	ATC TGC CAT GGT TAA TTC C

2.2.2.4 Agarose gel electrophoresis

DNA fragment sizes were analysed by agarose gel electrophoresis using 1% (w/v) agarose (Melford) in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.8) with 100 ng/mL ethidium bromide to visualise DNA. Electrophoresis was performed at 120V for ca. 30 min. DNA band size was compared to 1 kb DNA ladder or 2-Log DNA ladder (both NEB). Imaging was performed in a GelDoc-It UV transilluminator (Ultra-150 Violet Products, Cambridge, UK).

2.2.2.5 Site-directed mutagenesis by whole plasmid PCR and ligation

PCR mutagenesis was carried out in a 50 µL reaction volume using 1 U Q5 DNA Polymerase in Q5 reaction buffer with 200 µM dNTPs (all NEB), 0.5 µM forward and 0.5 µM reverse primers and 0.2 ng plasmid template DNA, in nuclease free water (Melford). Thermocycling was carried out using a Techne Flexigene thermal cycler with the following steps:

1. Initial denaturation: 98°C, 30 seconds
2. Denaturation: 98°C, 10 seconds
   Annealing: Specific temp., 30 seconds
   Elongation: 72°C, 30 seconds per kb
   } 30 cycles
3. Final extension: 72°C 2 minutes
   Final hold: 4°C

PCR products were analysed using agarose gel electrophoresis (2.2.2.4) and reactions that generated DNA at the expected sizes were purified with a QIAquick PCR purification kit (QIAGEN, Venlo, Netherlands). Purified linear plasmid (5 µL) was then end-phosphorylated by adding an equal volume of Quick Ligase buffer (QLB, NEB) and 10 U T4 polynucleotide kinase (T4 PNK, NEB) and incubating for 30 minutes at 37°C. After incubation, 2 µL QLB (to replenish depleted ATP) and 1 µL Quick Ligase (concentration not stated, NEB) were added for five minutes at room temperature, to re-circularise the

linear plasmid by ligating its blunt ends. Ligated circular plasmid was then purified with a MinElute Reaction Cleanup Kit (QIAGEN), eluting in 12 µL volume, and 5 µL used for transformation of *E. coli* DH5α cells.

2.2.2.6 Site-directed mutagenesis using ligase-free SLIM method

Amino acid Mutations required for cyt b_{562}^{ZnPP} were made according the Site-directed, Ligase-Independent Mutagenesis (SLIM) protocol[188]. The SLIM protocol uses two pairs of PCR primers in two separate reactions to amplify the two opposite DNA strands of a plasmid with long, complementary single stranded overhangs (“tails”) containing the desired mutations. Thus, when combined, the opposite strands can re-anneal to each other and spontaneously re-circularise via their sticky ends without ligase to reconstitute the original plasmid, incorporating the mutations included in the primers as shown in Table 2.1. Figure 2.3 shows the SLIM mutagenesis process. The re-annealed plasmid contains strand breaks (nicks) but this is repaired after transformation by host *E. coli* DNA ligases[189].

Two primer pairs were used to introduce the cyt b_{562}^{ZnPP} mutations into pBAD-cyt b_{562} in one step, as all mutations required for cyt b_{562}^{ZnPP} are within eight residues at the protein’s N-terminus. Primer pairs consist of standard primers called "short" (S), and long primers containing the mutations and the sticky end overhang "tail" for recircularisation (T). PCR amplification/mutagenesis was as described above (2.2.2.5). Amplification of PCR products at the expected sizes was confirmed using agarose gel electrophoresis (2.2.2.4).

Amplified PCR product DNA (12.5 µL) were added to 10 µL buffer-D concentrate (20 mM $MgCl_2$, 20 mM Tris-HCl (pH 8) and 5 mM DTT) with 20 U DpnI restriction enzyme (NEB) and incubated at 37°C to digest methylated template DNA. After one hour 30 µL buffer-H concentrate (300 mM NaCl, 50 mM Tris-HCl pH 9, 20 mM EDTA pH 8) was added and the mixture boiled for 3 min to denature DpnI and stop digestion. The PCR products were then hybridised by two cycles of 5 min at 65°C then 15 min at 30°C to allow hybridisation and non-covalent recircularisation via their sticky end overhangs. Circular plasmid was then purified using a MinElute Reaction Cleanup Kit, eluting in 12 µL volume, and 5 µL used for transformation of *E. coli* DH5α cells.

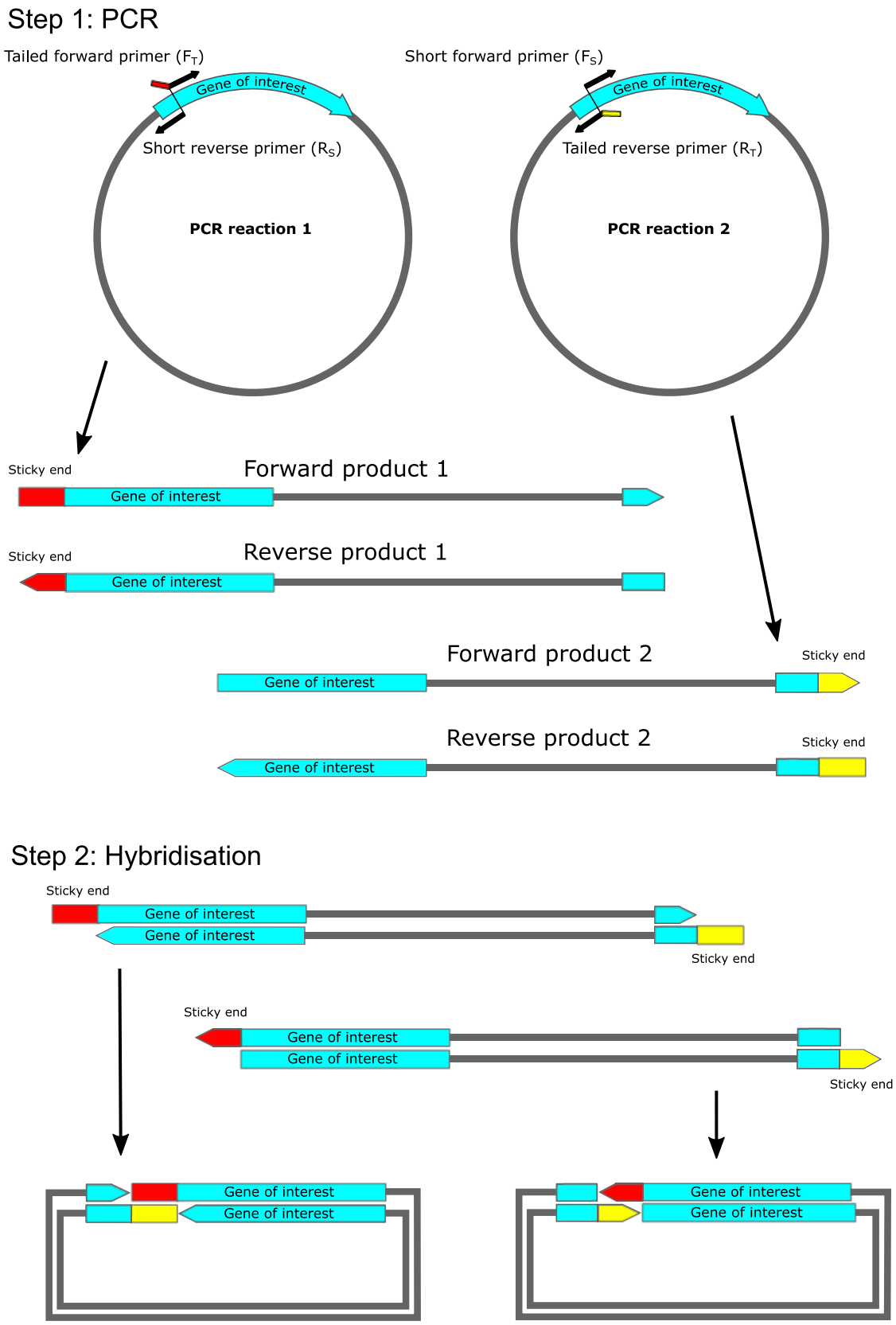


Figure 2.3 - Site-directed mutagenesis by SLIM method. **Step 1**: Two PCR reactions using primer pairs incorporating "tails" to produce products with complementary sticky end overhangs containing the desired mutations. **Step 2**: Both amplification products are mixed then melted and re-hybridised, spontaneously re-circularising via overhangs into nicked plasmids that can be used directly for transformation. The long-tailed SLIM primers enable insertions, deletions or incorporation of several amino acid mutations at once, if they are sufficiently close together.

2.2.2.7 Transformation and selection of electrocompetent *E. coli* cells

Recombinant plasmids were transformed into electrocompetent *E. coli* DH5α cells by electroporation. Resuspended competent cells (50 µL) were incubated with 5 µL ligated and purified plasmid on ice for 30 minutes, then placed into a chilled Gene Pulser/MicroPulser Electroporation Cuvette (0.2 cm gap, Bio-Rad, Watford, UK) and subject to a 5 ms pulse at 2.5 kV/cm with a Celljet electroporator (Eurogentec, Seraing, Belgium). The cells were resuspended in 450 µL SOC media and incubated for 1-3 hours at 37 °C with shaking, to recover. Cells were then spread on petri dishes containing LB agar supplemented with the relevant selection antibiotic, allowed to dry, then incubated overnight at 37 °C.

Individual bacterial colonies that grew on selective antibiotic agar were picked and used to inoculate several 10 mL LB broth cultures containing the same selection antibiotic and grown at 37 °C overnight to saturation. A 0.5 mL sample of each culture was archived at -80 °C in 25% v/v glycerol, and the rest centrifuged, the bacterial pellet lysed and plasmid DNA isolated using QIAprep Spin Miniprep kits (QIAGEN). Plasmid DNA from several different colonies was sent for Sanger sequencing (Eurofins) using the appropriate sequencing primers. Plasmid sequences were verified by alignment to the wild-type gene using Serial Cloner and those with the expected mutations were used for experiments, while plasmids with erroneous mutations were discarded.

2.2.3 Fast protein liquid chromatography (FPLC)

Protein purification by fast protein liquid chromatography (FPLC) was performed with a variety of FPLC columns on ÄKTApurifier, ÄKTAprime plus or ÄKTA pure systems (all Cytiva, Amersham, UK). Protein elution from columns was monitored by absorbance at 280 nm for most proteins, and at 418 nm for cyt b_{562}, its haem absorbance wavelength. Only the ÄKTApurifier system was able to monitor absorbance at wavelengths other than 280 nm. UV-visible absorbance monitoring was disabled during elution of AzF-containing proteins, to prevent photolysis of the azide group.

Protein samples were concentrated as required using Vivaspin 10 kDa or 30 kDa centrifugal concentrators (Sartorius, Göttingen, Germany), centrifuged at 2,800 ×*g*. Samples were desalted by buffer exchange using Sephadex G-25 PD10 gravity columns (Cytiva).

2.2.4 Recombinant cytochrome b_{562} production in *E. coli*

2.2.4.1 Cytochrome b_{562} expression from pBAD plasmids

To express cyt b_{562}^{WT} or cyt b_{562}^{ZnPP} using variants of the pBAD-cyt b_{562} expression plasmid, the plasmid was transformed into electrocompetent *E. coli* TOP10 cells by electroporation, as described in 2.2.2.7, using ca. 100 ng plasmid DNA. Transformed cells were spread on LB agar supplemented with carbenicillin (50 µg/mL) for selection of transformants. A single transformant colony was picked and used to inoculate a batch culture (100 mL – 2 L) of modified Studier autoinduction media (with carbenicillin) and grown to saturation overnight at 37 °C with shaking.

To express AzF-containing cyt b_{562} variants, the pBAD-cyt b_{562} plasmid containing TAG codon mutations at the desired location were co-transformed into *E. coli* TOP10 along with the pDULE-AzF plasmid that carries the engineered $tRNA_{CUA}$/aaRS pair required for AzF incorporation. Approx. 100 ng of both plasmids was incubated with cells and transformed as above. Cells were spread on LB agar containing carbenicillin (50 µg/mL) and tetracycline (25 µg/mL) to select for cells with both plasmids. Individual colonies were used to inoculate batch cultures (100 – 500 mL) as above but with carbenicillin, tetracycline and the nnAA AzF (1 mM) included the autoinduction media. These were grown to saturation overnight as previously described, but in the dark to protect the phenyl azide group from photolysis.

Cells were harvested by centrifugation at 5000 ×*g* for 15 minutes at 4 °C with a JLA 16.250 rotor in an Avanti J-E centrifuge (Beckman Coulter, High Wycombe, UK). Expression of cyt b_{562} is evident from the pink-red colour of the cell pellet, except for cyt b_{562}^{ZnPP}, which is colourless. Expression of cyt b_{562} was analysed using SDS-PAGE.

2.2.4.2 Bacterial lysis and lysate clarification

Cell pellets containing cytoplasmically expressed cyt b_{562} were resuspended in ca. 15 mL Tris buffer (50 mM Tris-HCl, pH 8) at 4 °C then lysed using a chilled French pressure cell at ca. 20,000 psi to disrupt the cells. The sample was centrifuged at 25,000 ×*g* for 15 minutes at 4 °C with a JA 25.50 rotor in an Avanti J-E centrifuge (Beckman Coulter) to remove insoluble cell debris and clarify the lysate.

2.2.4.3 Cytochrome b_{562} purification by ammonium sulphate precipitation

The cyt b_{562} gene in pBAD-cyt b_{562} lacks a polyhistidine affinity tag for IMAC purification so cell lysate containing cyt b_{562} was subject to partial purification by ammonium sulphate precipitation. This method exploits the high solubility of cyt b_{562} by dissolving high concentrations of ammonium sulphate ($(NH_4)_2SO_4$) into the lysate to precipitate other

proteins out of solution. Cyt b_{562} is one of relatively few proteins to remain soluble at very high salt concentrations.

Ammonium sulphate (30% w/v) was dissolved into cell lysate containing cyt b_{562} while stirring on ice for 10 minutes. The lysate was centrifuged at 25,000 ×*g* for 10 minutes at 4°C to sediment precipitated proteins, which were discarded. The supernatant ammonium sulphate concentration was then increased to 90% w/v, stirred on ice for 10 minutes, and centrifuged. The resulting protein pellet containing cyt b_{562} was red in colour, except for the colourless variant cyt b_{562}^{ZnPP}. This was re-dissolved in the minimum volume of Tris buffer (50 mM Tris-HCl, pH 8) for further purification.

2.2.4.4 Cytochrome b_{562} purification by ion exchange chromatography

Cyt b_{562} was purified using ion exchange chromatography (IEC). Proteins were thoroughly desalted using a PD10 column. Cyt b_{562} is negatively charged at pH 8, with isoelectric points (pI) of 6.5. Protein was flowed through a 5 mL HiTrap Q anion exchange column (Cytiva) to bind to the column resin. The column was washed with 5 column volumes of Tris buffer (50 mM Tris-HCl, pH 8) at 5 mL/min then bound protein was eluted with a 0-1 M sodium chloride (NaCl) gradient over 30 column volumes at 5 mL/min, collecting 10 mL fractions. Fractions were analysed by SDS-PAGE.

2.2.4.5 Cytochrome b_{562}purification by size exclusion chromatography

Cyt b_{562} variants were purified by size exclusion chromatography (SEC) using an FPLC system (2.2.3) and one of three Cytiva SEC columns, depending on protein size and sample volume. All columns were equilibrated in Tris buffer (50 mM Tris-HCl, pH 8). The columns were:

- HiLoad 16/600 Superdex 75 prep grade, 120 mL bed volume. Maximum sample volume 2 mL, flow rate 1.6 mL/min (equilibration) or 1 mL/min (purification). Maximum pressure 0.5 MPa. Separation range 3 kDa – 75 kDa. (S75 column)
- HiLoad 26/600 Superdex 200 prep grade, 320 mL bed volume. Maximum sample volume 5 mL, flow rate 4 mL/min (equilibration) or 1.6 mL/min (purification). Maximum pressure 0.5 MPa. Separation range 10 kDa – 600 kDa. (S200 column).

At least two rounds of SEC were required to purify cyt b_{562} following ammonium sulphate precipitation, first using the 320 mL 26/600 S200 column, then concentrating the eluted cyt b_{562} sample and re-purifying it with the 120 mL 16/600 S75 column. The red colour of cyt b_{562} allows the protein's progress through an SEC column to be monitored by eye to begin fraction collection. Eluted fractions were analysed by SDS-PAGE.

2.2.4.6 Haem cofactor extraction from cytochrome b_{562}

Haem was extracted from cyt b_{562} holo-protein by adjusting the pH of the sample to pH 2 to dissociate the haem from cyt b_{562}, yielding apo-protein[190]. The pH was adjusted by gradually adding 1M HCl, monitoring sample pH with ca. 2 mm thin strips of Whatman pH indicator paper (Cytiva). Free haem was then removed from the sample by organic extraction: an equal volume of ice-cold butanone (methyl ethyl ketone) was added and the sample was shaken thoroughly for 30 seconds, left to settle for 1 minute, then centrifuged at 16,000 ×*g* for 2 minutes in a Sigma 1-14 Microfuge (Sigma-Aldrich) to settle into two immiscible phases. Haem dissolved into the butanone upper phase, which was immediately removed and discarded. The aqueous lower phase containing cyt b_{562} apo-protein was then immediately exchanged into Tris buffer (50 mM Tris-HCl, pH 8) by gel filtration using a PD10 desalting column to remove residual butanone and return the pH to 8.

2.2.5 Recombinant BLIP-II production in *E. coli*

2.2.5.1 BLIP-II expression from pET plasmids

For proteins without AzF residues, *E. coli* BL21 (DE3) cells (NEB) were transformed with ca. 100 ng pET-BLIP-II plasmid by electroporation, then grown on LB agar with 30 µg/mL kanamycin for transformant selection. One colony was picked and used to inoculate a 2×YT batch culture (containing kanamycin) at 37 °C until reaching an optical density at λ = 600 nm (OD_{600}) of 0.4. IPTG was added to 1 mM to induce protein expression and the cells grown for a further 24 hours at 16 °C with 200rpm shaking.

Expression of AzF-containing proteins was as above, except the pET-BLIP-II expression plasmid was co-transformed along with the pEVOL-AzF plasmid (ca. 100 ng) that encodes the $tRNA_{CUA}$/aaRS pair for AzF incorporation. To select for co-transformants, agar and growth media were supplemented with 30 µg/mL kanamycin and 35 µg/mL chloramphenicol. At induction, growth media was supplemented with 1 mM IPTG, 0.02% (w/v) arabinose (for pEVOL-AzF induction) and 1 mM AzF. From induction, cultures were grown at or 16 °C for 24 hours, in the dark to protect the azide of AzF. Cells were harvested by centrifugation as previously and analysed by SDS-PAGE.

2.2.5.2 Bacterial lysis and lysate clarification

As cyt b_{562} in section 2.2.4.2.

2.2.5.3 BLIP-II purification by immobilised metal affinity chromatography

BLIP-II was purified by cobalt IMAC, either by FPLC or using a gravity-flow IMAC column. The BLIP-II construct had a C-terminal hexahistidine affinity tag (see Figure 2.2). Both column types were pre-equilibrated with binding buffer (50 mM Tris-HCl, 5 mM imidazole, pH 8) with a low concentration of imidazole to minimise non-specific binding of proteins to the cobalt IMAC resin.

Clarified, filtered cell lysate was passed through a HiTrap TALON 5 mL cobalt IMAC column (Cytiva), using an ÄKTA FPLC system (Cytiva) (2.2.3) at 2 mL/min, allowing affinity-tagged proteins to bind to the cobalt resin. The column was then washed with 5 column volumes binding buffer at 5 mL/min. Bound protein was eluted with a 0-100 mM imidazole gradient over 10 column volumes at 5 mL/min, with 5 mL elution fractions collected. Fractions were analysed by SDS-PAGE.

A simplified protocol was devised to purify polyhistidine-tagged proteins by cobalt affinity chromatography without using an FPLC system. This purification was carried out in a closed cabinet to keep the phenyl azide proteins in the dark and prevent photolysis of the light-sensitive azide group. A gravity flow column was filled with 5 mL HisPur cobalt resin (Thermo Fisher) and equilibrated with binding buffer. Clarified cell lysates were then flowed through the column and the column washed with 20 mL binding buffer. Bound proteins were eluted in one step with 20 mL elution buffer (Tris 50 mM, imidazole 100 mM, pH 8), then exchanged into Tris buffer (50 mM Tris-HCl, pH 8) using a PD10 desalting column. Fractions were analysed by SDS-PAGE.

2.2.5.4 BLIP-II purification by size exclusion chromatography

As cyt b_{562} in section 2.2.4.5 except elution monitored at λ_{abs} = 280 nm (S75 column).

2.2.6 Recombinant CTX-M β-lactamase production in *E. coli*

2.2.6.1 CTX-M expression from pET plasmids

As BLIP-II in section 2.2.5.1 except grown at room temperature overnight.

2.2.6.2 Bacterial lysis and lysate clarification

As cyt b_{562} in section 2.2.4.2.

2.2.6.3 CTX-M purification by immobilised metal affinity chromatography

As BLIP-II in section 2.2.5.3.

2.2.6.4 CTX-M purification by size exclusion chromatography

As cyt b_{562} in section 2.2.4.5 except elution monitored at λ_{abs} = 280 nm (S200 column).

2.2.7 Recombinant TEM β-lactamase production in *E. coli*

2.2.7.1 TEM expression from pET plasmids

As BLIP-II in section 2.2.5.1 except grown at room temperature overnight.

2.2.7.2 Periplasmic extraction of TEM

The TEM β-lactamase gene in pET-TEM1 has an N-terminal ompA signal sequence, so is periplasmically exported. It also lacks a polyhistidine tag for immobilised metal affinity chromatography (IMAC), so cells containing TEM were subject to periplasmic extraction by cold osmotic shock[191,192] to harvest only periplasmic proteins. Cell pellets were resuspended in 1/20 the original culture media volume of buffer-1 (20% w/v sucrose, 50 mM Tris-HCl, 1 mM sodium EDTA, pH 8) and rotated at room temperature for 10 minutes. Cells were centrifuged at 10,000 ×*g* for 10 minutes at 4 °C with a JA 25.50 rotor in an Avanti J-E centrifuge. The supernatant was discarded, and the cells resuspended in the same volume of buffer-2 (5 mM $MgCl_2$, ice-cold) and stirred gently on ice for 10 minutes. The sample was centrifuged again leaving the periplasmic contents including TEM in the supernatant fraction.

2.2.7.3 TEM purification by ion exchange chromatography

As cyt b_{562} in section 2.2.4.4.

2.2.7.4 TEM purification by size exclusion chromatography

As cyt b_{562} in section 2.2.4.5 except elution monitored at λ_{abs} = 280 nm (S75 column).

2.2.8 Protein analysis

2.2.8.1 SDS-PAGE and native PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out as per Laemmli *et al.* (1970)[193] to analyse protein expression, purification or modification. The mini-PROTEAN 3 gel electrophoresis system (Bio-Rad) was used with 0.75 mm thickness gels as described in Table 2.2. The acrylamide proportion of resolving gels was varied depending on target protein size, with a 12.5% gel used for BLIP-II, TEM and CTX-M (all ca. 30 kDa) and a 15% gel used to resolve cyt b_{562} (12 kDa).

Protein samples were prepared for analysis by adding sample loading buffer concentrate to a final concentration of 2% w/v SDS, 200 mM Tris-HCl pH 6.8, 0.04% w/v bromophenol

blue dye, 8% v/v glycerol and 10% v/v β-mercaptoethanol. Whole cell contents were prepared for analysis by centrifugation of cell culture at 16,000 ×*g* for 5 minutes, then resuspension of the cell pellet in sample loading buffer (composition as above) to an OD_{600} of 10. Samples were boiled in a water bath on a hot plate at 100 °C for 10 minutes then centrifuged at 16,000 ×*g* for 1 minute to sediment debris and aggregates.

Non-reducing SDS PAGE was performed as above, but with β-mercaptoethanol omitted from the sample buffer. BLUeye Pre-Stained Protein Ladder (Geneflow) was used for comparison of band sizes.

SDS-PAGE was carried out in electrophoresis buffer consisting of 25 mM Tris-HCl, 200 mM glycine and 0.1% (w/v) SDS, pH 8.3. Gels were run at 250 V, ca. 50 mA for ca. 30 minutes. Protein bands were visualised with UV light (for fluorescent labelled proteins) and/or stained with Coomassie blue stain (40% v/v methanol, 10% v/v glacial acetic acid and 0.15% w/v Coomassie blue R-250 dye). Gels were destained with either boiling water or destain solution (40% v/v methanol and 10% v/v glacial acetic acid).

Native PAGE is a similar method, that does not denature proteins and preserves complexes formed by non-covalent interactions. The preparation is as SDS-PAGE above but omits SDS and β-mercaptoethanol from all buffers and the sample is not boiled.

Table 2.2 Composition of SDS-PAGE resolving and stacking gels used.

	Resolving gel (cyt b_{562})	Resolving gel (other proteins)	Stacking gel
Acrylamide/bis-Acrylamide 37.5:1 (v/v)	15%	12.5%	5%
Tris-HCl* (mM)	375	375	65
Sodium dodecyl sulphate (SDS) (w/v)	0.1%	0.1%	0.2%
Ammonium persulphate (APS) (w/v)	0.05%	0.05%	0.1%
Tetramethylethylenediamine (TEMED) (v/v)	0.02 %	0.02 %	0.02 %

* Resolving gels made with 1.5 M Tris-HCl pH 8.8, stacking gel with 0.5 M Tris-HCl pH 6.8.

2.2.8.2 Determining protein concentration using Bio-Rad DC assay

Concentrations of new proteins or those without a known molar extinction coefficient (ε) were determined using a DC protein assay (Bio-Rad). The assay was calibrated using ten concentrations of bovine serum albumin (BSA, Melford) in a 96-well plate, to make a standard curve. Aliquots of purified protein of unknown concentration were used at 1×, 2×, 4×, 8× and 10× dilution. DC assay reagents were added as per the instructions and mixed on a shaker plate at room temperature for 30 minutes. Absorbance values at 750 nm were recorded for all samples using a CLARIOstar plate reader (BMG LABTECH) and plotted in MS Excel to calculate the protein concentration.

2.2.8.3 UV-visible absorbance spectroscopy

All absorbance readings and spectra were recorded on a Cary 60 UV-visible spectrophotometer (Agilent), using disposable polystyrene cuvettes (Starlab) for cell density OD_{600} readings and quartz cuvettes (Hellma) for spectral work. Spectra were recorded at a scan rate of 600 nm/min, usually from 200-700 nm.

The concentration of proteins or chemicals containing chromophores (e.g. porphyrins, dyes or chromogenic substrates) with known molar extinction coefficient (ε) were calculated using absorbance values measured by UV-vis, using a rearranged Beer-Lambert law equation (Equation 2.1).

Equation 2.1 Beer-Lambert law equation[194]

$A = \varepsilon c l$, rearranged to $c = \frac{A \cdot l}{\varepsilon}$

A = Absorbance (AU)
ε = Molar extinction coefficient ($M^{-1} \cdot cm^{-1}$)
c = Concentration (M)
l = Path length (cm)

2.2.8.4 Fluorescence spectroscopy

All fluorescence spectra were recorded on a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent) using 101.057-QS quartz cuvettes (Hellma), path length 5 × 5 mm. Spectra were recorded at a scan rate of 600 nm/min with an excitation/emission slit widths of 5 or 10 nm and photomultiplier voltage set to medium or high. Emission spectra were recorded with a fixed excitation wavelength near the ZnPP excitation maximum (λ_{ex} ~ 425 nm) from 500 nm to 700 nm at 1 nm intervals. Excitation spectra were recorded by monitoring a fixed emission wavelength near the ZnPP emission maximum (λ_{em} ~ 595 nm) with excitation from 300 nm to 500 nm at 1 nm intervals. To measure whole cell fluorescence, cells were centrifuged at 10,000 ×*g* for 10 minutes, then resuspended to OD_{600} 0.2 in Tris buffer (50 mM Tris-HCl, pH 8).

2.2.8.5 Enzyme kinetics - TEM^{WT}

The kinetic parameters k_{cat} and K_M of TEM^{WT} were determined by measuring the initial rate of substrate hydrolysis by 0.3 nM TEM at a range of substrate concentrations, then fitting the results to the Michaelis-Menten equation[195] (Equation 2.2) using GraphPad Prism software[196]. The substrate used was the chromogenic cephalosporin nitrocefin, which undergoes a shift in absorbance maximum (λ_{max}) on hydrolysis by β-lactamases, from 390 nm to 486 nm. Initial rate was measured as the increase in sample absorbance

at 486 nm over the first 30 seconds of the reaction, using 100 µL volume in a 105.250-QS quartz cuvette (Hellma), path length 1 cm, in a Cary 60 UV-vis (Agilent). Nitrocefin hydrolysis initial rates were measured between 0 – 400 µM.

Equation 2.2 Michaelis-Menten equation.

$$v = \frac{V_{max}[S]}{K_M + [S]}$$

v = Initial velocity of substrate hydrolysis (µM·second^{-1})
V_{max} = Maximum velocity of substrate hydrolysis (µM·second^{-1})
[S] = Substrate concentration (µM)
K_M = Michaelis constant (µM)

2.2.8.6 Inhibition assays - β-lactamase-BLIP-II

Assays to measure β-lactamase enzyme inhibition by BLIP-II proteins were performed as described in Brown and Palzkill (2010)[154]. Final enzyme concentrations used were TEM: 0.3 nM, CTX-M: 1 nM, in 50 mM sodium phosphate buffer (pH 7) with 0.1 mg/mL BSA (to minimise nonspecific binding). A range of increasing concentrations of BLIP-II were added to the enzyme and equilibrated for 30 minutes at room temperature. Binding and inhibition of β-lactamase by BLIP-II was measured by recording the initial rate (30 seconds) of nitrocefin hydrolysis (final concentration 40 µM) using 100 µL volume in a 105.250-QS quartz cuvette (Hellma), path length 1 cm, in a Cary 60 UV-vis spectrophotometer (Agilent). The resulting rates were plotted against BLIP-II concentration and fitted to the Morrison tight binding equation[197,198] (Equation 2.3) using GraphPad Prism software[196] to determine the apparent inhibition constant (K_i^{app}) of the BLIP-II variants. This method cannot not reveal the true K_i of the interaction due to its extremely slow dissociation rate (half-life of BLIP-II – TEM: ca. 3 days, BLIP-II – CTX-M: ca. 5 days) but was judged sufficient for comparing BLIP-II variants to BLIP-II^{WT}. Experiments measuring the true K_i (0.5 pM for TEM or CTX-M with BLIP-II^{WT}) can take as long as a week to perform[147].

Equation 2.3 Morrison tight binding equation

$$\mathrm{Q} = K_i \times \left[1 + \left(\frac{[S]}{K_M}\right)\right]$$

$$\mathrm{V} = V_0 \times \left(1 - \left\{\frac{\left[([E]_t + [I]_t + Q) - \sqrt{([E]_t + [I]_t + Q)^2 - 4[E]_t[I]_t}\right]}{2[E]_t}\right\}\right)$$

$[E]_t$ = Concentration of catalytic sites (μM)

$[I]_t$ = Inhibitor concentration (μM)

$[S]$ = Substrate concentration (μM)

K_M = Michaelis-Menten constant (μM)

K_i = inhibitor constant (μM)

V_0 = Enzyme velocity without inhibitor (μM·second^{-1})

V = Enzyme velocity (μM·second^{-1})

2.2.8.7 Cytochrome b_{562} porphyrin binding – absorbance assay

The porphyrin binding affinities of cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} were determined by titration of haem or ZnPP into apo-protein while monitoring the increase in intensity of absorbance features associated with the cyt b_{562}-porphyrin holo-protein complex. The protein was incubated for 15 minutes after each addition of porphyrin. Haem binding by cyt b_{562}^{WT} was monitored by absorbance at the 418 nm major peak of the holo-protein. The A_{418} values for the titration of haem into cyt b_{562}^{WT} were plotted against titrant volume added and fitted to Equation 2.4 published by Wang *et al*. (1992)[199] by adapting the equation into a format suitable for GraphPad Prism. This equation determines the dissociation constant (K_D) of a protein-ligand complex by titrating increasing volumes of a stock ligand into a fixed concentration of protein. This was not possible for cyt b_{562}^{ZnPP} as no binding-associated absorbance features were observed. ZnPP binding was investigated using fluorescence (next section, 2.2.8.8) due to excessive overlap between absorbance peaks of the free and protein-bound porphyrins.

Equation 2.4 Fitting absorbance, varying ligand and accounting for dilution, Wang *et al*. (1992).

$$[P]_t = \frac{[P]_0 V_0}{V_0 + V_L}$$

$$[L]_t = \frac{[L]_0 V_L}{V_0 + V_L}$$

$$\Delta A = \frac{C}{2}\left[([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}\right]$$

C = molar extinction coefficient ($\mu M^{-1} \cdot cm^{-1}$)
$[V]_0$ = Initial volume (µL)
$[V]_L$ = Volume of ligand stock added (µL)
$[P]_0$ = Initial protein concentration (µM)
$[P]_t$ = Protein concentration (µM)
$[L]_t$ = Ligand concentration (µM)
$[L]_0$ = Ligand stock concentration (µM)
ΔA = Absorbance (AU)
K_d = dissociation constant (µM)

2.2.8.8 Cytochrome b_{562} porphyrin binding – fluorescence assay

The difference in fluorescent properties between ZnPP and the cyt b_{562}-ZnPP complex was exploited to derive the ZnPP binding affinities of cyt b_{562}^{WT} and cyt b_{562}^{ZnPP}. ZnPP binding to cyt b_{562}^{WT} or cyt b_{562}^{ZnPP} were monitored by peak emission values at 597 nm (excitation at 430 nm) during titration of ZnPP stock of known concentration into the protein. The protein was incubated for 15 minutes after each addition of ZnPP stock. Emission values (Em_{597}) were plotted against volume of ZnPP stock added, as previously in section 2.2.8.7. To calculate the K_D of the cyt b_{562} – ZnPP binding interaction, the data was fitted to Equation 2.5[199–201], a modified version of Equation 2.4, using GraphPad Prism. This equation accounts for dilution and the reduced emission contribution from free ZnPP.

Equation 2.5 Fitting fluorescence emission, varying ligand and accounting for dilution.

$$[P]_t = \frac{[P]_0 V_0}{V_0 + V_L}$$

$$[L]_t = \frac{[L]_0 V_L}{V_0 + V_L}$$

$$[PL] = \left[\frac{[L]_t - [PL]}{[L]_{free}}\right] \times \frac{1}{2}\left[([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}\right]$$

$$\Delta F = F_B + \left(F_L \times [L]_{free}\right) + (F_R \times F_L \times [PL])$$

$[V]_0$ = Initial volume (µL)
$[V]_L$ = Volume of ligand stock added (µL)
$[P]_0$ = Initial protein concentration (µM)
$[L]_0$ = Ligand stock concentration (µM)
$[P]_t$ = Protein concentration (µM)
$[L]_t$ = Ligand concentration (µM)
$[PL]$ = Protein-ligand complex concentration (µM)
$[L_{free}]$ = Free ligand concentration (µM)
F_B = Background fluorescence (AU)
F_L = "Molar fluorescence" of free ligand (AU/µM)
F_R = Fluorescence ratio of ligand (unitless). $F_L \times F_R$ = bound PL complex "molar fluorescence"
ΔF = Emission (AU)
K_d = dissociation constant (µM)

2.2.8.9 Crystallography

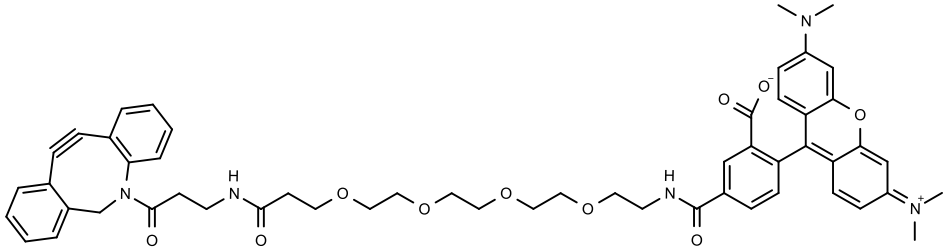
A crystal trial was set up to grow cyt b_{562}^{ZnPP} crystals in a 96-well vapour diffusion sitting-drop type plate with 1 µL protein at ca. 10 mg/mL and 1 µL buffer per well, with 100 µL buffer in each reservoir; the buffers used were from a JCSG-plus HT96 kit (Molecular Dimensions). The plates were incubated at 20 °C for several weeks and were examined by microscope with UV illumination. Crystalline objects were analysed at Diamond Light source using an X-ray beamline by Dr. Pierre Rizkallah.

2.2.9 Artificial post-translational modification of phenyl azide and cysteine protein variants

2.2.9.1 Covalent modification with DBCO-TAMRA

It was necessary to verify reactivity of the phenyl azide side chain of genetically-incorporated AzF residues. Though surface-accessible residues were selected for mutation to AzF, it is possible that unforeseen factors during protein expression could cause the residue to become inaccessible or with reduced reactivity. Also, the phenyl azide is prone to inactivation by photolysis if not stored correctly; ambient light can convert the phenyl azide to a nitrene radical, with the radical becoming reduced to a primary amine.

The strained-ring alkyne fluorophore DBCO-TAMRA (Click Chemistry Tools) (Scheme 2.1) was used to selectively react with proteins containing an available AzF residue. The dye label was incubated with proteins of interest at fivefold molar excess compared to the target protein. Samples were incubated at room temperature, rotating, for one hour then analysed by SDS-PAGE. Fluorescent bands visible on the gel with UV illumination represented dye-labelled proteins. Protein band intensity was compared to excess dye band intensity using ImageJ software[202] to estimate the proportion of protein that reacted (labelling efficiency). Gels were also stained with Coomassie blue to compare total protein content and visualise the protein marker ladder.



Scheme 2.1 DBCO-(PEG4)-TAMRA.

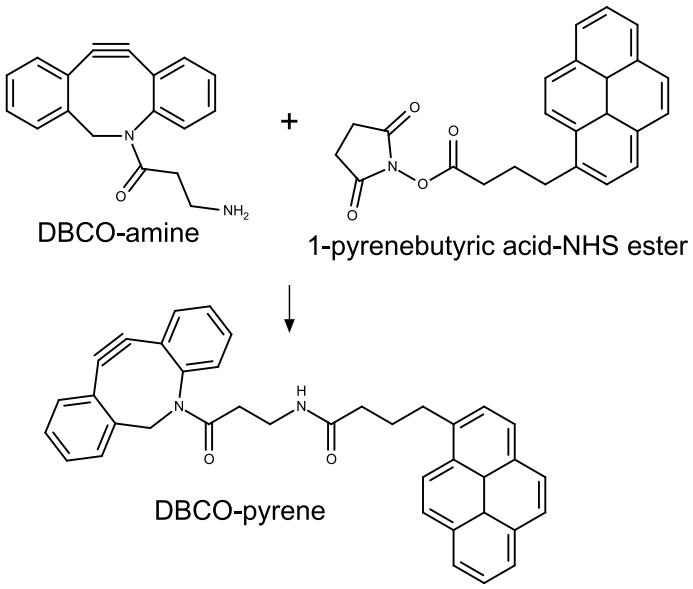
2.2.9.2 Covalent modification with DBCO-pyrene

To non-covalently functionalise NTFET devices with proteins containing AzF, a linker molecule was synthesised to adsorb to the CNT surface by π-π stacking. This molecule consisted of a pyrene moiety at one end to adsorb to CNTs, and a DBCO strained-ring alkyne at the other end to react with the AzF group on the protein via SPAAC.

DBCO-pyrene was synthesised by reacting DBCO-amine (Click Chemistry Tools) with 1-pyrenebutyric acid-N-Hydroxysuccinimide ester (NHS ester, Sigma-Aldrich) in equimolar proportions in dimethylformamide (DMF, Sigma-Aldrich) at room temperature

overnight (Scheme 2.2). Any remaining unreacted NHS groups were then blocked by addition of 5% v/v ethanolamine.

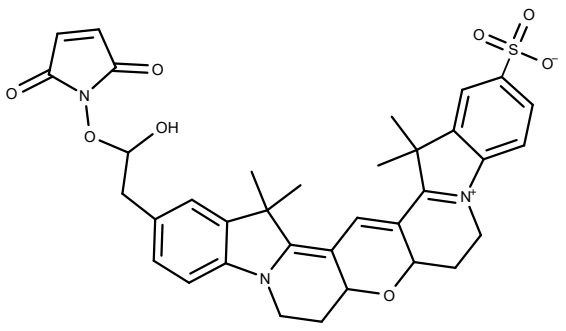
The CNT channels of NTFET devices were pre-coated with DBCO-pyrene linker molecules, then the AzF-containing proteins of interest incubated with the coated CNTs. This results in functionalisation of the CNTs with the protein in defined orientations via a SPAAC reaction (see section 2.2.10).



Scheme 2.2 Synthesis of DBCO-pyrene linker molecule from DBCO-amine and 1-pyrenebutyric acid-NHS ester.

2.2.9.3 Covalent modification with Cy3B-maleimide

The engineered protein CTX-M^{86Cys} was labelled with the reactive fluorescent dye Cy3B-maleimide (Cytiva) (Scheme 2.3) to enable analysis of BLIP-II – CTX-M complex formation on CNT surfaces. CTX-M^{86Cys} has a mutant cysteine residue added at the protein surface suitable for labelling with a maleimide moiety. Cysteine-maleimide chemistry reacts quickly, so CTX-M^{86Cys} was incubated with a twofold excess of Cy3B-maleimide, for five minutes at room temperature. Excess unreacted Cy3B-maleimide dye was then quickly removed from the sample using a PD10 desalting column (Cytiva).



Scheme 2.3 Cy3B-maleimide

2.2.10 Construction and analysis of protein functionalised NTFET devices

This work was carried in collaboration with Dr. Matteo Palma's group at the Department of Chemistry, Queen Mary, University of London with proteins designed and produced in sections 2.2.5 and 2.2.7. Device functionalisation and analysis was performed by Xinzhao Xu.

2.2.10.1 Fabrication of NTFET nanodevices

NTFET devices were based on a p-doped silicon wafer substrate with a 400 nm grown SiO_2 coating. These were coated with 5 nm of chromium, then 40 nm of gold by electron beam evaporation. Ten pairs of nanoscale gold electrodes, each with 300 nm gaps and with pairs separated by 50 µm, were fabricated on each chip by electron beam lithography, connected to contact pads for interfacing with electronic equipment (Figure 2.4). These substrate/electrodes were made as described in Xinzhao Xu *et al*. (2018)[112] by Johnas Eklöf-Österberg and Kasper Moth-Poulsen at Chalmers University of Technology, Gothenburg, Sweden.

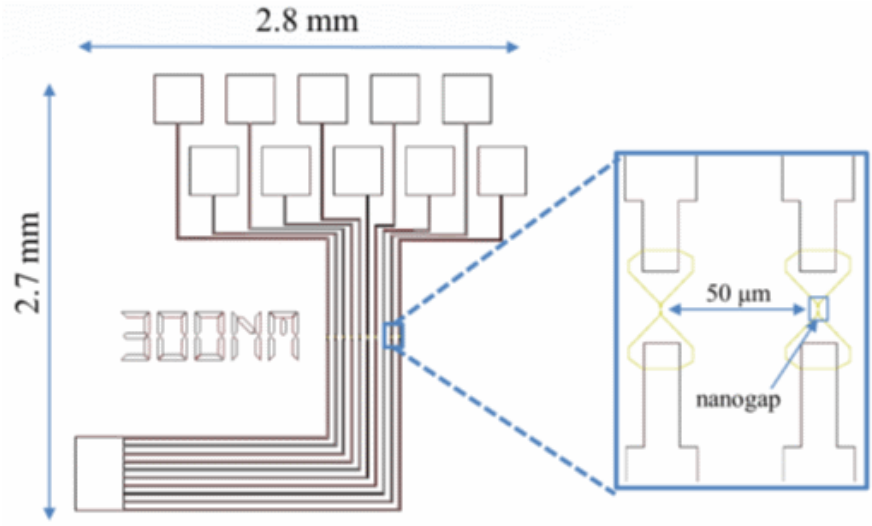


Figure 2.4 Silicon substrate with 300 nm gap between gold electrodes to be bridged by CNTs. Used as basis of NTFET devices. From Xinzhao Xu *et al*. (2018)[112] with permission.

To construct the transistor channels of the NTFET devices, the 300 nm gaps between electrodes were bridged by bundles of immobilised S-SWCNTs. These were assembled by dielectrophoresis (DEP), the phenomenon where a non-uniform electric field can be used to move and align molecules[203]. 0.1 mg 95% semiconducting SWCNT (Nanointegris Technologies) were solubilised by sonication for 1 hour in 0.5 mL 1% (w/v) SDS, the CNT solution was then centrifuged for 1 hour to remove aggregates and insoluble materials. The CNT stock solution was diluted 100-fold and 5 µL was drop-cast on the device at the electrode gaps. An AC voltage of 3 V at 400 kHz was applied across the electrodes for 10 seconds to align CNTs between the electrodes. The chip was then

washed gently with water to remove surfactant and dried with a stream of N_2 gas. Electrical measurements were taken across each electrode to confirm immobilisation of CNTs across the gaps, using a PS-100 probe station (Lakeshore) with a Keithley 4200SCS semiconducting parameter analyser (Tektronix). I-V characteristics were measured in bias sweeping mode (-0.5 V to 0.5 V) applied to source electrode with a grounded drain. Transfer characteristics were measured in gate bias sweeping mode (-15 V to 15 V), monitoring the source-drain current (I_{sd}) against gate voltage (V_g).

2.2.10.2 Functionalisation of NTFETs with BLIP-II via DBCO-pyrene linkers

Before functionalisation of the devices with AzF-containing proteins, the NTFET channels were functionalised with DBCO reactive groups to react with the azide groups on protein surfaces by SPAAC. The CNT channels were coated with a 3:1 mixture of 1-pyrenebutanol (Sigma-Aldrich) and DBCO-pyrene (synthesised as in 2.2.9.2) to functionalise the channel with DBCO reactive groups spaced out along the channel, to avoid crowding of conjugated proteins. The pyrenebutanol/DBCO-pyrene mixture was applied to the NTFETs for 1 hour for the pyrene groups to adsorb to CNTs by π-π stacking and washed with isopropanol then Dulbecco's phosphate-buffered saline (DPBS).

Four different BLIP-II AzF-containing variants were conjugated to the DBCO reactive groups coating the CNT channels: BLIP-II^{41AzF}, BLIP-II^{43AzF}, BLIP-II^{49AzF} and BLIP-II^{213AzF}, sampling different conjugation orientations. A 20 µL drop of BLIP-II in DPBS at 200 nM was cast on each device, then incubated at room temperature for 3 hours for the azides to react via SPAAC with the immobilised DBCO groups. The protein solution was then rinsed off thoroughly with a stream of ddH_2O to remove any non-specifically bound proteins from the CNTs and dried with N_2 gas.

2.2.10.3 Analysis of NTFET devices by atomic force microscopy

Tapping mode AFM topographical data was recorded in air for the electrode gap and CNT channel region of each NTFET chip, to assess the functionalisation of CNTs with BLIP-II. Changes in the height of surface features such as the CNT channel bundle or individual CNTs were thought to be the result of decoration with BLIP-II proteins. The AFM equipment used was a Dimension Icon AFM with ScanAsyst Air tips (Bruker, Coventry, UK). AFM data were analysed with NanoScope Analysis software (Bruker) by using Gwyddion[204]. An overview of AFM is given in chapter 3.2.

2.2.10.4 Real-time detection of TEM^{WT} with NTFET devices

Real-time channel conductivity measurements were performed for the four NTFET devices functionalised with different orientations of BLIP-II, to assess whether TEM^{WT} binding to immobilised BLIP-II could be detected in the current signal. A fixed bias of 0.1 V was applied across the source and drain electrodes of the devices, with the back gate electrode grounded, and the current monitored.

Measurement/detection was performed after allowing the current response to baseline for several minutes, with the addition of TEM^{WT} in DPBS buffer of high ionic strength to resemble physiologically relevant conditions. Later experiments were performed in steroid-free serum (MP Biomedicals, Eschwege, Germany), diluted tenfold with DPBS to mimic a physiological fluid with a complex population of proteins. First just DPBS, then increasing concentrations of TEM^{WT} were applied to all BLIP-II-functionalised NTFET devices to measure signal response. As a negative control, the response of an NTFET device functionalised with only pyrenebutanol was tested by applying TEM^{WT}.

2.2.11 Photo-functionalisation of CNTs with AzF-containing proteins

This work was carried out by me at Cardiff University School of Physics with some data gathered by Suzanne Thomas and Rebecca Gwyther.

2.2.11.1 Dispersion and adsorption of CNTs on silicon surfaces

CNTs were adsorbed on silicon surfaces as a level substrate for CNT-protein functionalisation by photoattachment and subsequent analysis by atomic force microscopy (AFM, see 2.2.11.3). Single-walled semiconducting CNTs (S-SWCNTs) were used, supplied by Sigma-Aldrich; these were grown by chemical vapour deposition and had an average diameter of 2 nm. CNTs were solubilised for dispersion in water by sonication on ice in a 1% (w/v) SDS solution using a Soniprep 150 probe sonicator (MSE, Heathfield, UK), using full power until suspended. CNT solution was centrifuged at 16,000 ×*g* before each use to remove CNT aggregates.

The substrate consisted of ca. 5 mm^2 cleaved squares of silicon wafer with a top layer of SiO_2, with electrodes etched from a gold surface layer by lithography (Figure 2.5). These were offcuts from previous electrical experiments, and the raised gold features proved useful as reference points for navigating the surface using AFM. The substrates were heated in acetone for 5 minutes at 75 °C to remove residual photoresist coating, then soaked in isopropanol for 5 minutes at 75 °C to remove any remaining acetone. The S-SWCNT solution was drop-cast onto the substrate surface and dried on a hot plate at

100 °C, then the substrates washed with a jet of ddH_2O to remove SDS and debris then soaked in ethanol for 1 hour to dissolve residual SDS.

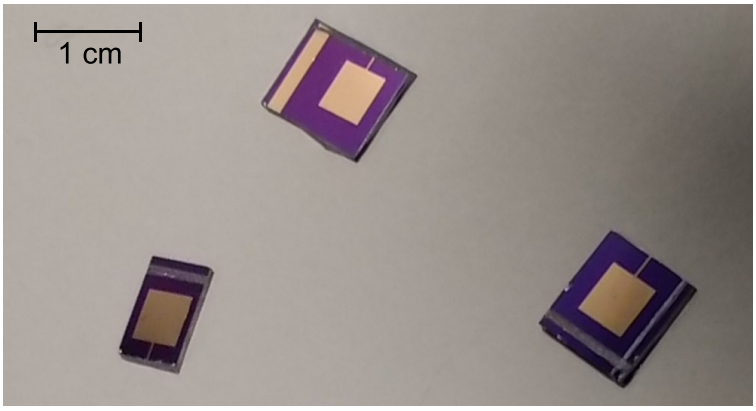


Figure 2.5 Silicon wafer with gold electrodes. Used as a substrate for CNT drop-casting. Gold electrodes were a reference point to re-locate the same CNTs using AFM.

2.2.11.2 UV photoattachment of proteins to carbon nanotubes

CNTs adsorbed on silicon substrate were functionalised with engineered AzF-containing proteins in defined orientations by UV-nitrene cycloaddition. A 15 µL drop of protein in Tris buffer (50 mM Tris-HCl, pH 8) at 100 nM was drop-cast onto the substrate surface and the phenyl azide group of AzF was photoactivated to phenylnitrene by UV irradiation to react with the CNT π-electron system by nitrene cycloaddition, forming an aziridine link. The sample was irradiated with UV-B light (λ = 305 nm, intensity 18 $W \cdot m^{-2}$, LED) for 5 minutes. The protein droplet was then washed off for 1 minute with a ddH_2O jet to remove unbound proteins and the substrate dried with a stream of nitrogen gas (N_2).

CNTs shown to be functionalised with BLIP-II^{41AzF} by AFM analysis were further incubated for 10 minutes with a 15 µL drop of TEM^{WT} at 100 µM, to form protein-protein complexes on the CNT surface. Excess TEM^{WT} was rinsed off as previously and the CNT sample analysed again by AFM.

2.2.11.3 Analysis of CNT constructs by atomic force microscopy

Nanoscale topographic data for CNTs and proteins on silicon substrates were measured in air under ambient conditions using tapping mode AFM on a Veeco Multimode Nanoscope III AFM (Bruker) with polysilicon probes (Nanosensors, Neuchâtel, Switzerland). Data were recorded at the maximum available raster scanning resolution of 512 × 512 pixels at 1 Hz. The extent of CNT functionalisation with BLIP-II, TEM and cyt b_{562} proteins was assessed using AFM data to examine the height and morphology of surface features. The appearance of globular objects of increased height on CNT surfaces indicate the presence of immobilised proteins.

CNT samples that showed clear functionalisation with BLIP-II on AFM imaging were then incubated with TEM^{WT} and imaged again, using the surface features of the substrate and patterns of surface CNTs to find and re-image the same CNTs. The aim was to find evidence of TEM^{WT} complexed to surface-immobilised BLIP-II by the appearance of new globular objects or increased height of existing objects. Averages of 8 height profiles were taken across objects to detect increases in height. Data was analysed using Gwyddion software[204] and MS Excel.

3 Integration of engineered proteins into a carbon nanotube based field-effect transistor sensing device

Work in this chapter contributed to two papers, one where I am joint 1st author.

Freeley, M., Worthy, H. L. *et al*. Site-Specific One-to-One Click Coupling of Single Proteins to Individual Carbon Nanotubes: A Single-Molecule Approach. *Journal of the American Chemical Society* **139**, 17834–17840 (2017).

Xu, X., **Bowen, B. J.** *et al*. Tuning Electrostatic Gating of Semi-Conducting Carbon Nanotubes by Controlling Protein Orientation. **Publication pending**. (2020).

3.1 Introduction

This chapter focuses on the design, engineering, production and characterisation of a range of engineered proteins of interest for use in bottom-up assembly of nanoscale systems. These proteins were then intimately interfaced with semiconducting single-walled carbon nanotubes (S-SWCNTs), constructing biosensing devices to detect the binding of a target protein analyte with high specificity and sensitivity. CNTs are considered one-dimensional conductors, with aspect ratios > 100:1 and diameters comparable in scale to proteins, typically 0.7-10 nm[205]. The unique properties of S-SWCNTs make them an attractive basis for biochemical sensing devices, especially proteins, due to: 1) the highly sensitive modulation of S-SWCNT electrical conductivity in response to variations in the electrostatic environment[206]; 2) the high surface area to volume ratio of single-walled CNTs with all conduction occurring at the surface[207]; 3) the ease of incorporating CNTs into nanoelectronic components. Electric current conducted through a S-SWCNT can be modulated by charged molecules within one Debye length of the CNT surface. A Debye length (or radius) (λ_D) expresses the range of the electrostatic effect of charge carriers, with their electric potential decreasing by 1/*e* with each Debye length. The Debye length is reduced in a solution containing ionic solutes by electrostatic screening from charged ions. In an ionic electrolyte solution at physiological ionic strength, the Debye radius is ca. 0.7 nm from the CNT surface. Charged species within 1 λ_D of a CNT can increase or decrease the available charge carriers in a CNT, affecting the CNT conductance. Negatively-charged species within 1 λ_D of a p-type S-SWCNT (in which the mobile charge carrier is electron holes) cause accumulation of holes, increasing conductance; positive charges within 1 λ_D of a p-type S-SWCNT surface deplete local hole concentration, reducing conductance[105,208]. The opposite is true for n-type CNTs. This property is analogous to semiconductor voltage gating and is known as electrostatic gating, and has enabled semiconducting CNTs to

form the channels of chemiresistor-type nanotube field effect transistors (NTFET) in chemical sensing devices[109,110,119] (see introduction, chapter 1.2.2). In these devices, charged molecules near the CNT channel surface provide gating for the NTFET, and a gate electrode is not required; changes in current between the source and drain electrodes form the detection signal.

Protein surfaces are decorated with patterns of electrostatically charged residues unique to that protein species. Exposing different protein surfaces to a S-SWCNT can potentially have different effects on CNT conductivity based on the protein's geometric position relative to the CNT. Thus, changes in protein orientation or conformation can be detected by changes in current signal through the CNT. Protein-NTFET based biosensors have been demonstrated previously[209], with Hatada *et al*.[119] having electrically sensed conformational change in a protein's charged surface on substrate binding, when conjugated on a CNT channel. Such biosensors have functionalised CNTs with proteins conjugated using intrinsic amino acid chemistries, in this case succinimide-amine reactions that link the protein to the CNT via nonspecific amine groups on the protein surface. This yields CNTs functionalised with proteins in a random range of orientations, meaning subtle electrical signal changes through the CNT are likely to be lost to the averaging effect of a wide range of binding orientations at the CNT surface. The Philip G. Collins group at UC Irvine[117,126,210] achieved defined orientation of proteins on CNTs for electrical biosensing using engineered single cysteine attachment points, but this is not a universal general approach as many proteins have one or more native cysteine residues. In a protein-functionalised NTFET, defined orientation of proteins at the CNT channel surface is important to achieving consistent signal changes through the CNT because different surfaces of a protein are likely to have different electrostatic effects on CNT conductivity. If proteins are bound to the CNT in a range of random orientations, electrostatic effects on CNT conductance will be an average of all possible orientations (i.e. an ensemble signal), with some orientations probably cancelling out others. This arrangement for a protein NTFET is likely to result in a poor signal-to-noise ratio compared to an NTFET functionalised with proteins all conjugated in the same orientation and presenting a consistent protein surface to the CNT (Figure 3.1).

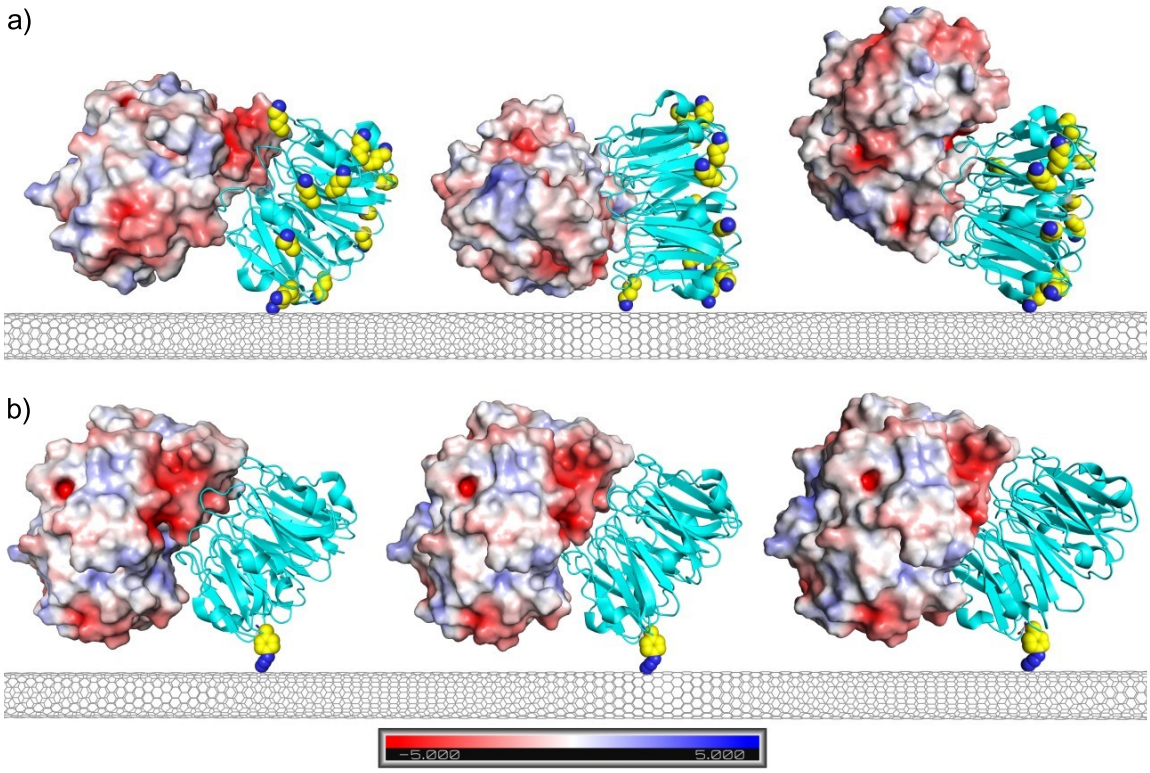
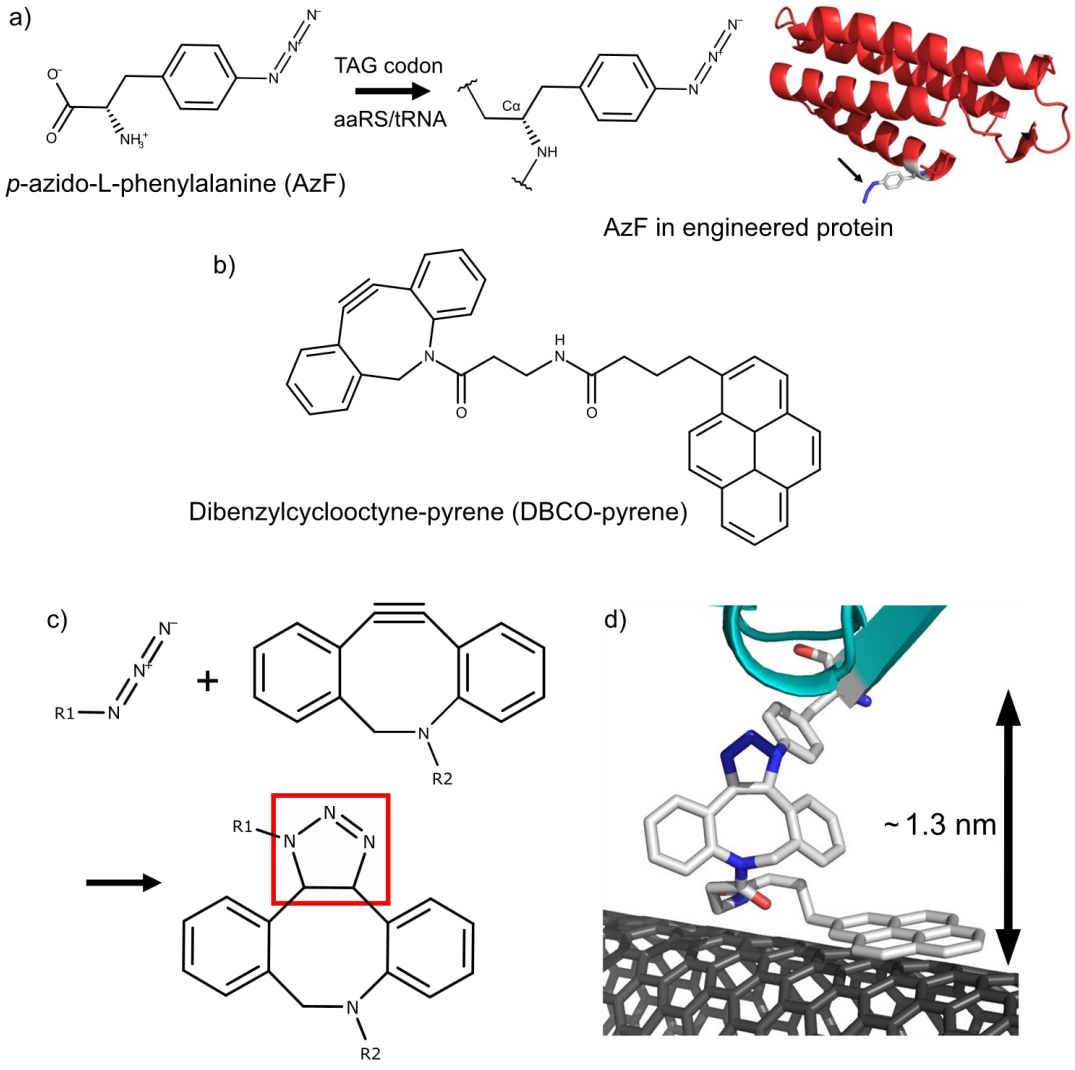


Figure 3.1 Significance of defined protein orientation on CNT surface. a) binding protein (cyan) is conjugated to CNT surface using intrinsic lysine chemistry and is interfaced with the CNT in a random range of orientations due to 12 possible conjugation residues (yellow/blue). This presents the target analyte protein to the CNT in a random range of orientations, resulting in a signal that is an average of all sampled surfaces. b) Binding protein is conjugated to CNT via a single bioorthogonal reactive handle (yellow/blue), all proteins are oriented in the same way and analyte proteins are bound with a consistent electrostatic surface adjacent to the CNT. This results in an optimal signal that represents single-molecule characteristics. Surface charge was calculated using APBS[211] electrostatics in PyMOL[11]. The relative scale of charge distribution by colour in shown in the bar. Charges: red - negative; blue – positive (units -5 to +5 $k_b \cdot T \cdot e_c^{-1}$). PDB: 1JTD[36].

Here I plan to address the challenges of defined and designed protein attachment to CNTs using genetically-incorporated, non-natural, bioorthogonal chemistry to provide a single unique reactive handle on each protein. The bioorthogonal reactive handle used is the phenyl azide side chain of the non-natural amino acid (nnAA) *p*-azido-L-phenylalanine (AzF), genetically incorporated into our engineered proteins at designed locations by using a reprogrammed amber stop (TAG) codon approach. This system uses an engineered tRNA and amino acid tRNA synthetase (aaRS) to insert AzF into a nascent peptide during translation in response to a TAG codon (see chapter 1.1.3)[79]. The phenyl azide group of AzF adds useful new chemistry to proteins, enabling the targeted labelling or immobilisation of a protein at one specific residue via the phenyl azide and eliminating off-target binding at other residues. Protein engineering allows us to place a TAG codon at defined positions in a gene encoding our protein of interest. The system allows targeted placement of the phenyl azide reactive group in the protein structure, defining the orientation of an entire protein population immobilised on a CNT surface. The key reaction used in this chapter for oriented immobilisation of proteins on

CNTs is strain-promoted azide-alkyne cycloaddition (SPAAC) click chemistry: the formation of a covalent triazole linkage between the azide and an alkyne target (see chapter 0). Proteins were covalently linked to the intermediate alkyne molecule DBCO-pyrene using SPAAC to facilitate non-covalent adsorption of proteins to the CNT surface by π-π stacking of pyrene (Scheme 3.1).

The nnAA AzF also has other interesting chemical properties, including photoactivatable nitrene crosslinking chemistry that can be used to directly conjugate proteins to CNTs via covalent aziridine bonds. This approach is discussed in more detail in chapter 4 of this thesis, with the work published in Bioconjugate Chemistry (2019)[212].



Scheme 3.1 Interfacing protein to CNT via pyrene conjugated by SPAAC click chemistry. a) AzF with phenyl azide side chain is incorporated in protein in response to genetically-encoded TAG codon in presence of engineered aaRS and tRNA. b) DBCO-pyrene: pyrene adsorbs to CNT surface, strained-ring alkyne DBCO reacts specifically with azide in bioorthogonal reaction. c) SPAAC reaction of azide with DBCO forms covalent triazole bond (red box). d) Three-dimensional render of AzF-triazole-pyrene linker bridging BLIP-II (cyan) and CNT (grey). Linker model conformation optimised computationally by Alexander Nemukhin.

The aim of functionalising NTFETs with a population of consistently-oriented proteins was to create a sensing device capable of detecting a specific protein-protein interaction. The CNT channel of the NTFET device was functionalised with immobilised binding proteins, oriented to trap an analyte protein close to the CNT surface, presenting a consistent electrostatic surface of the analyte to the CNT for sensing. The target proteins of interest for our NTFET sensing devices are class A serine β-lactamases, bacterial enzymes that are the most common agents of antimicrobial resistance (AMR; see chapter 1.3) and a major challenge to health systems. Rapid detection of β-lactamases in a patient sample using a point-of-care electronic sensor would be a major improvement on current β-lactamase detection protocols, e.g. culture of patient samples on a range of antibiotic media, which takes several days, or PCR or DNA sequencing methods, all of which require a dedicated laboratory, most probably offsite. This could help the current AMR crisis by reducing unnecessary antibiotic prescription. The NTFET β-lactamase sensors described here are functionalised with high-affinity β-lactamase binding proteins (BLIP-II), immobilised in a defined orientation on the CNT surface to bind and trap a wide range of class A β-lactamases from solution close to the CNT surface in consistent orientation. The electrostatic surface of a β-lactamase bound within one Debye length of the CNT induces a shift in conductivity related to the charge of the protein surface adjacent to the S-SWCNT, detectable as a change in current. The high specificity of protein-protein interactions ensure that only target β-lactamases bind to the functionalised CNTs. BLIP-II has evolved high binding specificity for class A β-lactamases, and binding affinity among the strongest known protein-protein interactions with a dissociation constant (K_D) for TEM and CTX-M β-lactamases of 0.5 pM[147]. BLIP-II can be considered a universal antibody against class A β-lactamases as it binds a wide variety of β-lactamases with varying degrees of affinity. BLIP-II has the advantage of being only ca. 4 nm high, which is less than half the height of antibodies such as immunoglobulin G (> 8.5 nm)[213]; the small size of BLIP-II is especially important as the target β-lactamase must be bound as close as possible to the CNT surface.

This chapter aimed to construct NTFET biosensors functionalised with pyrene-linked BLIP-II proteins in defined orientations via engineered bioorthogonal AzF residues. These devices were analysed using atomic force microscopy (AFM) to verify immobilisation of BLIP-II on CNT channel surfaces. TEM β-lactamase was applied to functionalised devices to analyse electrical current effects caused by TEM binding to BLIP-II at the CNT surface. Devices were functionalised with several different variants of BLIP-II via AzF residues located at distinct positions at the protein, to sample different orientations of the BLIP-II-TEM complex at the CNT surface. Electrical effects of these different electrostatic surfaces on conduction through the CNT channel were analysed.

It was found that NTFETs functionalised with different orientations of BLIP-II on the CNT surface showed differing current responses on the addition of TEM^{WT}. This suggested that the CNT channel sampled different electrostatic surfaces of TEM^{WT} as dictated by the orientation of BLIP-II, and CNT conductance was modulated.

3.2 Atomic force microscopy

In chapters 3 and 4 of this thesis, protein functionalisation of CNTs was assessed using tapping mode atomic force microscopy (AFM) to record topographic data from CNTs adsorbed onto flat surfaces. AFM is a form of scanning probe microscopy that can resolve objects in three dimensions down to the sub-nanometre scale, far below the resolution limit of light microscopy (approx. 200 nm). Topographic data is recorded by scanning the probe at the tip of an AFM cantilever across the sample in one of three modes: non-contact mode which does not touch the sample but is affected by intermolecular forces; contact mode which "drags" the probe across the sample; and tapping mode in which the cantilever tip is oscillated at a fixed frequency, tapping against the surface of the sample. AFM cantilevers are fabricated from crystalline silicon wafer by nanoscale photolithography and are oscillated by a piezoelectric transducer. The forces between the tip of the cantilever are measured by the deflection of the cantilever, which is measured by a laser reflected by the cantilever onto a photosensor (**Figure 3.2**a). Tapping mode is preferable for analysis of "soft" materials like proteins, which tend to be damaged by dragging of the tip in contact mode. The oscillation of the tip tends to cause plastic deformation of the proteins, which can result in under-measurement of true protein height (z-axis). AFM is preferable to similarly high resolution microscopic techniques such as scanning electron microscopy for imaging proteins, as AFM is less destructive and repeated scans of a sample can be made without significantly damaging proteins. However, AFM measurement of biomolecules can be inaccurate, as they are typically smaller than the effective interaction area between feature and AFM tip[214]. This is for reasons that are speculated to be related to dehydration of biomolecules[215], mechanical deformation by the tapping action of the AFM tip[216], or a combination of all of the above. However, the high spatial resolution makes AFM one of the few methods of directly observing the morphology of individual CNT-protein complexes. Apparent physical parameters reported here may not reflect true dimensions of the constructs but are useful for comparing these experimental samples.

AFM data in chapter 3 of this thesis was gathered with a Bruker Dimension Icon AFM by Mark Freeley and Xinzhao Xu at Queen Mary, University of London; AFM data in chapter 4 was gathered with a Veeco Multimode Nanoscope III AFM at Cardiff School of Physics by me, Suzanne Thomas or Rebecca Gwyther. Equipment is shown in Figure 3.2b-c.

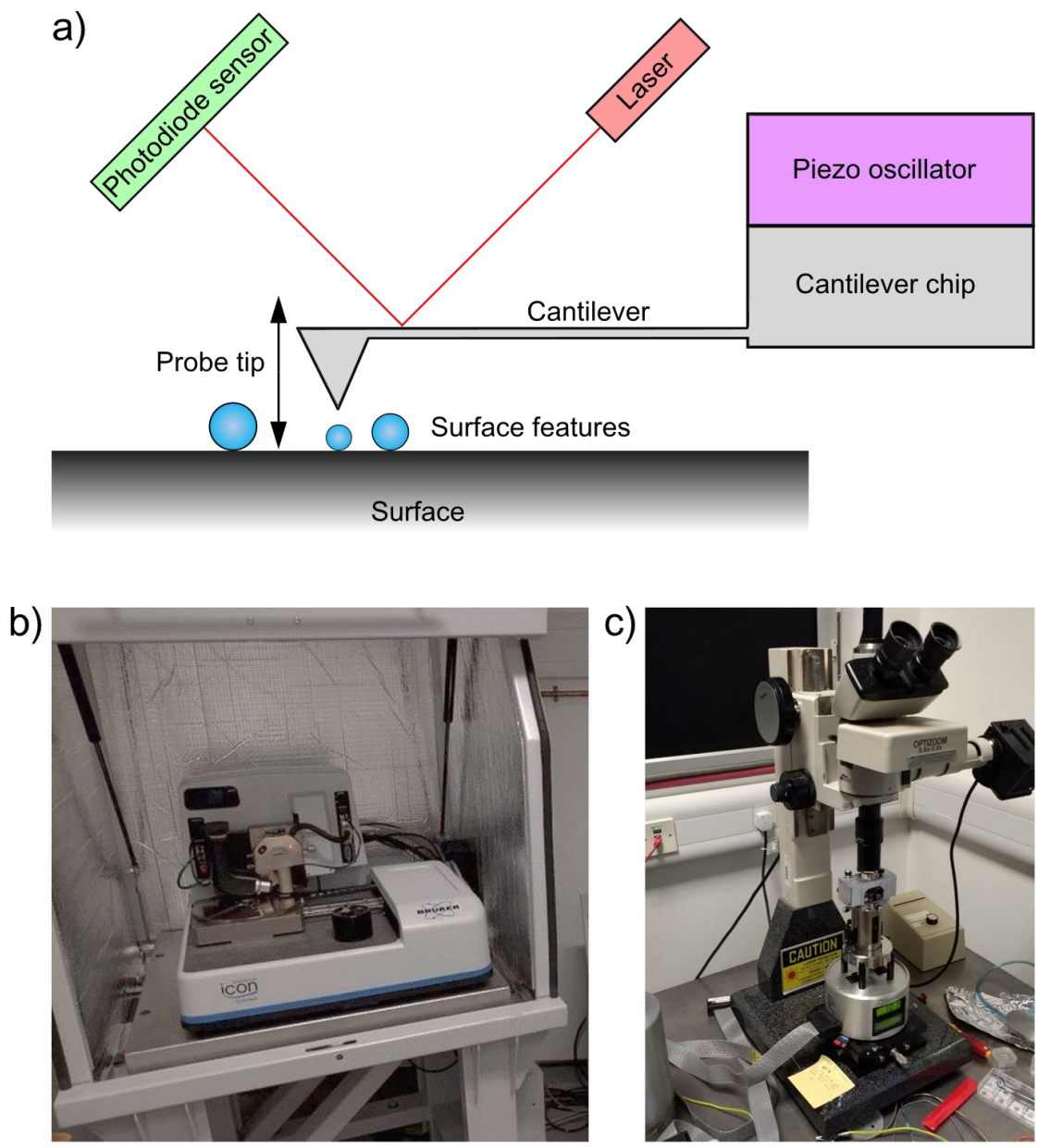


Figure 3.2 Atomic force microscopes. a) General scheme of an AFM. The cantilever is oscillated by a piezo transducer and the probe tip interacts with the sample surface. A laser beam is reflected by the cantilever onto a photodiode sensor, recording deflection of the cantilever. b) Bruker Dimension Icon AFM at Queen Mary, University of London (image: Mark Freeley). c) Veeco Multimode Nanoscope III AFM at Cardiff School of Physics.

3.3 Results and discussion

3.3.1 Production of TEM^{WT}

3.3.1.1 TEM^{WT} expression

To express TEM^{WT} for downstream use as the target protein for our NTFET sensors, *Escherichia coli* BL21 (DE3) cells were transformed by electroporation with the pET-TEM1 plasmid, induced, and grown overnight. Samples of growth cultures were tested for expression of active TEM^{WT} by addition of the chromogenic cephalosporin substrate

nitrocefin: samples with active enzyme immediately changed from yellow to red as TEM^{WT} hydrolysed the amide of the β-lactam ring of nitrocefin. The TEM^{WT} gene in pET-TEM1 carries an N-terminal ompA signal sequence[217] and is secreted to *E. coli* periplasm (Figure 3.3). The periplasmic contents of TEM^{WT} expressing cells were extracted[191,192] (see chapter 2.2.7.2) and analysed by SDS-PAGE, showing overexpression of TEM^{WT} in the periplasm (Figure 3.3).

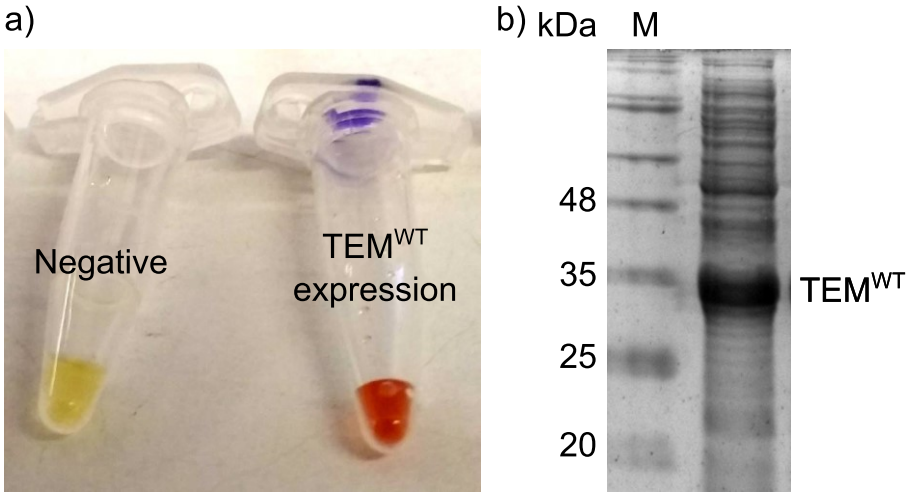


Figure 3.3 Expression of TEM^{WT} in *E. coli*. a) Colourimetric test for TEM^{WT} expression using nitrocefin. Nitrocefin turns red in presence of β-lactamases. b) Overexpression of TEM^{WT} in periplasmic extract. M: molecular weight marker ladder.

3.3.1.2 TEM purification

As the TEM^{WT} gene in pET-TEM1 lacks a polyhistidine affinity tag, purification by IMAC was not possible. Anion exchange chromatography (IEC) was attempted as the isoelectric point (pI) of TEM^{WT} is 5.4 as calculated using the ProtParam online server (https://web.expasy.org/protparam[218]), and is predicted to bind to an anion exchange column in Tris buffer, pH 8 (binding buffer). Extracted periplasmic contents containing TEM^{WT} were diluted into the binding buffer and passed through a HiTrap Q anion exchange column (Cytiva) using an ÄKTApurifier FPLC system (Cytiva). SDS-PAGE analysis showed that TEM^{WT} was present in the flow-through fraction and had not bound to the anion exchange matrix. However, many contaminating periplasmic proteins did bind to the column and were thus removed from the sample (Figure 3.4). The flow-through containing TEM was concentrated to 2 mL and further purified by size exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 pg column (S75, Cytiva). Elution fractions from SEC were analysed by SDS-PAGE: TEM^{WT} eluted at V_e 60-70 mL and was estimated to be ≥ 99% pure using ImageJ band densitometry software[202,219]. Purified TEM^{WT} was used for inhibition assays with BLIP-II variants and for binding experiments on CNTs functionalised with BLIP-II AzF variants.

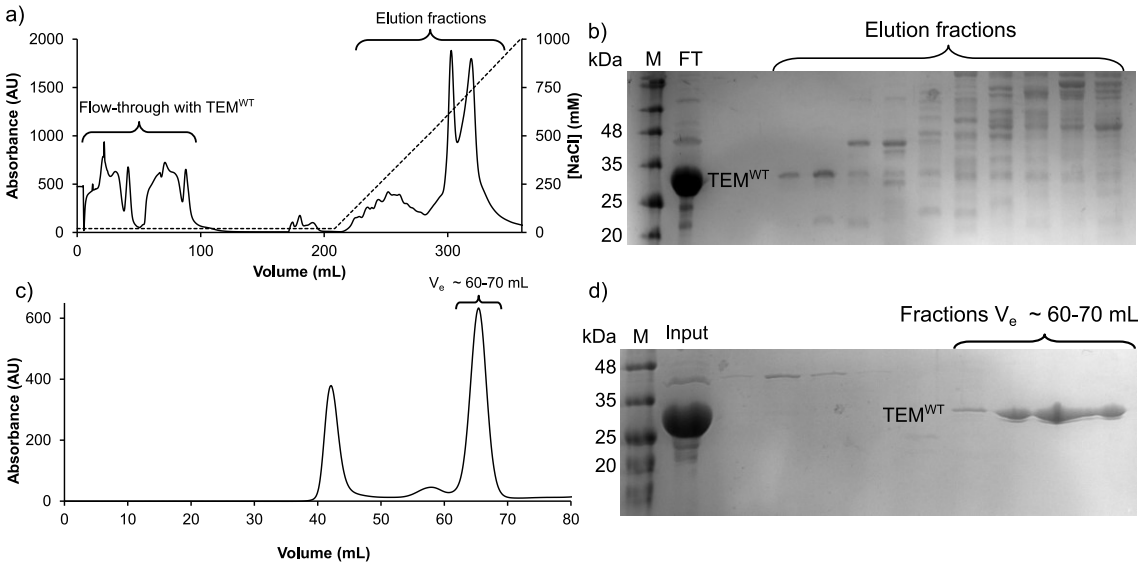


Figure 3.4 Purification of TEM^{WT} from periplasm. a) 280 nm absorbance trace (solid line) from anion exchange chromatography showing flow-through of unbound protein containing TEM at V_e 0-100 mL, and contaminating proteins eluted at V_e 200-350 mL with increasing [NaCl] gradient (dotted line). b) SDS-PAGE analysis of fractions from IEC. FT: flow-through, containing TEM^{WT}. Remaining lanes: fractions from elution by NaCl gradient, containing contaminating proteins. c) 280 nm absorbance trace from size exclusion chromatography (S75 column) showing elution of TEM at V_e 60-70 mL after injection onto the column. d) SDS-PAGE analysis of fractions from SEC. Input: the original sample. 60-70 mL fractions: containing pure TEM. M: molecular weight marker ladder.

3.3.2 Characterisation of TEM^{WT} enzyme activity

To confirm that purified TEM^{WT} was fully functional, its kinetic parameters were determined experimentally measuring hydrolysis of the colorimetric cephalosporin substrate nitrocefin. The initial reaction rates for the first 30 seconds of nitrocefin hydrolysis by 0.3 nM TEM^{WT} were recorded at 10 substrate concentrations from 5-400 μM, recording data for three biological replicates. These results were fitted to the Michaelis-Menten equation[195] using nonlinear regression in GraphPad Prism software[196] (Figure 3.5). The calculated kinetic parameters are shown in Table 3.1. The parameters were close to previously published values for hydrolysis of nitrocefin by TEM^{WT}: Michaelis constant (K_M) was determined to be 175 ± 25 μM, consistent with published values between 30 and 220 μM[220,221]; the turnover number (k_{cat}) was 1087 ± 73 s^{-1}, also consistent with the published values between 714 and 1188 s^{-1} [220,222]. This showed that TEM^{WT} has the expected functional characteristics.

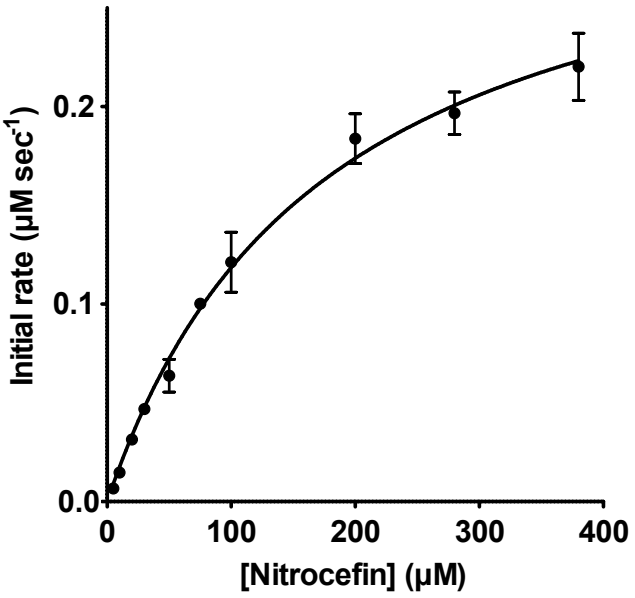


Figure 3.5 TEM^{WT} nitrocefin catalysis. Michaelis-Menten plot of initial rates of nitrocefin hydrolysis by 0.3 nM TEM^{WT}. Data was recorded for three biological replicates. Error bars: standard error for three biological replicates.

Table 3.1 Kinetic parameters calculated for TEM hydrolysis of nitrocefin with standard error of the mean. Parameters including goodness of fit values were calculated by fitting results to the Michalis-Menten equation[195] using nonlinear regression in GraphPad Prism[196]. Published values for comparison.

	TEM^{WT}	Stojanoski *et al.*[220]	Zafaralla *et al.*[221]	Fryszczyn *et al.*[222]
k_{cat} (s^{-1}) (±SE)	1087 ± 73	714 ± 80	920	1188 ± 133
K_M (µM) (±SE)	175 ± 25	30 ± 10	220	83 ± 16
s_{yx} [a]	0.015			
Pseudo-R^2 [b]	0.97			

[a] Standard error of the estimate, in units of Y.
[b] Pseudo-R^2 for nonlinear regression as per method 1 in Kvålseth (1985)[223]

3.3.3 BLIP-II variants: Design and production

3.3.3.1 Design of BLIP-II phenyl azide mutants

BLIP-II is a 29 kDa bacterial β-lactamase inhibition protein that originates in the soil bacterium *Streptomyces clavuligerus*. It has a seven-bladed β-propeller fold and specifically binds and inhibits class A β-lactamases with femtomolar affinity, sterically obstructing the β-lactamase catalytic site[147] (Figure 3.6). BLIP-II was considered an ideal recognition protein for the target β-lactamases in an NTFET sensing device because of this specificity and affinity; measured dissociation constants for BLIP-II interactions with class A β-lactamases range from 80 fM for the carbapenemase KPC[148] to 540 fM for the extended-spectrum β-lactamase (ESBL) CTX-M[147]. This extremely tight binding will aid sensing of low concentrations of β-lactamases at the CNT surface, since BLIP-II-β-

lactamase complex formation is essentially permanent over usual experimental timescales, with the half-lives of these complexes measured in days to weeks[147,148].

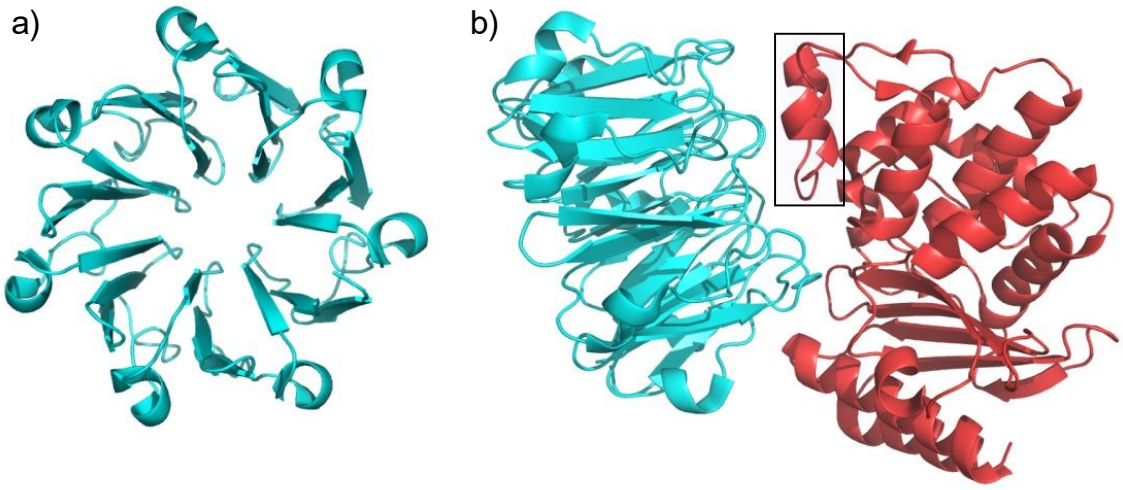


Figure 3.6 BLIP-II structure. a) BLIP-II seven bladed β-propeller fold. b) BLIP-II complexed with TEM β-lactamase (red) with the BLIP-II-binding loop-helix indicated. PDB: 1JTD[36]

Two crystal structures are available of BLIP-II in complex with β-lactamases, one with TEM (PDB: 1JTD[36]) and one with Bla1 (PDB: 3QHY[147]). When overlaid, both structures align closely, with the conformation of the BLIP-II binding residues N112, D131, Y191 and W269 conserved, forming hydrogen bonds to TEM/Bla1 at their conserved V103, E110 and K111 residues (Figure 3.7). Results from alanine scanning mutagenesis found that BLIP-II W269 is a key residue for binding and inhibition, with mutation of this residue reducing β-lactamase affinity between 400-fold (CTX-M) to 18,000-fold (TEM)[147]. Aligning the crystal structures of two further β-lactamases, CTX-M (PDB: 1YLT[37]) and KPC (PDB: 2OV5[224]), reveals that the BLIP-II binding loop-helix is conserved in the same conformation among these four β-lactamases. The loop-helix domain (residues 99-114) also shows sequence identity of 31% and sequence homology of 69% (Figure 3.7, inset. Alignment by Clustal Omega[225]). Conserved residues V103, E110 and K111 form hydrogen bonds to BLIP-II and P107 appears to constrain the loop-helix structure, while the role of L102 is unclear but mutation of this residue to alanine is known to reduce affinity ca. 20-fold[222]. This data suggests that BLIP-II binds to these four β-lactamases in the same way, via the loop-helix domain, and probably other class A β-lactamases too. This means that an NTFET functionalised with BLIP-II in an orientation that binds TEM will likely bind other class A β-lactamases.

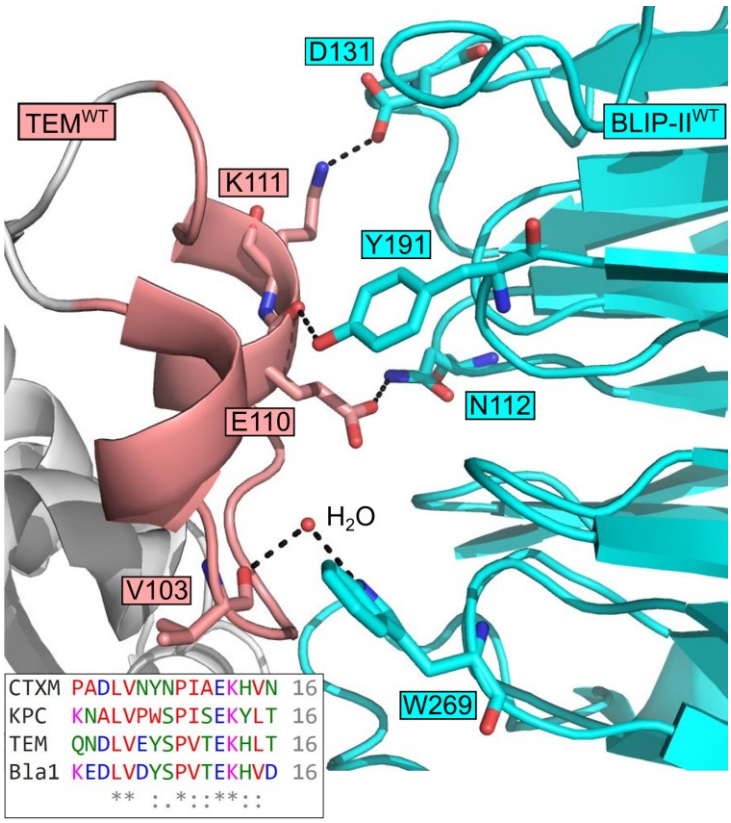


Figure 3.7 Binding interactions between BLIP-II^{WT} and TEM^{WT}. Cyan: BLIP-II^{WT}; Pink: TEM^{WT} loop-helix binding domain; Dotted lines: hydrogen bonds. The loop-helix domain is probably the BLIP-II binding site in all class A β-lactamases, with the loop conformation, residues and interactions highlighted here conserved between TEM, Bla1, CTX-M and KPC. Inset: Sequence alignment showing homologous loop-helix residues in these four β-lactamases (*Identical in all four, alignment: CLUSTAL[225]). PDB: 1JTD[36].

As these class A β-lactamases share close structural and sequence homology and appear to all bind BLIP-II via the highly conserved loop-helix motif, TEM was selected as the target enzyme when designing the BLIP-II functionalised NTFET devices. TEM is the most common and widely characterised of the class A β-lactamases[226]. The electrostatic surface topography of TEM has many discrete regions of positive and negative electrostatic charge and is likely to influence local charge carrier concentration and thus conductivity of a S-SWCNT if bound in close proximity to its surface, i.e. within one Debye length. The charged face of TEM presented to the CNT surface depends on the bound orientation of TEM, which is dictated by BLIP-II binding orientation on the CNT surface. Functionalising different CNT channels with BLIP-II via different residues will result in sampling different electrostatic regions of TEM at the CNT surface, affecting CNT conduction in distinct, measurable ways. The electrostatic topography of TEM also differs from other class A β-lactamases, so the binding of different β-lactamases to BLIP-II functionalised CNTs may cause distinct conductive effects. It is hoped to construct a sensor that can not only detect class A β-lactamases, but also potentially discern between different types (Figure 3.8).

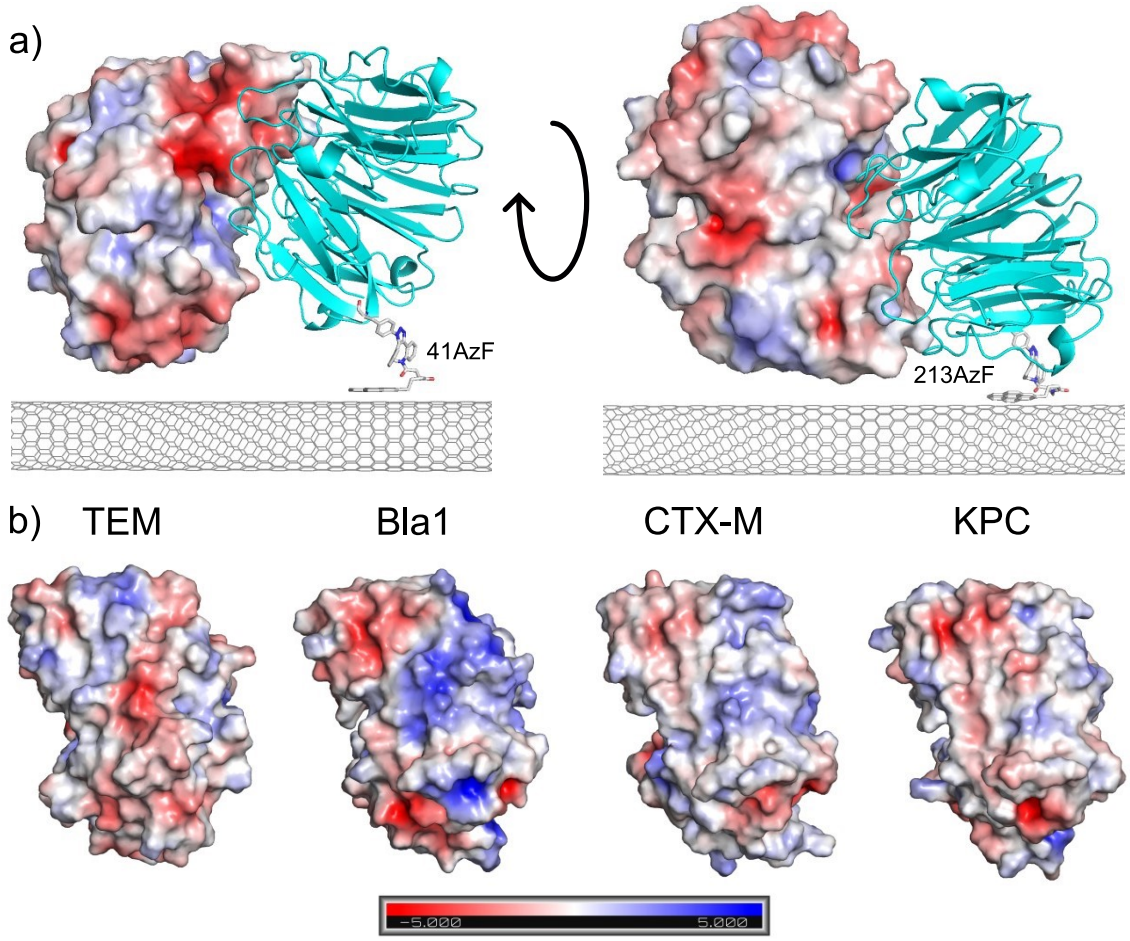


Figure 3.8 Orientation of BLIP-II-β-lactamase complexes at the CNT surface via conjugated pyrene π-π stacking. The proteins have unique patterns of surface charges that can influence the charge carrier concentration of an intimately interfaced CNT and thus its conductivity. a) By changing conjugation point of BLIP-II to the CNT, different electrostatic surfaces of TEM can be sampled. BLIP-II (cyan) adsorbed to CNT via AzF-triazole-pyrene at residue 41: bound TEM face is within range of CNT. BLIP-II conjugated via residue 213: BLIP-II-TEM complex is rotated approx. 180° relative to CNT, presenting different surface of TEM to CNT surface (possible conformations pictured). b) Electrostatic surface charge topographies of four common β-lactamases in the same orientation showing their respective surface charge distributions. The relative scale of charge distribution by colour in shown in the bar. Charges: red - negative; blue – positive (units -5 to +5 $k_b \cdot T \cdot e_c^{-1}$). PDB: 1JTD[36], 3QHY[147], 2OV5[224], 1YLT[37].

Variants of BLIP-II were designed *in silico* with a phenyl azide reactive handle positioned at different points around the protein surface by mutating the native residues to AzF. Surface residues were selected for mutation to AzF by examining the three-dimensional structure of the BLIP-II^{WT}-TEM^{WT} complex (PDB: 1JTD[36]) using PyMOL software[11]. A key consideration in the design process was to produce several different AzF mutants to immobilise BLIP-II in distinct orientations on a CNT surface, so as to bind TEM close to the CNT surface. Mutant residues were also chosen for their solvent accessibility for binding TEM^{WT} and positioned away from the known protein-protein interface so that TEM^{WT} binding will not be hampered. Important β-lactamase binding residues were avoided, informed by data from Brown *et al.* (2013)[147], which used alanine scanning mutagenesis to identify the key BLIP-II^{WT} residues for Class A β-lactamase affinity and

specificity, and Fryszczyn *et al.* (2014)[222] which used similar methods to identify the key BLIP-II binding residues of TEM^{WT}. Seven BLIP-II surface residues were initially selected for mutation to AzF: residues selected for CNT interfacing were at positions A41, S43, D120, D159 and T213 (Figure 3.9). Two other BLIP-II variants were designed, initially to attempt UV-mediated photocrosslinking between BLIP-II and TEM by placing an AzF residue in the TEM binding surface of BLIP-II, with the phenyl azide close to reactive side chains of a bound TEM molecule such as phenylalanine or lysine: these residues were G49 and N50. Variant BLIP-II^{49AzF} was instead used to conjugate BLIP-II face-down on CNTs, via the β-lactamase binding site. CNTs functionalised with BLIP-II^{49AzF} will not bind TEM and were useful as a negative control in CNT binding and conductivity experiments. The AzF mutation in BLIP-II^{49AzF} was made at the periphery of the β-lactamase binding surface[222] to minimise attenuation of TEM binding.

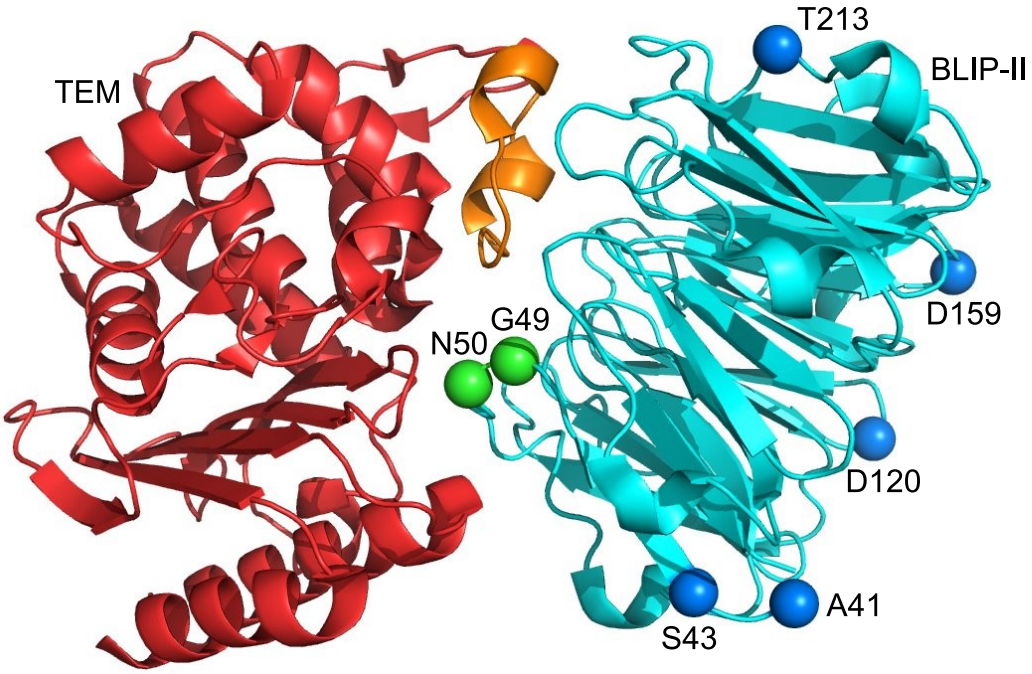


Figure 3.9 BLIP-II residues selected for mutation to AzF. Blue: mutations designed to conjugate BLIP-II to CNTs in a range of orientations. Green: mutations designed to photocrosslink to TEM, also used to conjugate BLIP-II to CNT via binding site, as a non-binding control. Orange: BLIP-II binding loop-helix on TEM. PDB: 1JTD[36].

Each of the mutants was modelled using the Phyre2 web portal (www.sbg.bio.ic.ac.uk/phyre2)[13], to estimate the impact of amino acid substitution on the protein fold. Mutation to tyrosine was used as a substitute for AzF, as the non-natural amino acid is unavailable in Phyre2. Output structures were aligned to the BLIP-II^{WT} crystal structure (PDB: 1JTD) to highlight predicted conformational changes, with two mutations predicted to affect β-lactamase binding by obstructing the binding site. G49Y/AzF, replacing glycine at the binding site, was predicted to displace the β-lactamase binding residue R286 to the protein interior (Figure 3.10a-b). However

mutation of R286 to alanine has previously been shown to have minimal impact on β-lactamase binding affinity (2-4 fold increase in K_D[147]) so the BLIP-II^{49AzF} mutation was kept. N50Y/AzF was found to replace an asparagine that is directly involved in TEM binding, so it was decided not to make this BLIP-II^{50AzF} variant.

The Phyre2-optimised structures of all BLIP-II variants were fitted to TEM^{WT} by docking simulation using the ClusPro web app (https://cluspro.bu.edu), to screen for TEM^{WT} binding issues such as major steric clashes. ClusPro generates ca. 100,000 protein-protein docking positions and scores them using balanced energy coefficients[14]. 1000 of the lowest-energy docked conformations are grouped into clusters of "neighbours" within 9 Å around a structure at the cluster centre. Clusters with the most members are considered to represent the most probable docked position. Again, tyrosine was used in place of AzF as the nnAA is not compatible with ClusPro. All designed BLIP-II mutants showed their largest cluster size and lowest binding energy in the standard BLIP-II^{WT}-TEM^{WT} binding orientation (Table 3.2), including BLIP-II^{49AzF} as shown in Figure 3.10c. Alignment of docked structures showed the most probable binding orientations for all variants closely match the BLIP-II^{WT}-TEM^{WT} complex, suggesting that all mutants retain wild-type TEM^{WT} binding properties.

Table 3.2 ClusPro docking simulation results for the Phyre2-optimised BLIP-II AzF variants with TEM^{WT}. Tyrosine was substituted for AzF due to limitations of the server. The native-type binding orientation of the BLIP-II-TEM^{WT} complex was most probable for all BLIP-II variants, with the most members per cluster (higher is better) and lowest binding energy score (kcal/mol, lower is better). Scores are shown for these and for the next best scoring orientations for comparison.

BLIP-II^{WT}	**Cluster members**	**Lowest energy (kcal/mol)**
Native	114	-896
Next best	85	-754
BLIP-II^{41Y}		
Native	111	-852
Next best	82	-741
BLIP-II^{43Y}		
Native	111	-906
Next best	76	-789
BLIP-II^{49Y}		
Native	103	-906
Next best	68	-803
BLIP-II^{120Y}		
Native	116	-893
Next best	79	-754
BLIP-II^{159Y}		
Native	116	-902
Next best	80	-768
BLIP-II^{213Y}		
Native	108	-815
Next best	103	-725

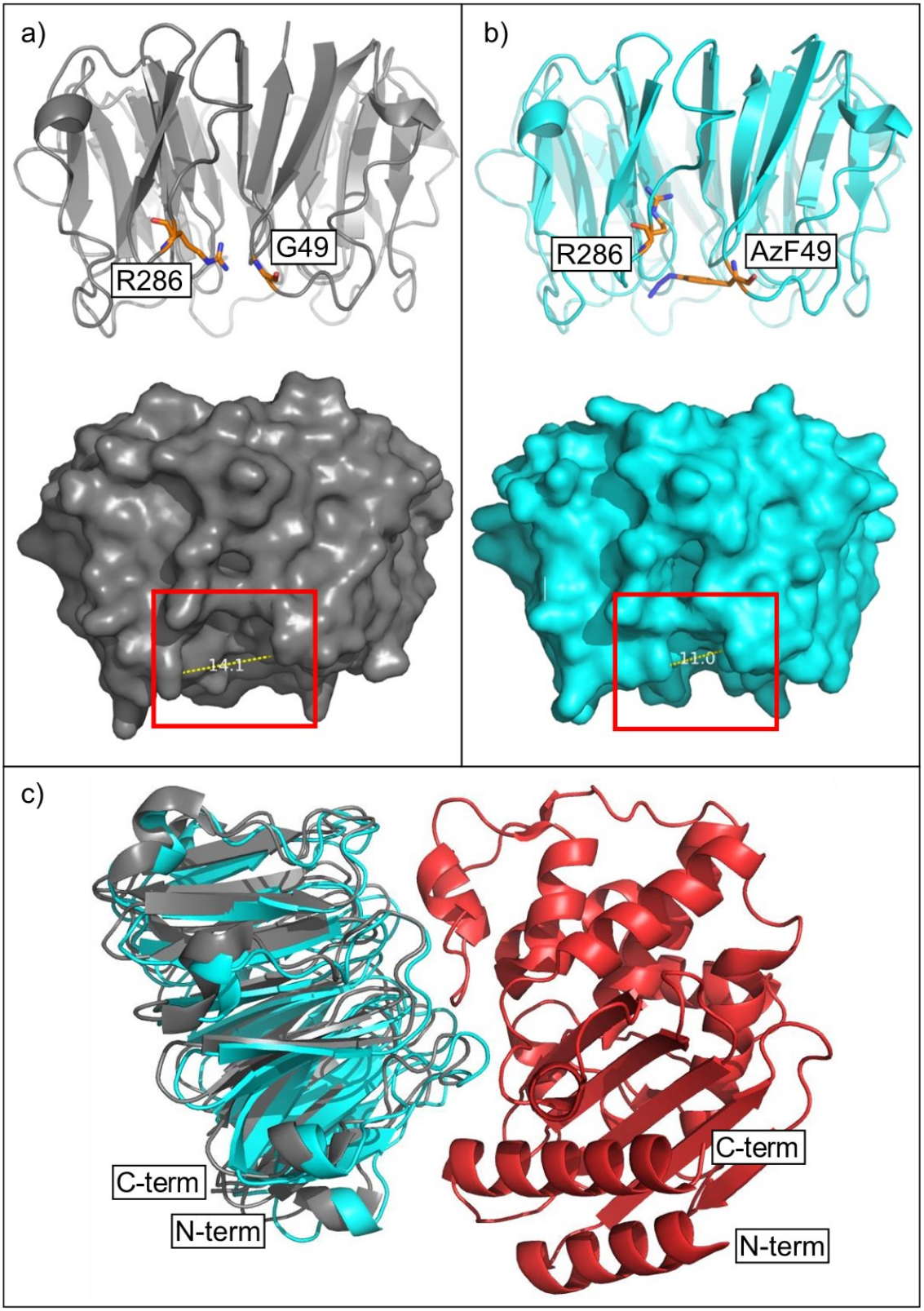
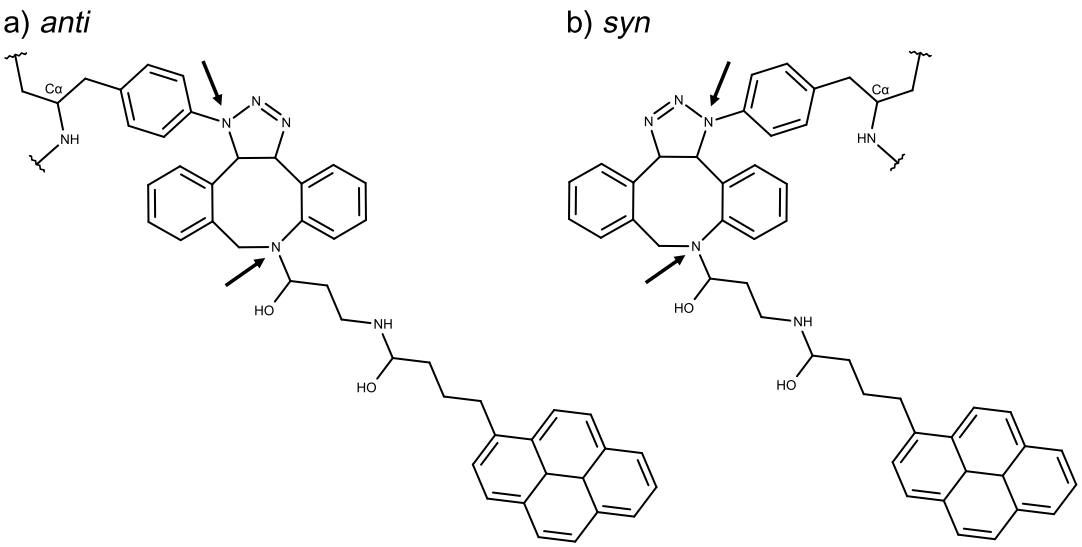


Figure 3.10 Possible conformational changes caused by BLIP-II^{49AzF} mutation. Binding sites indicated by red boxes. Grey: BLIP-II^{WT}; cyan: BLIP-II^{49AzF}; red: TEM^{WT}. a) BLIP-II^{WT} β-lactamase binding site width 14 Å; R286 is a surface residue and forms a hydrogen bond to TEM Y105. b) Predicted structure of BLIP-II^{49AzF}: the large AzF side chain is predicted to displace R286 to the protein interior away from its binding position and part of the binding cavity is narrowed to 11 Å. c) Predicted BLIP-II^{49AzF}-TEM^{WT} binding conformation overlaid on BLIP-II^{WT}-TEM^{WT} complex. PDB: 1JTD[36].

The *anti* form of the triazole-DBCO linker was used for modelling the AzF-triazole-DBCO-pyrene structure formed by the SPAAC reaction between genetically-encoded AzF on the protein and DBCO-pyrene on CNTs (Scheme 3.2). This regioisomer was used to construct these models because *in silico* geometry optimisation and energy minimisation simulations of the AzF-triazole-DBCO structure at the surface of the protein sfGFP predicted it is the more likely to form. The simulations, performed by Harley Worthy, found that the *anti*-isomer relaxed to a lower final potential energy of *circa* -1250 kJ/mol, compared to -1150 kJ/mol for the *syn*-isomer[227]. Quantum chemistry calculations were carried out by Prof Alexander Nemukhin, Moscow State University, to optimise the three-dimensional geometry of the AzF-triazole-DBCO-pyrene bridge; this predicted eight possible low-energy conformer structures. The conformer with the lowest free energy and a conformation compatible with the protein-CNT interface was integrated into the BLIP-II protein structure 1JTD[36] at the designed target residues to construct the final 3D models used in this chapter.



Scheme 3.2 The two possible regioisomers of the AzF-triazole-DBCO-pyrene linker construct. a) The *anti*-isomer was considered more thermodynamically probable to form, with a predicted potential energy of *circa* -1250 kJ/mol. b) The *syn*-isomer was less likely to form, with a predicted potential energy of *circa* -1150 kJ/mol.

The models were built using the PyMOL mutagenesis wizard to replace target residues in 1JTD with AzF, which was then joined to the optimised AzF-triazole-DBCO-pyrene bridge structure using the FUSE command. For each target residue, the conformation of the linker was manually adjusted by rotation of the AzF side chain bonds to move the linker to a position where it was accessible to adsorb to a S-SWCNT model, making no steric clashes with the protein. Rendered models illustrating possible orientations of each mutant adsorbed on a CNT surface were modelled and are shown in Figure 3.11. The six designed BLIP-II AzF mutants BLIP-II^{41AzF}, BLIP-II^{43AzF}, BLIP-II^{49AzF}, BLIP-II^{120AzF},

BLIP-II^{159AzF} and BLIP-II^{213AzF} were taken forward to the molecular biology stage, being made by site-directed mutagenesis of the BLIP-II^{WT} gene.

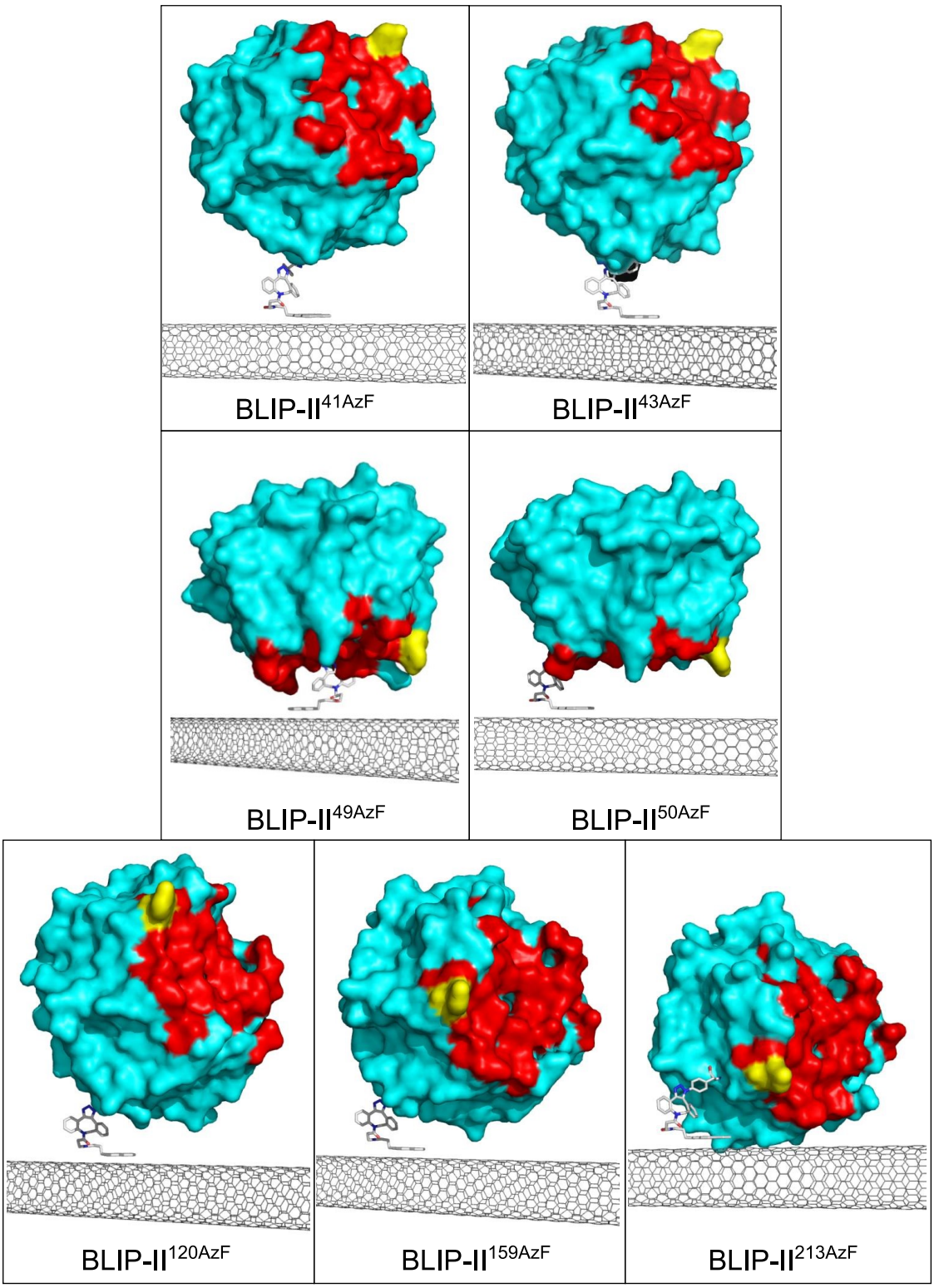


Figure 3.11 Possible orientations of BLIP-II AzF mutants attached to CNT surface via covalently conjugated pyrene. BLIP-II^{49AzF} and BLIP-II^{50AzF} will adsorb to CNT surfaces with their binding site sterically obstructed by the CNT. Yellow: Binding residue Y208 highlighted to contrast binding site orientations. Red: Other TEM binding residues. PDB: 1JTD[36]

3.3.3.2 Incorporation of designed mutations into BLIP-II by site-directed mutagenesis

The mutations discussed in the previous section were made using whole-plasmid PCR site-directed mutagenesis followed by ligation, to mutate the codons for the selected amino acid resides to TAG for AzF incorporation, as described in chapter 2.2.2.5. The original pET-BLIP-II[154] plasmid that was used to express BLIP-II^{WT}, and as a template for production of all BLIP-II mutants, was kindly supplied by Prof. Timothy Palzkill at Baylor College of Medicine, Houston, TX, USA.

Primer design for BLIP-II mutagenesis was difficult due to the high GC content of the gene (59%) requiring high annealing temperatures and causing primers to bind non-specifically during PCR. This was exacerbated by the repeating structure of the BLIP-II gene, which consists of seven repeated units of three β-strands and a small α-helix, in a β-propeller arrangement. This is a source of many potential off-target binding sites for PCR primers throughout the gene and designing acceptable primers to make TAG mutations in the selected locations required much trial and error and optimisation of primer sequences using Amplifx software[187] to simulate PCR results. All mutagenesis primers are listed in chapter 2.2.2.3. PCR mutagenesis was initially attempted using annealing temperatures for each primer as calculated by the NEB T_m Calculator web app for Q5 polymerase (https://tmcalculator.neb.com). Amplification was unsuccessful using these parameters, so T_m was lowered until amplification was observed. Amplification of BLIP-II mutant genes was successful using annealing temperatures between 56-64 °C, around 5-10 °C lower than originally recommended. NEB GC enhancer was used in all successful PCR reactions due to the high GC content of the BLIP-II gene.

PCR product was analysed by agarose gel electrophoresis to confirm amplification of pET-BLIP-II at the correct size (6.1 kpb) (Figure 3.12). Amplified plasmids were ligated and transformed into *E. coli* DH5α cells by electroporation. Cells were grown on kanamycin agar to select pET-BLIP-II variant transformants, with 10s to 100s of colonies observed on agar plates for each transformation. Colonies (up to four) from each transformation were sequenced to confirm the desired mutations and that no secondary mutations were present. Sequencing revealed cases of deletions and duplications of gene segments, due to the repeating nature of the BLIP-II gene, and absent or incomplete mutagenesis in some colonies but at least one instance of correct mutant generation was achieved for variants BLIP-II^{41AzF}, BLIP-II^{43AzF}, BLIP-II^{49AzF} and BLIP-II^{213AzF}. Bacteria containing verified TAG mutants of BLIP-II were cryopreserved in 25% glycerol at -80 °C for archiving purposes. Correct mutagenesis was not achieved for variants BLIP-II^{120AzF} and BLIP-II^{159AzF} despite attempts to optimise PCR conditions. The

TAG mutation was incorporated in the BLIP-II^{120AzF} gene, but all colonies sequenced showed single nucleotide deletions in the 5′ untranslated region, and repeating sections within the gene. All sequencing results for the BLIP-II^{159AzF} gene showed that the TAG mutation either did not occur, or partially occurred (TAC instead of TAG), despite repetition and fine-tuning of annealing temperatures and times. Successful mutation of these variants would require redesign of the mutagenesis primers, but it was decided to abandon BLIP-II^{120AzF} and BLIP-II^{159AzF} and instead express the four correctly-mutated BLIP-II AzF variants BLIP-II^{41AzF}, BLIP-II^{43AzF}, BLIP-II^{49AzF} and BLIP-II^{213AzF}.

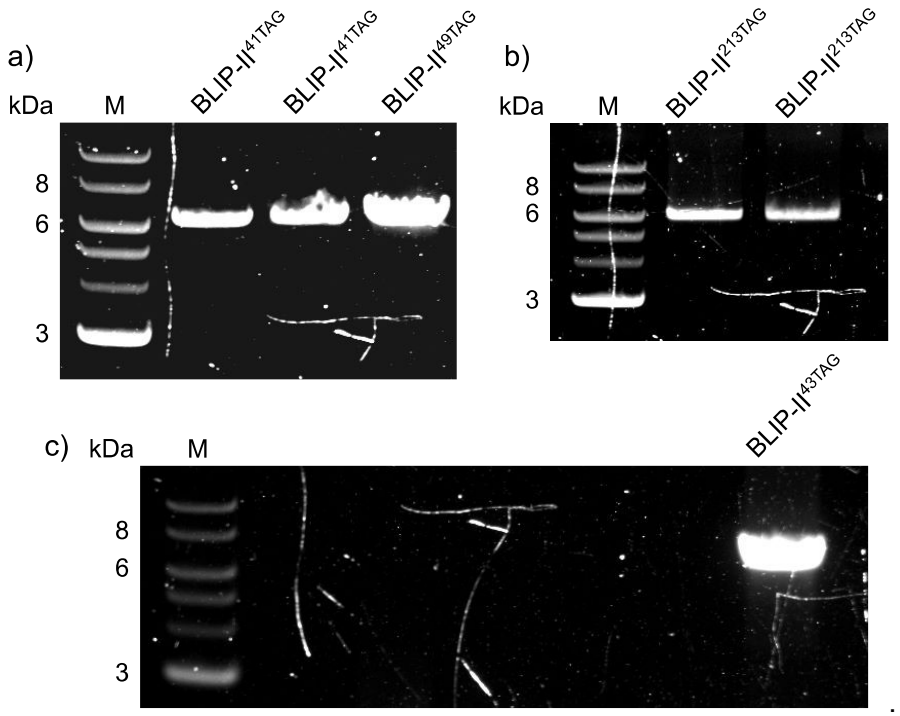


Figure 3.12 PCR mutagenesis of BLIP-II. Amplified products from successful site-directed whole plasmid PCR mutagenesis, placing TAG codon mutations for AzF incorporation into the BLIP-II gene in the pET-BLIP-II vector (6.1 kbp). a) Mutagenesis of BLIP-II^{41TAG} and BLIP-II^{49TAG} variants. b) BLIP-II^{213TAG} variant. c) BLIP-II^{43TAG} variant. Empty lanes are from other, unsuccessful PCR attempts. M: kilobase pair DNA marker ladder.

3.3.3.3 BLIP-II expression

To express BLIP-II^{WT}, *E. coli* BL21 (DE3) cells were transformed with the pET-BLIP-II plasmid, induced with IPTG and grown overnight at 37 °C. BLIP-II AzF variants were also co-transformed with the pEVOL-AzF plasmid[79] that encodes the aaRS and tRNA pair required for incorporation of AzF during protein synthesis at the ribosome in response to the TAG codon (see introduction, 1.1.3). Small samples of overnight cell culture were lysed using a freeze-thaw method and tested for BLIP-II^{WT} expression: TEM^{WT} was added to the lysed pET-BLIP-II cell culture and incubated for ten minutes, then nitrocefin was added to assess TEM^{WT} inhibition by BLIP-II^{WT}. The control sample lacking BLIP-II turned red indicating no inhibition of TEM^{WT}, while lysate from cells grown

with pET-BLIP-II inhibited TEM^{WT} hydrolysis of nitrocefin, remaining yellow as shown in Figure 3.13a. This result indicated expression of functional BLIP-II^{WT}. This nitrocefin assay was used to test subsequent *E. coli* cultures for expression of all BLIP-II AzF variants as it was found to be fast and convenient. Expression of variants BLIP-II^{41AzF}, BLIP-II^{43AzF}, BLIP-II^{49AzF} and BLIP-II^{213AzF} in *E. coli* was confirmed using this colorimetric inhibition test.

In the absence of the non-natural amino acid AzF in cell growth media, the TAG codon in mutant genes encoding BLIP-II operates as a standard amber stop codon, causing premature termination and generating a truncated protein. Truncation is likely to result in an incomplete or improperly folded, inactive protein that is subject to aggregation and degradation. To verify incorporation of AzF into the mutant BLIP-II variants, *E. coli* containing the pET-BLIP-II variant plasmids and pEVOL-AzF were grown and induced in media containing either 1) no AzF or 2) 1 mM AzF. Samples of cell lysates were analysed by SDS-PAGE to assess BLIP-II mutant expression: lysate from cells grown with AzF showed a prominent BLIP-II sized band at the expected size of ca. 30 kDa, while those grown without AzF lacked this band, confirming successful AzF incorporation and expression of BLIP-II AzF variants. Figure 3.13b shows a representative example with variants BLIP-II^{41AzF} and BLIP-II^{49AzF}.

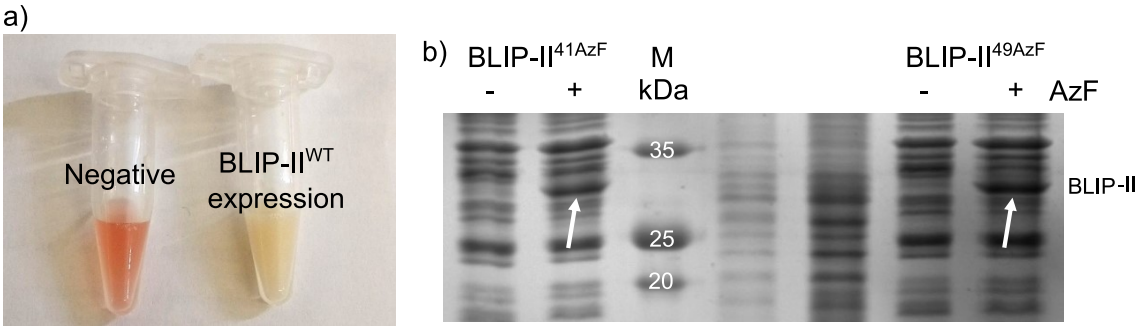


Figure 3.13 Representative examples of BLIP-II expression tests. a) Colourimetric test for BLIP-II^{WT} expression in *E. coli* cells. Left (-): Nitrocefin, TEM^{WT} and *E. coli* cells: the mixture turned red indicating nitrocefin hydrolysis. Right (+): Nitrocefin, TEM^{WT} and BLIP-II^{WT}-expressing *E. coli* cells: no colour change due to inhibition of TEM by BLIP-II. b) BLIP-II AzF mutant expression in *E. coli*. BLIP-II mutant plasmids A41TAG and G49TAG grown with AzF (+) and without (-). BLIP-II (arrowed) is expressed only in the presence of AzF showing that the amino acid is being incorporated. Unlabelled lanes contained a failed expression. M: molecular weight marker ladder.

3.3.3.4 BLIP-II purification

The pET-BLIP-II plasmid encodes a variant of BLIP-II^{WT} with a C-terminal hexahistidine tag (His-tag) for purification using immobilised metal affinity chromatography (IMAC) using nickel or cobalt charged gel beads. Initially an ÄKTAprime plus FPLC system (Cytiva) was used to purify BLIP-II^{WT}, eluting the protein from a HiTrap TALON cobalt

IMAC column (Cytiva) with an increasing imidazole gradient. Elution fractions were analysed by SDS-PAGE (Figure 3.14) and those containing a BLIP-II^{WT} sized protein at ca. 30 kDa were concentrated then further purified by size exclusion chromatography (SEC) (Figure 3.15), that will also remove any residual imidazole from the sample.

For subsequent purifications of the BLIP-II AzF variants, a simplified protocol was developed using a home-made gravity IMAC column containing HisPur cobalt agarose (Thermo Fisher), with elution in a single step. SDS-PAGE analyses of elution fractions from this cobalt column are shown in Figure 3.16. This method was faster and more convenient than using FPLC, and yielded BLIP-II at purity ≥ 99% in most cases, measured by band densitometry using ImageJ software[202,219]. BLIP-II variants that eluted from the home-made IMAC column with some contaminating proteins were further purified using SEC. If SEC was not required, imidazole was removed using a PD10 desalting column (Cytiva). The home-made IMAC cobalt column also had the advantage of being small enough to set up in an enclosed cabinet, to shield phenyl azide side chains from photolysis by ambient light.

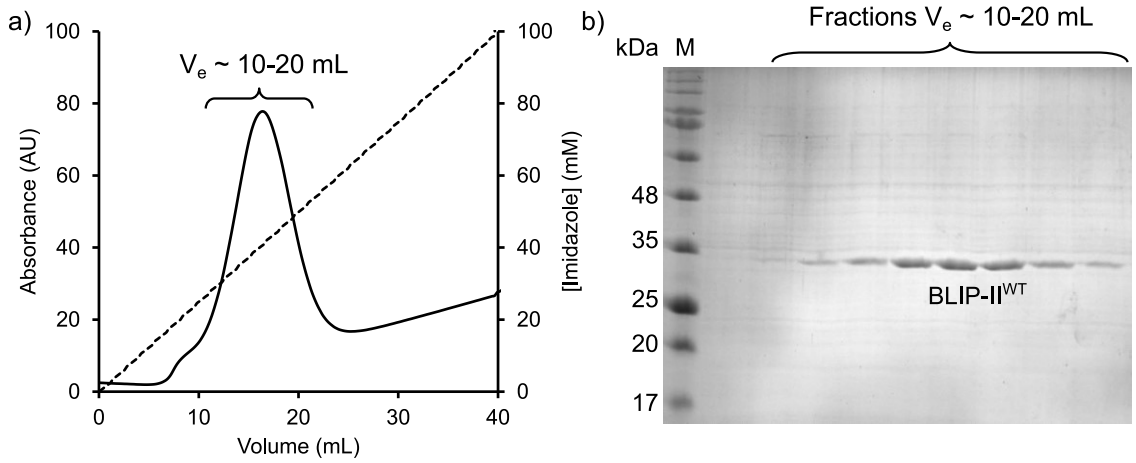


Figure 3.14 Cobalt affinity purification of BLIP-II^{WT}. a) The 280 nm absorbance peak (solid line, left y-axis) represents elution of BLIP-II^{WT} between V_e 10-20 mL with an increasing imidazole gradient (dotted line, right y-axis). b) SDS-PAGE analysis of fractions from elution of BLIP-II from cobalt affinity column. M: molecular weight marker ladder.

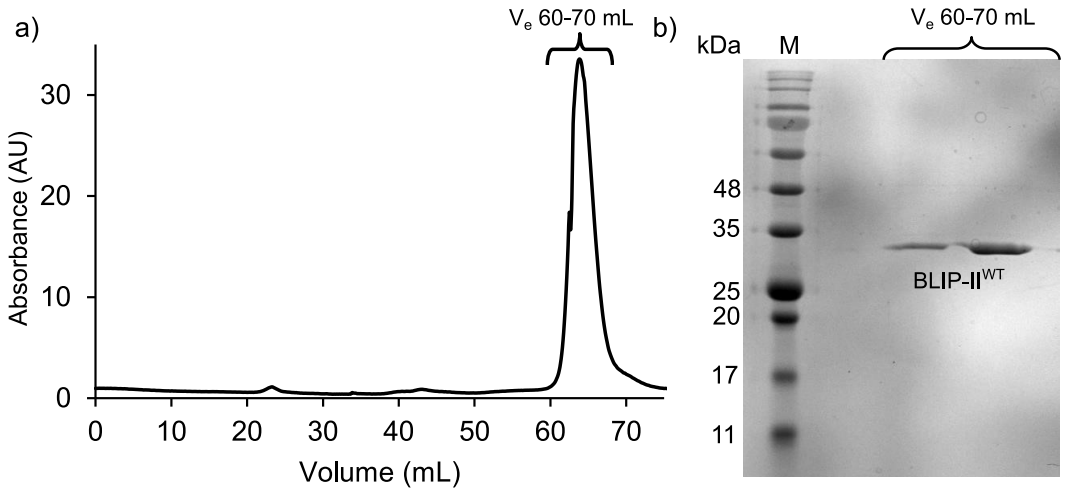


Figure 3.15 a) Size exclusion purification of BLIP-II^{WT} (S75 column). The 280 nm absorbance peak represents elution of BLIP-II^{WT} at V_e 60-70 mL. b) SDS-PAGE analysis of elution fractions. M: molecular weight marker ladder.

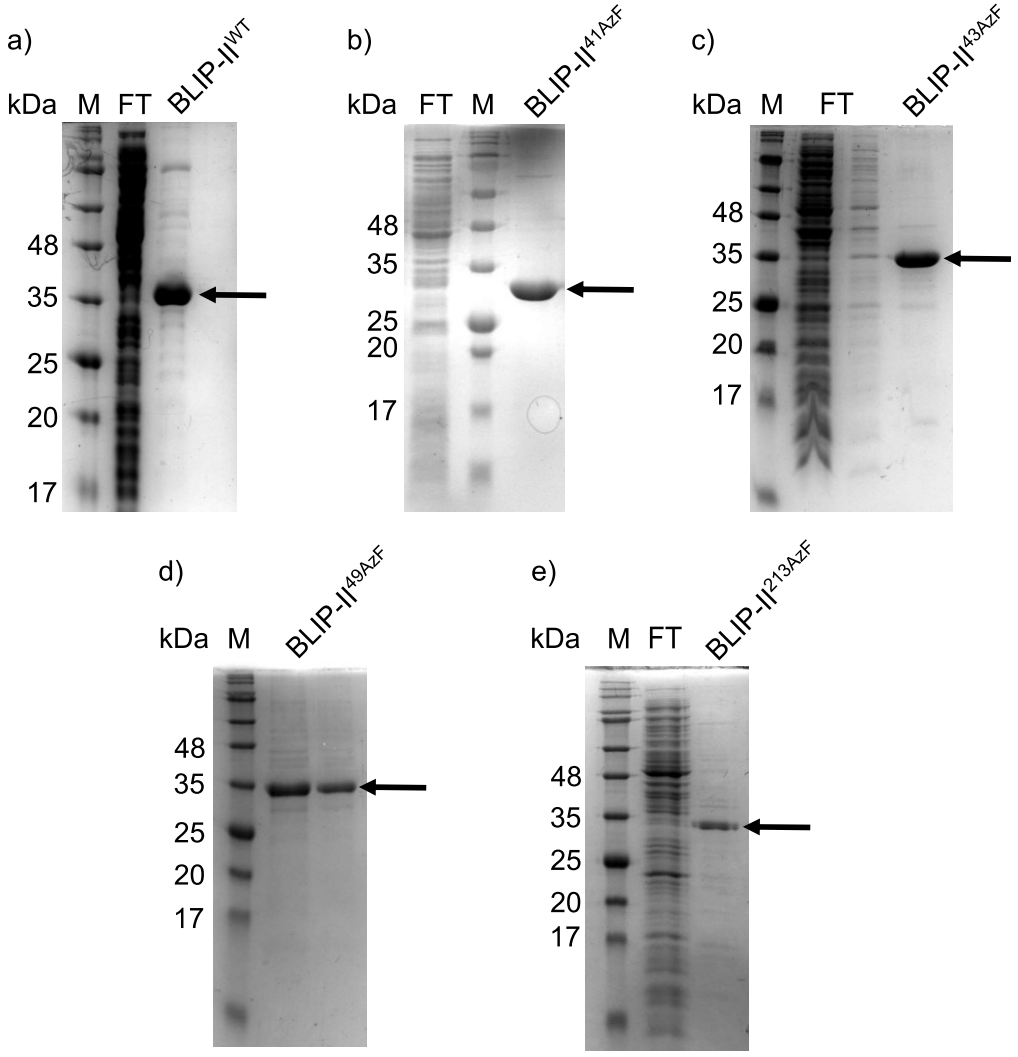


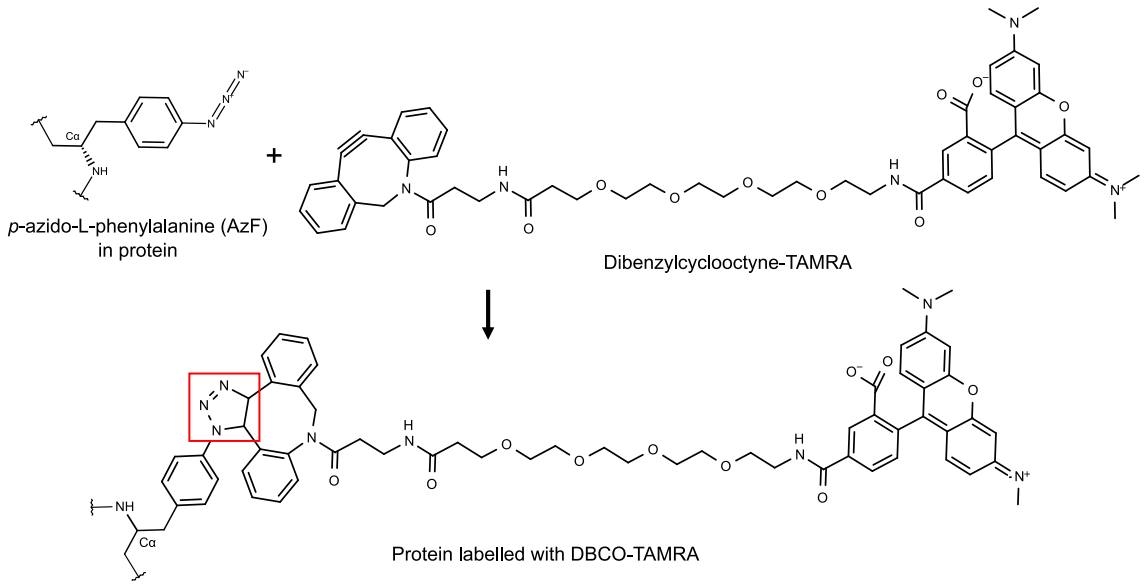
Figure 3.16 BLIP-II variants: one-step IMAC elution fractions. a) BLIP-II^{WT} b) BLIP-II^{41AzF} c) BLIP-II^{43AzF} d) BLIP-II^{49AzF} e) BLIP-II^{213AzF}. BLIP-II variants were purified to typically ≥ 99% purity using a home-made cobalt affinity gravity flow column, with one-step elution (BLIP-II band arrowed). FT: unbound proteins in flow-through fractions. M: molecular weight marker ladder.

3.3.4 BLIP-II variants: Modification and characterisation

3.3.4.1 BLIP-II phenyl azide click chemistry

Before the new BLIP-II AzF variants were used for CNT-interfacing, the presence and reactivity of their genetically-encoded AzF residues were assessed: BLIP-II^{41AzF}, BLIP-II^{43AzF,} BLIP-II^{49AzF} and BLIP-II^{213AzF} were labelled with the azide-specific strained-ring alkyne fluorophore DBCO-TAMRA (structure and reaction scheme with AzF in Scheme 3.3). DBCO-TAMRA dye was incubated with BLIP-II AzF variants for one hour at fivefold molar excess over protein. BLIP-II^{WT} and TEM^{WT} were also incubated with DBCO-TAMRA as negative controls: these should not react. Samples were then analysed by SDS-PAGE and the gel was imaged by UV illumination to visualise fluorescent labelled proteins. All BLIP-II AzF variants reacted with DBCO-TAMRA and were labelled, visible as fluorescent bands at ca. 35 kDa, while BLIP-II^{WT} and TEM^{WT} did not react (Figure 3.17), this demonstrated that the labelling was specific to the azide groups of the mutants.

To estimate the efficiency of DBCO-TAMRA labelling of the BLIP-II AzF variants, the four proteins were again incubated with a fivefold molar excess of DBCO-TAMRA. Samples were analysed by SDS-PAGE with a shorter electrophoresis run time to prevent the excess unreacted DBCO-TAMRA fully migrating out of the gel. While this left a large fluorescent band near the bottom of the gel, it did allow a ratio of BLIP-II AzF reacted versus unreacted fluorescence species to be estimated (Figure 3.18a). ImageJ was used to estimate SPAAC reaction efficiency from SDS-PAGE images, comparing the band intensity of the unreacted DBCO-TAMRA bands to the intensity of the protein-dye bands. Images with minimal exposure time were used to control for saturation of signal from the bands. ImageJ plots the pixel density of each SDS-PAGE gel lane as an X-Y curve with each band appearing as a peak of the curve. The area under these peaks can be measured (in arbitrary units) to give a numerical value for the intensity of each band, enabling bands to be quantified for comparison (Figure 3.18b). Band intensity values for the labelled proteins were divided by the values for the unbound excess dye bands, then multiplied by five to account for the fivefold excess of DBCO-TAMRA used. This gave the percentage of protein that had become fluorescently labelled by SPAAC. All samples tested contained a high population of proteins with functional azide groups required for downstream applications. Estimates of SPAAC reaction efficiency for the four BLIP-II AzF variants produced were BLIP-II^{41AzF} = 51%, BLIP-II^{43AzF} = 44%, BLIP-II^{49AzF} = 55% and BLIP-II^{213AzF} = 62%. These figures are consistent with published values for azide-alkyne labelling of proteins[228].



Scheme 3.3 SPAAC click reaction for bioorthogonal labelling of AzF residues with DBCO-TAMRA fluorophore. The azide of an AzF residue in the protein of interest reacts with the alkyne in the strained ring of DBCO to form a stable covalent triazole bond (red box) to the TAMRA fluorophore.

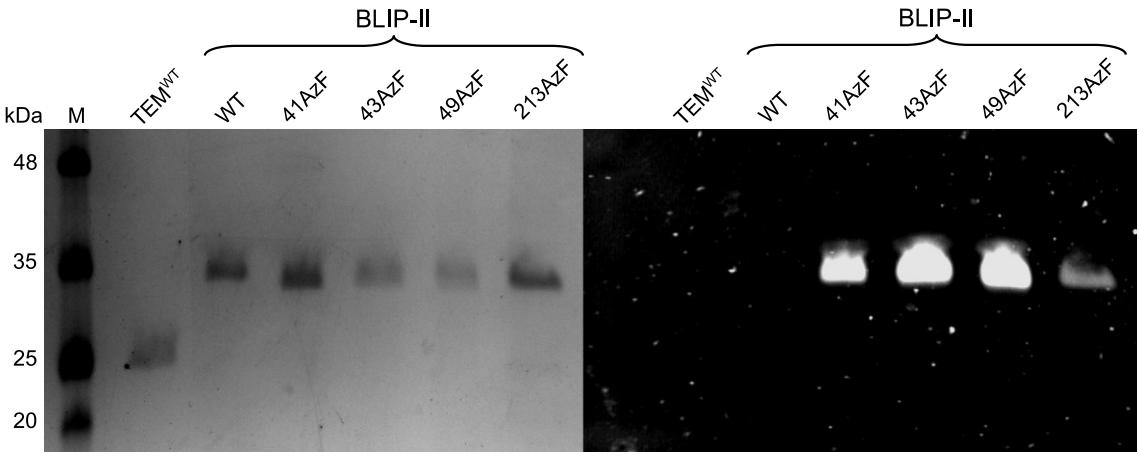


Figure 3.17 Fluorescent SPAAC labelling of BLIP-II AzF mutants. Gel imaged by staining (left) and UV illumination (right), showing TEM^{WT}, BLIP-II^{WT} and four BLIP-II AzF mutants. Only AzF mutants were labelled with the alkyne fluorescent dye, showing that their phenyl azide groups are functional. M: molecular weight marker ladder.

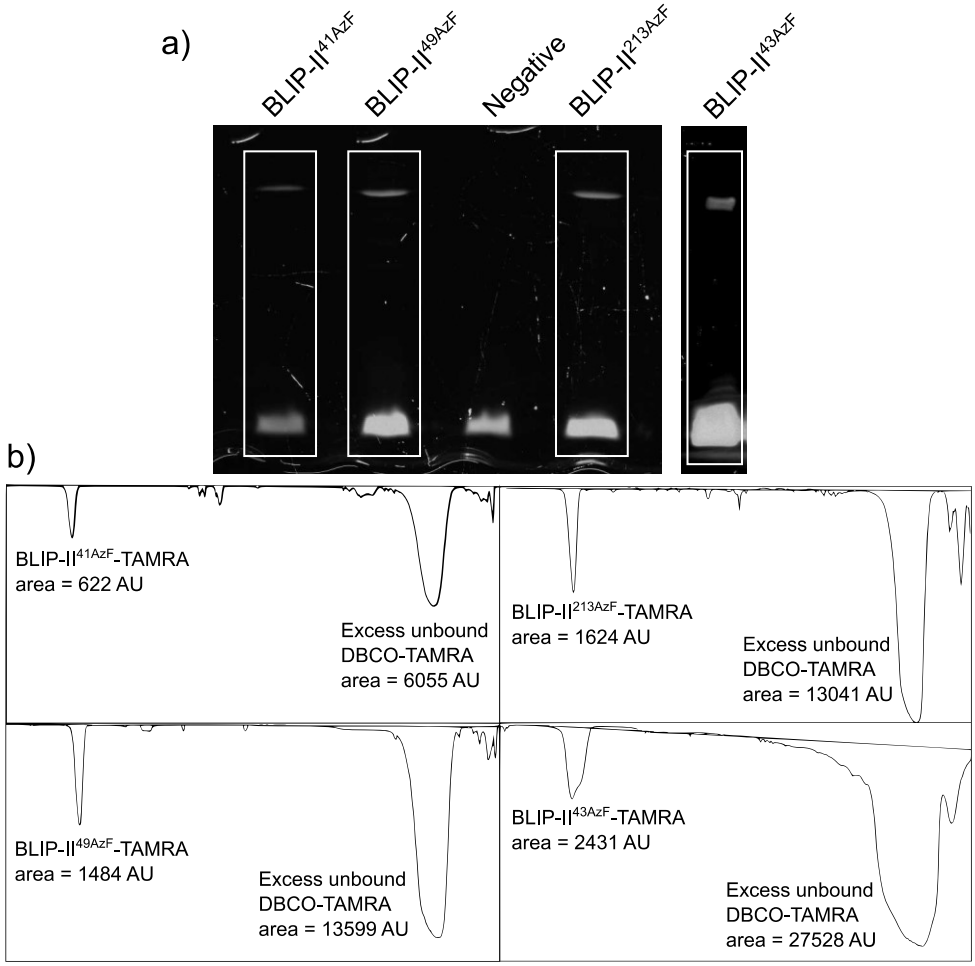


Figure 3.18 Estimation of BLIP-II AzF variants SPAAC reaction efficiency with fivefold molar excess of DBCO-TAMRA dye. a) UV-illuminated SDS-PAGE gels showing BLIP-II^{41AzF}, BLIP-II^{49AzF}, BLIP-II^{213AzF} and BLIP-II^{43AzF} labelled with DBCO-TAMRA dye (upper bands) and excess unreacted dye (lower bands). Band intensities were compared to estimate labelling efficiency (boxed areas). b) ImageJ analysis of SDS-PAGE gel. Each lane on gel was plotted as a profile to compare band densities, displayed as inverse peaks. ImageJ does not label axes and values are arbitrary for comparison purposes only. The area above each curve was measured using integration, to estimate the comparative intensity of bands for TAMRA-labelled protein vs. unbound excess DBCO-TAMRA. These values were used to calculate estimated labelling efficiency.

3.3.4.2 BLIP-II - TEM binding and inhibition

One of the reasons BLIP-II was chosen for NTFET biosensor construction is the extremely high binding affinity of BLIP-II for class A β-lactamases. The mutations of BLIP-II residues to AzF were introduced away from the β-lactamase binding site so were not expected to reduce TEM^{WT} binding affinity. The exception of course is BLIP-II^{49AzF} which is adjacent to the binding site; the variant was intended to conjugate to CNT surfaces face-down via the binding site (see previous Figure 3.11), providing a negative control that can functionalise CNTs but then is unable to bind TEM^{WT} due to the obstructed binding site. This is a useful control as it can differentiate between CNT conduction effects caused by specific TEM^{WT}-BLIP-II binding at the CNT surface and

those caused by nonspecific interactions e.g. adsorption of TEM^{WT} on CNTs or non-specific BLIP-II-TEM^{WT} interactions.

To compare the TEM^{WT}-binding properties of $BLIP\text{-}II^{41AzF}$ with $BLIP\text{-}II^{WT}$, both proteins were incubated with a 1.5-fold molar excess of TEM^{WT} then analysed using native (non-denaturing) PAGE. This electrophoresis method does not use detergent or high temperatures to unfold proteins and allows protein complexes to remain bound by non-covalent interactions; protein-protein complexes appear as higher molecular weight bands on the gel. Both $BLIP\text{-}II^{WT}$ and $BLIP\text{-}II^{41AzF}$ formed complexes with TEM^{WT}, evidenced by the formation of a band of higher molecular weight than either individual protein; the absence of bands equivalent to BLIP-II implies full complexation. This implies that all $BLIP\text{-}II^{WT}$ and $BLIP\text{-}II^{41AzF}$ bound to TEM^{WT} (Figure 3.19).

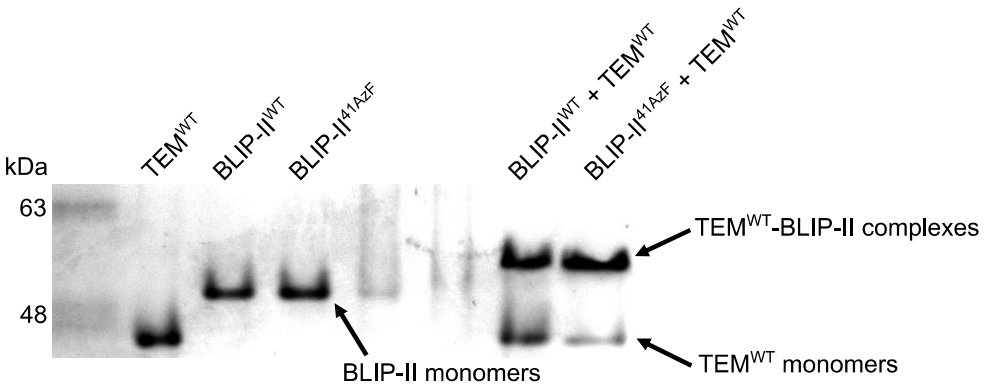


Figure 3.19 Native PAGE gel analysing TEM-BLIP-II binding. A band of higher molecular weight than either protein forms when $BLIP\text{-}II^{WT}$ or $BLIP\text{-}II^{41AzF}$ and TEM^{WT} are incubated together, indicating complex formation. BLIP-II monomer is absent in these samples, suggesting all BLIP-II had formed complexes with TEM. M: molecular weight marker ladder (molecular masses inaccurate due to non-denaturing conditions).

To quantitatively determine the effect of the four mutations on BLIP-II's ability to bind (and thus inhibit) TEM^{WT}, enzyme inhibition assays were used to measure the apparent inhibition constants (K_i^{app}) compared to $BLIP\text{-}II^{WT}$. K_i is equivalent to dissociation constant (K_D) but in the context of enzyme inhibition. Absolute binding parameters for BLIP-II-β-lactamase interactions are difficult to determine by normal methods as their dissociation constant is so low (80-540 fM[147,148]); working at protein concentrations in this range are impractical for our requirements so a comparative analysis with $BLIP^{WT}$ will allow an apparent K_i to be determined. This will allow the direct effect of the mutation on binding affinity to be assessed. The $BLIP\text{-}II^{WT}$-TEM^{WT} dissociation constant (K_D) has previously been determined to be 480 fM[147] using separately measured association and dissociation rate constants. However, these experiments require up to one week incubation of BLIP-II-TEM^{WT} complexes in an excess of the inactive TEM^{166Ala} mutant and was considered unnecessary for our purpose of comparing the variants to $BLIP\text{-}II^{WT}$.

The assay measured the initial rate of nitrocefin hydrolysis by TEM^{WT}, by recording the absorbance increase on hydrolysis of the β-lactam ring amide bond. Initial rates were recorded using 0.3 nM TEM with increasing concentrations of BLIP-II until full inhibition had been achieved for all BLIP-II variants (Figure 3.20). The titration values were plotted and fitted to the Morrison Equation (Equation 2.3, chapter 2.2.8.6)[197] using GraphPad Prism software[196] to estimate the K_i^{app} for each interaction. Results are shown in Table 3.3.

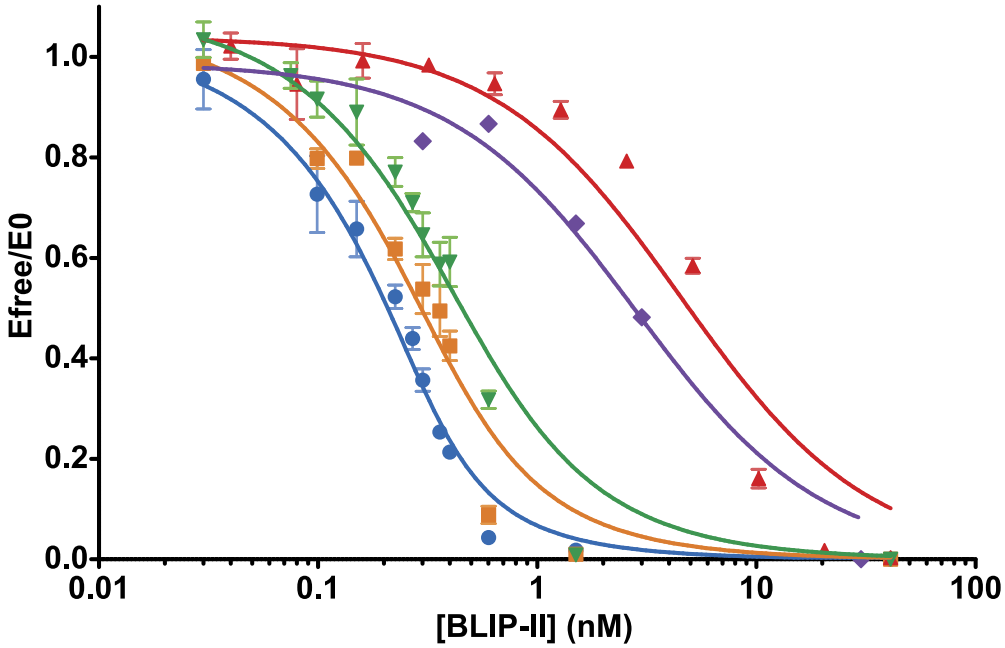


Figure 3.20 Inhibition of TEM^{WT} by BLIP-II variants. In order of binding affinity, left to right: Blue: BLIP-II^{WT}. Orange: BLIP-II^{41AzF}. Green: BLIP-II^{213AzF}. Purple: BLIP-II^{43AzF}. Red: BLIP-II^{49AzF}. Values were fitted to the Morrison tight binding equation to calculate K_i^{app} (Table 3.3). **Efree/E0** is the proportion of free enzyme remaining. Error bars are SEM of three biological replicates.

Table 3.3 Apparent dissociation constants (K_i^{app}) of TEM^{WT} with BLIP-II variants with standard error of the mean. Parameters including goodness of fit values were calculated by fitting results to the Morrison tight binding equation using nonlinear regression in GraphPad Prism.

BLIP-II	BLIP-II^{WT}	BLIP-II^{41AzF}	BLIP-II^{43AzF}	BLIP-II^{49AzF}	BLIP-II^{213AzF}
K_i^{app} (pM) (±SE)	32 ± 4	78 ± 12	1724 ± 378	2839 ± 406	157 ± 20
s_{yx} [a]	0.06	0.08	0.06	0.10	0.08
Pseudo-R^2 [b]	0.96	0.94	0.98	0.93	0.93
95% CI	23 to 41	53 to 102	676 to 2772 [c]	2014 to 3665	116 to 198

[a] Standard error of the estimate, in units of Y.
[b] Pseudo-R^2 for nonlinear regression as per method 1 in Kvålseth (1985)[223].
[c] Wide confidence interval due to low number of data points.

The apparent TEM inhibitory constant (K_i^{app}) for BLIP-II^{WT} was calculated at 32 ± 4 pM. This is higher (lower affinity) than the previously reported K_D of 0.5 pM[147]. All four AzF variants of BLIP-II retained tight TEM binding and inhibitory function, with apparent inhibition constants ranging over two orders of magnitude between 78 pM to 2.8 nM. While this appears a wide variation, even the lowest affinity BLIP-II variant here retains potent TEM binding and inhibitory function, by the standards of most protein-protein interactions[229] and small molecule inhibitors[230].

The BLIP-II^{41AzF} variant (K_i^{app} = 78 ± 12 pM) inhibited TEM with a similar affinity as BLIP-II^{WT}. This mutation is located at the extreme N-terminus of BLIP-II and while the crystal structure of the protein, PDB:3QHY[155], shows that it forms two hydrogen bonds to residues G65 and V66, it is still the most disordered part of the protein with backbone temperature factors (B-factors) ca. 40-50 Å^2 [155]. This disorder probably explains why mutating A41 to AzF does not significantly impact TEM binding.

The apparent inhibition constant of BLIP-II^{43AzF} was calculated to be K_i^{app} = 1.72 ± 0.38 nM, *circa* fifty-fold lower affinity than wild type. This estimate is less certain than the other variants as data from a preliminary inhibition assay with very few data points was used. This assay was unable to be repeated with BLIP-II^{43AzF} as more protein could not be made because European suppliers of AzF suspended sales for most of 2019 as AzF was found to be explosive[231]. While the BLIP-II^{43AzF} mutation was designed to avoid residues that would disrupt protein stability or function, its attenuated binding affinity suggests some negative effect on native conformation. Residue S43 is located close to the N-terminus at the start of β-strand 1 of the β-propeller structure. In BLIP-II^{WT}, the backbone carbonyl and amide groups of S43 are hydrogen bonded to the backbone of residue K310 near the C-terminus, forming part of a β-sheet. It is unclear why the BLIP-II^{43AzF} mutation should impact binding affinity in this way; the only direct link from residue 43 to the β-lactamase binding site is that 43 is at the far end of β-strand 1 from a loop containing four residues implicated in TEM^{WT} binding (N50, D52, W53, T57)[147]. Mutation of the polar hydroxymethyl group of serine 43 to the much larger phenyl azide side chain of AzF appears to cause some disruption to BLIP-II that was not predicted during *in silico* design. It is possible the mutation causes some unknown change in the molecular dynamics of BLIP-II^{43AzF}, transmitted through β-strand 1 to affect the TEM^{WT} binding characteristics of the binding site. However, this unforeseen reduction in binding affinity may be useful for functionalising CNTs with a lower affinity BLIP-II variant. This could enable the sensing of on-off binding effects instead of the one-way binding events typical of BLIP-II^{WT} and the low picomolar binding affinity BLIP-II AzF variants.

The variant BLIP-II^{49AzF} had the lowest TEM^{WT} binding affinity of the four analysed (K_i^{app} = 2.84 ± 0.4 nM), about a hundred-fold lower than BLIP-II^{WT}. However, this still represents tighter binding than was expected, the G49 mutation is near the BLIP-II-TEM binding interface because it was originally designed to crosslink the proteins by UV mediated nitrene addition to the nearby tyrosyl ring of TEM Y105[232]. This position was selected to avoid residues previously shown to be important to β-lactamase binding affinity[147], but the addition of a large, negatively charged AzF phenyl azide side chain in place of glycine near the binding surface was still likely to affect binding interactions. Residue G49 is located at the start of a TEM binding loop containing N50, D52, W53 and T57 and the bulk of the large phenyl azide chain may have distorted the conformation of this region of the binding face. This is likely to be the cause of the higher K_i^{app} recorded for BLIP-II^{49AzF}. It is important to note that though BLIP-II^{49AzF} affinity for TEM^{WT} is a hundred-fold lower than BLIP-II^{WT}, the K_i^{app} of 2.8 nM still represents an extremely tight interaction. The most important β-lactamase binding residues in BLIP-II^{WT} form a ring of aromatic residues that surround the loop-helix domain of TEM^{WT} when bound: Y73, W152, Y191, F230 and W269 (Figure 3.21). Mutation of any of these residues to alanine reduces TEM^{WT} affinity by 200 - 18,000-fold[147]. A likely explanation for the retention of TEM^{WT} binding by BLIP-II^{49AzF} is that the bulky AzF mutation at residue 49 is distant enough from the core β-lactamase binding residues to not disable binding. The high affinity and broad range of specificity of BLIP-II-β-lactamase binding is a result of multiple BLIP-II residues working in concert so the loss of a few of these interactions does not completely extinguish binding affinity.

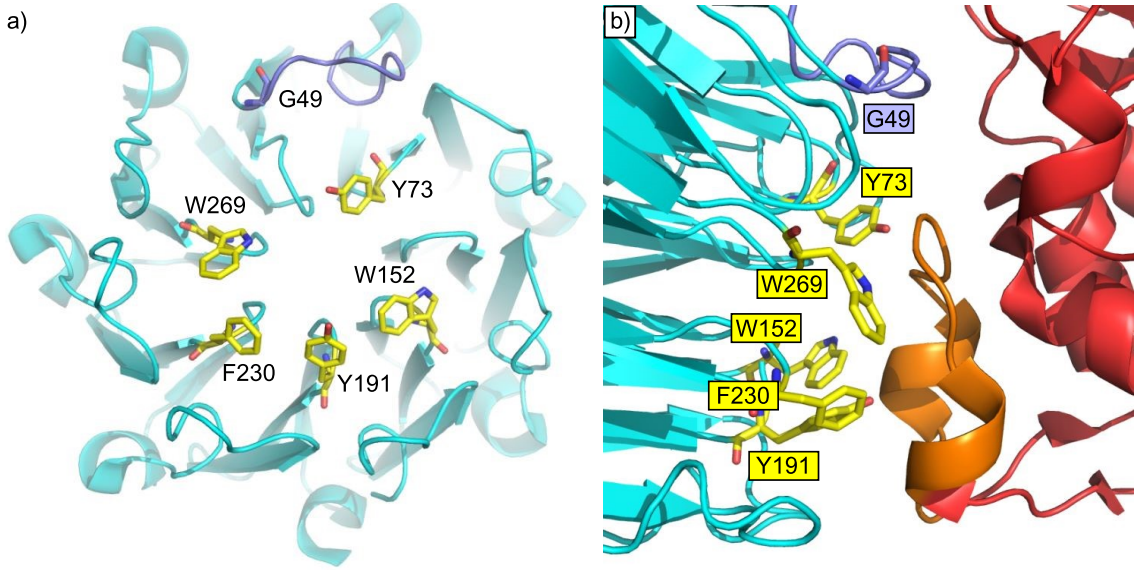


Figure 3.21 Location of BLIP-II^{WT} (cyan) residue G49 and its neighbouring binding loop (purple) relative to main binding interfaces. a) BLIP-II^{WT} residue G49 is relatively distant from the ring of aromatic residues most crucial to TEM^{WT} binding (yellow). b) BLIP-II^{WT}-TEM^{WT} key binding interface (red: TEM^{WT}; orange: TEM^{WT} loop-helix binding domain). Residue 49 is removed from this key interface, possibly explaining the retention of binding observed for BLIP-II^{49AzF}.

Analysis of the BLIP-II^{213AzF} variant showed TEM^{WT} binding affinity similar to BLIP-II^{WT}, with an apparent dissociation constant of 157 ± 20 nM. This mutation is on the opposite side of BLIP-II to A41/S43 (see Figure 3.9), positioned with its side chain facing outwards on a comparatively ordered surface loop (B-factor ca. 13-20 Å^2)[155] that has minimal interaction with TEM^{WT}. The retention of extremely tight TEM binding for this variant suggests that the BLIP-II^{213AzF} mutation does not significantly affect folding into the native conformation or inherent binding.

The four phenyl azide BLIP-II variants described here were all confirmed to have high binding affinity for TEM, with the BLIP-II^{41AzF} and BLIP-II^{213AzF} variants showing essentially the same affinity as BLIP-II^{WT}. These variants have their phenyl azide reactive handles at opposite sides of the protein, enabling the binding of TEM^{WT} in two distinct orientations at the CNT surface to sample two discrete electrostatic faces. BLIP-II^{43AzF} offers approximately the same CNT conjugation orientation as BLIP-II^{41AzF}, but with a TEM^{WT} binding affinity 22-fold lower. This variant may prove useful to avoid sensor saturation at higher β-lactamase concentrations. BLIP-II^{49AzF} offers a variant capable of conjugating to CNT surfaces, then being incapable of binding TEM^{WT} due to a blocked binding site. This will provide a useful confirmation that any observed conductivity changes through the NTFET are caused by specific BLIP-II-TEM^{WT} interactions.

3.3.5 Interfacing engineered β-lactamase inhibitor protein BLIP-II with carbon nanotubes via conjugated intermediates for binding event sensing

Functionalisation of CNTs with AzF variant BLIP-II proteins was undertaken, in collaboration with the Palma lab at QMUL, London; AFM and electrical raw data in this section was recorded by Xinzhao Xu. Data was analysed by both myself and Xinzhao Xu. This approach exploits the SPAAC click chemistry properties of the AzF phenyl azide to conjugate BLIP-II to an alkyne-bearing intermediate DBCO-pyrene adsorbed to the surface of CNTs (see Scheme 3.1). This method has the advantage of not requiring the introduction of defects into pristine CNTs, as occurs when oxidising CNTs to provide carboxyl reactive groups on the CNT sidewall. Oxidation of CNTs with nitric acid or similar harsh oxidisers creates vacancies in the graphitic carbon network which affect conductivity, with limited control over the extent of this damage[233]. Pyrene adducts such as 1-aminopyrene (1-AP)[234] or pyrene-*N*-hydroxysuccinimide ester (pyrene-NHS-ester)[125] have been used for functionalising CNTs for at least twenty years by adsorption to the surface, to both improve CNT water solubility and provide reactive groups for conjugating molecules to the surface. However, when linking proteins to pyrene-NHS-ester coated CNTs, the covalent linkage is made to any primary amine on the protein,

offering no control over the final orientation of the protein on the CNT. Using a genetically-encoded AzF as the conjugation point of the protein to a DBCO-pyrene coated CNT overcomes this problem by enabling a single, defined point of attachment to the CNT surface for every protein in the system. The aim was to conjugate BLIP-II proteins to the CNT surface through a defined and consistent linkage position, with the objective of detecting consistent signal changes in the current through the CNT on binding of β-lactamase partner proteins near the surface. A CNT-based β-lactamase sensing device functionalised with a population of BLIP-II proteins bound in a wide variety of orientations would likely give an ensemble signal due to averaging of the mixed population, with subtle binding-related signal features lost in the noise.

3.3.5.1 Sidewall functionalisation of NTFET devices with BLIP-II AzF variants via DBCO-pyrene click chemistry

Before attempting to functionalise CNTs with BLIP-II-AzF-pyrene proteins, tests were carried out to determine whether conjugation to DBCO-pyrene would degrade the binding affinity of BLIP-II variants for TEM^{WT}. $BLIP\text{-}II^{41AzF}$ and $BLIP\text{-}II^{213AzF}$ were incubated with a tenfold molar excess of DBCO-pyrene for one hour then analysed by SDS-PAGE, imaging with both Coomassie staining and UV illumination, as pyrene is fluorescent. Labelling of BLIP-II variants with pyrene was confirmed by a weakly fluorescent blue band matching the size of BLIP-II (ca. 30 kDa, Figure 3.22a); labelled variants were called $BLIP\text{-}II^{41Pyr}$ and $BLIP\text{-}II^{213Pyr}$. To confirm the pyrene-labelled BLIP-II variants were still able to bind TEM^{WT}, they were incubated with equimolar TEM^{WT} for ten minutes then analysed by native PAGE (non-denaturing) to assess binding. Both $BLIP\text{-}II^{41Pyr}$ and $BLIP\text{-}II^{213Pyr}$ formed complexes with TEM^{WT} (Figure 3.22b), as seen previously with unlabelled BLIP-II (section 3.3.4.2), showing that DBCO-pyrene conjugation had not significantly reduced BLIP-II variants' TEM^{WT} binding properties. Absorbance spectra were taken for both proteins after desalting, the spectra showed characteristic absorbance features typical of both BLIP-II (single peak at 280 nm) and pyrene (triple peak between 300-350 nm[235]), further supporting successful conjugation.

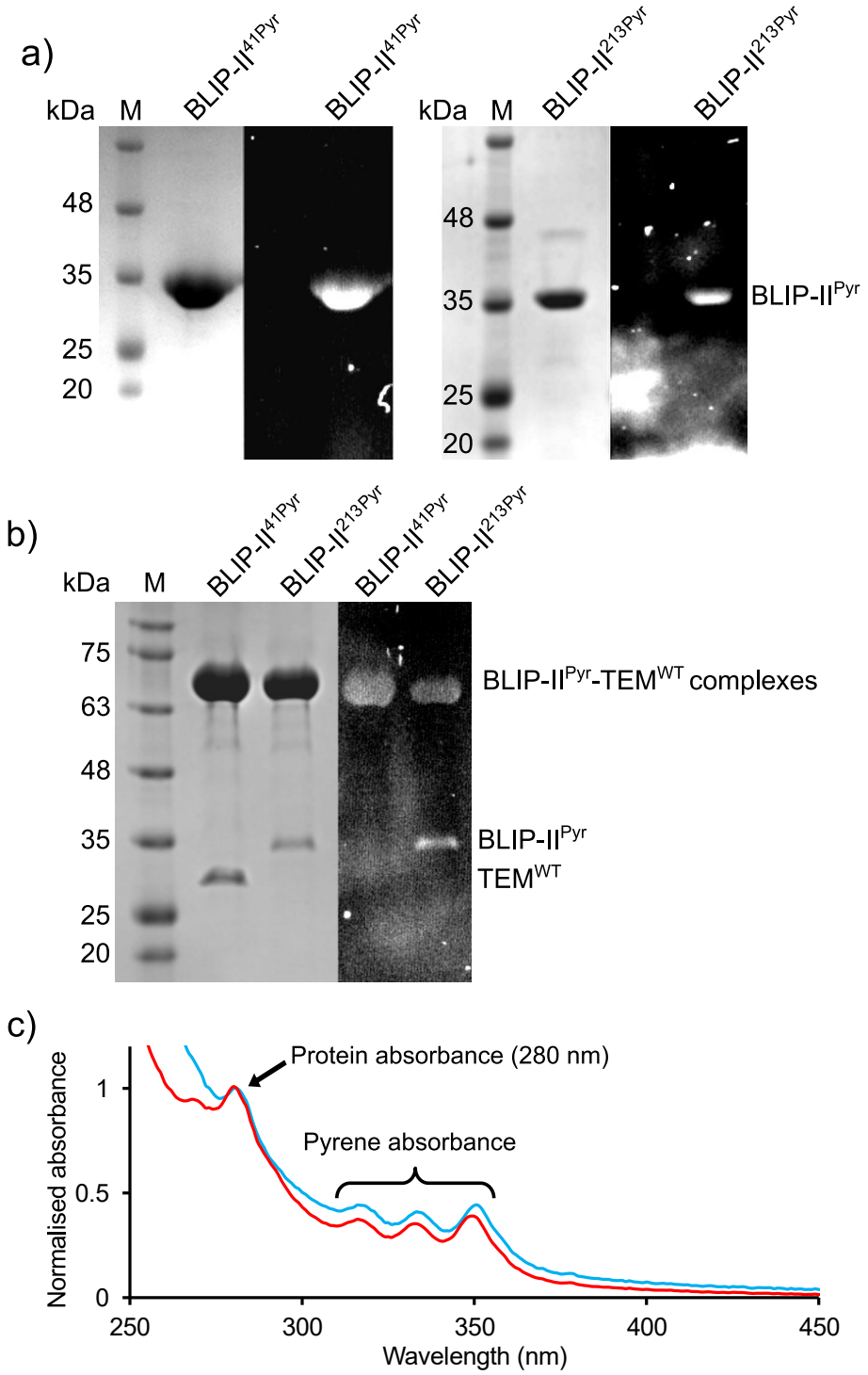


Figure 3.22 Conjugation of BLIP-II AzF variants to DBCO-pyrene. a) SDS-PAGE analysis of BLIP-II^{41Pyr} and BLIP-II^{213Pyr} after incubation with an excess of DBCO-pyrene, imaged using Coomassie stain (left) and UV illumination (right). Faint fluorescent bands indicate pyrene labelling. b) Native PAGE of pyrene-labelled BLIP-II AzF variants after incubation with equimolar TEM^{WT}. BLIP-II-TEM complexes were formed, confirming binding affinity. M: molecular weight marker ladder (not accurate for native PAGE). c) Absorbance spectra of BLIP-II^{41Pyr} (red) and BLIP-II^{213Pyr} (blue) with typical absorbance features of both BLIP-II and pyrene.

NTFET devices were fabricated and functionalised by Xinzhao Xu at QMUL: small bundles of p-type semiconducting SWCNTs were immobilised by dielectrophoresis bridging a 300 nm gap between gold source and drain electrodes on a p-doped silicon

wafer (ConScience AB) to form the transistor channel, as described in methods, 2.2.10.1 (Figure 3.23). As CNTs and DBCO-pyrene are highly nonpolar species, CNTs were functionalised with DBCO-pyrene in a nonpolar solution (dimethyl formamide, DMF) to coat the CNTs before removal of solvent and drying. To limit the number of BLIP-II proteins immobilised on each CNT and leave gaps for β-lactamase analytes to bind to immobilised BLIP-II proteins, a ratio of 3:1 unreactive pyrenebutanol to DBCO-pyrene was used. Pyrenebutanol has a similar structure to DBCO-pyrene but does not contain a bioorthogonal alkyne group for the protein-housed azide. This functionalisation provides the DBCO groups at the CNT surface for conjugation of BLIP-II AzF variants, while also improving water solubility of the CNTs (Figure 3.24). BLIP-II AzF variants were later incubated with the DBCO-functionalised CNTs in aqueous buffer to bind in defined orientations via SPAAC. This approach was preferred to labelling the BLIP-II AzF variants with pyrene first, then incubating with bare CNTs, as it kept the proteins in aqueous solution away from contact with DMF and excess DBCO-pyrene. Devices were also made as negative controls for non-specific protein attachment though functionalisation with pyrenebutanol only.

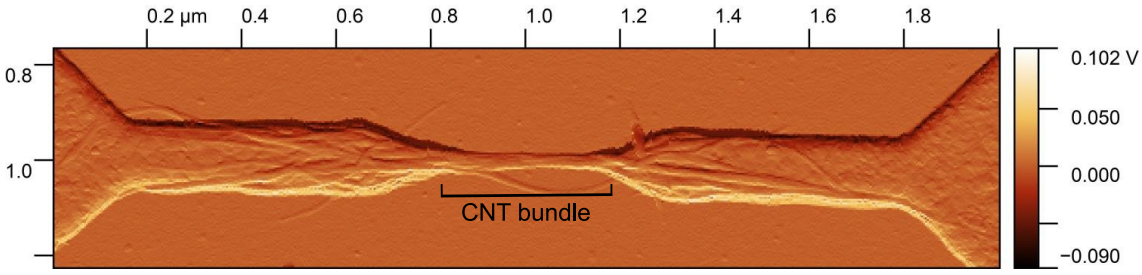


Figure 3.23 CNT bundle bridging two electrodes on silicon wafer. Transistor channel for NTFET experiments, peak force error mode AFM image, before functionalisation with BLIP-II AzF proteins.

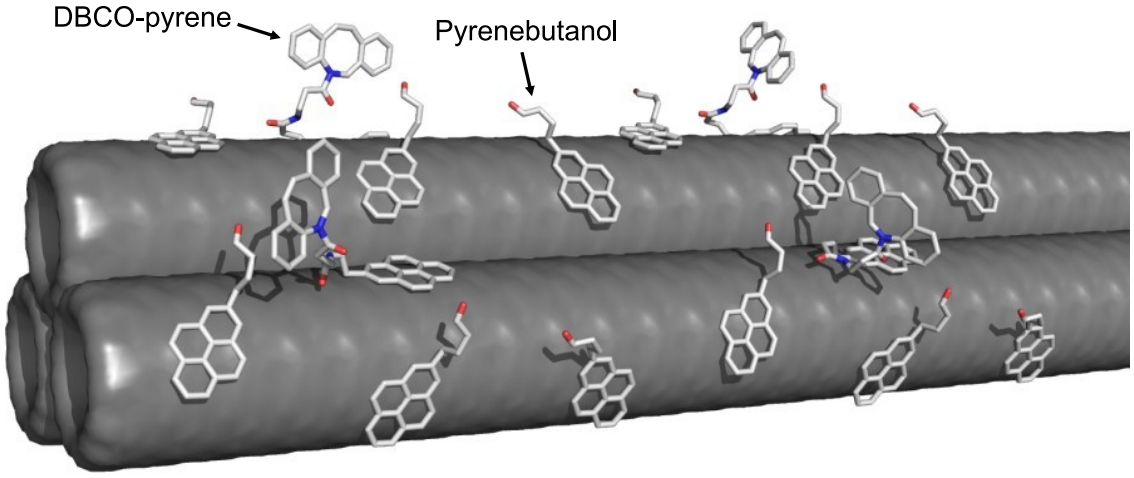


Figure 3.24 Representation of a 3-CNT bundle functionalised with a 3:1 ratio of pyrenebutanol and DBCO-pyrene adsorbed on the surface.

DBCO-pyrene functionalised NTFET devices were examined using AFM, before and after incubation with BLIP-II AzF proteins, to confirm protein attachment through changes in height profiles. Immobilised CNT bundles forming the NTFET channel between electrodes consistently ranged from ca. 9-16 nm high suggesting they were composed of ca. 10-20 CNTs. To verify that the channel CNT bundles had been functionalised, AFM data from before and after protein incubation were compared. Individual proteins were difficult to distinguish on the channel CNT bundle due to its size, but some single CNTs were visible, attached to the edges of electrodes: these appeared decorated with proteins for all variants (examples: Figure 3.25). The number of proteins were counted for 13 visible CNTs and the mean result was calculated at 26 proteins per micrometre of CNT. To estimate the number of proteins per device (i.e. per channel bundle), the surface area of CNT bundles were compared to individual CNTs. CNT bundles containing 10 or 20 CNTs were modelled on a flat surface using PyMOL and their solvent-accessible surface area measured using the "get_area" command. It was calculated that a 10-CNT bundle has ca. 5.5 times the available surface area of an individual CNT on a flat surface and a 20-CNT bundle has ca. 9.3 times the surface area. The gap between gold electrodes in the NTFET devices is 300nm. The above coverage value of 26 proteins per µm of CNT was multiplied by the increased area for the bundles (5.5 or 9.3), then multiplied by 0.3 to give the approximate number of proteins per 300 nm channel bundle. The number of proteins attached to each device is thus estimated to be ca. 43 to 73 proteins.

Eight transverse height profiles were taken across the same sections of each channel CNT bundle, averaged and plotted (Figure 3.26). The average channel CNT bundle heights increased by between 3 and 6 nm after incubation with the BLIP-II AzF proteins, further supporting protein functionalisation of CNT bundles. This height increase was observed with all four BLIP-II AzF variants used (BLIP-II^{41AzF}, BLIP-II^{43AzF}, BLIP-II^{49AzF} and BLIP-II^{213AzF}); functionalised devices were all used for electrical measurement experiments with the aim of TEM^{WT} detection. CNT bundles functionalised with pyrenebutanol instead of DBCO-pyrene as negative controls were not found to increase in height after protein incubation.

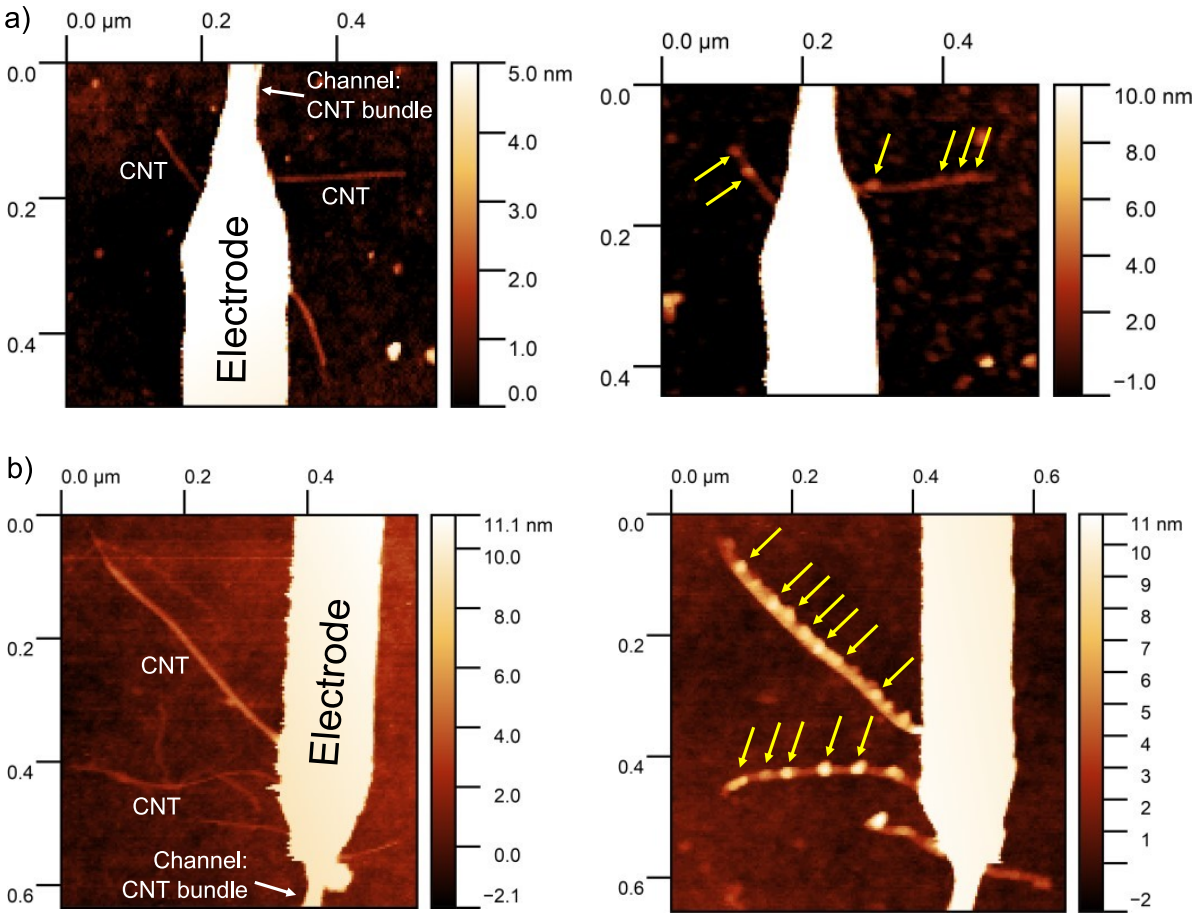


Figure 3.25 Peripheral CNTs attached to the NTFET electrodes/channel. a) CNTs before and after device functionalisation with BLIP-II^{41AzF}. b) CNTs before and after device functionalisation with BLIP-II^{213AzF}. The appearance of proteins (yellow arrows) on individual CNT surfaces after functionalisation is evident. These peripheral CNTs should not affect the signal through the device as they do not contact the opposite electrode.

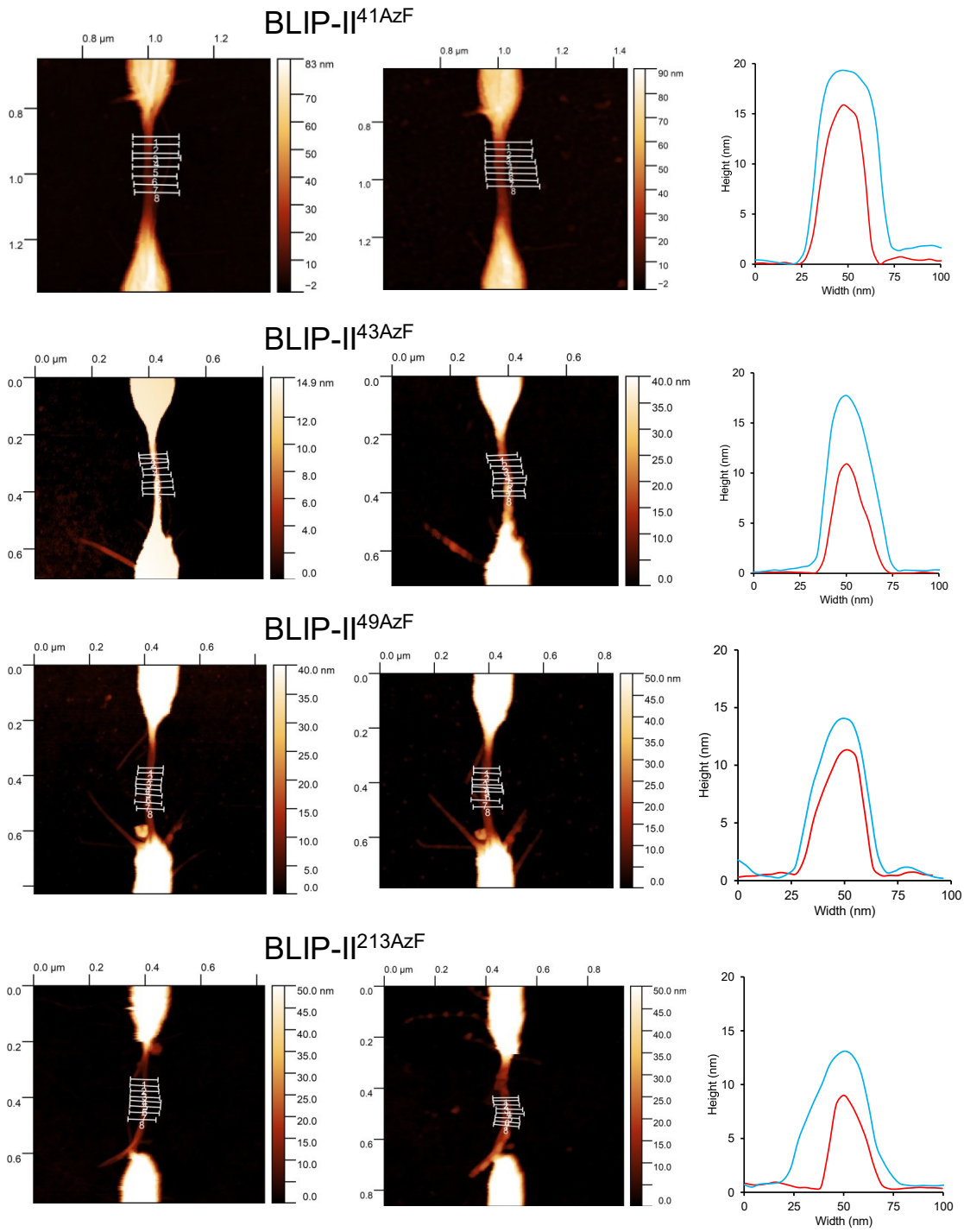


Figure 3.26 Functionalisation of NTFET channel CNT bundles with BLIP-II AzF proteins. Transverse height profiles across DBCO-pyrene functionalised CNT bundles before (left AFM images) and after (right AFM images) protein functionalisation. Average bundle heights increased between 3-6 nm for all variants after incubation with proteins (plots, red – before; blue – after).

3.3.5.2 Electrostatic gating of CNT conductance through β-lactamase binding

After successful fabrication and functionalisation of four CNT devices with different BLIP-II proteins linked in various orientations via their engineered AzF residues, these devices were used for electrical tests to determine if TEM^{WT} binding to the immobilised BLIP-II proteins had measurable effects on conductance through the devices.

Conduction parameters had been recorded for the NTFET devices before functionalisation with either pyrenebutanol/DBCO-pyrene or proteins, to verify the CNT bundles were conducting and making good contact with the electrodes. Devices consisted of a p-doped silicon wafer with SiO_2 insulating top layer, and a p-type semiconducting S-SWCNT bundle (ca. 10-20 CNTs) bridged across a 300 nm gap between two gold electrodes (source and drain) by dielectrophoresis. An example I-V plot was recorded using bias sweeping mode (DC -0.5 V to 0.5V) with a semiconducting parameter analyser, applied to source with a grounded drain, the results were typical for an S-SWCNT bundle of this type (Figure 3.27a). Transfer characteristics were measured in gate bias sweeping mode (-15 V to 15 V), monitoring the source-drain current (I_{sd}) against gate voltage (V_g). Negative gating voltage was shown to increase conductance and positive gate voltage reduced conductance, as is expected for a p-type S-SWCNT (Figure 3.27b).

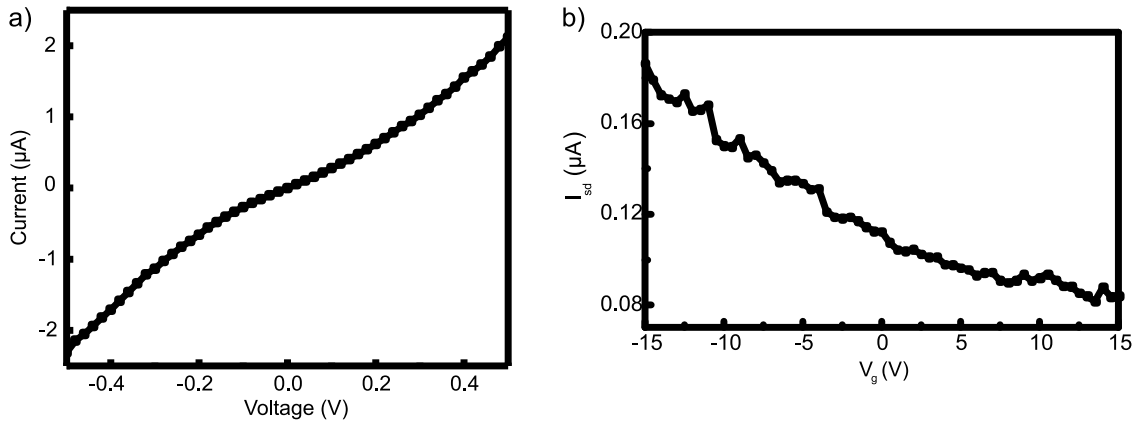


Figure 3.27 Representative electrical tests of non-functionalised CNT NTFET device (V_{sd}=0.1V). a) I-V plot. b) Transfer characteristics.

In a preliminary test, real-time current measurements were recorded for three NTFET devices: a negative control (functionalised with pyrenebutanol, no protein functionalisation), a device functionalised with $BLIP\text{-}II^{49AzF}$ and a device functionalised with $BLIP\text{-}II^{213AzF}$. For each, a bias of 0.1 V was applied across the source and drain electrodes, and the gate electrode grounded, 0.1 V was used because higher voltages have been found to damage CNT bundles. Initial current through the devices were recorded for several minutes to establish background, with different devices showing

stable currents between 0.2 and 1 μA, presumably depending on the number of CNTs making up the channel bundle. A 5 μL drop of Dulbecco's phosphate-buffered saline (DPBS) was then added to the devices to see if this affected conduction: DPBS had no effect on current though any of the devices. A 5 μL drop of 10 nM TEM^{WT} solution was then cast on each device, and the current monitored for several more minutes. No effect was seen on current through the negative control device, as was expected, or through the BLIP-II^{49AzF}-functionalised device; this result was also as expected because the BLIP-II^{49AzF} variant conjugates to the CNT surface via its TEM^{WT}-binding face thus BLIP-II-TEM complexes should not form. The device functionalised with BLIP-II^{213AzF} showed a ca. 10% drop in current shortly after TEM^{WT} solution was drop-cast on the device, indicating binding had occurred that was specific to the BLIP-II^{213AzF}-TEM^{WT} interaction and that current was influenced (Figure 3.28).

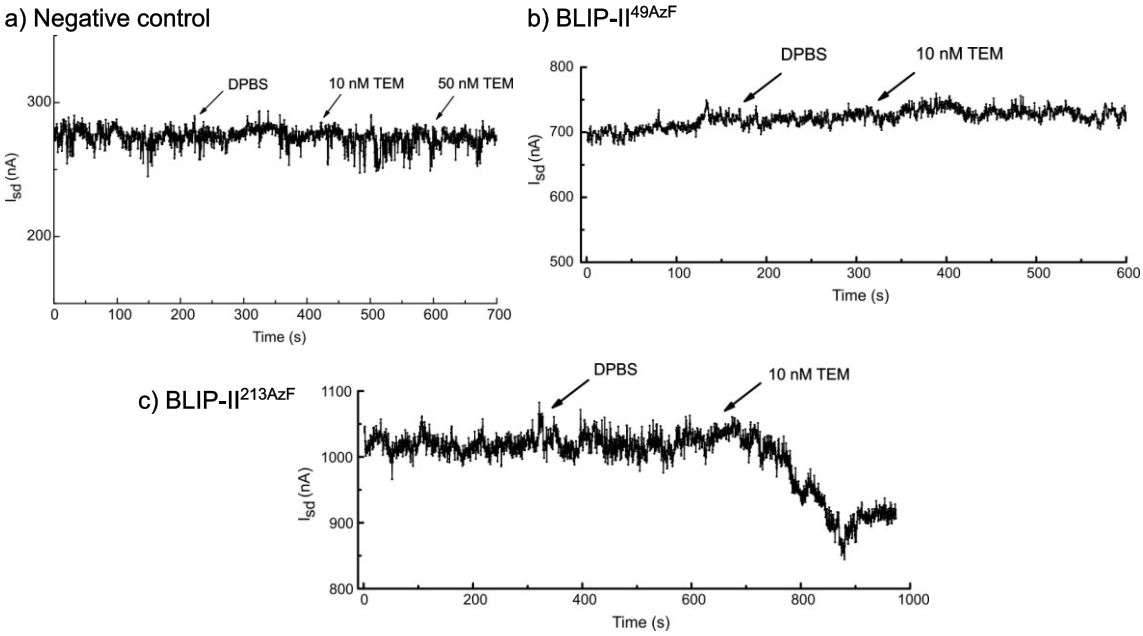


Figure 3.28 Preliminary device responses on addition of DPBS and TEM^{WT}. a) Pyrenebutanol-functionalised NTFET device (no protein). b) Device functionalised with BLIP-II^{49AzF}. c) Device functionalised with BLIP-II^{213AzF} showing 10% ± 1% current drop on addition of 10 nM TEM^{WT}.

Further conduction tests were carried out on four newly-fabricated NTFET devices, functionalised with BLIP-II^{41AzF}, BLIP-$II^{43AzF,}$ BLI.P-II^{49AzF} or BLIP-II^{213AzF}, using similar methods and parameters. In this experiment, the current through each device was recorded and increasing concentrations of TEM^{WT} were applied to the devices at intervals of several minutes (Figure 3.29).

During this experiment, the device functionalised with BLIP-II^{49AzF} showed no current response, consistent with the previous measurement, even with increasing concentrations of TEM^{WT} up to 1 μM. The BLIP-II^{41AzF} device showed increases in current at each TEM^{WT} concentration, in distinct signal steps after each concentration increase,

up to a 12.4% (SD ± 1.91%, 3 biological replicates) current increase with 1 μM TEM^{WT}. The current increase steps became less distinct with increasing TEM^{WT}, possibly as BLIP-II^{41AzF} sites on the CNT channel became saturated. The BLIP-II^{213AzF} device also showed discrete steps in signal, with the current reducing in sharp drops at each increase in TEM^{WT} concentration, down to an 8.6% ± 2.55% reduction in current with 1 μM TEM^{WT}, with the steps becoming less distinct at higher TEM^{WT} concentrations, similarly to BLIP-II^{41AzF}. The BLIP-II^{43AzF} device responded to TEM^{WT} concentration with a gradual increase in current up to a maximum increase of ca. 8% with 1 μM TEM^{WT}, with small, transient current spikes following application of TEM^{WT} to the device. It is thought the more gradual increase in current with transient spikes may reflect the lower binding affinity of the BLIP-II^{43AzF} variant, compared to the than rapid, stepped current changes observed with the higher affinity variants. Binding of TEM^{WT} by BLIP-II^{41AzF} and BLIP-II^{213AzF}-functionalised CNTs is essentially permanent on these timescales, while the lower-affinity BLIP-II^{43AzF} may be displaying on-off binding kinetics. This effect could prove to be a useful property when constructing NTFET devices, as binding proteins immobilised on the device may quickly become saturated by trace quantities of analyte if their binding affinity is too high.

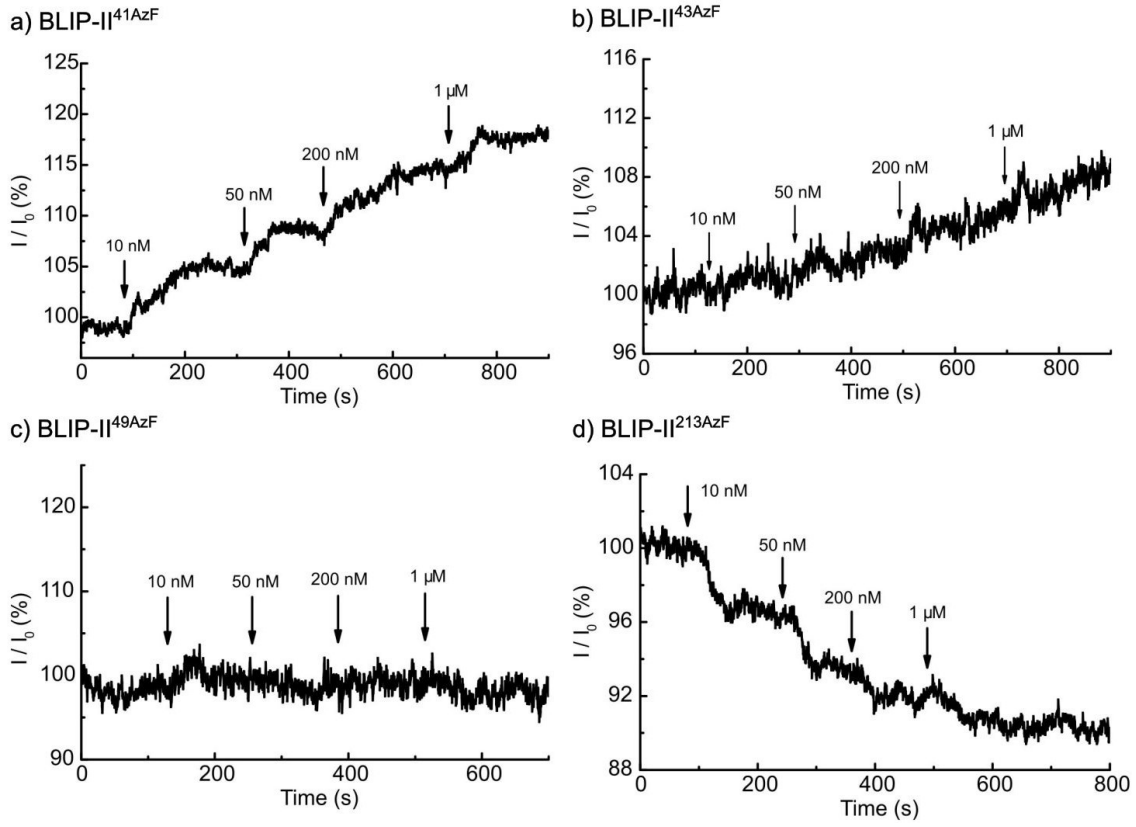


Figure 3.29 Current responses on addition of TEM^{WT} to BLIP-II-functionalised NTFET devices. Increasing concentrations of TEM^{WT} show distinct responses in current. a) Device functionalised with BLIP-II^{41AzF}. b) BLIP-II^{43AzF}. c) BLIP-II^{49AzF}. d) BLIP-II^{213AzF}. Traces are typical and representative of at least 10 repeats. I/I_0 is current given as a percentage of current at time zero, to account for varying current magnitude due to differing numbers of CNTs in different channel bundles.

These results show that binding interactions between BLIP-II AzF proteins immobilised on the NTFET channel surface and target protein TEM^{WT} in solution were clearly discernible as current signal changes through the devices. These effects were consistent across at least ten biological replicates. Three of the four devices functionalised with AzF variants of BLIP-II showed signal responses to the presence of TEM^{WT}, while the variant designed to be inactive (BLIP-II^{49AzF}) showed no response. The responses of BLIP-II^{41AzF} and BLIP-II^{213AzF} devices are of particular interest because their signal changes were discrete and rapid, and they caused opposite effects on conductivity through the CNT channel. The results represent specific recognition of a non-labelled TEM^{WT} β-lactamase by signal modulation caused by electrostatic gating of an NTFET channel composed of semiconducting CNTs. The more gradual increase observed in the signal response of the BLIP-II^{43AzF} device is also interesting, as it may be an effect of this variant's lower TEM^{WT} binding affinity. BLIP-II^{43AzF} shows small spikes in current after addition of TEM, which then settle back to the gradual increase; this is may represent rapid on-off TEM binding in line with reduced binding affinity. It is possible signal noise may be concealing more distinct on-off binding events caused by this variant's lower affinity.

Clinical biological fluid samples have complex compositions containing thousands of proteins and other molecules, each with potential for non-specific binding to, or interference with our BLIP-II functionalised devices. To assess the practicality of using the NTFET device with complex biological mixtures, a conduction experiment was repeated using a BLIP-II^{41AzF} functionalised NTFET immersed in steroid-free serum (10%, diluted in DPBS) (MP Biomedicals). In the absence of TEM, no conductance signal changes were observed; very little non-specific binding was evident. A similar pattern of current increases was observed as previously, in response to addition of TEM^{WT} (also dissolved in 10% serum). Stepped current increases occurred at increasing TEM^{WT} concentration, up to a ca. 25% current increase recorded in response to 1 μM TEM^{WT} (Figure 3.30). This result is further evidence of the BLIP-II functionalised NTFET devices' high specificity to the TEM^{WT} target protein, even in solution containing high concentrations of other proteins.

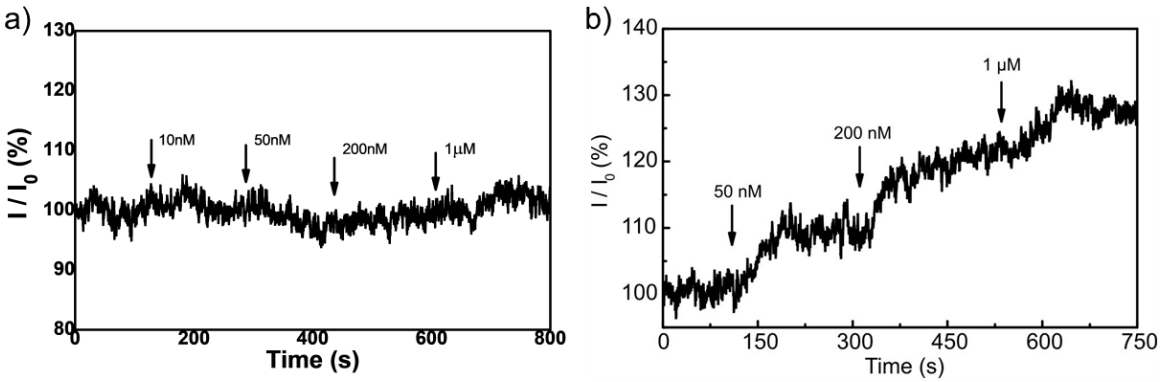


Figure 3.30 Current responses on addition of TEM^{WT} in biological fluid (FBS/DPBS mixture) to NTFET device. a) Negative control: device not functionalised with BLIP-II. b) BLIP-II^{41AzF}-functionalised device. Increasing concentrations of TEM^{WT} show current increase up to +25%. I/I_0 is current given as a percentage of current at time zero, to account for varying current magnitude due to differing numbers of CNTs in different channel bundles.

The models built previously, described in section 3.3.3.1 were used to rationalise the observed results by aligning the BLIP-II-TEM^{WT} complexes to CNT surfaces in sterically feasible conformations, via the geometry-optimised AzF-triazole-pyrene linker model. The electrostatic character of TEM^{WT} surfaces adjacent to the NTFET channel were examined to explain the observed changes in CNT conductance. In our devices, p-type S-SWCNT bundle channels were used, which have positive charge carriers (electron holes). The primary mechanisms of detection in p-type NTFET biosensors are electrostatic gating and Schottky barrier effects. A Schottky barrier is the interface between a semiconductor (S-SWCNT in this case) and a metal (gold electrode) and biomolecules adsorbed on the metal at this junction can modulate current through the device. It is not thought that this mechanism was involved in our conduction results, as the non-functionalised and BLIP-II^{49AzF}-functionalised devices showed no response to TEM^{WT}. Electrostatic gating is modulation of the conductivity of the CNT channel bundle by the electrostatic effects of charged molecules near the CNT surfaces. Relatively small electrostatic charges within the one Debye length range of the CNT surface can induce a screening charge in the CNT, affecting the charge carrier population and in effect doping the CNT. As all conduction in the hollow CNTs happens at the surface, adsorption of charged proteins can have profound effects on the conductivity of the channel[207]. The range of these interactions is one Debye length, which in physiological conditions (such as in DPBS, ionic strength 162 mM, as used here) is ca. 0.7 nm due to electric field screening from counter-charged ions in solution. The conductance of this p-type NTFET channel was expected to decrease when sampling a positively charged surface of TEM^{WT} and increase when sampling a negatively charged surface of TEM^{WT105}. The close proximity of a partial opposite charge stabilises a higher local concentration of charge carriers in the CNT channel[106], in this case positive holes, an effect called hole-doping. Current increase was observed for the NTFET functionalised with BLIP-II^{41AzF}

and BLIP-II^{43AzF}, implying these variants sample a negative section of TEM^{WT}. Current decrease was observed for the NTFET functionalised with BLIP-II^{213AzF}, implying this variant samples a positive section of TEM^{WT}.

Adjusting the BLIP-II AzF variants' structural models by rotation of AzF side chain bonds into sterically feasible conformations showed that the differently-oriented BLIP-II-TEM^{WT} complexes can present different electrostatic faces of TEM^{WT} to the CNT surface, in a manner consistent with our electrical measurements. NTFETs functionalised with BLIP-II^{41AzF} and BLIP-II^{43AzF} can both present the same negative patch of TEM^{WT} consisting of residues E28, D35 and D38 within 0.7 nm of the CNT, as these AzF mutations are close together in the BLIP-II (Figure 3.31 and Figure 3.32, electrostatic surface by ABPS[211]). This conformation appears the most likely cause of the increase in current observed through these NTFET devices, with lower binding affinity a possible reason for the lack of discrete current steps seen for the BLIP-II^{43AzF} device. Examination of the structural model of TEM^{WT} bound to a BLIP-II^{213AzF} functionalised CNT suggests that a cause for the current decrease observed during the conductance experiments could be a positively charged patch of TEM^{WT} consisting of R93, R94 and H96 being brought within range of the CNT surface (Figure 3.34).

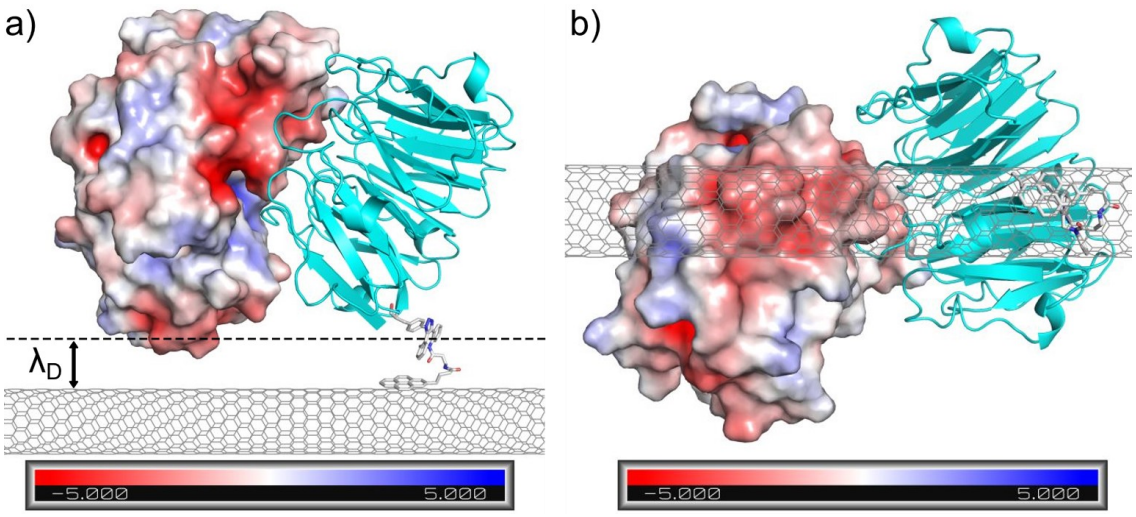


Figure 3.31 Electrostatic properties of TEM^{WT} CNT-adjacent surface when bound to BLIP-II^{41AzF}-functionalised CNT. Possible conformation. Red: negative charge, blue: positive charge (units -5 to +5 $k_b \cdot T \cdot e_c^{-1}$); cyan: BLIP-II^{41AzF}. a) BLIP-II^{41AzF} presents a more negatively charged face of TEM^{WT} to the CNT. λ_D: Debye length from CNT surface (ca. 0.7 nm). b) View from "underside", showing TEM^{WT} surface charges adjacent to CNT. PDB: 1JTD[36].

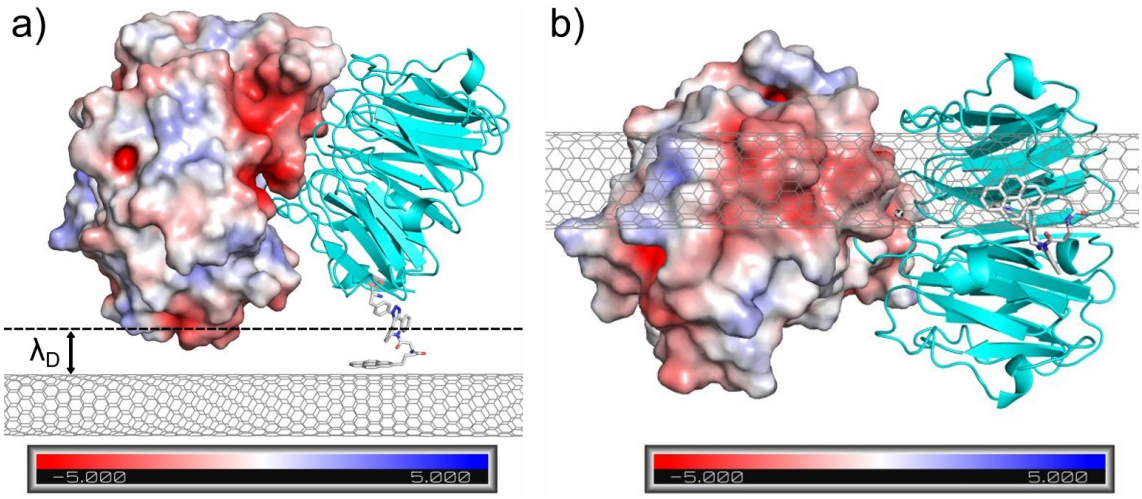


Figure 3.32 Electrostatic properties of TEM^{WT} CNT-adjacent surface when bound to BLIP-II^{43AzF}-functionalised CNT. Possible conformation. Red: negative charge, blue: positive charge (units -5 to +5 $k_b \cdot T \cdot e_c^{-1}$); cyan: BLIP-II^{43AzF}. a) BLIP-II^{43AzF} presents a more negatively charged face of TEM^{WT} to the CNT. λ_D: Debye length from CNT surface (ca. 0.7 nm). b) View from "underside", showing TEM^{WT} surface charges adjacent to CNT. PDB: 1JTD[36].

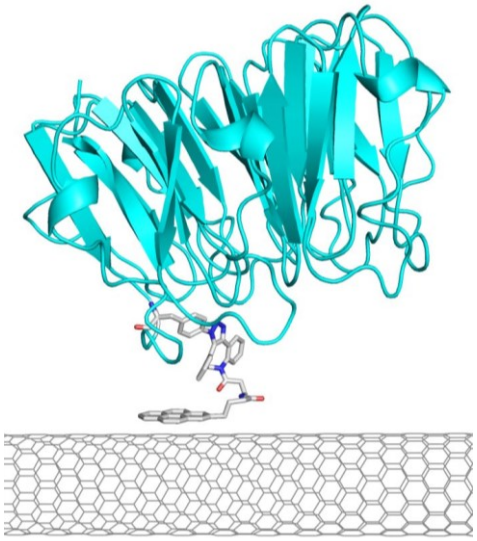


Figure 3.33 BLIP-II^{49AzF}-functionalised CNT. Possible conformation. β-lactamase binding site is obstructed by bound CNT and no current response was observed. PDB: 1JTD[36].

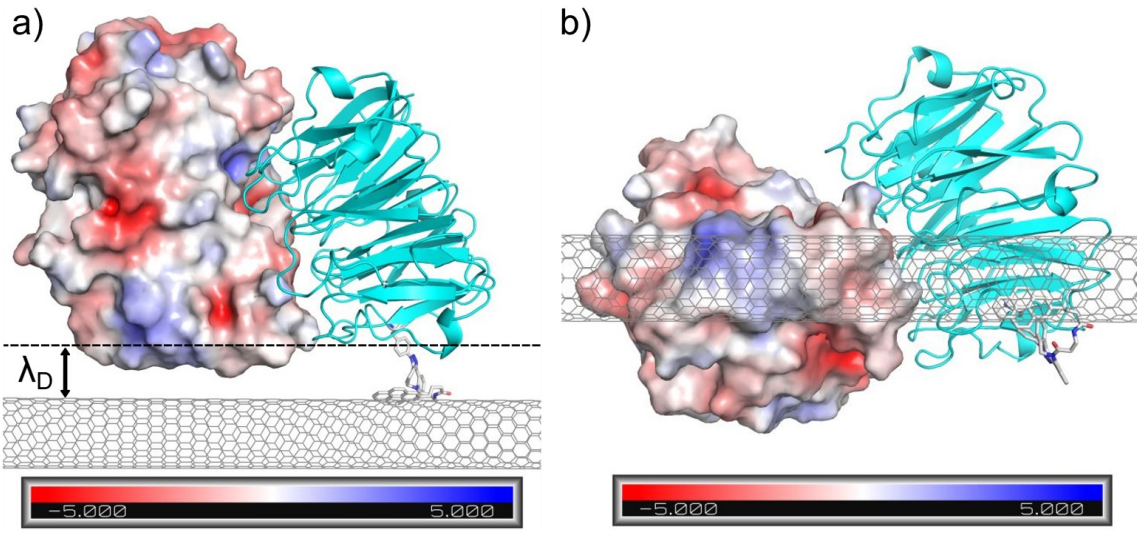


Figure 3.34 Electrostatic properties of TEM^{WT} CNT-adjacent surface when bound to BLIP-II^{213AzF}-functionalised CNT. Possible conformation. Red: negative charge, blue: positive charge (units -5 to +5 $k_b \cdot T \cdot e_c^{-1}$); cyan: BLIP-II^{213AzF}. a) BLIP-II^{213AzF} presents a more positively charged face of TEM^{WT} to the CNT. λ_D: Debye length from CNT surface (ca. 0.7 nm). b) View from "underside", showing TEM^{WT} surface charges adjacent to CNT. PDB: 1JTD[36].

3.4 Conclusions

The observed results demonstrated that our BLIP-II-functionalised NTFET biosensors fulfilled the design aim by specifically binding target TEM β-lactamases at the CNT channel surface, causing distinct effects on CNT conductivity, dependent on the specifically engineered orientation of BLIP-II proteins on the CNT surface. The non-natural azide chemistry of the genetically-incorporated AzF residue enabled the bioorthogonal SPAAC reaction that was central to achieving defined orientation of BLIP-II on the CNT channel, via a DBCO-pyrene linker. The use of non-covalent π-π stacking via pyrene to functionalise CNTs removed the need to introduce CNT sidewall defects by carboxylation to provide covalent attachment points. Sidewall defects adversely affect the conductive behaviour of one-dimensional conductors such as CNTs by introducing gaps in the CNT delocalised π-network, causing electron scattering[236], so attachment of protein by pyrene adsorption was preferred.

Other approaches exist to specifically conjugate moieties such as our pyrene linker to proteins of interest, with advantages and disadvantages of their own. Such techniques include protein affinity tags such as SNAP-tag[237] and HaloTag[21], which conjugate with high specificity to synthesised target molecules which can include pyrene or any other moiety. However, these are large fusion proteins, approximately the same size as BLIP-II (20-30 kDa) and can only be localised at the N- or C-termini of the protein. For functionalisation of CNTs, a large tag protein such as these may obstruct native folding or β-lactamase binding and would probably position BLIP-II too far from the CNT so

binding to the would be beyond the required Debye length to elicit the required response. The simplest approach to inserting a unique reactive handle for conjugation would be to mutate the chosen attachment residue to cysteine for conjugation via a maleimide-bearing moiety; maleimide-thiol conjugation reactions are rapid and form stable covalent bonds. This approach was used by the Philip G. Collins group at UC Irvine to functionalise a CNTs with single lysozyme[117,126] or protein kinase A[210] protein in a predetermined orientation for electrical measurement of single enzyme function. This approach uses only the endogenous translation machinery of the cell, without requiring engineered components like the tRNA and aaRS needed for AzF incorporation, likely improving protein expression levels. However, conjugation via an engineered cysteine is not practical for all proteins as many contain native free surface cysteine residues and thus other potential binding sites; Collins and colleagues had to remove a cysteine from lysozyme before they could introduce the reactive cysteine at their preferred residue. While BLIP-II lacks any native cysteines, a downside of this approach is that engineered surface cysteines are prone to off-target reactions such as spontaneous formation of protein homodimers via disulphide bonds, or off-target conjugation to other cysteine-containing proteins. This can be prevented by including reducing agents such as dithiothreitol (DTT) in buffers, but this would need be removed prior to maleimide conjugation, and is likely to break any native disulphide bridges that exist in the protein, reducing protein stability[238].

Of existing conjugation strategies, none offers the combination of small size, specificity and biorthogonality of azide-alkyne cycloaddition, with AzF being approx. the same size as a natural amino acid, reducing structural perturbation, and because neither azide nor alkyne exist in biological systems. AzF SPAAC click chemistry also has the advantage of being among the more popular genetically encoded nnAA conjugation systems, with a large body of past research, and generally easy availability of compatible reagents. Furthermore, as I will demonstrate in the next chapter, AzF offers other routes to functionalisation of CNTs to create even more intimate interactions and covalent complexes.

The most common traditional approach to specific protein binding is to create monoclonal antibodies that bind the protein of interest, but this approach is of limited use for a CNT-based biosensor as antibodies are too large (10-15 nm) to bind the target protein within the effective range of an NTFET, one Debye screening length, or 0.7 nm in fluid of physiological ionic strength[239]. Our recognition and binding protein BLIP-II has the advantage of being only 3-4 nm in size and having evolved extreme affinity for a range of class A β-lactamases.

A closely-related area of CNT-based biomolecular sensing employs DNA[106] or short peptide[206] aptamers as the protein recognition and binding elements on an NTFET channel surface. This approach has advantages as these short oligonucleotides or peptides can be synthesised relatively cheaply and quickly compared to the cost of design, expression and purification of whole proteins, especially those containing non-natural amino acids. However, aptamers generally cannot offer the extreme levels of specificity or affinity that millions of years of natural selection has conferred on protein-protein interactions. For a biosensing application, especially with a diagnostic purpose e.g. blood testing, it is crucial that recognition elements can selectively bind their target protein against a background of thousands of other proteins, many at much higher concentration. Even the best-designed aptamers have target selectivity and binding affinity inferior to evolved protein-protein interactions, of which BLIP-II-β-lactamase binding is among the strongest, with a typical K_D about 1000-fold lower than the most sensitive peptide aptamers[206]. For this reason, engineered proteins such as ours appear well suited to electronic biosensing applications requiring the highest sensitivity and specificity.

Biosensors of a similar NTFET design have been constructed previously, including by Choi *et al*. (2013), who functionalised CNTs with single molecules of lysozyme with different charged mutations[117], and Hatada *et al*. (2019), who functionalised an NTFET channel with the bacterial protein SocA for detection of fructosyl valine, a biomarker of long-term high levels of blood glucose, by binding-induced conformational change in the charged surface of the protein[119]. However, this device was conjugated to the CNT channel via less specific NHS ester-amine chemistry, which results in a wide range of different protein orientations as the conjugation point on each protein is essentially random. This may be the cause of the slower and less pronounced steps in conductance/resistivity binding response compared to our data (Figure 3.35).

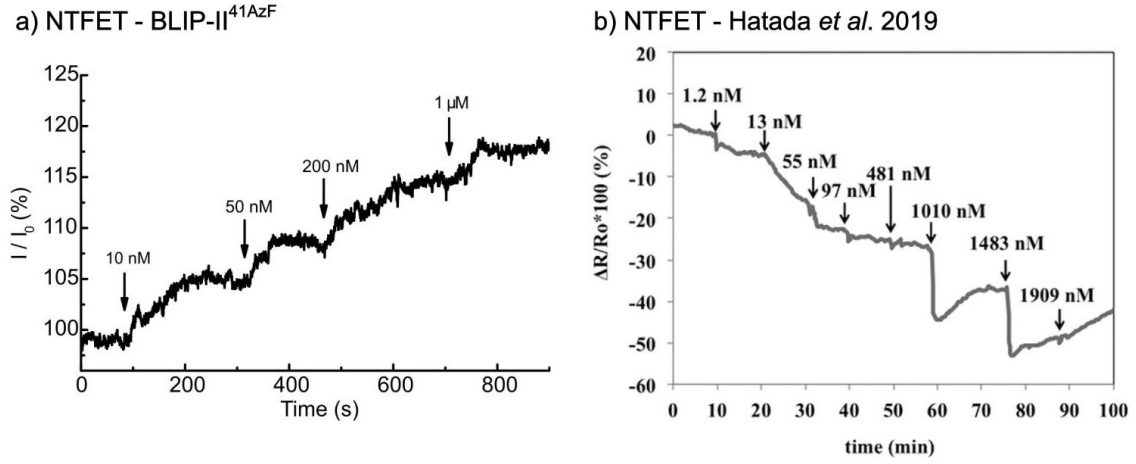


Figure 3.35 Changes in current/resistivity response due to binding in NTFET biosensors. a) Current increase in our BLIP-II^{41AzF}-functionalised NTFET in response to TEM^{WT} binding. b) Resistivity decrease in SocA-functionalised NTFET in response to fructosyl valine binding (Hatada *et al*., 2019)[119].

It is expected that the differences in signal response observed from differing orientations of the TEM^{WT} at the channel surface will also apply to different class A β-lactamase species bound at the CNT surface, since different β-lactamases show marked differences in surface electrostatic charge (see Figure 3.8b). Work is ongoing with QMUL to differentiate between β-lactamase types using our NTFET system, with signal response between TEM^{WT}, CTX-M^{WT} and KPC-2^{WT}, three of the most common causes of bacterial β-lactam resistance, being compared. BLIP-II has extremely high and specific affinity for all these class A β-lactamases, each with unique surface charge topographies that likely to induce differential gating of NTFET channel conduction properties that can be discerned from changes in electrical signal. Our system demonstrates potential to sample different proteins, different binding affinities and importantly, different electrostatic surfaces of a protein, suggesting that local charges of a bound analyte at the CNT surface are more important to electrostatic gating than the analyte's net charge.

Our system of functionalising NTFET biosensors with highly specific recognition/binding proteins via engineered AzF mutations has potential for detection of many other binding proteins. Of particular current interest would be an NTFET biosensor functionalised with the human ACE2 receptor protein via an engineered AzF residue, which may be able to specifically bind and detect the SARS-CoV-2 virus spike protein S1 subunit receptor binding domain (RBD) as a diagnostic test for infection with the virus. Mammalian expression systems already exist for incorporation of AzF into proteins[240] like human ACE2 that are unsuitable for recombinant expression in bacteria.

4 Site-specific photochemical covalent attachment of proteins to carbon nanotube side walls

Work in this chapter contributed to a published paper where I am a joint 1st author:

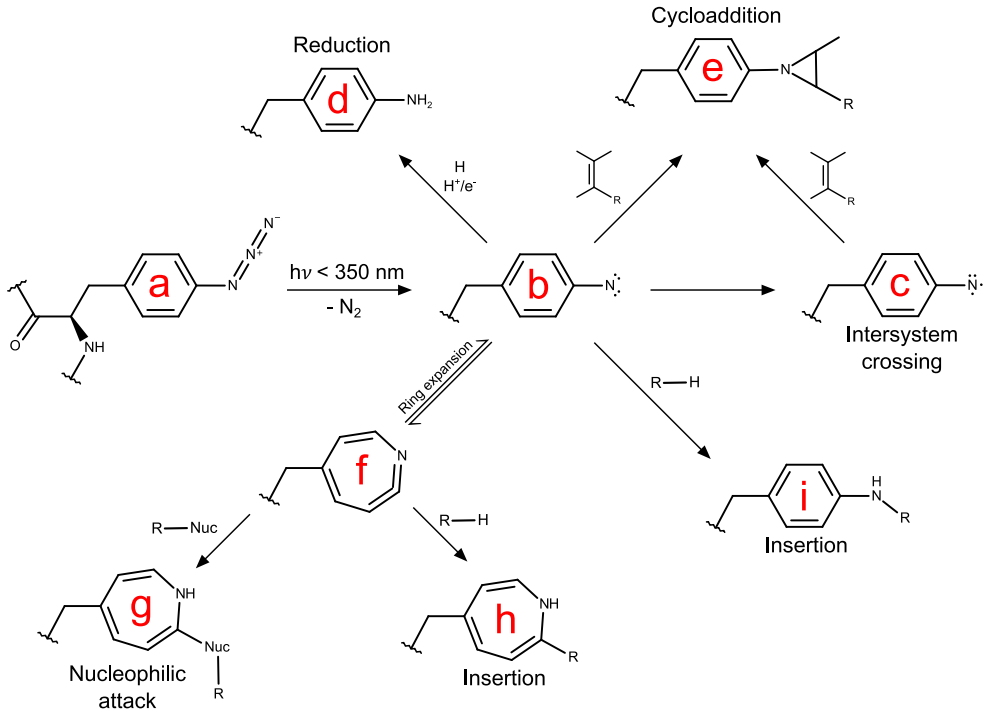
Thomas S.K., Jamieson W.D., Gwyther R.E.A., **Bowen, B.J.** *et al*. Site-Specific Protein Photochemical Covalent Attachment to Carbon Nanotube Side Walls and Its Electronic Impact on Single Molecule Function. *Bioconjugate Chemistry* **31**, 584–594 (2020).

4.1 Introduction

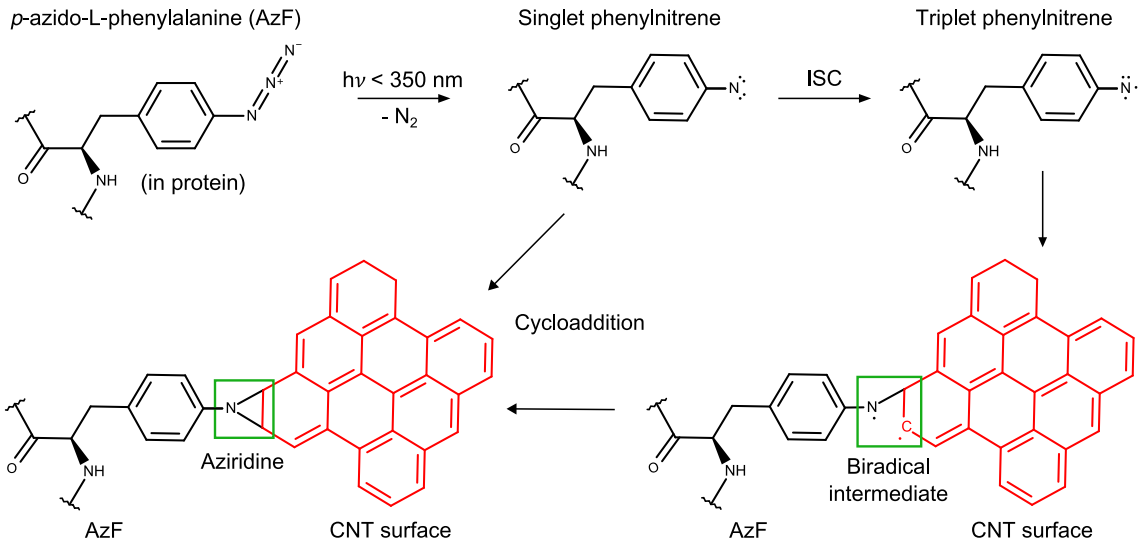
Semiconducting single-walled carbon nanotubes (S-SWCNTs) have great potential as the basis for new protein-based nanoscale electronic sensing devices due to their small size (ca. 1-2 nm), biocompatibility, high surface area to volume ratio (especially single-walled CNTs[206]) and their variable electrical resistivity in different chemical environments. CNTs are highly sensitive to local conditions due to all carbon atoms being at the CNT surface and accessible to the environment[207]. Chapter 3 of this thesis demonstrated the construction of protein-functionalised CNT-based nanotube field-effect transistor (NTFET) biosensing devices. Variants of the β-lactamase binding protein BLIP-II were intimately interfaced with CNT surfaces in different orientations, via residues at defined locations on the protein surface. This system used the genetically-incorporated non-natural amino acid (nnAA) *p*-azido-L-phenylalanine (AzF) as a bioorthogonal reactive handle. AzF has a phenyl azide side chain which was used to attach a linker molecule to BLIP-II via bioorthogonal click chemistry, for noncovalent adsorption to CNT surfaces via an attached pyrene moiety. Subsequent binding of target β-lactamases to immobilised BLIP-II was then detected by their electrostatic effects on CNT conductivity.

This chapter also aims to functionalise CNT surfaces with engineered proteins of interest via a genetically-incorporated AzF reactive handle, but via a different chemical approach: direct covalent conjugation to the surface of CNTs using the photochemical properties of AzF. At high temperatures (> 400 °C [241]), or on photoactivation with UV-B light (λ < 350 nm), the phenyl azide side chain of AzF releases molecular nitrogen (N_2), yielding a short-lived and highly reactive singlet phenylnitrene radical[120]. Many possible reactive routes exist for the phenylnitrene radical including intersystem crossing to a triplet phenylnitrene radical[120], reduction to an amine group[87], expansion of the phenyl ring to a seven membered ketenimine[120], insertion into C-H bonds to form an amide bond[242] or addition to an electron-rich target such as a C=C double bond[121] (Scheme 4.1). For our purpose of functionalising CNTs, the relevant reaction is nitrene [2+1] cycloaddition to

CNT sidewall which occurs with either the singlet or triplet nitrene radical species, through slightly different reaction routes[83,122]. Either route results in a covalent triangular aziridine ring between the AzF nitrene of the protein and two carbons of the CNT graphitic lattice (Scheme 4.2).



Scheme 4.1 Photochemistry of the *p*-azido-L-phenylalanine (a) phenyl azide group. On exposure to UV-B light (< 350 nm) phenyl azide exudes nitrogen (N_2) to become a singlet phenylnitrene (b). This radical can convert to a triplet phenylnitrene (c) by intersystem crossing or become protonated to a nitrenium ion (NH^+) or a phenyl amine (d). The radicals b and c can both react with a double bond (nitrene cycloaddition) to form a triangular aziridine (e) ring. Radical b can undergo ring expansion to a seven-membered ketenimine (f) ring which can be attacked by a nucleophile (g) or insert into σ-bonds (h). Radical b is also able to insert into C-H bonds to form a phenyl amide (i).



Scheme 4.2 Photochemical reactions of AzF during functionalisation of CNT. UV-B irradiation at λ < 350 nm converts phenyl azide to singlet phenylnitrene radical (molecular N_2 extruded) which can undergo intersystem crossing to triplet phenylnitrene. Either nitrene undergoes cycloaddition resulting in a covalent aziridine ring anchoring protein to CNT surface.

Nitrenes have been used previously to functionalise CNTs with a range of moieties for conjugation or to aid solubility, however the conditions traditionally used to activate the nitrene (160 °C, dichlorobenzene) are unsuitable for proteins[122,243]. The more recent UV-irradiation method of activating the azide reactive group is more compatible with proteins, and has been previously used by to activate genetically incorporated phenyl azides[87,88]. Our group has previously demonstrated the functionalisation of graphene sheets with proteins via aziridine linkages by photoactivation of engineered AzF residues; three different proteins (GFP, TEM β-lactamase and cytochrome b_{562}) were conjugated to the surface of graphene by this method[90]. This work intends to intimately link proteins to CNT surfaces, while demonstrating retention of their native function.

The aziridine bond provides a direct link between protein and CNT so is much shorter than the triazole-pyrene linker used to conjugate protein to CNTs non-covalently in chapter 3, offering the possibility of more intimate integration between protein and CNT. This method also avoids the creation of vacancies in the graphitic lattice of CNTs that are typical of other covalent functionalisation strategies such as acid oxidation. This prevents the undesirable mechanical and electrical effects of introducing gaps into the CNT matrix. The geometry-optimised models produced for the triazole-pyrene linker (see chapter 3) suggest the distance between CNT and α-carbon of AzF is likely to be at least 1.3 nm for triazole-pyrene linked proteins and approx. 0.8 nm for aziridine-linked proteins (Figure 4.1). More intimate CNT interfacing may improve electrical detection of bound target proteins by bringing their charged electrostatic surfaces closer to the CNT channel surface of an NTFET biosensor. The range of electrostatic interactions (Debye screening length, λ_D) in solutions of physiological ionic strength is ca. 0.7 nm so closer binding could improve the signal-to-noise ratio achieved in a protein-CNT based FET biosensor (see chapter 3 for details of electrostatic gating of NTFET biosensors).

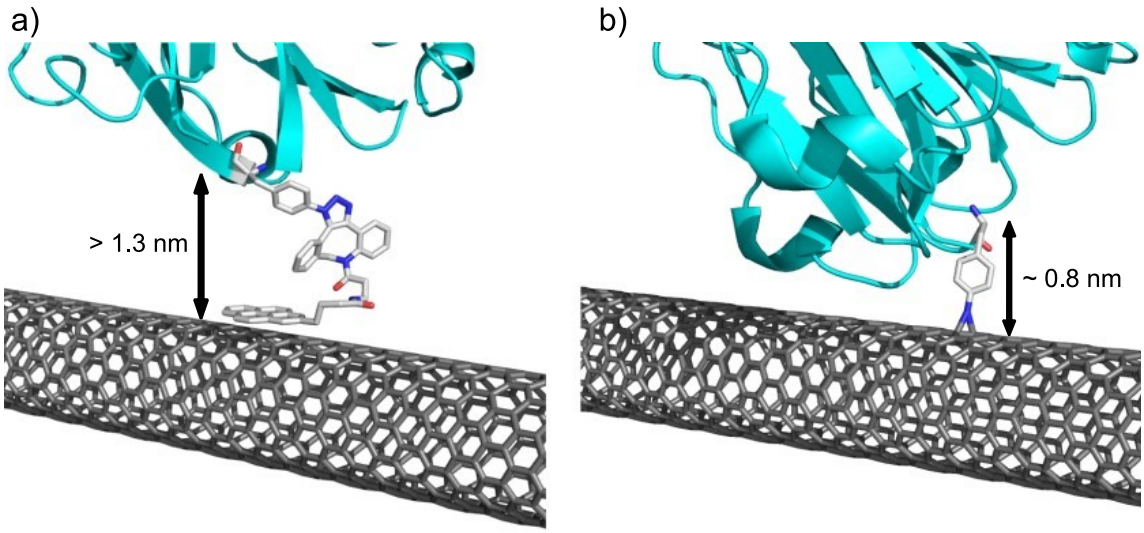


Figure 4.1 Comparison of CNT BLIP-II protein functionalisation strategies via engineered *p*-azido-L-phenylalanine (AzF) residue. Cyan: BLIP-II^{41AzF}. a) Noncovalent π-π stacking of a DBCO-pyrene linker molecule conjugated to AzF by SPAAC click chemistry. b) Direct photochemical covalent bonding by nitrene [2+1] cycloaddition to the CNT π-bond system PDB 1JTD[36]. Modelled in PyMOL[11].

Variants of three different proteins were made for functional integration with CNTs via engineered, genetically incorporated AzF reactive handles, to assess the practicality of CNT functionalisation by UV-nitrene cycloaddition and to compare it with the DBCO-pyrene approach used in chapter 3. The location of AzF conjugation points on the protein surfaces were selected to present the proteins in specific spatial orientations relative to the CNT surface when assembled.

CNTs were functionalised by bioorthogonal bottom-up assembly with the BLIP-II^{41AzF} and BLIP-II^{213AzF} variants as used in chapter 3. Briefly, BLIP-II binds to a wide range of class A serine β-lactamases[147] with extremely high-affinity. β-lactamases are bacterial enzymes that deactivate β-lactam antibiotics and as common agents of antimicrobial resistance (AMR, see chapter 1.3), are attractive target analytes for bioelectronic sensing. These BLIP-II AzF variants were expressed and purified then used to functionalise S-SWCNTs by UV-nitrene photoattachment via genetically-incorporated AzF residues (Figure 4.1b). CNT-linked BLIP-II proteins were designed to present their β-lactamase binding surface to solution and bind β-lactamases within sensing range of the CNT. BLIP-II – β-lactamase complexes were self-assembled by design via protein-protein binding interactions at the CNT surface. Two of the most clinically-important β-lactamases[244,245] were used for this bottom-up assembly: TEM^{WT} for analysis by atomic force microscopy (AFM), and a fluorescent-labelled variant of CTX-M for analysis using total internal reflection fluorescence (TIRF) microscopy.

To further demonstrate the wide applicability of this method of functional integration of proteins with CNTs, other engineered proteins were created to functionalise CNT

surfaces. Using the same methods as with BLIP-II, CNTs were functionalised with an engineered AzF-containing variant of TEM β-lactamase, TEM^{87AzF}.

The third protein produced for functionalising CNTs was the small bacterial electron transport haemprotein cytochrome b_{562}. Cyt b_{562} has a number of properties that make it an interesting subject for integration with carbon nanomaterials. Our lab has previously integrated engineered cyt b_{562} variants into nanosystems, including immobilising the proteins on graphene sheets via this photochemical approach[90]. Cyt b_{562} has been conjugated on gold (III) surfaces by our group for measurement of single protein molecule electron transfer[169], and in different orientations, via engineered cysteine residues; single protein orientation was discernible using scanning tunnelling microscopy (STM) where cyt b_{562} acted as a single molecule bridge between tip and gold substrate[170]. For the work described here, two variants of cyt b_{562} were produced, cyt b_{562}^{SA} (short axis) and cyt b_{562}^{LA} (long axis). These variants have engineered AzF residues located at different positions to orient the proteins differently relative to the CNT surface (Figure 4.2). The variants also contain engineered cysteine residues at the opposite side of the protein and are known to form homodimers via disulphide bonds between cysteines (see section 4.2.4). CNTs were functionalised with cyt b_{562}^{SA} and cyt b_{562}^{LA} then analysed to measure binding orientation and dimeric/monomeric state using AFM.

CNT-protein constructs were analysed and imaged using AFM at Cardiff School of Physics and Astronomy, with the help of Suzanne Thomas and Adam Beachey and by TIRF microscopy at Cardiff School of Pharmacy and Pharmaceutical Sciences with the help of David Jamieson. Some protein-CNT experiments were performed with Rebecca Gwyther.

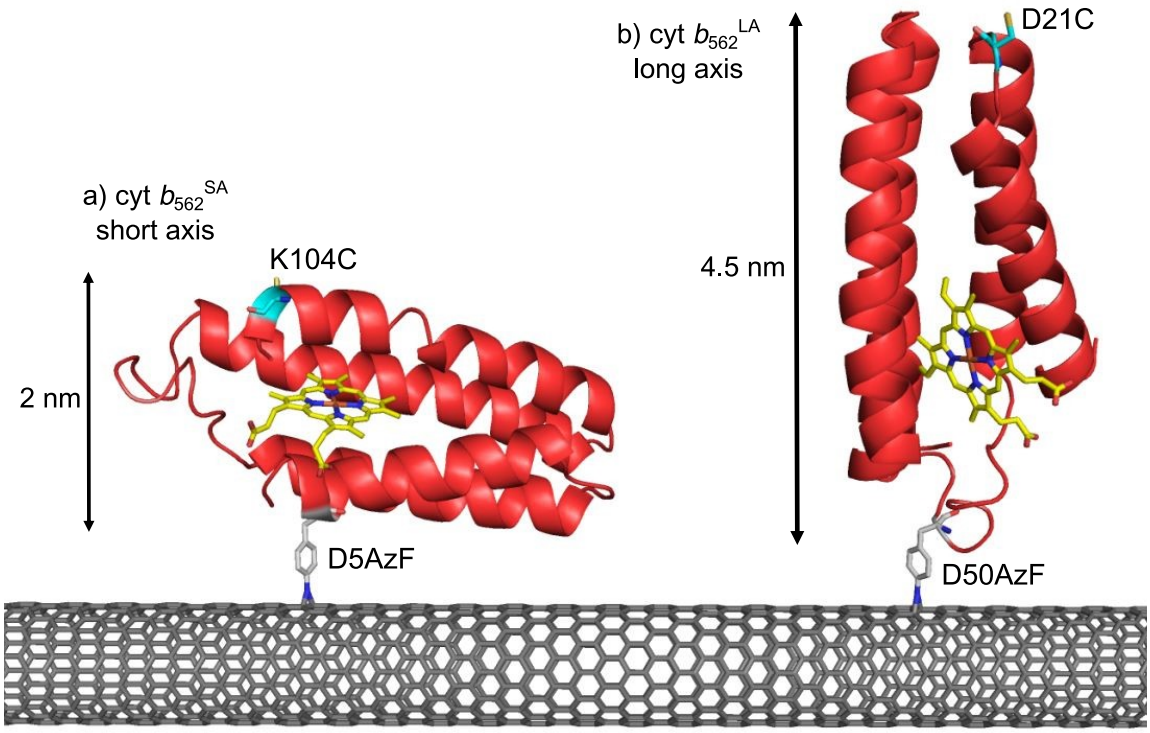


Figure 4.2 Possible orientations of cyt b_{562}^{SA} and cyt b_{562}^{LA} when conjugated to CNT by UV-nitrene cycloaddition via engineered AzF residues. "Short axis" variant is attached via D5AzF residue (grey), with a second engineered Cys residue (K104C) positioned at the opposite side of the protein (cyan) for further conjugation. "Long-axis" variant is attached via D50AzF (grey) with engineered Cys at position D21C (cyan).

4.2 Results and discussion

4.2.1 Functionalisation of CNTs with BLIP-II^{41AzF} via nitrene photoaddition for TEM β-lactamase binding and analysis using AFM

4.2.1.1 Production of BLIP-II^{41AzF} and TEM^{WT} proteins

Production of BLIP-II and TEM^{WT} proteins is described in chapter 3.

4.2.1.2 CNT sidewall functionalisation with BLIP-II^{41AzF} via photoattachment and AFM analysis

The AzF-containing variant BLIP-II^{41AzF} was functionally integrated with CNT surfaces by photochemical attachment by [2+1] nitrene photoaddition as described in Methods, 2.2.11.2. CNTs were analysed using tapping mode AFM in air (see chapter 3.2) to assess sidewall functionalisation. All AFM data was analysed and measured using Gwyddion software[204].

AFM topographic data for CNTs incubated with BLIP-II^{41AzF} but without UV-B photoactivation were compared to bare CNTs. The morphology of CNTs incubated with non-UV-activated BLIP-II^{41AzF} matched that of bare CNTs. Both appeared relatively

smooth with consistent longitudinal height profiles of ca. 1.0-1.6 nm above the surface (Figure 4.3). The AFM data showed no evidence of non-specific protein binding to CNTs or to the silicon surface, as has been observed previously when functionalising carbon nanomaterials with proteins[90]. This result confirms that BLIP-II^{41AzF} does not bind to CNTs without UV irradiation and is removed during washing.

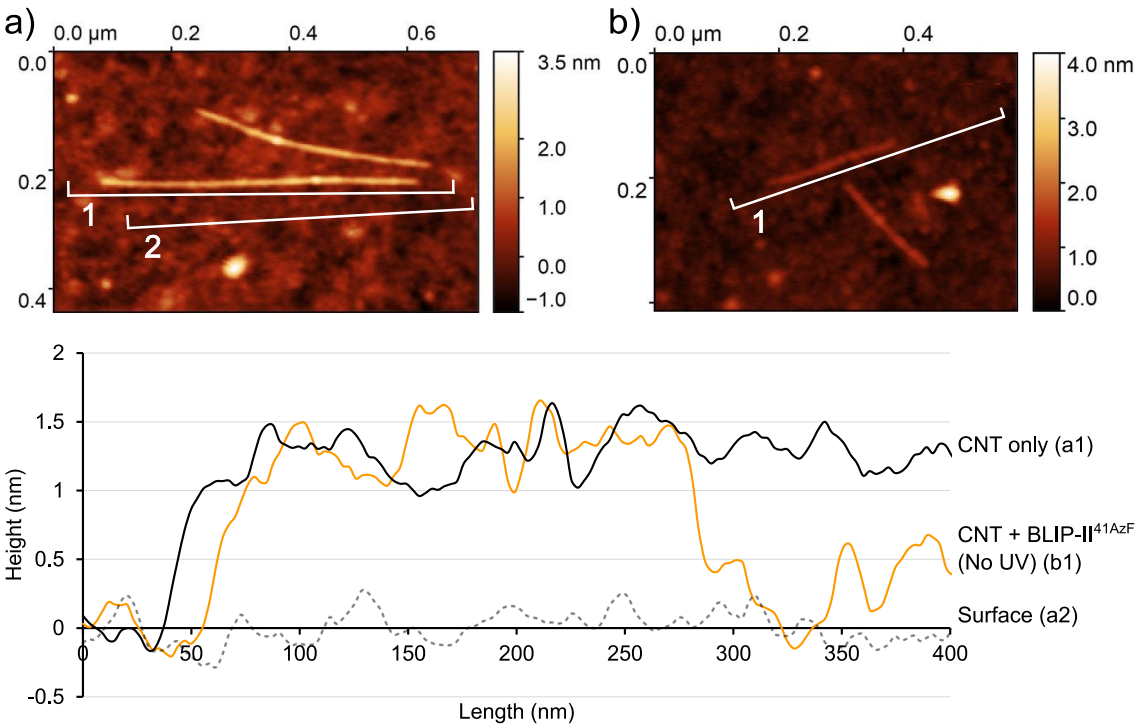


Figure 4.3 AFM images of longitudinal CNT height profiles. a) Height profiles along a single bare CNT (1-1.5 nm, black line, **1**), and along the silicon surface (grey, dashed line, **2**) with no protein present. b) Profile along a single CNT after incubation with BLIP-II^{41AzF} without UV-B exposure (orange line, **1**), height matches bare CNT (1-1.6 nm) and no sign of decoration with globular proteins.

AFM topographic data for CNTs incubated with BLIP-II^{41AzF} with and without UV irradiation were analysed and compared. Non-photoactivated protein-CNT samples appeared smooth, like images of the bare CNTs, while CNTs incubated with BLIP-II^{41AzF} under UV-B light appeared decorated with globular objects of increased height, with a similar dispersion as observed in chapter 3 (ca. 20-30 per nanometre). AFM data rendered in three-dimensional representation illustrates periodic height increases along the length of the CNT (Figure 4.4). Further analysis was performed to measure the dimensions of the CNTs and objects decorating the CNTs. Eight transverse height profiles were measured across CNTs at ca. 90° and averaged to give height values. The average height of CNTs incubated with BLIP-II^{41AzF} without UV activation was 1.59 nm (SD = 0.14 nm), consistent with the height of bare CNTs as supplied. The CNTs incubated with UV-activated BLIP-II^{41AzF} were measured across the decorating objects, and between the objects across the CNT itself: the average height of globular objects

was 3.57 nm (SD = 0.47 nm) and the CNT average height was 1.91 nm (SD = 0.13 nm). The objects are thought to be proteins permanently immobilised via covalent aziridine bonds, separated by regions of bare CNT (Figure 4.4).

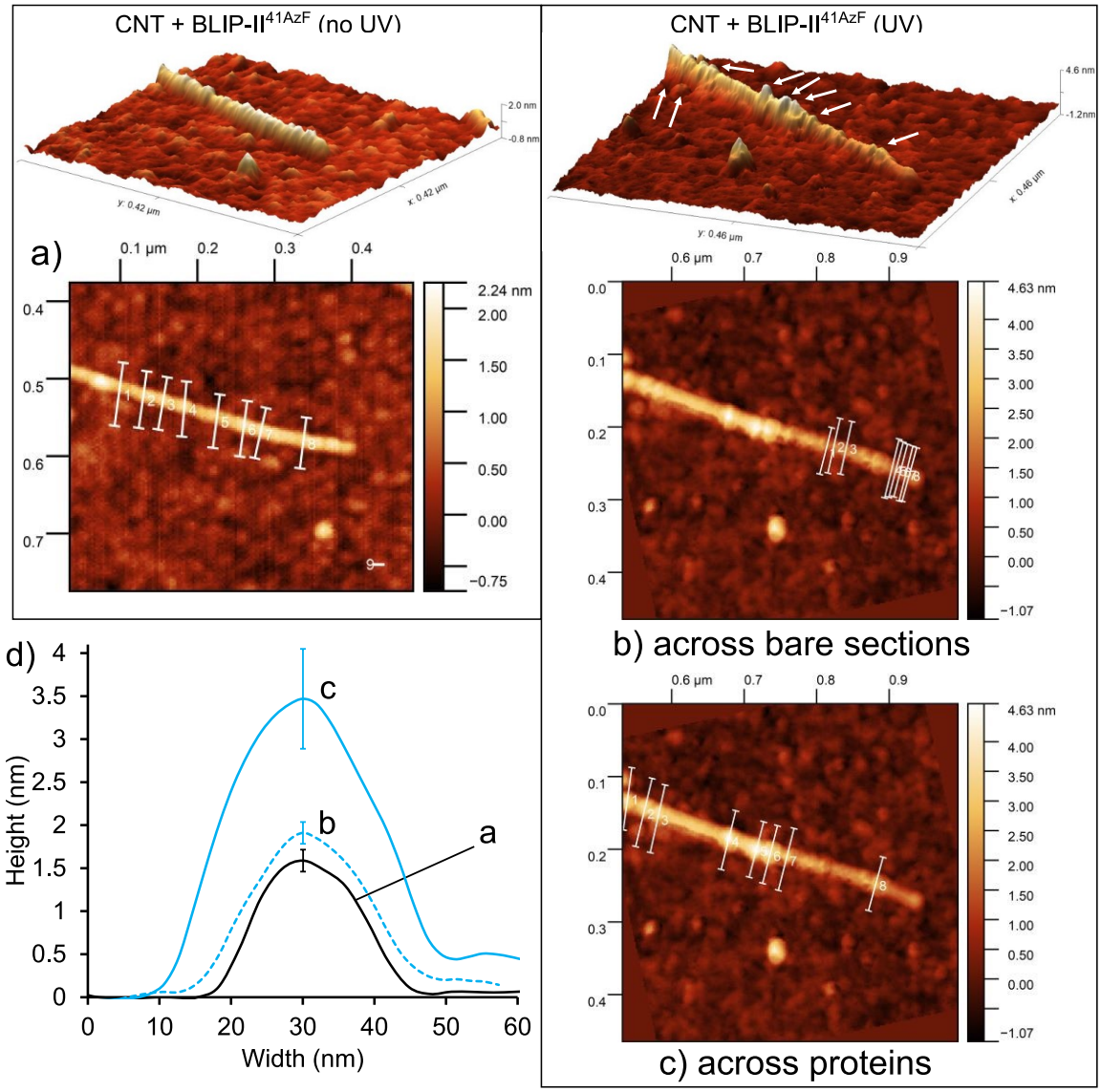


Figure 4.4 CNT-BLIP-II^{41AzF} functionalisation AFM data. a) CNT incubated with BLIP-II^{41AzF} without UV-B photoactivation, showing a relatively level CNT surface with average height ca. **1.59 nm** (SD = 0.14 nm). b) CNT functionalised with BLIP-II^{41AzF} by photochemical attachment, measured across bare sections of CNT, average height **1.91 nm** (SD = 0.13 nm). c) The same CNT as b), measured across proteins (arrowed on 3D render), average height **3.57 nm** (SD = 0.47 nm). d) Mean height profiles across **a-c;** measurements n=8, SD bars shown for highest point, measurements shown on AFM images.

The maximal height of CNT-immobilised BLIP-II^{41AzF} proteins varied around a mean height of 3.57 nm with a relatively large SD of 0.47 nm, suggesting that the proteins may be conjugated in a range of random positions around the circumference of the CNT surface, as shown in Figure 4.5, however the location of the AzF attachment point at residue 41 means that β-lactamase binding sites are likely to remain accessible to

solution. This is a crucial requirement for the intended goal of trapping target β-lactamase proteins on the CNT surface by protein-protein interactions.

The height increase of 1.6-1.9 nm on BLIP-II^{41AzF} immobilisation is less than the expected difference of 3-5 nm due to issues with analysing biological macromolecules using AFM (see chapter 3.2 for overview of AFM). All proteins analysed by AFM during this work were subject to this level of underestimation.

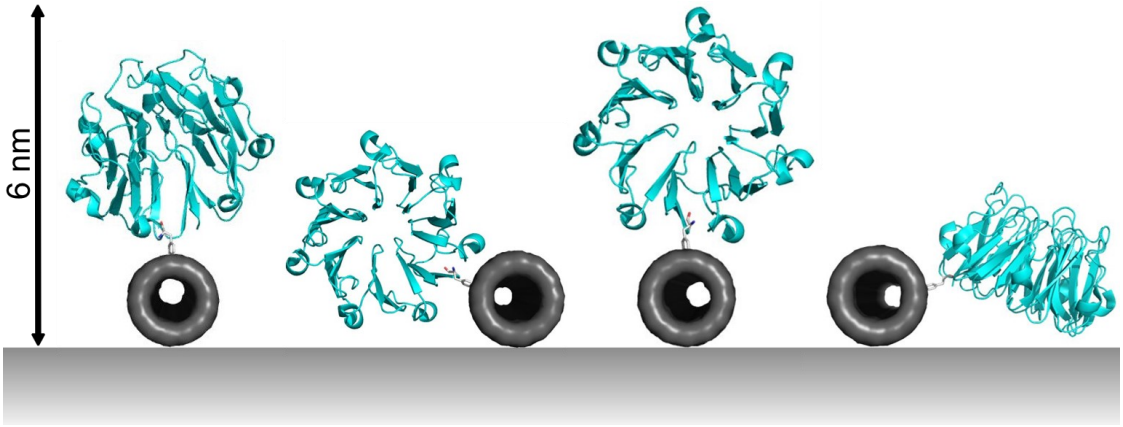


Figure 4.5 CNT functionalised with BLIP-II^{41AzF} on flat surface, four possible geometries. The protein is likely to sample many conformations on the CNT surface, accounting for the variation in heights observed by AFM, up to a maximum height of ca. 6 nm (real) An average of ca. 3.5 nm was measured by AFM. PDB 1JTD

4.2.1.3 Binding TEM to BLIP-II^{41AzF}-functionalised CNTs and AFM analysis

For the AzF photoattachment strategy to be useful, the immobilised proteins on CNT surfaces must retain their native fold and function. Contact with the hydrophobic surfaces of CNTs and similar graphitic materials such as graphene can induce partial unfolding of proteins and affect their function[246]. Following successful construction of BLIP-II^{41AzF}-functionalised CNTs, experiments were performed to assess whether β-lactamase binding of BLIP-II^{41AzF} was still functional. To assess if the BLIP-II still retained its binding capacity after attachment to CNTs, TEM^{WT} was added and the resulting change in height was measured by AFM.

BLIP-II^{41AzF}-functionalised CNTs on silicon wafer surfaces were first analysed with AFM to assess functionalisation, then incubated with TEM^{WT} solution (100 nM) to allow BLIP-II^{41AzF}-TEM^{WT} complexes to form on CNTs. The CNTs were washed and dried, then the same individual CNTs were found using reference points on the silicon surface and re-imaged using AFM. Topographical data were compared from before and after incubation with TEM^{WT} to assess formation of CNT-BLIP-II^{41AzF}-TEM^{WT} complexes. Eight transverse height profiles were measured: across bare CNTs, across bare sections of BLIP-II^{41AzF}-functionalised CNTs, and across proteins immobilised on CNTs; these profiles were averaged and plotted to compare average heights of features on the CNT. Bare CNTs

and the bare sections of functionalised CNTs showed consistent average heights of *circa* 1.6 nm (Figure 4.6) and the average height across immobilised BLIP-II^{41AzF} proteins was 2.46 nm (SD = 0.07 nm). After incubation with TEM^{WT} the average height across protein features on the BLIP-II^{41AzF}-functionalised CNTs increased to 3.04 nm (SD = 0.18 nm). This height increase suggested binding of TEM^{WT} to immobilised BLIP-II^{41AzF}, as was planned. However, in the AFM data recorded for the BLIP-II^{41AzF}-functionalised CNT (Figure 4.6c), proteins appeared flattened and fuzzy, with increased apparent width compared to the previous data. This was thought to be caused by setting the tapping mode amplitude set point of the AFM cantilever too low, which increases force between the cantilever tip and sample and can deform the sample, especially with relatively soft topographic features such as proteins[247].

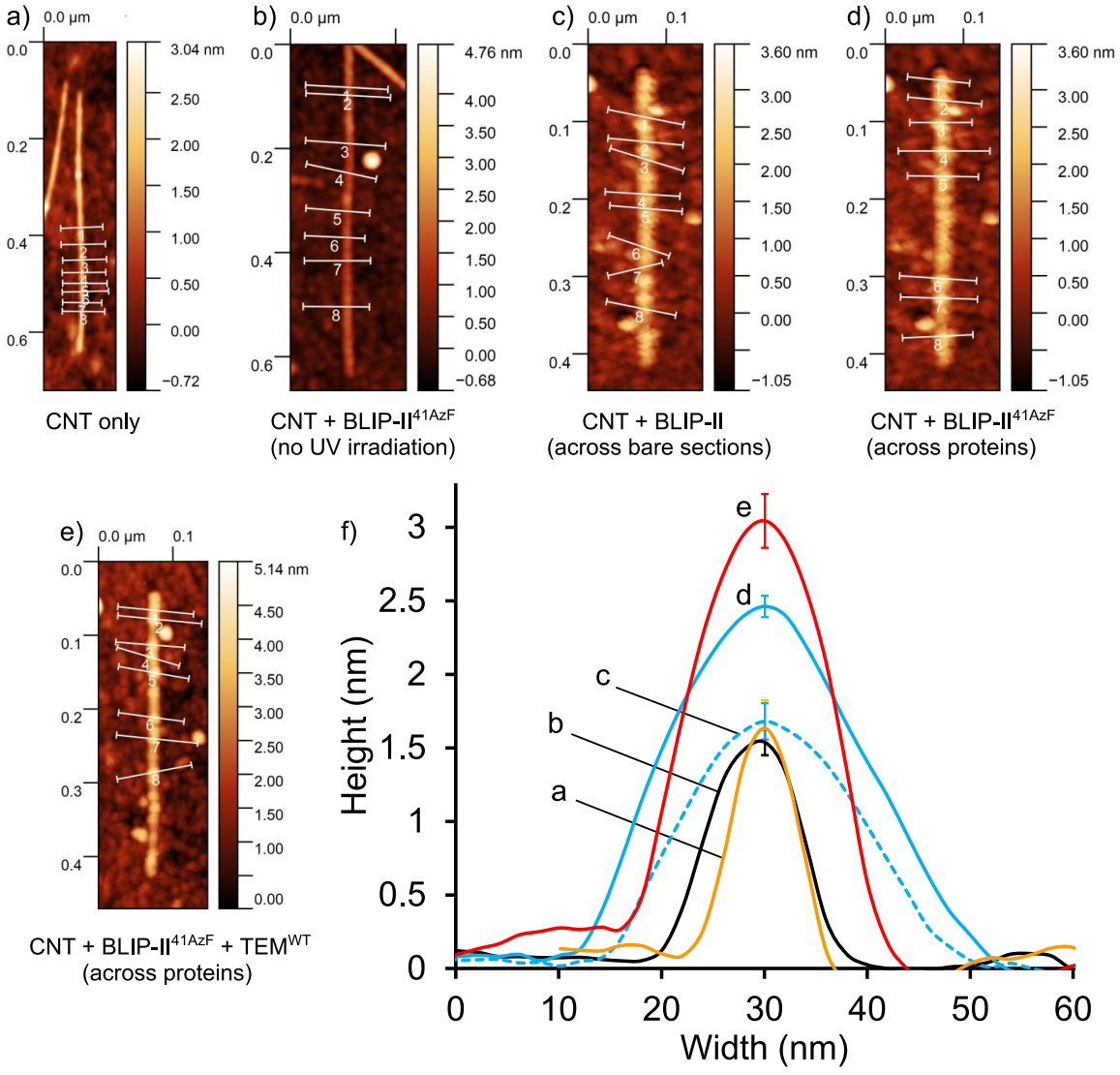


Figure 4.6 Transverse CNT AFM height profiles during functionalisation with BLIP-II^{41AzF} and TEM^{WT}. a) Bare CNT: height **1.63 nm** (SD = 0.15 nm). b) After incubation with BLIP-II^{41AzF} without UV-B irradiation: **1.54 nm** (SD = 0.09 nm). c) After incubation with BLIP-II^{41AzF} with UV-B irradiation, CNTs became decorated with globular protein-like objects. CNT height measured between the objects were **1.68 nm** high (SD = 0.12 nm. d) Height measured across the objects was **2.46 nm** (SD = 0.07 nm). e) After incubation with TEM^{WT}, height increased to **3.04 nm** (SD

= 0.18 nm), suggesting CNT-BLIP^{41AzF}-TEM^{WT} complex formation. f) Mean height profiles across **a-e;** measurements n=8, SD bars shown for highest point, measurements shown on AFM images. The same individual CNT was analysed for **c)**, **d)** and **e)**.

Because the CNT-protein AFM data showed some deformation of BLIP-II^{41AzF} due to excessive force from the AFM tip, the experiment was repeated, with data being gathered by Suzanne Thomas. Similar techniques were used for the repeated experiments, with topographical data recorded before and after photochemical functionalisation of CNTs with BLIP-II^{41AzF}, and again after incubation with TEM^{WT} to assess protein-protein complex formation. Longitudinal and transverse height profiles were taken from this data and analysed (Figure 4.7, Figure 4.8). CNTs used for this experiment were from a different batch and were slightly larger in diameter, measuring ca. 2 nm high. This amount of variation between CNT batches was typical during this work, and sourcing a more consistent supply of CNTs would be of importance if fabricating CNT-based devices in quantity. After functionalisation with BLIP-II^{41AzF}, average protein height from transverse profile analysis was 3.04 nm (SD = 0.18 nm) and after binding of TEM^{WT}, average protein heights increased to ca. 3.66 nm (SD = 0.11 nm). The height increases of ca. 1 nm on BLIP-II^{41AzF} functionalisation and ca. 0.6 nm for TEM^{WT} binding are consistent with the previous findings, confirming repeatability of the process. Some possible conformations are shown for BLIP-II^{41AzF}-TEM^{WT} complexes on CNTs (Figure 4.9). Most possible conformations of the complex show increased height compared to BLIP-II^{41AzF} alone, consistent with height differences observed in the AFM data, considering the inherent underestimation of features by AFM. The findings of all the experiments with UV-mediated nitrene attachment of proteins to CNTs indicated that it was possible to specifically functionalise CNT sidewalls with engineered proteins via their non-natural AzF residue, leaving binding function intact. These could then form complexes with a target protein on the CNT surface via protein-protein interactions, potentially forming an engineered receptor for a biomolecular analyte, and the basis of a sensing device. The functionalisation and target protein binding processes were also repeatable, showing consistent CNT morphologies and measurements.

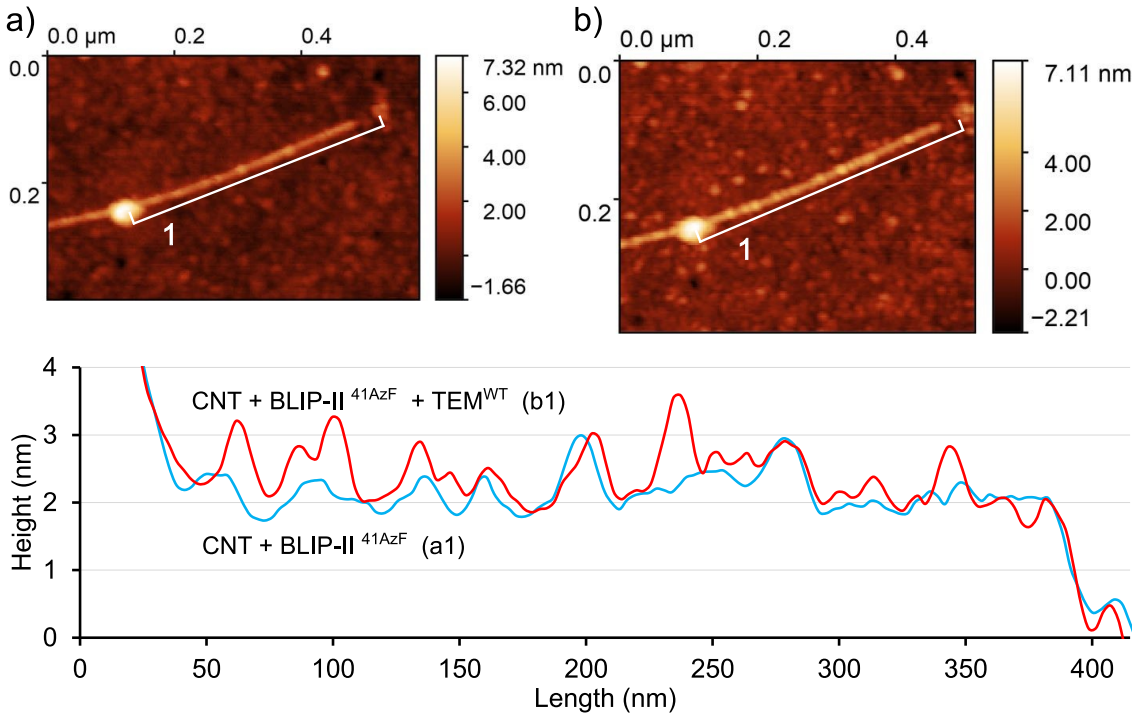


Figure 4.7 Longitudinal CNT height profiles during functionalisation with BLIP-II^{41AzF} and TEM^{WT}. a) Profile along single CNT after functionalisation with BLIP-II^{41AzF} (blue line). CNT appears decorated with globular objects with heights up to 3 nm. b) The same CNT after incubation with TEM^{WT}. Profile shows further distinct peaks reaching 3.6 nm (red line), indicating likely BLIP-II^{41AzF}-TEM^{WT} complexes formed on CNT surface. A more quantitative analysis of multiple transverse height profiles is in **Figure 4.8**.

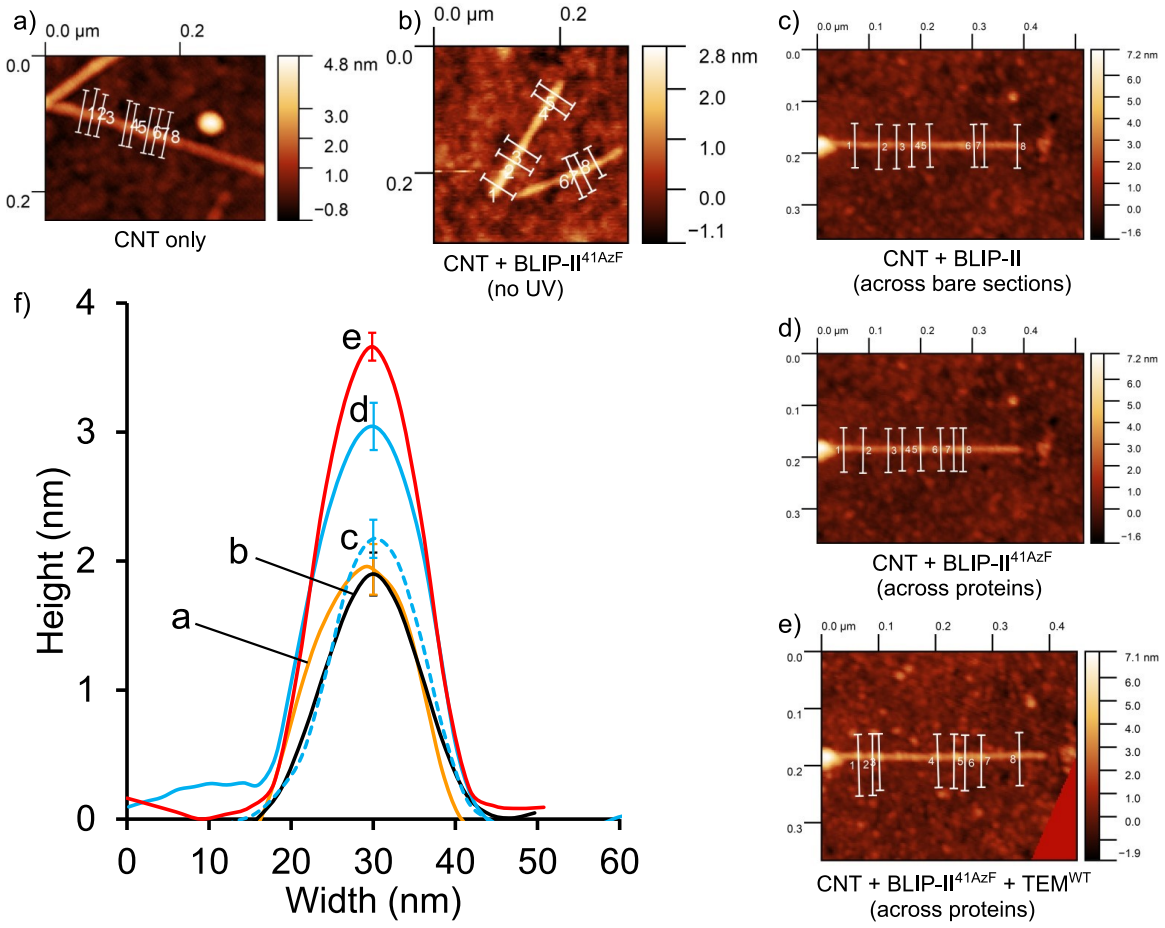


Figure 4.8 Transverse CNT AFM height profiles during functionalisation with BLIP-II^{41AzF} and TEM^{WT}. Repeated experiment. a) Bare CNT: height **1.94 nm** (SD = 0.08 nm). b) After incubation with BLIP-II^{41AzF} without UV-B irradiation: **1.90 nm** (SD = 0.07 nm). c) After incubation with BLIP-II^{41AzF} with UV-B irradiation, CNTs became decorated with globular protein-like objects. CNT height measured between the objects were **2.17 nm** high (SD = 0.15 nm. d) Height measured across the objects was **3.04 nm** (SD = 0.18 nm). e) After incubation with TEM^{WT}, height increased to **3.66 nm** (SD = 0.11 nm), suggesting CNT-BLIP^{41AzF}-TEM^{WT} complex formation. f) Mean height profiles across **a-e;** measurements n=8, SD bars shown for highest point, measurements shown on AFM images. The same individual CNT was analysed for **c)**, **d)** and **e)**.

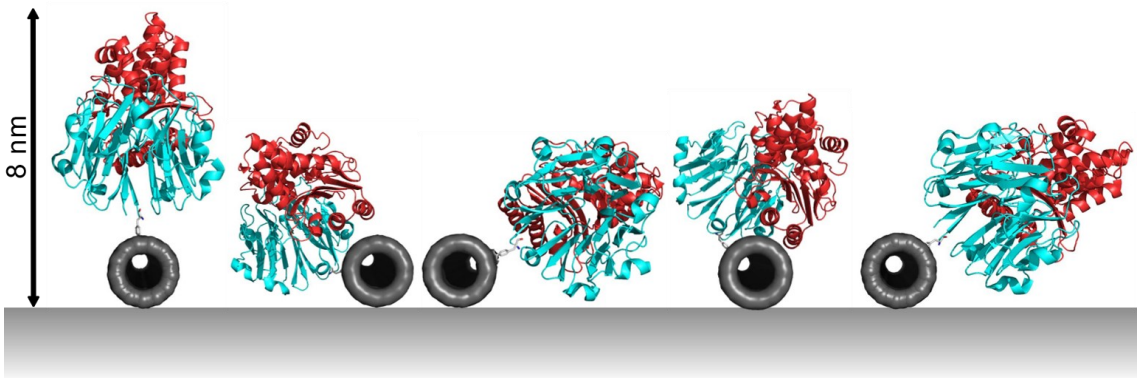


Figure 4.9 BLIP-II^{41AzF}-functionalised CNT with TEM^{WT} bound by protein-protein interactions. BLIP-II^{41AzF} (cyan), TEM^{WT} (red). The complex on CNT surface is shown in five possible conformations. Real height of the CNT-complex can be up to ca. 8 nm, with an average of 3.66 nm (SD = 0.11 nm) measured by AFM. PDB 1JTD.

4.2.2 Functionalisation of CNTs with BLIP-II^{213AzF} via nitrene photoaddition for CTX-M β-lactamase binding and analysis using TIRF microscopy

4.2.2.1 Production of BLIP-II^{213AzF} protein

Production of BLIP-II^{213AzF} is described in chapter 3.3.3.

4.2.2.2 CNT sidewall functionalisation with BLIP-II^{213AzF} via photoattachment

In addition to analysis of CNT-BLIP-II-β-lactamase complexes using AFM, it was decided to further analyse these binding interactions using TIRF microscopy. TIRF can offer single-protein resolution imaging and the technique can definitively detect fluorescent-labelled proteins at the CNT surface. For functionalisation with BLIP-II proteins and TIRF analysis, CNTs were adsorbed onto glass cover slips instead of silicon wafers, to allow optical imaging. CNTs were functionalised with the AzF-containing variant BLIP-II^{213AzF} by UV-nitrene photoattachment to the CNT surface, as described previously in section 4.2.1.2. The variant BLIP-II^{213AzF} was used in order to assess the β-lactamase-binding ability of BLIP-II in a different orientation, as it had successfully been used to functionalise CNTs using the pyrene linker method in chapter 3.

4.2.2.3 Design of CTX-M^{86Cys} variant

To examine the broad-spectrum β-lactamase binding potential of BLIP-II-functionalised CNTs a second β-lactamase, CTX-M, was engineered for covalent labelling with a fluorescent marker molecule to enable analysis by TIRF microscopy. In this case, the protein was designed to be labelled via an engineered cysteine residue instead of an engineered AzF. Cysteine-maleimide chemistry was chosen for fluorophore labelling to avoid risk of CTX-M proteins photoattaching directly to CNTs via an AzF residue activated by ambient light. This reaction is faster and more efficient than SPAAC, reaching high labelling efficiency in a few minutes at room temperature compared to ≤ 60% efficiency in one hour using SPAAC[228]. Proteins containing only the 20 canonical proteinogenic amino acids tend to express at higher quantity than those with engineered AzF mutations, as they do not rely on exogenous engineered components for translation; also, AzF-containing proteins must be kept in dark conditions to protect azide groups from photolysis.

CTX-M is a class A serine β-lactamase that shares close structural homology with TEM: amino acid sequence identity 38%, similarity 55% (via EMBOSS[143]). CTX-M was chosen for this work for two reasons: 1) CTX-M is a very common agent of AMR in pathogenic bacteria[245] and thus of practical interest as a target analyte in future sensing devices (as with TEM). 2) CTX-M contains just one native cysteine buried at the protein core while

TEM contains two cysteines: CTX-M was thought to have less potential for formation of mis-paired disulphides leading to a misfolded protein, compared to TEM. The BLIP-II-binding interactions of CTX-M have not been established and no crystal structure is available for the complex, so the structure of CTX-M (PDB 3QHY[147]) was aligned and compared to the available β-lactamase-BLIP-II complex structures using PyMOL[11]: TEM with BLIP-II (PDB 1JTD[36]) and Bla1 with BLIP-II (PDB 1YLT[37]). These three β-lactamases all bind BLIP-II with extreme affinity (K_D: CTX-M = 0.5 pM, TEM = 0.5 pM, Bla1 = 0.3 pM[147]) and share close structural homology, and both TEM and Bla1 interact with BLIP-II through a common set of conserved residues, so it was assumed that CTX-M also binds BLIP-II in the same way (Figure 4.10). The variant CTX-M^{86Cys} was designed to replace residue T86 with a surface cysteine for fluorophore labelling, at the opposite side of the protein from the assumed BLIP-II binding site. This location was chosen to minimise impact on CTX-M enzyme and binding functions by being as far as possible from the active/binding sites and was also located on a relatively unstructured loop (B-factor ca. 20 Å^2), avoiding residues involved in secondary structural features (Figure 4.11).

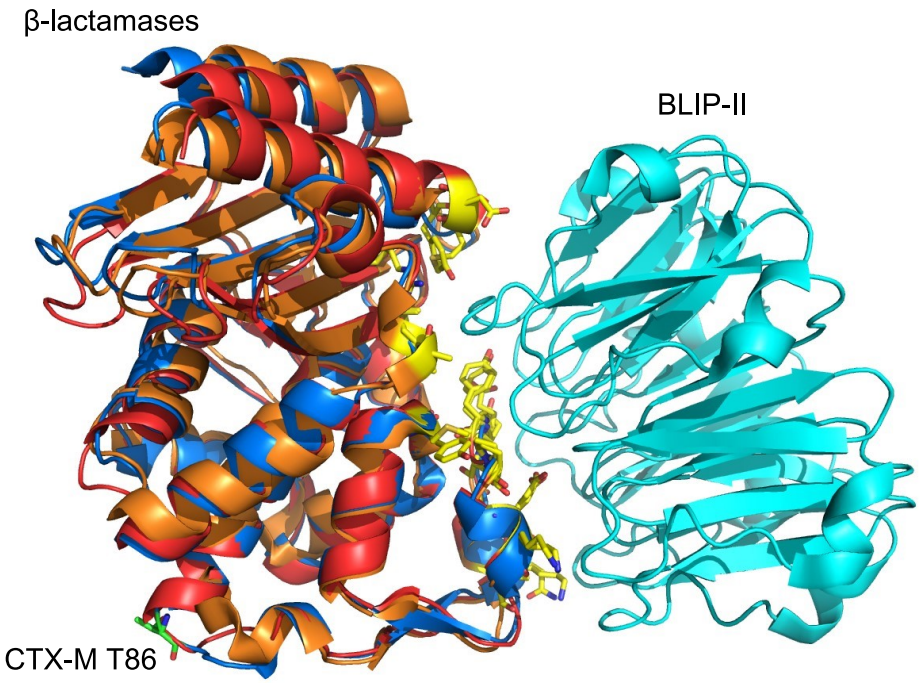


Figure 4.10 Structures of three class A β-lactamases superimposed, bound to BLIP-II (cyan). Blue: CTX-M; red: TEM; orange; Bla1. These β-lactamases share a highly conserved fold and extremely high affinity for BLIP-II (0.3-0.5 pM). Conserved BLIP-II-interacting residues of TEM and Bla1 are shown in yellow. It is assumed that this common BLIP-II-binding interface is also shared by CTX-M. Residue mutated to Cys for fluorescent labelling of CTX-M (T86) is shown in green. PDB 1JTD[36], 3QHY[147], 1YLT[37].

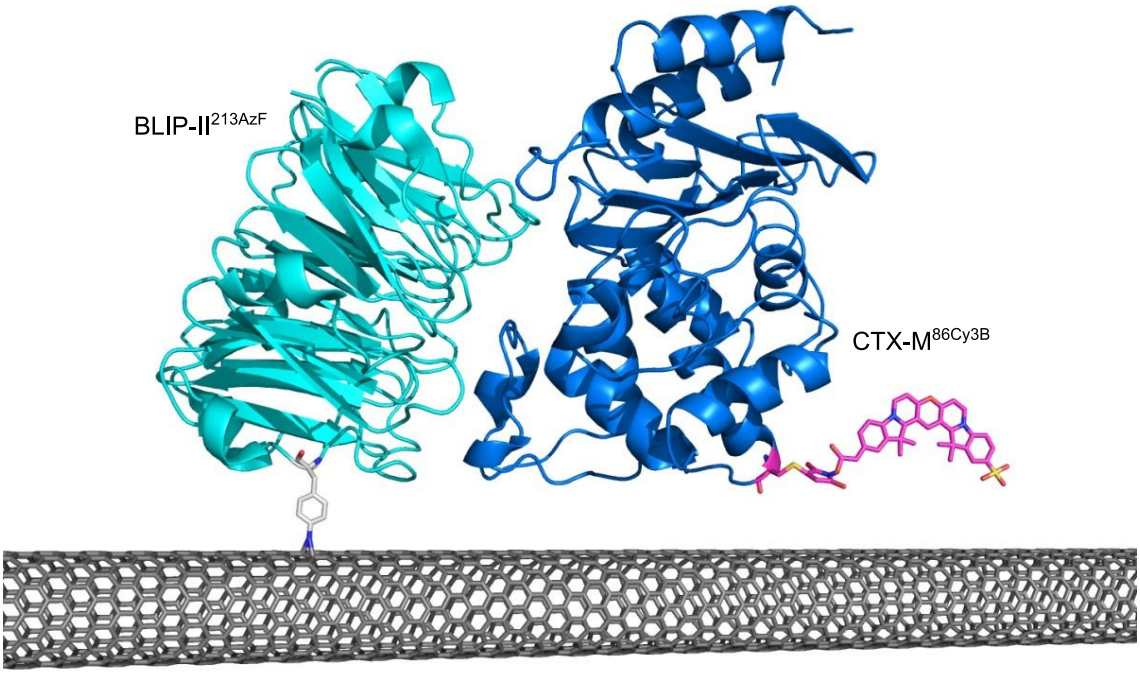


Figure 4.11 Projected structure of BLIP-II^{213AzF} immobilised on CNT and bound to fluorescent-labelled CTX-M^{86Cy3B}. Light grey: UV-mediated covalent BLIP-II^{213AzF}-aziridine link to CNT. Magenta: Cy3B fluorophore covalently added to CTX-M^{86Cys}. This construct was designed for analysis of complex formation on CNTs at single-molecule resolution using TIRF microscopy. Made using PDB 1JTD[36], 1YLT[37] by manual editing with ChemAxon MarvinSketch[248] and PyMOL.

4.2.2.4 CTX-M^{86Cys} variant molecular biology

The CTX-M^{86Cys} variant of the CTX-M gene was made by PCR whole-plasmid site-directed mutagenesis as described previously, to insert a TGC codon at residue 86. The pET-CTXM14 plasmid containing the wild-type CTX-M gene, and as a template for mutagenesis, was supplied by Prof. Timothy Palzkill[184]. Similar PCR mutagenesis conditions were used as with BLIP-II variants (chapter 3), a primer annealing temperature of 60 °C was used. PCR mutagenesis and sequencing primers are listed in chapter 2.2.2.3.

PCR products from pET-CTXM14 mutagenesis were analysed by agarose gel electrophoresis and a band was visible at the expected size of *ca.* 6 kpb (Figure 4.12a). Amplified plasmids were ligated and transformed into *E. coli* BL21 (DE3) cells by electroporation. Cells were grown on kanamycin agar to select transformants, resulting in 10-100 colonies per variant. Because previous mutagenesis attempts sometimes produced repeats or deletions within the gene of interest, five colonies were picked from each transformation and screened by PCR to confirm gene size. Amplified fragments from each colony were analysed by agarose gel electrophoresis (Figure 4.12b) to compare fragment size. Colonies that produced gene fragments matching CTX-M^{WT} size (ca. 1.2 kbp) were sequenced to check that the intended mutation was made and that

the rest of the gene fragment matched the wild-type gene. Bacteria containing the successfully mutated CTX-M variant plasmids were archived at -80 °C in 25% glycerol.

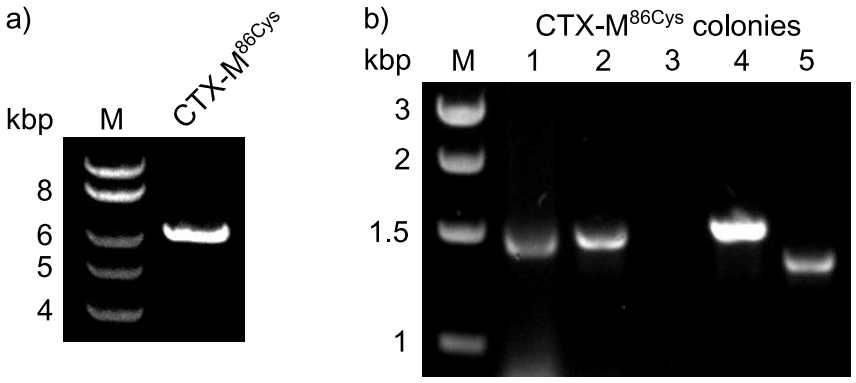


Figure 4.12 Creation of CTX-M^{86Cys} variant by whole-plasmid site-directed mutagenesis. a) Whole pET-CTXM14 plasmid from PCR mutagenesis. b) Colony PCR amplification with T7 primers of *E. coli* colonies after transformation with the ligated CTX-M^{86Cys} mutant plasmid. Colonies that produced gene fragments were sequenced; one contained the intended mutation. M: kilobase pair DNA marker ladder.

4.2.2.5 CTX-M expression and purification

CTX-M^{WT} and the CTX-M^{86Cys} variant were transformed, induced and grown in *E. coli* BL21 (DE3) cells by the same methods described previously for TEM^{WT} and BLIP-II^{WT} in chapter 3. Cell cultures were tested for CTX-M expression by addition of the chromogenic cephalosporin substrate nitrocefin: if a β-lactamase is present, nitrocefin is hydrolysed, turning from yellow to red. Cell cultures testing positive for CTX-M were lysed then centrifuged at 25,000 ×*g* to clarify the lysate. These CTX-M constructs are expressed into the cell cytoplasm and have N-terminal hexahistidine tags (His-tag) for purification. CTX-M^{WT} and CTX-M^{86Cys} were purified from cell lysate using an IMAC gravity flow column containing HisPur Cobalt agarose (Thermo Fisher) with single-step elution by 100 mM imidazole, as described previously for BLIP-II variants (chapter 3). Fractions from IMAC elution were analysed by SDS-PAGE and found to be relatively pure (Figure 4.13), the fractions were concentrated and further purified and desalted by size exclusion chromatography using a Superdex 200 26/600 column (Cytiva). Elution fractions were analysed by SDS-PAGE (Figure 4.14) and purity was estimated using ImageJ densitometry software. Final samples of CTX-M^{WT} and CTX-M^{86Cys} were estimated at ≥ 99% pure.

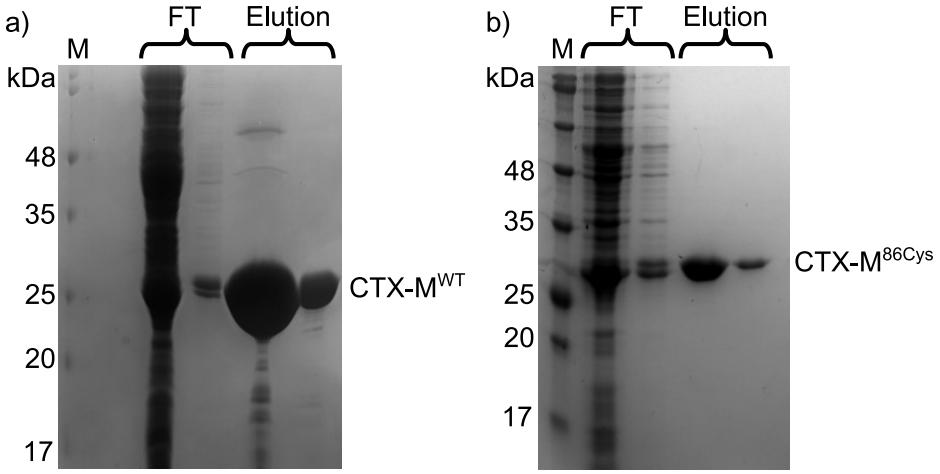


Figure 4.13 Purification of CTX-M^{WT} and CTX-M^{86Cys}. Fractions from cobalt IMAC purification from cell lysate for CTX-M^{WT} and CTX-M^{86Cys}. FT: flow-through containing unbound proteins. M: molecular weight marker ladder.

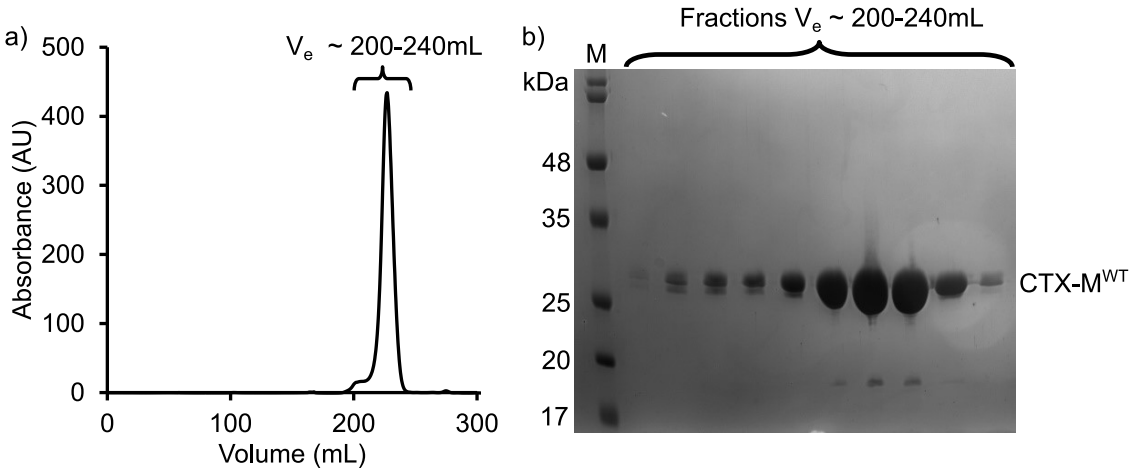


Figure 4.14 Purification of CTX-M^{WT}. a) Elution of purified CTX-M^{WT} from Superdex 200 26/600 size exclusion column between 200-240 mL volume. b) Fractions from SEC showing ≥ 99% pure CTX-M^{WT} eluted at V_e 200-240 mL. M: molecular weight marker ladder.

4.2.2.6 Fluorescent labelling of CTX-M^{86Cys} variant

The purpose of the CTX-M^{86Cys} variant was to bind to BLIP-II-functionalised CNTs in order to capture single-molecule resolution data for BLIP-II-CTX-M complexes on CNT surfaces using TIRF microscopy. This data will allow us to analyse the extent of CNT decoration with BLIP-II-β-lactamase complexes by a second method, in addition to the AFM analysis in section 4.2.1.3. This approach should provide further evidence of β-lactamase binding to immobilised BLIP-II^{213AzF} at the CNT surface as BLIP-II^{213AzF} is not fluorescent and CTX-M^{86Cys} should not itself bind to CNT surfaces; any fluorescent decoration of CNTs should be attributable to complex formation.

For this, CTX-M^{86Cys} needed to be specifically labelled with a suitable fluorophore, in this case the rigidified trimethine cyanine dye Cy3B-maleimide (see chapter 2.2.9.3), to form

the construct CTX-M^{86Cy3B}. The dye carries a maleimide reactive group for covalent conjugation to the thiol group of the CTX-M^{86Cys} via formation of a thioester bond. For labelling, CTX-M^{WT} and CTX-M^{86Cys} were incubated with a twofold excess of Cy3B-maleimide, for five minutes at room temperature, then Cy3B-maleimide dye was quickly removed by gel filtration (PD10 mini column, Cytiva). This was to minimise any labelling of the internal, buried Cys residue of CTX-M^{WT} or CTX-M^{86Cys}, while labelling the surface Cys of CTX-M^{86Cys}. These samples were analysed by SDS-PAGE, with sample boiling time reduced to one minute to minimise unfolding and labelling of the buried cysteines. CTX-M^{WT} showed no fluorescent labelling under UV illumination, while CTX-M^{86Cys} was fluorescent (**Figure 4.15**), indicating specific labelling of CTX-M^{86Cys} Cys86 residue was achieved. The Cy3B-labelled CTX-M^{86Cys} variant was thereafter called CTX-M^{86Cy3B}.

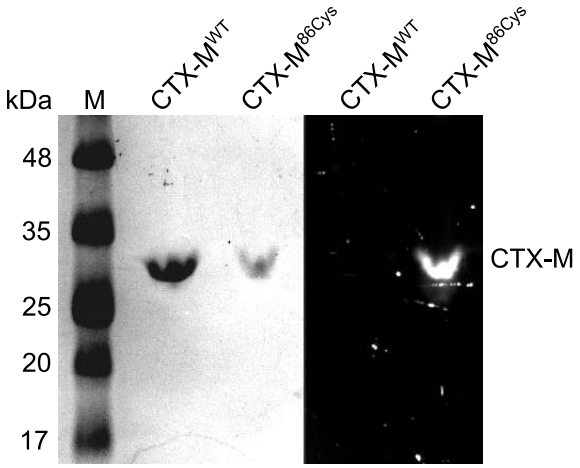


Figure 4.15 Specific fluorescent labelling of the engineered Cys86 of CTX-M^{86Cys} with Cy3B-maleimide. Left: Coomassie stain, right: UV illumination.

4.2.2.7 Characterisation of CTX-M^{86Cys} variant catalytic function

Before using CTX-M^{86Cy3B} in protein binding experiments, it was necessary to confirm that the labelled protein was still functional after covalent modification with the Cy3B fluorophore. A catalytic assay was performed to compare the function of the dye-labelled protein with unlabelled CTX-M^{86Cys} and CTX-M^{WT} to assess the impact of labelling on function. Hydrolysis of the chromogenic cephalosporin substrate nitrocefin at 40 μM by the CTX-M (1 nM) variants was monitored over 15 minutes. Hydrolysis of nitrocefin by a β-lactamase results in a shift in its absorbance maximum from 390 nm to 486 nm, so the increase in absorbance at λ_{486} was monitored. The absorbance peak of Cy3B has minimal overlap with nitrocefin, with its maximum at 558 nm.

The enzymatic function of CTX-M^{WT} and CTX-M^{86Cys} were comparable, with initial nitrocefin hydrolysis rates of 5.56 and 5.18 μM/min/nM (micromolar nitrocefin hydrolysed per minute, per nanomolar enzyme) respectively; after dye labelling, the catalytic function of CTX-M^{86Cy3B} was approximately halved, initial rate 2.42 μM/min/nM (Figure 4.16a,

Table 4.1). There appears no obvious reason for the large reduction in catalytic function on dye labelling, Cy3B is a large fluorophore but is situated far from the active site of CTX-M^{86Cy3B} on the opposite side of the protein. The closest Cy3B could possibly approach the active site residues S70, K73 or E166 is still more than 15 Å distant (**Figure 4.16**b), so could not be directly blocking the β-lactam binding site. Residue 86Cys is however at the distal end of an α-helix from residue K73 which is thought to be an important component of the enzyme proton shuttle system during β-lactam hydrolysis[249]; the large, nonpolar fluorophore may in some way disrupt the stability or conformation of this region. Despite the reduction in catalysis, the enzyme remained functional.

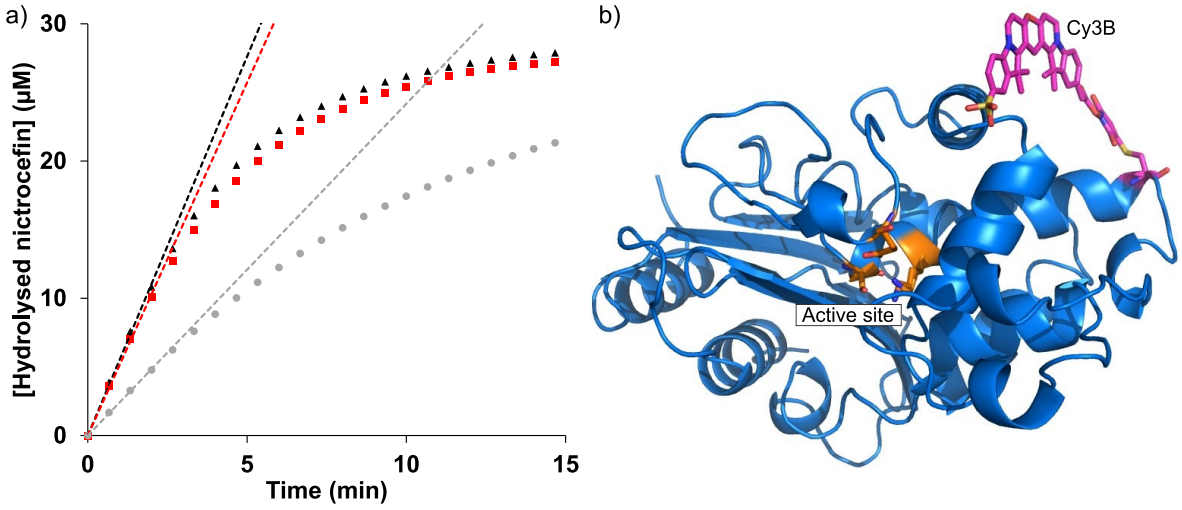


Figure 4.16 Effect of fluorophore labelling on CTXM^{86Cys} catalysis. a) CTX-M^{WT} (black triangles, initial rate shown as dashed line), CTX-M^{86Cys} (red squares & dashed line) and CTX-M^{86Cy3B} (grey circles & dashed line). The negligible difference in catalytic function between CTX-M^{WT} and CTX-M^{86Cys} shows this variant is fully functional. CTX-M^{86Cy3B} has compromised catalytic function, with initial rate halved after labelling. b) Theoretical structure of CTX-M^{86Cy3B} with the Cy3B fluorophore (magenta) at closest allowable distance (> 15 Å) to active site residues (orange). PDB 1YLT. Structure made manually with MarvinSketch and PyMOL.

Table 4.1 Nitrocefin catalysis initial rates for CTX-M variants: CTX-M^{WT} and CTX-M^{86Cys} had comparable function, while CTX-M^{86Cy3B} had less than half of the catalytic function. Final concentration of nitrocefin was 40 μM.

	CTX-M^{WT}	**CTX-M^{86Cys}**	**CTX-M^{86Cy3B}**
Initial rate (mol/min/mmol enzyme)	5.56	5.18	2.42
Goodness of fit (R^2)	> 0.99	> 0.99	> 0.99

4.2.2.8 BLIP-II - CTX-M binding and inhibition

Before attempting CNT functionalisation with BLIP-II^{213AzF} and subsequent protein-protein binding with CXT-M^{86Cy3B}, binding affinity of the pair was assessed using inhibition assays as described previously in chapter 3 and methods, chapter 2. Apparent inhibition

constants (K_i^{app}) were calculated for inhibition of CTX-M^{WT}, CTX-M^{86Cys} and CTX-M^{86Cy3B} by BLIP-II^{WT} and BLIP-II^{213AzF}. Nitrocefin hydrolysis initial rates were taken with 1 nM CTX-M, then BLIP-II was titrated in until full inhibition was reached. The values were fitted to the Morrison Equation in Prism (see methods, chapter 2) to estimate the K_i^{app} for each CTX-M - BLIP-II pair (**Figure 4.17**, **Table 4.2**). The apparent inhibitory constant for BLIP-II^{WT} with CTX-M^{WT} was calculated at K_i^{app} = 5 ± 3 pM (published value 0.5 pM[154]). The shortcomings of the method we used are known (see chapter 3) but it was judged adequate to estimate relative binding/inhibitory function between the engineered proteins. The affinity between BLIP-II^{WT} and CTX-M^{86Cys} was similar at K_i^{app} = 18 ± 5 suggesting the 86Cys mutation had little effect on binding, as was intended during design. This mutation was deliberately placed far from the binding face on a comparatively less structured loop to minimise disruption to CTX-M native conformation (see section 4.2.2.3).

Affinity between the AzF mutant BLIP-II^{213AzF} and CTX-M^{86Cys} was lower, with a K_i^{app} = 195 ± 17 pM, consistent with the value recorded for BLIP-II^{213AzF} with TEM^{WT} previously (157 ± 20 pM). After labelling with the fluorescent dye, CTX-M^{86Cy3B} had similar affinity for BLIP-II^{213AzF}, with K_i^{app} = 113 ± 15, showing that the dye had no effect on binding affinity.

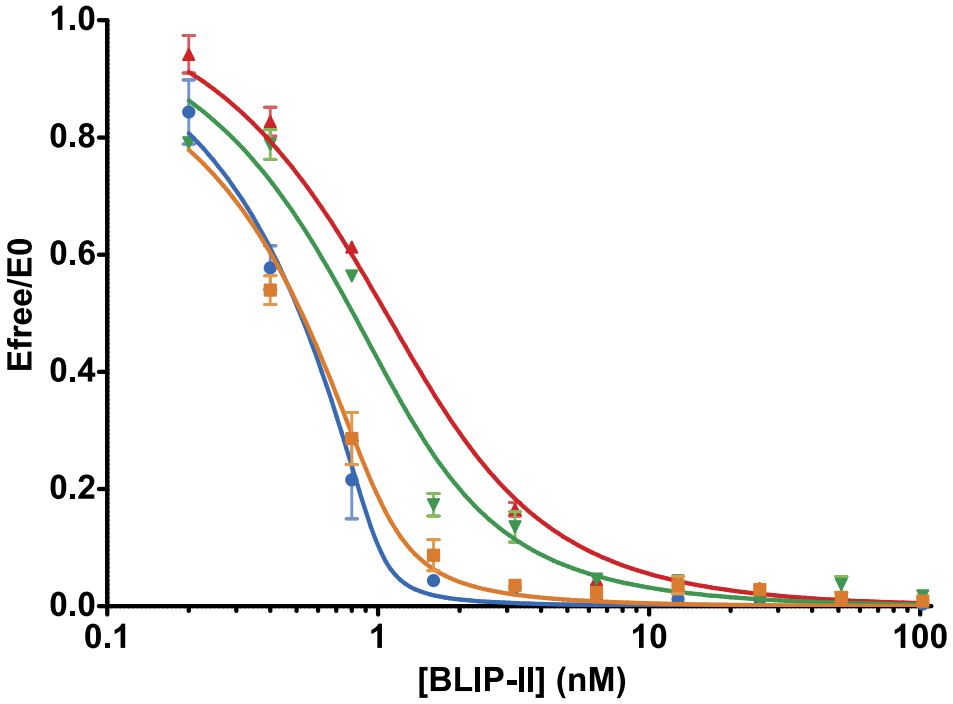


Figure 4.17 Plot of CTX-M nitrocefin hydrolysis inhibition by BLIP-II variants. In order of binding affinity, left to right: Blue: CTX-M^{WT} with BLIP-II^{WT}. Orange: CTX-M^{86Cys} with BLIP-II^{WT}. Green: CTX-M^{86Cy3B} (dye labelled) with BLIP-II^{213AzF}. Red: CTX-M^{86Cys} with BLIP-II^{213AzF}. Values were fitted to the Morrison tight binding equation to calculate K_i^{app} (Table 4.2). Error bars are SEM of three biological replicates.

Table 4.2 Apparent dissociation constants (K_i^{app}) of BLIP-II^{213AzF} and CTX-M^{86Cys}, with and without Cy3B dye labelling. BLIP-II^{WT} with CTX-M^{WT} is also given for comparison. Parameters including goodness of fit values were calculated by fitting results to the Morrison tight binding equation in GraphPad Prism.

	CTX-M^{WT} & BLIP-II^{WT}	**CTX-M^{86Cys} & BLIP-II^{WT}**	**CTX-M^{86Cys} & BLIP-II^{213AzF}**	**CTX-M^{86Cy3B} & BLIP-II^{213AzF}**
K_i^{app} (pM) (±SE)	5 ± 3	18 ± 5	195 ± 17	113 ± 15
s_{yx} [a]	0.04	0.04	0.03	0.05
Pseudo-R^2 [b]	0.99	0.99	0.99	0.98
95% CI	0 to 11	8 to 28	160 to 230	82 to 144

[a] Standard error of the estimate, in units of Y.
[b] Pseudo-R^2 for nonlinear regression as per method 1 in Kvålseth (1985)[223].

4.2.2.9 CNT sidewall functionalisation with BLIP-II^{213AzF} via photoattachment

CNTs were functionalised with BLIP-II^{213AzF} as described for BLIP-II^{41AzF} in section 4.2.1.2.

4.2.2.10 TIRF imaging of BLIP-II functionalised CNTs complexed with fluorescent CTX-M^{86Cy3B}

Experiments were performed with Rebecca Gwyther (Jones lab) and Dr. David Jamieson (Cardiff School of Pharmacy) to use TIRF microscopy to analyse CNTs functionalised with BLIP-II^{213AzF} then incubated with CTX-M86^{Cy3B} to form complexes. Our lab had recently used TIRF to analyse binding of engineered GFP to CNTs down to single-molecule resolution[163], so TIRF was considered a useful way to further confirm formation of CNT-BLIP-II-β-lactamase complexes. The fluorescently-labelled CTX-M^{Cy3B} was made for this purpose. Methods of functionalisation were to be the same as for AFM analysis (section 4.2.1), followed by imaging using the TIRF microscope's 532 nm laser to excite the Cy3B fluorophore covalently attached to the engineered cysteine at CTX-M residue 86.

The TIRF imaging was not successful, initial trials with bare CNTs drop-cast onto glass slides revealed a very strong background fluorescence signal when excited at available laser wavelengths. This background was found to be a fluorescent contaminant in the SDS used to disperse and solubilise the CNTs in water (**Figure 4.18**a). More vigorous washing and soaking failed to reduce this contamination to acceptable levels, so ultra-purified SDS was sourced from Sigma-Aldrich: this also contained fluorescent contaminants, but to a lesser extent. More CNTs were solubilised, drop-cast on glass slides and washed very thoroughly, then incubated with CTX-M^{Cy3B} for 5 minutes to

assess potential non-specific binding of the protein to CNTs. It was found that CNTs had become completely coated with CTX-M^{Cy3B} (Figure 4.18b). As the closely-related β-lactamase TEM^{WT} had showed no evidence of this kind of non-specific binding to CNTs during the AFM experiments it was thought that the covalently-linked Cy3B fluorophore was probably binding to the CNTs. Being a large, planar fluorophore with an extensive delocalised π-electron system it is likely that Cy3B can interface with CNTs via π-π stacking as has been well established with pyrene (chapter 3). This effect complicated the efforts to image our CNT-protein constructs via TIRF and no time was left to continue.

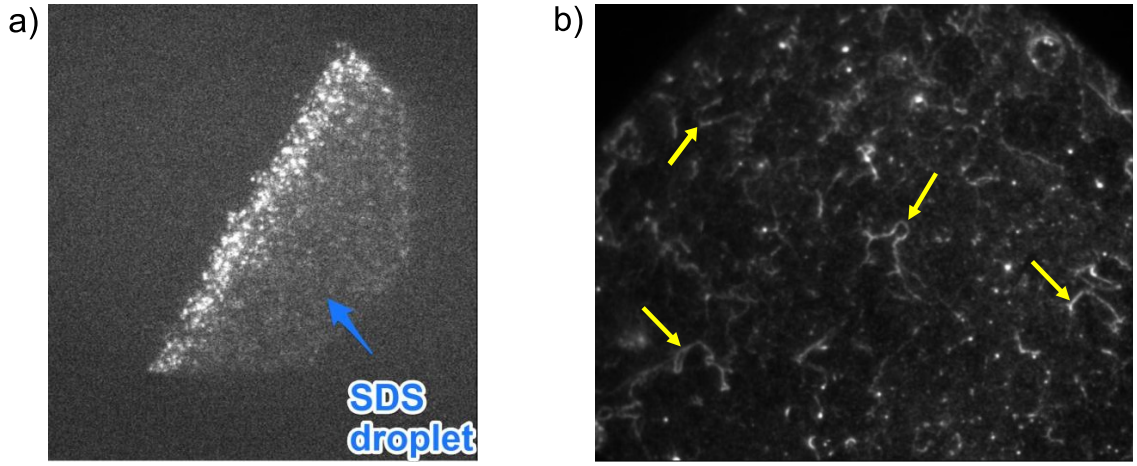


Figure 4.18 TIRF microscopy experiments with CNTs. a) Fluorescent contamination in SDS used for dispersion and solubilisation of CNTs, giving excessively noisy background. b) Image following incubation of non-functionalised CNTs with CTX-M^{86Cy3B} and thorough washing, intended to be a negative control. The fluorescent-labelled proteins bound non-specifically, appearing to coat CNTs (arrowed).

4.2.3 Functionalisation of CNTs with TEM^{87AzF} via nitrene photoaddition and analysis using AFM

4.2.3.1 Design of TEM^{87AzF} phenyl azide mutant

TEM is a class A serine β-lactamase enzyme that inactivates β-lactam antibiotics by hydrolysis of the β-lactam ring amide; it is among the most common agents of antimicrobial resistance (AMR) found in bacterial populations. The clinical relevance of TEM also means that the enzyme has been thoroughly characterised, and much information is available about its structure and function[250]. This section discusses the covalent functionalisation of CNTs with TEM by photochemical [2+1] nitrene cycloaddition, using the same methods as with BLIP-II (section 4.2.1), to demonstrate the wide applicability of this oriented conjugation approach with other proteins. The same approach was taken when designing TEM^{87AzF} as previously with CTX-M^{86Cys} (section 4.2.2) and the BLIP-II AzF variants (chapter 3), using the crystal structure of the TEM-BLIP-II complex (PDB 1JTD) to inform residue choice. Preserving the native function of

the subject protein is essential when integrating proteins into nanoscale systems. TEM has previously retained enzyme activity when immobilised on, or inside CNTs[251] and we aim to construct CNT components populated with enzymatically-active TEM in a consistent orientation, towards a goal of better understanding immobilised enzyme conformational effects on CNT signal transduction.

A surface-accessible TEM residue was selected for mutation to AzF and conjugation to the CNT surface, with the aim of avoiding areas of the protein that are 1) important to catalytic function and 2) integral to TEM secondary structure and fold, to prevent disruption of protein structure and function. The residue selected for mutation to AzF was G87, which is at the same position in TEM as the Cys86 mutation made to CTX-M^{86Cys} in section 4.2.2. This location was intended to present the β-lactam catalytic site outwards to solution to maintain TEM catalytic function when immobilised on CNTs. G87 is at the outer surface of the all-α-helical domain of TEM, forming part of a relatively unstructured loop at the far end of an α-helix from the active site residues S70 and K73 (**Figure 4.19**)[252]. This position was judged to be far enough away from the active site so the G87AzF mutation was not expected to impact TEM structure or function significantly.

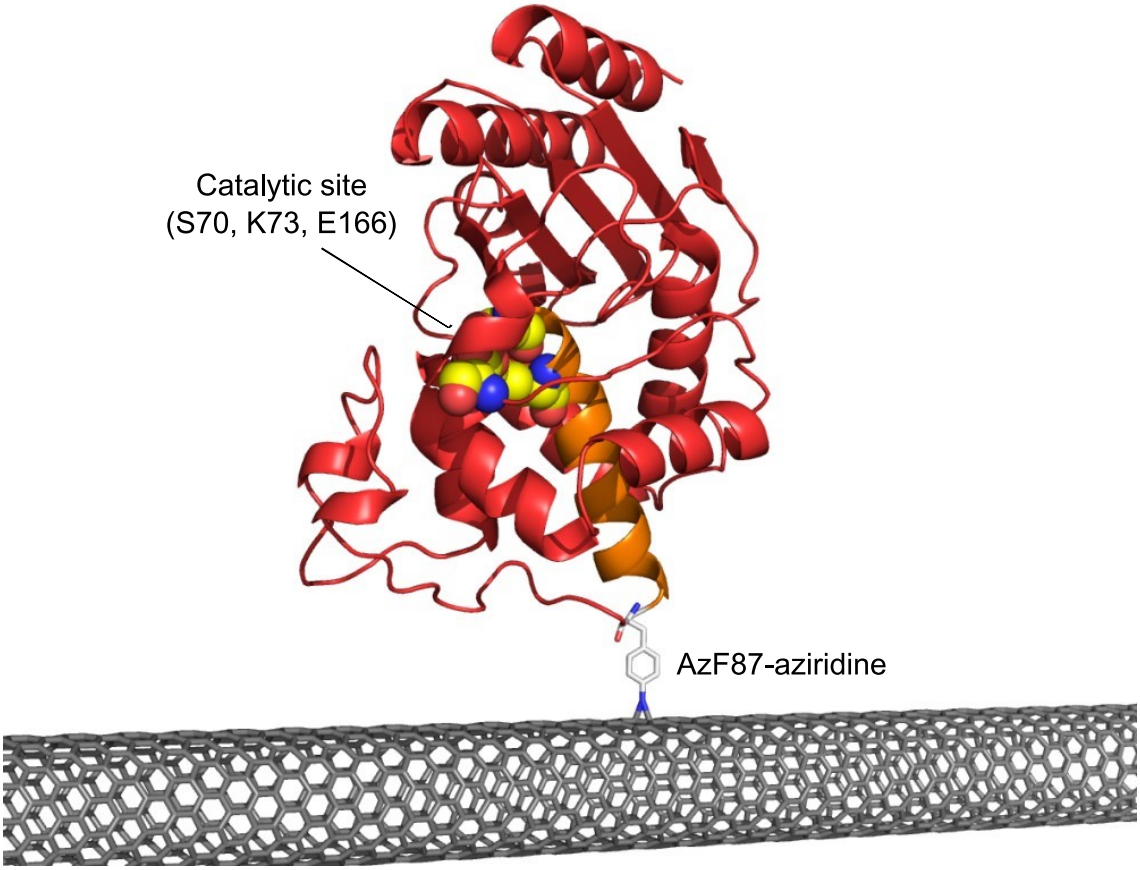


Figure 4.19 Location of AzF87 mutation in TEM^{87AzF} conjugated to CNT (projected structure). Yellow: TEM catalytic site; light grey: AzF87; orange: α-helix between AzF87 and catalytic site. The mutation was designed to conjugate TEM close to the CNT with catalytic and binding sites facing outwards. PDB 1JTD[36].

4.2.3.2 TEM molecular biology

The pET-TEM1 plasmid that was used to express TEM^{WT}, and as a template for site-directed mutagenesis, was originally made by Sosa-Peinado *et al.*[185].

The plasmid containing the mutant gene designated TEM^{87TAG}, which contains a TAG codon substitution for incorporation of AzF, was made by whole-plasmid PCR mutagenesis. Similar PCR conditions were used as described previously in chapter 3. A primer annealing temperature of 60 °C was used to make the TEM^{87AzF} mutations, PCR mutagenesis primers are listed in chapter 2. PCR product was analysed by agarose gel electrophoresis to confirm amplification of pET-TEM1, with the predicted size being 7 kpb (**Figure 4.20**). Amplified plasmids were ligated and transformed into *E. coli* (DH5α) cells by electroporation, yielding > 10 colonies. Colonies were sequenced to confirm the required mutation was incorporated and bacteria containing TEM^{87TAG} were cryopreserved in 25% glycerol at -80 °C for archiving.

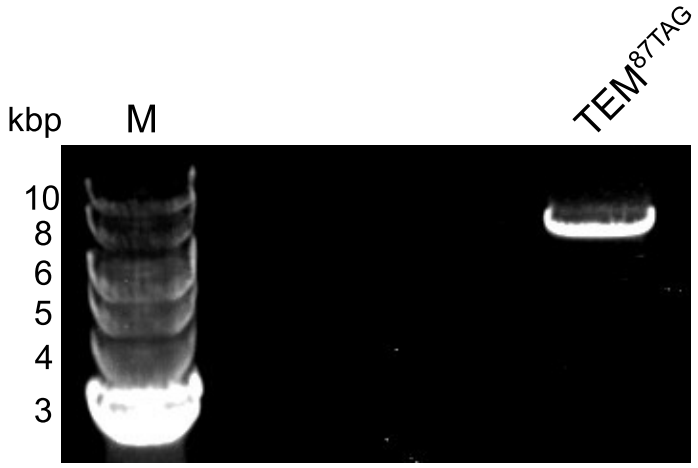


Figure 4.20 PCR product from TEM mutagenesis. The TAG codon mutation was inserted into the TEM gene in the pET-TEM1 plasmid (7 kbp) for expression of TEM AzF proteins. Empty lanes contain other, unsuccessful amplifications. M: kilobase pair DNA marker ladder.

4.2.3.3 TEM^{87AzF} expression and purification

The pET-TEM1 plasmid containing the TEM^{87TAG} gene was co-transformed into *E. coli* (BL21) with the pEVOL-AzF plasmid[79] that encodes the cellular machinery required for AzF incorporation (see methods, chapter 2). Expression quantity of TEM^{87AzF} was lower than TEM^{WT}, as is typical for proteins that rely on the engineered tRNA/aaRS system. TEM^{87AzF} was purified using the same FPLC methods as TEM^{WT} in chapter 3, but the UV lamp of the ÄKTApurifier FPLC system was disabled to protect the phenyl azide group from photolysis, so no absorbance trace is available. TEM^{87AzF} elution was detected in elution fractions by SDS-PAGE analysis. It was found that unlike TEM^{WT}, TEM^{87AzF} bound to the anion exchange column, eluting at low salt concentration (< 50 mM NaCl), the protein was then further purified and desalted by size exclusion chromatography then

analysed by SDS-PAGE (**Figure 4.21**). This yielded a small amount of pure TEM^{87AzF}, though enough for the planned enzymology and CNT binding experiments.

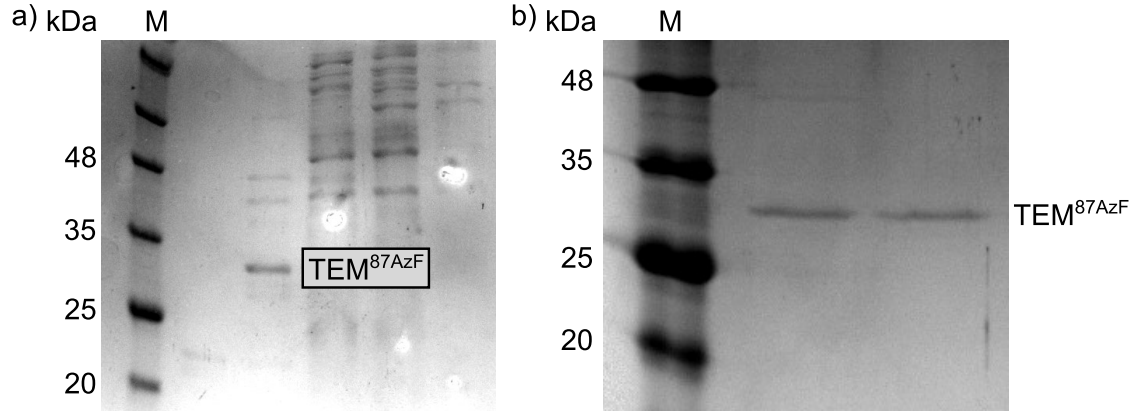


Figure 4.21 Purification of TEM^{87AzF}. a) SDS-PAGE showing fractions from elution of TEM^{87AzF} from anion exchange column. b) Purified and desalted TEM^{87AzF} after size exclusion chromatography. M: molecular weight marker ladder.

4.2.3.4 TEM^{87AzF} phenyl azide click chemistry

The presence and reactivity of the genetically-encoded AzF residue in TEM^{87AzF} was examined by SPAAC labelling with the azide-specific strained-ring alkyne fluorophore DBCO-TAMRA (method and reaction scheme, chapter 2.2.9.1). A sample containing TEM^{87AzF} was incubated with DBCO-TAMRA for 1 hour then analysed by SDS-PAGE and the gel was imaged by UV illumination to visualise fluorescent labelled proteins. TEM^{87AzF} was labelled with DBCO-TAMRA and was visible as fluorescent bands just above 25 kDa, this confirmed that AzF was incorporated in TEM^{87AzF} and was reactive.

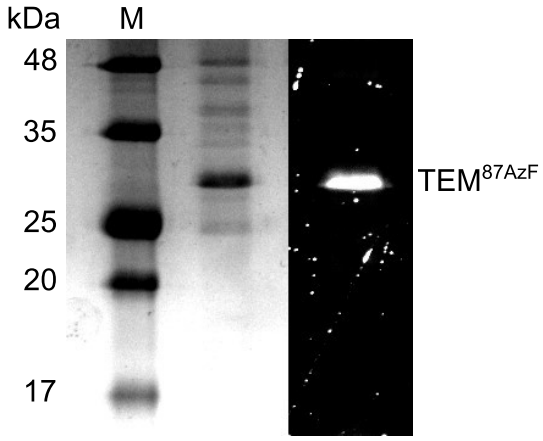


Figure 4.22 Fluorescent labelling of TEM^{87AzF} with azide-specific DBCO-TAMRA. Left: Coomassie stain; right: UV illumination. This protein sample was relatively impure and was further purified.

4.2.3.5 Characterisation of TEM^{87AzF} enzyme activity

To assess the effect of the TEM^{87AzF} mutation on protein function, the enzyme activity of this variant was analysed. Photoactivation of the AzF azide to nitrene for [2+1]

cycloaddition to the CNT requires a five minute irradiation with an ultraviolet LED (intensity 18 $W \cdot m^{-2}$) and the enzyme must remain functional when immobilised on the CNT. To establish whether UV-B irradiation damages TEM^{87AzF} enzyme activity, an enzyme assay was performed to compare the activity of TEM^{87AzF} (before and after 30 minutes UV-B irradiation), with TEM^{WT}.

The catalytic function of TEM^{87AzF} and TEM^{WT} were tested by recording the hydrolysis of 40 μM nitrocefin over 20 minutes by TEM^{WT} and TEM^{87AzF} at 0.3 nM. Nitrocefin hydrolysis was monitored at 486 nm as described previously in chapter 3. The experiment was also repeated for TEM^{87AzF} following 30 minutes exposure to UV-B irradiation using a lamp with λ_{max} ca. 300 nm at 1 cm distance, to discover whether photolysis of the TEM^{87AzF} phenyl azide would affect enzyme function. It was found that TEM^{WT} and TEM^{87AzF} had comparable catalytic function, with initial rates of 5.14 and 4.74 mol/min/mmol enzyme respectively. After TEM^{87AzF} UV-B exposure enzyme activity remains essentially the same (4.54 mol/min/mmol enzyme) (Figure 4.23, Table 4.3). As TEM^{87AzF} was found to function normally after this high dose of UV-B radiation, it was judged suitable for conjugation to CNTs by UV-mediated nitrene cycloaddition, which requires just five minutes of similar intensity UV-B exposure.

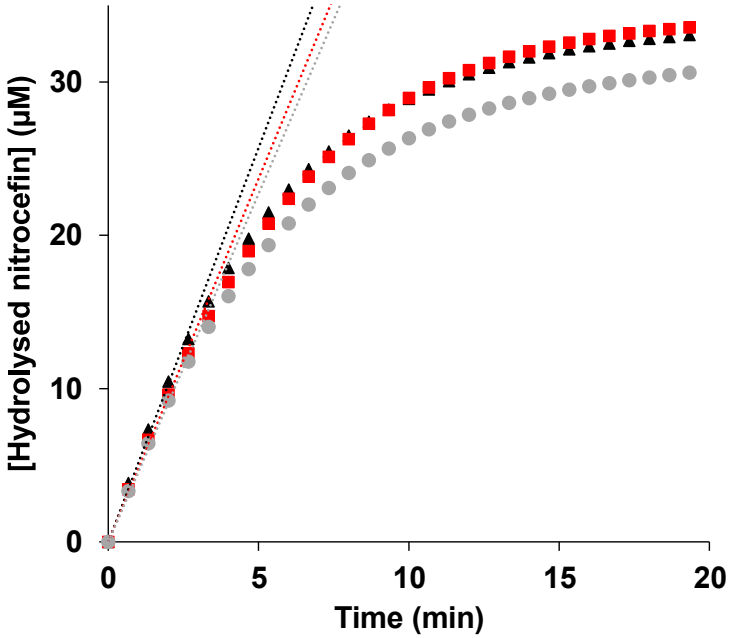


Figure 4.23 Nitrocefin hydrolysis by TEM variants. TEM^{WT} (black triangles, initial rate shown as dashed line), TEM^{87AzF} (red squares & dashed line) and TEM^{87AzF} after 30 minutes UV-B irradiation (grey circles & dashed line). All showed comparable enzyme function. See Table 4.3 below for initial rate values.

Table 4.3 Nitrocefin catalysis initial rates for TEM variants: TEM^{WT} and TEM^{87AzF} had comparable function, including after 30 minutes UV lysis of the AzF group of TEM^{87AzF}. Final concentration of nitrocefin was 40 μM.

	TEM^{WT}	**TEM^{87AzF}**	**TEM^{87AzF} (UV)**
Initial rate (mol/min/mmol enzyme)	5.14	4.74	4.54
Goodness of fit (R^2)	> 0.99	> 0.99	> 0.99

4.2.3.6 CNT sidewall functionalisation with TEM^{87AzF} via photoattachment

Functionalisation of S-SWCNTs with TEM^{87AzF} was carried out using the same methods as described in section 4.2.1.2, but CNTs were adsorbed on glass microscopy cover slips instead of silicon wafer. AFM data in this section were collected by Suzanne Thomas (Cardiff physics). Our lab has previously immobilised TEM via a genetically-incorporated AzF to graphene sheet[90], but S-SWCNTs offer a superior basis for protein integration into nanoscale electronic components due to their filamentous shape, high surface area and semiconducting properties.

To assess potential non-specific binding of TEM^{87AzF} to CNTs by other routes than photoactivated nitrene attachment, CNTs on cover slips were incubated with TEM^{87AzF} (100 nM, 15 μL) in dark conditions then washed thoroughly with water and dried with compressed nitrogen. AFM images of bare CNTs were compared to CNTs incubated with TEM^{87AzF} in the dark, neither CNT sample showed any sign of protein decoration (Figure 4.24a-b). This sample in Figure 4.24b contained mostly bundles of CNTs and the short CNTs shown were the best individual CNT images available. These results demonstrated that TEM^{87AzF} does not bind to CNTs non-specifically. AFM analysis of samples that were UV-irradiated during incubation showed CNTs decorated with protein-like globular objects, with 3D rendering clearly showing the objects higher than the CNT (Figure 4.24c). This was taken as evidence that the CNTs were functionalised with TEM^{87AzF} proteins, as observed previously with $BLIP\text{-}II^{41AzF}$ in section 4.2.1.2.

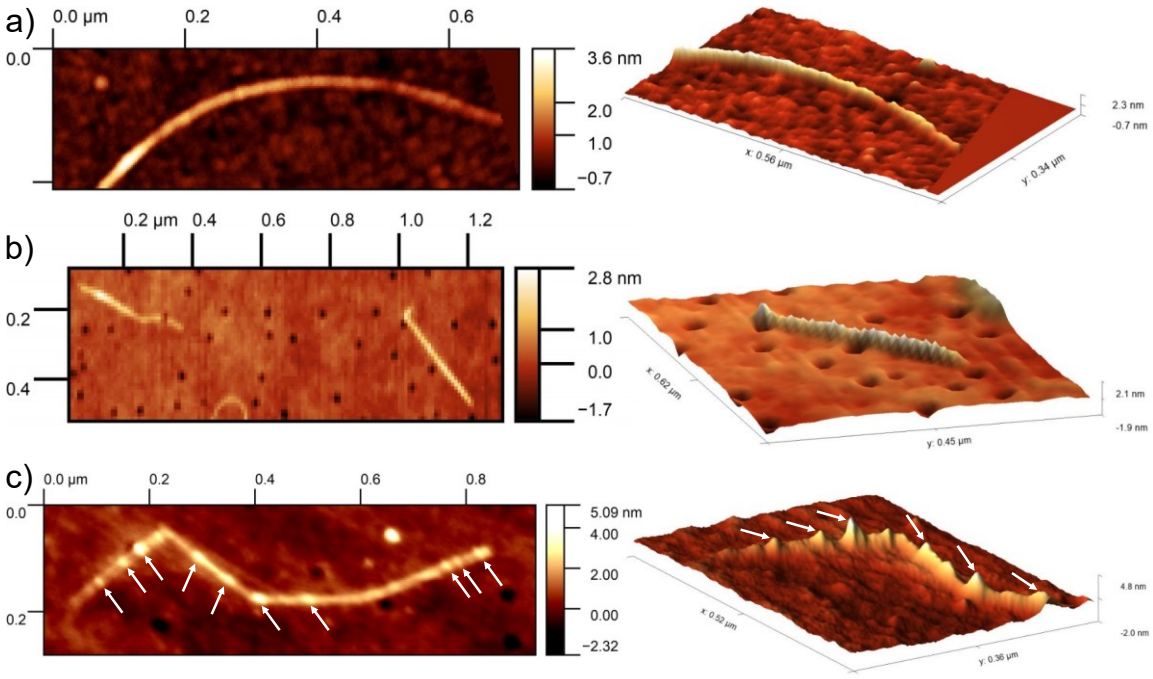


Figure 4.24 CNT functionalisation with TEM^{87AzF}. AFM height topography in 2D and 3D. a) Bare CNT, no protein – no decoration. b) CNTs incubated with TEM^{87AzF} without photoactivation – no decoration. This sample consisted mostly of bundled CNTs and these short CNTs were among the few individuals imaged. c) CNT incubated with TEM^{87AzF} with UV-B irradiation – shows decoration with globular objects (arrowed).

Using AFM data, eight transverse height profiles were taken across a CNT from the non-photoactivated sample and averaged to give the mean height of the bare CNT (1.97 nm, SD = 0.1 nm). Equivalent measurements were taken across the bare sections of a TEM^{87AzF} functionalised CNT (1.95 nm, SD = 0.19 nm), and across the immobilised TEM^{87AzF} proteins (3.09 nm SD = 0.5 nm) (Figure 4.25).

The CNTs of the non-photoactivated sample and the bare CNT sections of the photoactivated sample both had mean heights of ca. 2 nm. The height of TEM^{87AzF} proteins varied around an average of 3 nm. This ca. 1 nm height increase at the raised objects matches the increase observed when CNTs were functionalised with the similarly-sized protein BLIP-II^{41AzF} (ca. 1 nm) in section 4.2.1.2. These results give confidence that CNTs were being specifically functionalised with TEM^{87AzF} via nitrene photoactivation of the AzF residue and aziridine bond formation, in the same way as was observed for BLIP-II^{41AzF}. The variation in heights recorded for the proteins probably reflect a range of possible binding orientations to the CNT surface, with some examples shown in **Figure 4.26**. However, due to placement of the engineered AzF conjugation residue, most TEM^{87AzF} proteins should bind the CNT with their catalytic sites accessible to solution.

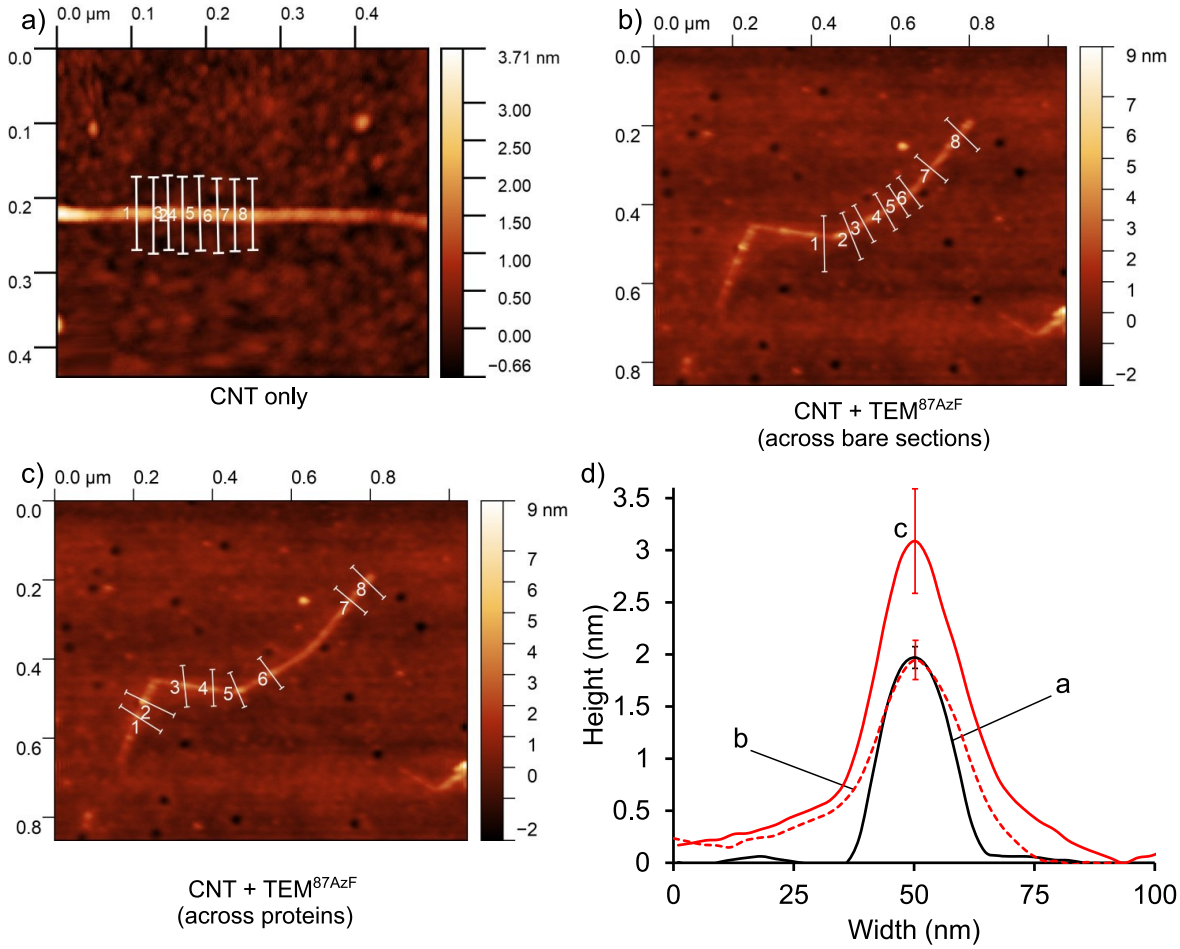


Figure 4.25 Transverse CNT AFM height profiles during functionalisation with TEM^{87AzF}. a) Bare CNT: height **1.97 nm** (SD = 0.1 nm). b) After incubation with TEM^{87AzF} with UV-B irradiation, CNTs became decorated with globular protein-like objects. CNT height measured between the objects were **1.95 nm** high (SD = 0.19 nm. c) Height measured across the objects was **3.09 nm** (SD = 0.5 nm). d) Mean height profiles across **a-c;** measurements n=8, SD bars shown for highest point, measurements shown on AFM images. The same individual CNT was analysed for **b)** and **c)**.

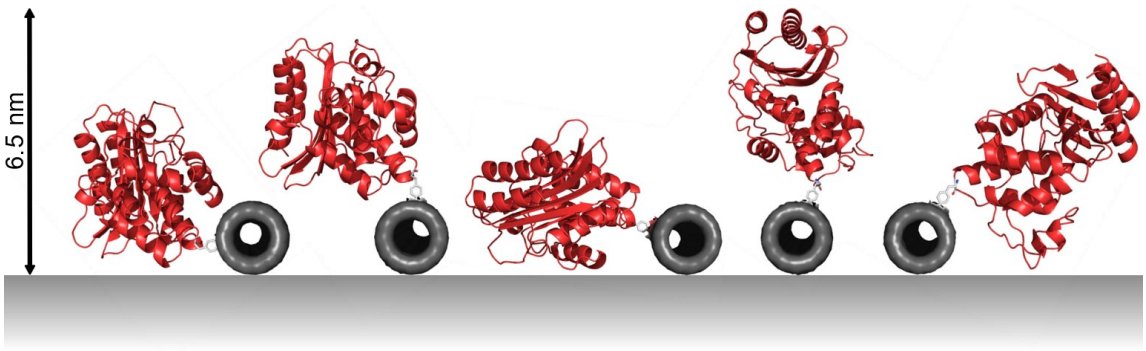


Figure 4.26 CNT functionalised with TEM^{87AzF} on flat surface, five possible conformations. The protein is likely to sample many conformations on the CNT surface, accounting for the variation in heights observed by AFM, up to a maximum height of ca. 6.5 nm (real). An average of ca. 3 nm was measured by AFM. PDB 1BTL.

4.2.4 Functionalisation of CNTs with cytochrome b_{562} variants via nitrene photoaddition and analysis using AFM

4.2.4.1 Design of cyt b_{562} phenyl azide/cysteine variants

Design and mutagenesis of the cyt b_{562}^{SA} (short axis) and cyt b_{562}^{LA} (long axis) variants used here were carried out previously by Samuel Reddington[253] (PhD thesis available at http://orca.cf.ac.uk/56226). The cyt b_{562}^{SA} and cyt b_{562}^{LA} variants were selected as they show high expression and have two different reactive side chains (AzF = phenyl azide and Cys = thiol) at opposite sides of the protein that can be used for labelling or to interface with CNTs (Figure 4.27). Since cyt b_{562}^{WT} contains no native cysteine residues, adding cysteines to the mutants enables targeted conjugation to the variant's single thiol group. Cyt b_{562}^{SA} and cyt b_{562}^{LA} are interesting subjects for functionalising CNTs as they have an approximately cylindrical shape (dimensions 5×2 nm) and are likely to conjugate to CNTs in different orientations. The variants should conjugate parallel (cyt b_{562}^{SA}) or at right angles (cyt b_{562}^{LA}) to CNTs, due to different positions of the AzF residues (see Figure 4.2). A height difference of ca. 2-3 nm between the species is potentially detectable using AFM. Cyt b_{562}^{SA} and cyt b_{562}^{LA} have previously been shown to spontaneously dimerise via their single cysteine residues in non-reducing buffer[253]. We aimed to distinguish a height difference between monomeric and dimeric species on CNT surfaces using AFM. Dimers are expected to take a range of conformations on the CNT surface, with cyt b_{562}^{SA} likely to be more conformationally restricted as its reactive handles are located within α-helices, with little peptide backbone movement possible. Cyt b_{562}^{LA} is likely to have more conformational freedom because its AzF and Cys reactive residues are on less structured end-loops. Possible CNT-dimer conformations were built by manual modelling of cyt b_{562} structure (PDB 256B) in PyMOL, with AzF and aziridines included as outlined in chapter 2.2.1.

Dimeric cyt b_{562} Cys/AzF variants may prove useful for a future aim of making self-assembling artificial electron transport systems, modelled after natural examples like the respiratory electron transport chain. Cyt b_{562} has previously been shown capable of electron transport over 2 nm by the Jones research group[169] and this distance could be extended further by interfacing dimeric cyt b_{562} on CNT surfaces. Electron transport could then be tuned by locating the inserted reactive handles at different residues relative to the haem group.

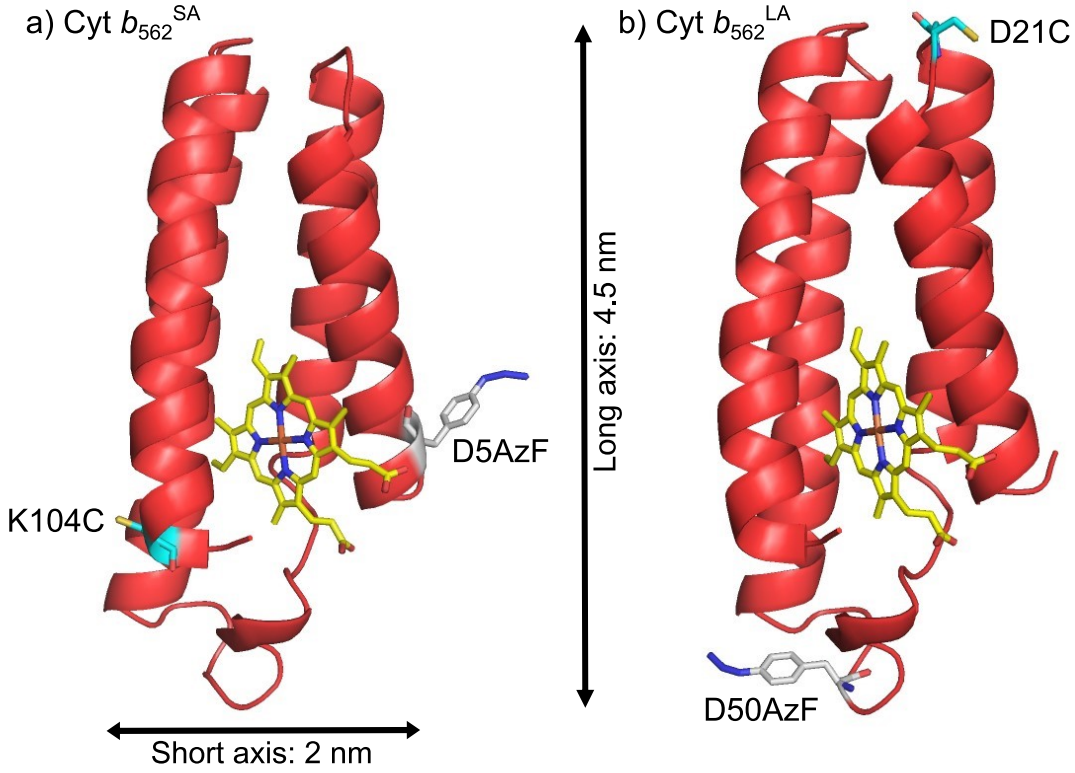


Figure 4.27 Locations of cyt b_{562} AzF and Cys mutations. a) cyt b_{562}^{SA}: the "short axis" variant, has two reactive handles on α-helices 1 and 4, on opposite sides of the protein (azide and thiol), either side of the haem cofactor. b) cyt b_{562}^{LA}: the "long axis" variant, has the same reactive handles on end-loops at opposite ends of the protein. Protein dimensions approx. cylindrical, 5×2 nm. PDB 256B[38].

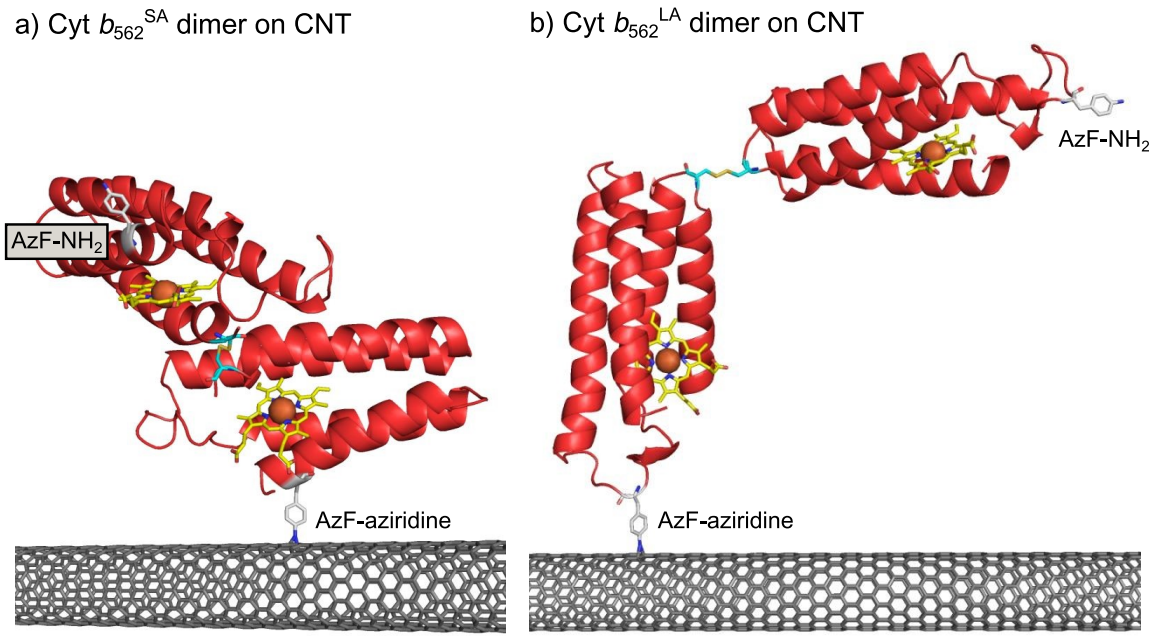


Figure 4.28 Possible conformations of CNTs functionalised with cyt b_{562}^{SA} and cyt b_{562}^{LA} dimers. Cyt b_{562}^{LA} is likely to have more conformational freedom than cyt b_{562}^{SA} as the engineered AzF and Cys residues are on less-structured end-loops. Phenylnitrenes that do not conjugate are expected to reduce to amines. PDB 256B[38].

4.2.4.2 Cytochrome b_{562}^{SA} and cyt b_{562}^{LA} expression

The *E. coli* DH5α cells containing the pBAD-cyt b_{562} plasmid variants were grown overnight then plasmid DNA extracted. Plasmid DNA for each mutant was sequenced and verified by alignment to the wild type gene. The pBAD-cyt b_{562} TAG mutant plasmids were co-transformed into electrocompetent *E. coli* TOP10 cells (NEB) along with the pDULE plasmid encoding the tRNA/aaRS pair required for AzF incorporation in response to a TAG codon. Single colonies of dual-transformed cells were used to inoculate 0.5 L batch cultures overnight using a modified Studier auto-induction protocol[177]. Media was supplemented with 50 µg/mL carbenicillin and 25 µg/mL tetracycline for dual plasmid transformant selection and 0.05% arabinose to induce cyt b_{562} expression. The nnAA AzF was added to growth media at 1 mM. Precautions were taken after adding AzF to keep the cultures in the dark, to prevent photolysis of the phenyl azide side chain. Expression of cyt b_{562} variants was evident from the strong red colour of haem-bound holo protein, visible in the cell pellets (Figure 4.29).

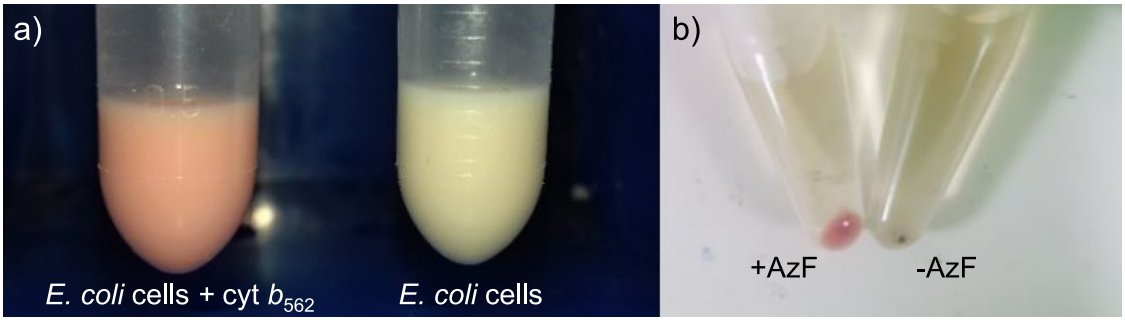


Figure 4.29 Expression of cyt b_{562} in *E. coli*. a) Cell pellets resuspended in buffer. Left: cells expressing haem-bound holo-cyt b_{562}, with an obvious red colour. Right: cells not expressing cyt b_{562}. b) *E. coli* cell pellet colours from expression of a representative variant, cyt b_{562}^{71AzF} with (+) and without (-) AzF in the growth media, showing that AzF is being incorporated into the protein. Photo credit (b): Tomas Evans.

4.2.4.3 Cytochrome b_{562}^{SA} and cyt b_{562}^{LA} purification

The cyt b_{562}^{SA} and cyt b_{562}^{LA} constructs used here do not have a polyhistidine affinity tag (His-tag) to simplify purification because His-tags have previously been found to interfere with haem binding through iron coordination[254], so these proteins must be purified by other means. From previous lab experience, cyt b_{562} variants were purified by ammonium sulphate precipitation and size exclusion chromatography (see methods chapter 2.2.4). The high solubility of cyt b_{562} in aqueous solution means that most contaminating proteins will precipitate at lower ammonium sulphate concentration than cyt b_{562}, and the small size of cyt b_{562} (12 kDa) aids purification by size exclusion chromatography. Care was taken to avoid exposing cyt $b_{562}^{SA/LA}$ variants to light as far as possible during purification to protect the photosensitive phenyl azide group. During size exclusion chromatography absorbance was monitored at the haem Soret maximum of 418 nm only, to avoid

photolysis of the AzF azide by UV-B exposure (λ < 350 nm). Cyt b_{562}^{SA} and cyt b_{562}^{LA} were found to elute at ca. 48 mL with two sharp A_{418} elution peaks of different intensities (Figure 4.30a). The double peak was attributed to formation of cyt b_{562} homodimers via their surface exposed cysteines, with dimers eluting in the latter peak, as has been previously documented[253]. Elution fractions were analysed by SDS-PAGE (Figure 4.25b). Fractions containing cyt b_{562}^{SA} and cyt b_{562}^{LA} were further purified using a Superdex 200 16/600 SEC column. No chromatograms are available as an ÄKTAprime system was used, with its 280 nm UV-B lamp disconnected to protect phenyl azide groups. Purification was tracked by observing the red protein in the SEC column. SDS-PAGE and ImageJ analysis was used to estimate protein purity at ≥ 98% (Figure 4.31).

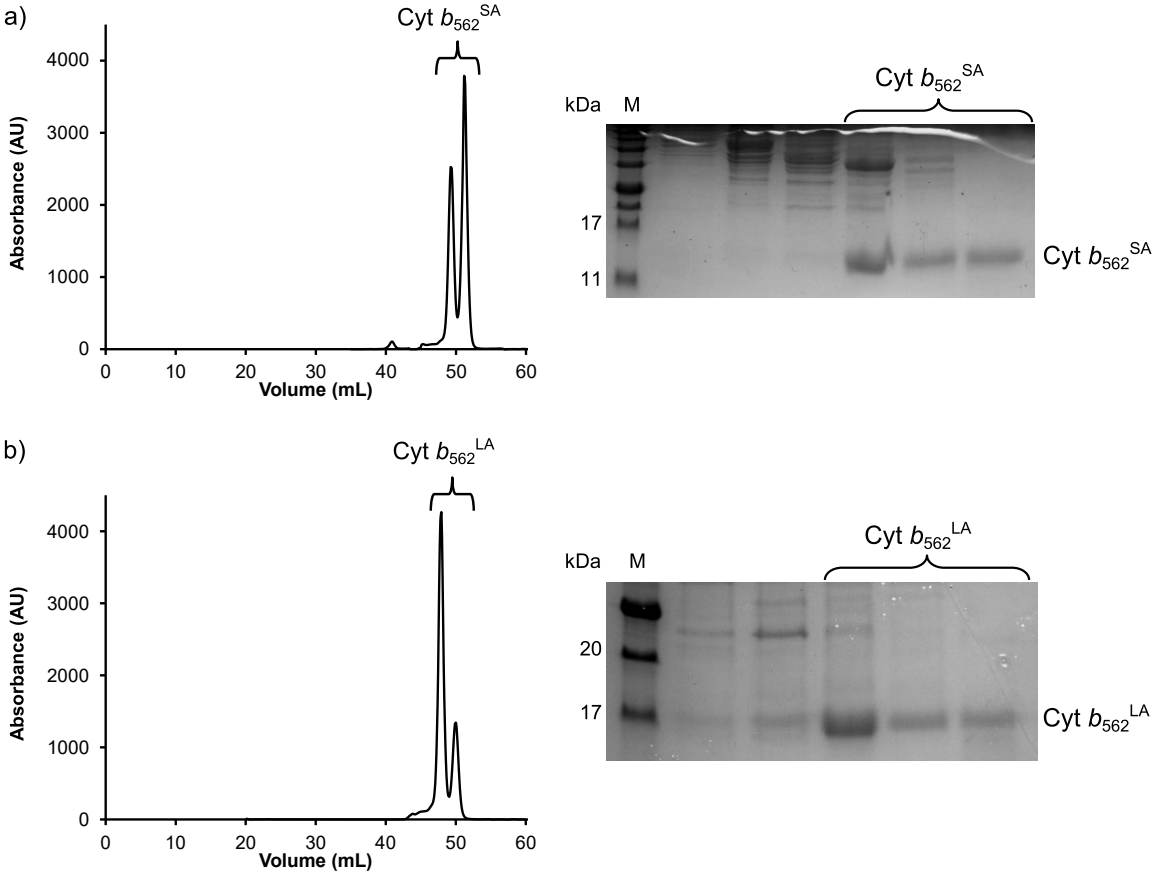


Figure 4.30 Purification of cyt b_{562}^{SA} and cyt b_{562}^{LA} by size exclusion chromatography (Superdex 200 16/600). Absorbance monitored at 418 nm. These cyt b_{562} variants are present in monomer and dimer form with separate elution peaks, which are pooled. a) Chromatogram and reducing SDS-PGE analysis for cyt b_{562}^{SA}. b) Chromatogram and reducing SDS-PGE analysis for cyt b_{562}^{LA}. Both variants elute at V_e ca. 48 mL volume after injection onto column. M: molecular weight marker ladder.

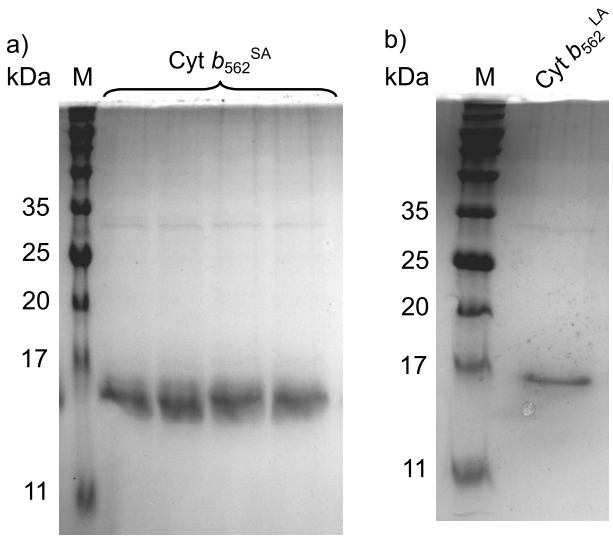


Figure 4.31 Second SEC purification of reduced, monomeric cyt b_{562}^{SA} and cyt b_{562}^{LA}. a) Cyt b_{562}^{SA}. b) Cyt b_{562}^{LA}. Both purified to ≥ 98% homogeneity as estimated using ImageJ.

4.2.4.4 Cytochrome b_{562}^{SA} and cyt b_{562}^{LA} dimerisation and phenyl azide reactivity

Elution chromatograms from purification of cyt b_{562}^{SA} and cyt b_{562}^{LA} showed evidence of two species that have previously been documented for these cyt b_{562} AzF/Cys variants: monomer and homodimer[253]. When concentrated in a non-reducing environment, cyt b_{562}^{SA} and cyt b_{562}^{LA} spontaneously form dimers via disulphide bonds between their cysteine residues, unlike wild type cyt b_{562} which does not dimerise by any means, even at millimolar concentrations[255]. Samples of purified cyt b_{562}^{SA} and cyt b_{562}^{LA} were concentrated to 100 µM and incubated at room temperature for three hours to encourage dimerisation. Analysis with SDS-PAGE and ImageJ showed approx. 25% of cyt b_{562}^{SA} and all cyt b_{562}^{LA} was dimerised. After allowing cyt b_{562}^{SA} to incubate at room temperature overnight, nearly all the protein had dimerised (Figure 4.32).

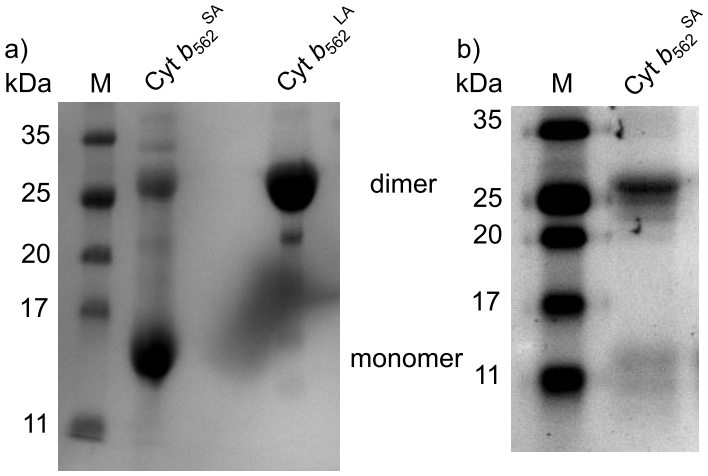


Figure 4.32 Dimerisation of cyt b_{562} phenyl azide/cysteine variants. Non-reducing SDS-PAGE. a) After concentration to 100 µM and three hours' incubation, cyt b_{562}^{SA} reached ca. 25% dimer and cyt b_{562}^{LA} was entirely dimerised. b) Cyt b_{562}^{SA} was incubated at room temperature overnight, becoming almost fully dimerised.

To quantitatively estimate SPAAC reactivity of their AzF phenyl azide groups, cyt b_{562}^{SA} and cyt b_{562}^{LA} were incubated with a threefold molar excess of the strained alkyne-linked fluorescent dye DBCO-TAMRA for one hour. The reducing agent dithiothreitol (DTT) was included to separate dimers. The proportion of protein labelled by SPAAC was analysed by SDS-PAGE, gel images were taken with UV illumination to visualise dye-labelled protein bands. A small amount of dimer remained in both samples, visible on the gel as two faint bands of higher molecular weight than the intense monomer band (**Figure 4.33**). It is thought that the two dimer bands represent dimers that are labelled with either one or two DBCO-TAMRA molecules. Gel images were quantified using ImageJ, total intensity of labelled proteins was compared to intensity of the unreacted dye at the bottom of the gels. Approximately 59% of cyt b_{562}^{SA} had been labelled with DBCO-TAMRA via SPAAC, while ca. 37% of cyt b_{562}^{LA} had been labelled. These efficiencies are comparable to those observed previously with BLIP-II AzF variants (chapter 3) and published data for SPAAC conjugation[228]. A sizeable proportion of each cyt b_{562} variant was reactive for downstream CNT functionalisation experiments.

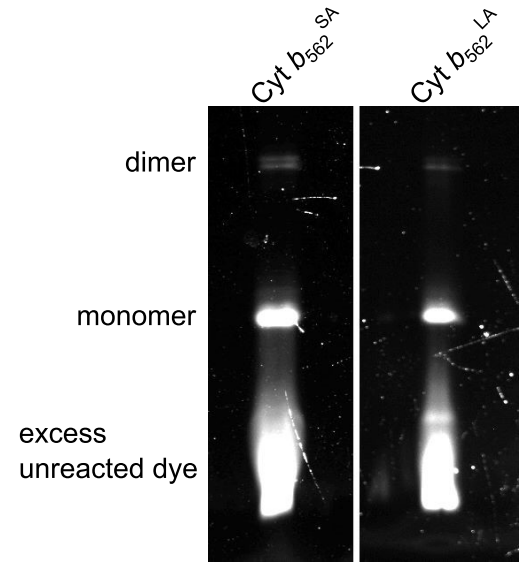


Figure 4.33 Estimation of cyt b_{562}^{SA} and cyt b_{562}^{LA} phenyl azide reactivity. Proteins were incubated with a threefold molar excess of DBCO-TAMRA dye to assess SPAAC reactivity of the proteins' AzF residue azide groups. DTT was included to separate dimers into monomers, but a small proportion of dimer remains as two small bands of higher MW.

4.2.4.5 Cytochrome b_{562} spectral characterisation

The spectral characteristics of cyt b_{562}^{SA} and cyt b_{562}^{LA} were analysed using absorbance spectroscopy. The intended CNT photoattachment experiments required the protein to withstand UV-B irradiation (λ_{max} ca. 305 nm, intensity 18 $W \cdot m^{-2}$) without significant radiation damage. UV-B exposure can cause various types of protein damage by excitation of aromatic amino acids and induction of reactive oxygen species (ROS) such as singlet oxygen[256]. These effects include aggregation, bityrosine crosslinking,

destruction of side chain groups and unfolding[257], all can cause functional inactivation of the protein. The structure of cyt b_{562} is key to its haem-binding function, and UV-induced structural and chemical changes are likely to cause dissociation of haem. This would be detectable as a change in the spectral features of cyt b_{562}, such as a reduction in intensity of the protein-bound haem Soret band at 418 nm (or 427 nm with the iron centre reduced to Fe^{2+}), likely coupled with an increase in absorbance at the free haem Soret wavelength (397 nm).

The effects of UV-B exposure on cyt b_{562}^{SA} and cyt b_{562}^{LA} dimers were examined by recording absorbance spectra for both protein variants before and after UV-B irradiation in both non-reducing and reducing conditions (0.5 mM DTT, 15 mins at room temperature). The DTT would cause three effects: 1) reduction of the haem iron centre from Fe^{3+} to Fe^{2+}; 2) reduction of disulphides to thiols, splitting cyt b_{562} dimers to monomers; 3) reduction of the cyt b_{562} AzF phenyl azide groups to phenyl amines, though this would be insignificant on these timescales[258], and UV-B irradiation will reduce the AzF anyway.

After 30 minutes UV-B irradiation, the cyt b_{562}^{SA} Soret maxima showed significantly lower intensity, down 32% in oxidised form and 20% in reduced form. Cyt b_{562}^{LA} Soret intensity fell by 14% (oxidised) and 6% (reduced) after irradiation (Figure 4.34). This result shows that some haem dissociation was caused by UV-B exposure, implying protein unfolding due to radiation damage. The effect was more pronounced for the cyt b_{562}^{SA} variant, suggesting that the D5AzF and/or K104C short-axis mutations may have some detrimental effect on protein stability or haem binding affinity, which are closely linked in cyt b_{562}. The axial ligand side chains that coordinate the haem iron centre are M7 and H102, neighbours of residues D5AzF and K104C. However, there is no obvious reason why mutation of D5 and K104 would affect protein stability or haem binding as their side chains face outward and do not interact with haem or the axial ligands.

For our purpose of CNT photoattachment, proteins require only 5 minutes of UV-B exposure to photoactivate the azide group of AzF to a nitrene radical for [2+1] nitrene cycloaddition, so any damage to cyt b_{562} during photoattachment should be minimal.

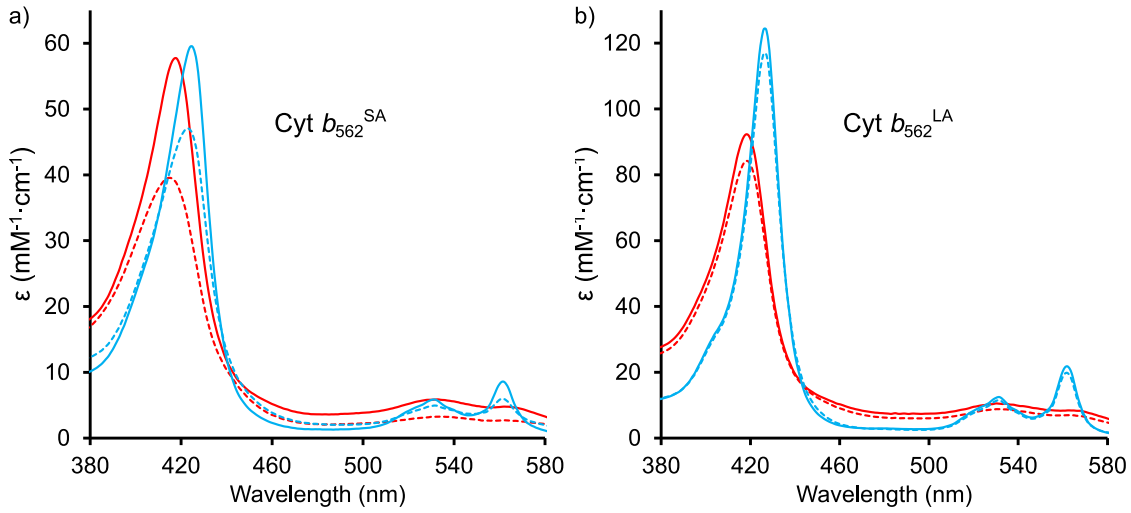


Figure 4.34 Absorbance spectra for cyt b_{562} variants in oxidised (red) and reduced (blue) states before (solid lines) and after 30 min UV-B irradiation (dashed lines). a) Cyt b_{562}^{SA} spectra showing 32% and 20% reductions in oxidised and reduced Soret maxima. b) Cyt b_{562}^{LA} spectra showing 14% and 6% reductions in oxidised and reduced Soret maxima. These results suggest protein degradation from UV-B irradiation.

4.2.4.6 CNT sidewall functionalisation with cytochrome b_{562} variants via photoattachment

Experiments were performed to investigate the physical dimensions of engineered cyt b_{562} dimers and monomers when conjugated to CNT surfaces, using AFM. It was expected that the dimeric forms of cyt b_{562}^{SA} or cyt b_{562}^{LA} immobilised on the CNT surface could be differentiated from monomeric cyt b_{562}^{SA} or cyt b_{562}^{LA} by apparent height. AFM data were gathered, and protein height histograms plotted by Suzanne Thomas (Cardiff School of Physics and Astronomy).

Samples of homodimeric cyt b_{562}^{SA} or cyt b_{562}^{LA} at 100 nM were used to functionalise CNTs by UV-nitrene photoaddition, as described previously in section 4.2.1.2. As was previously observed for the other AzF-containing proteins, AFM data for individual and small bundles of CNTs incubated with cyt b_{562}^{SA} or cyt b_{562}^{LA} in dark conditions showed no decoration with proteins (Figure 4.35a and Figure 4.36a), indicating that the proteins did not bind non-specifically to CNTs. CNT samples incubated with cyt b_{562}^{SA} or cyt b_{562}^{LA} and UV-B irradiation were found to be decorated with proteins at a similar dispersion as observed during BLIP-II functionalisation previously, ca. 20-30 proteins per µm of CNT (Figure 4.35b and Figure 4.36b).

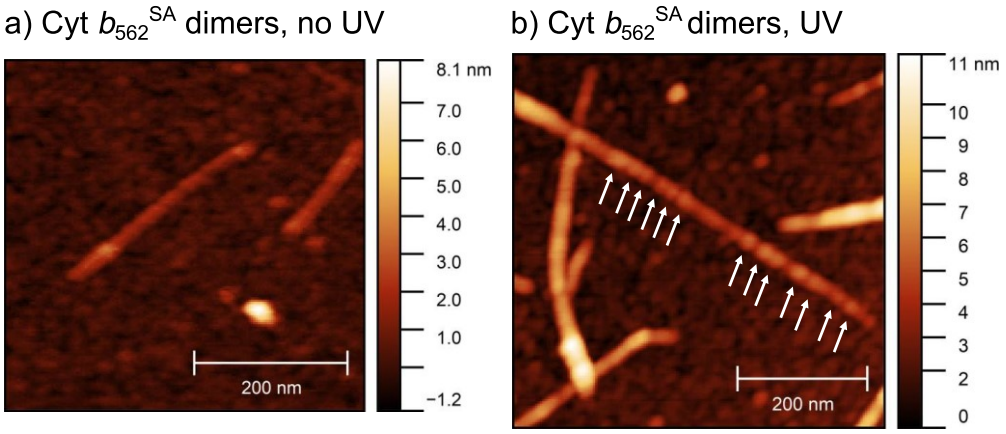


Figure 4.35 AFM images from functionalisation of CNTs with cyt b_{562}^{SA} dimers. a) Cyt b_{562}^{SA} dimers incubated with CNTs without UV-B exposure showed no decoration with proteins. b) Cyt b_{562}^{SA} dimers incubated with CNTs with UV-B exposure were decorated with proteins (arrowed).

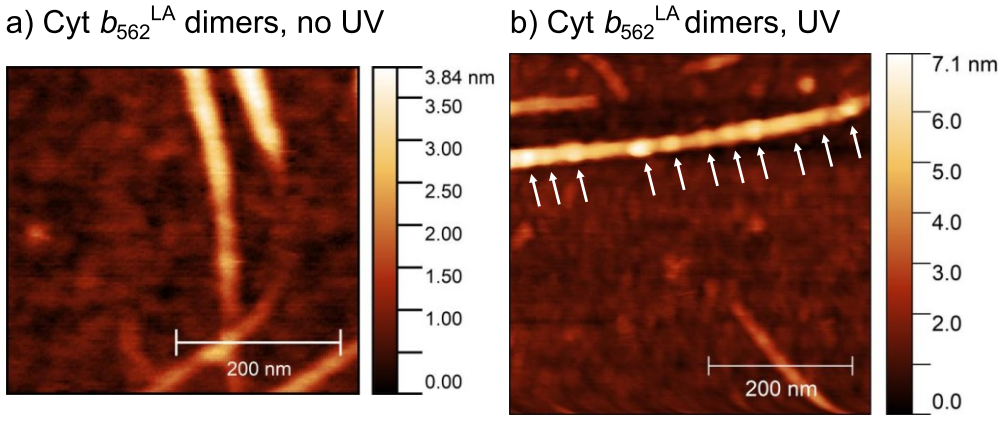


Figure 4.36 AFM images from functionalisation of CNTs with cyt b_{562}^{LA} dimers. a) Cyt b_{562}^{LA} dimers incubated with CNTs without UV-B exposure showed no decoration with proteins. b) Cyt b_{562}^{LA} dimers incubated with CNTs with UV-B exposure were decorated with proteins (arrowed).

CNTs functionalised with populations of dimeric cyt b_{562}^{SA} and cyt b_{562}^{LA} were further incubated in a 0.5 mM solution of the reducing agent DTT for ten minutes, then washed and dried. This was intended to reduce the disulphide bonds joining cyt b_{562} homodimers, splitting them back to monomers. To measure any change in apparent protein height between dimer and monomer forms of cyt b_{562}^{SA} or cyt b_{562}^{LA} on the CNT surface, the samples were again analysed by AFM. Using the AFM data, height profiles were measured for 50 individual cyt b_{562}^{SA} or cyt b_{562}^{LA} protein features on the CNT surface, before and after incubation with DTT, and CNT height was subtracted. These protein heights were plotted against their frequency of observation as histograms, with recorded heights split into bins.

The distribution of heights of CNT-immobilised cyt b_{562}^{SA} proteins shifted following reduction of inter-protein disulphide bonds with DTT. This implies that reduction of the disulphide changed the proteins functionalising the CNTs. The frequency of cyt b_{562}^{SA} topographic features in the > 2 nm height range decreased after treatment with DTT,

while frequencies in the < 2 nm range increased (Figure 4.37a). The mean feature height reduced by 37% from 2.54 nm (SD = 0.88 nm) to 1.59 nm (SD = 0.65 nm). The frequency of cyt b_{562}^{LA} features showed little change after reduction of the inter-protein disulphide bond with DTT (Figure 4.37b). The mean feature height increased by 12% from 1.72 nm (SD = 0.53 nm) to 1.96 nm (SD = 0.54 nm).

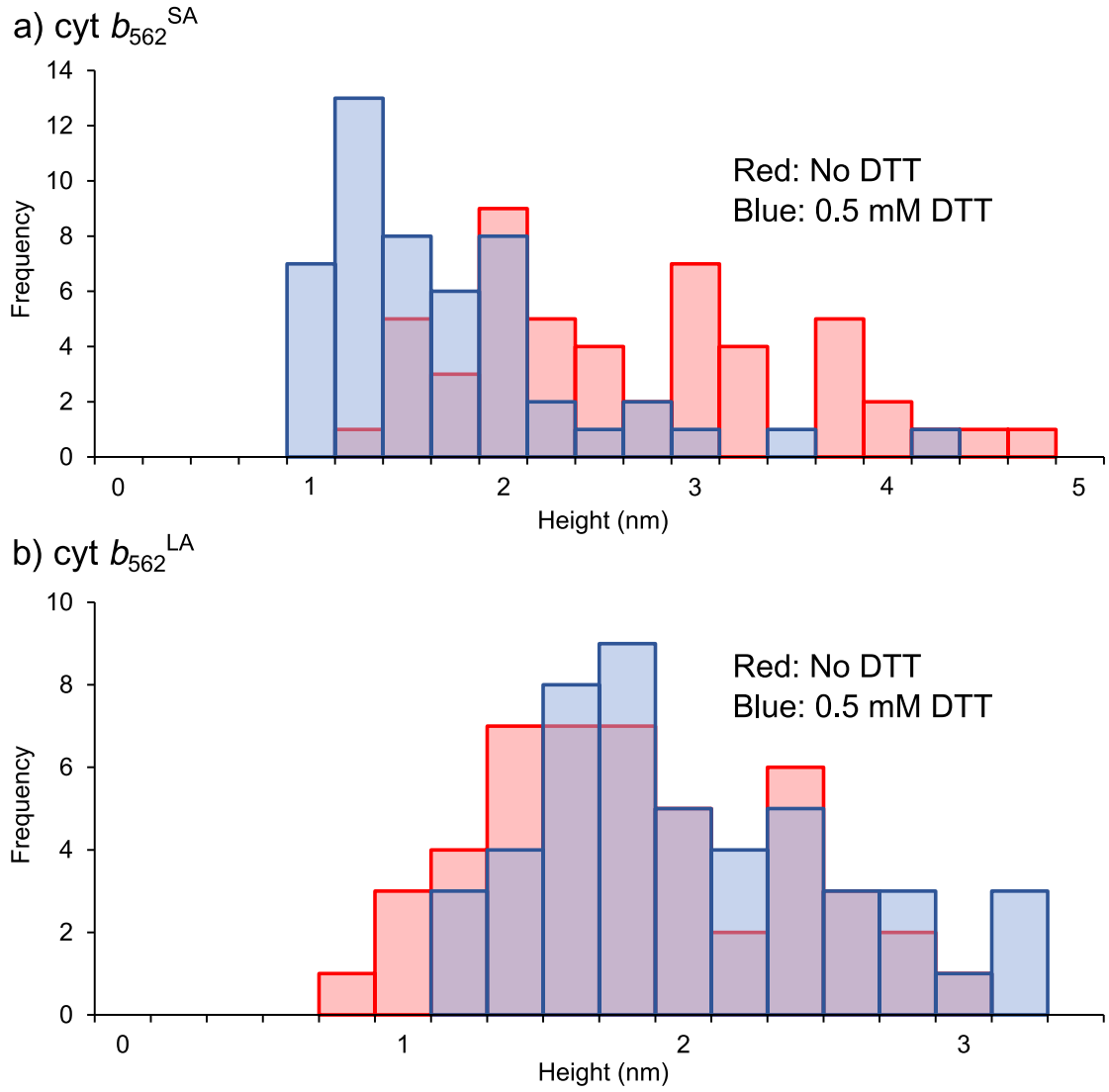


Figure 4.37 Cyt b_{562} height frequency on CNTs/CNT bundles. The height of 50 individual protein features were measured and the CNT or CNT bundle height subtracted to estimate protein height in nm. Observed heights were grouped into bins and plotted using Excel. a) Overlaid height distributions of 50 cyt b_{562}^{SA} proteins on CNTs without DTT (red) and with 0.5 mM DTT (blue). b) Overlaid height distributions of 50 cyt b_{562}^{LA} proteins on CNTs without DTT (red) and with 0.5 mM DTT (blue). Raw data: Suzanne Thomas.

The height distributions for cyt b_{562}^{SA} are consistent with a population of cyt b_{562}^{SA} dimers conjugated in a range of orientations up to a maximum height of ca. 5 nm. After reduction with DTT, it is likely that most dimers would be split, leaving CNTs decorated with mostly monomeric cyt b_{562}^{SA} species up to a maximum height of ca. 3 nm, as modelled in Figure 4.38. This model was made by docking two cyt b_{562} monomer structures at residue 104

using ClusPro[14] (https://cluspro.bu.edu/). The most probable conformation structure output by ClusPro was edited with PyMOL, to join the cysteine thiols into a disulphide bridge and conjugate the AzF to a CNT model via an aziridine bond. The apparent heights measured before and after reduction match the modelled dimensions of the dimer and monomer.

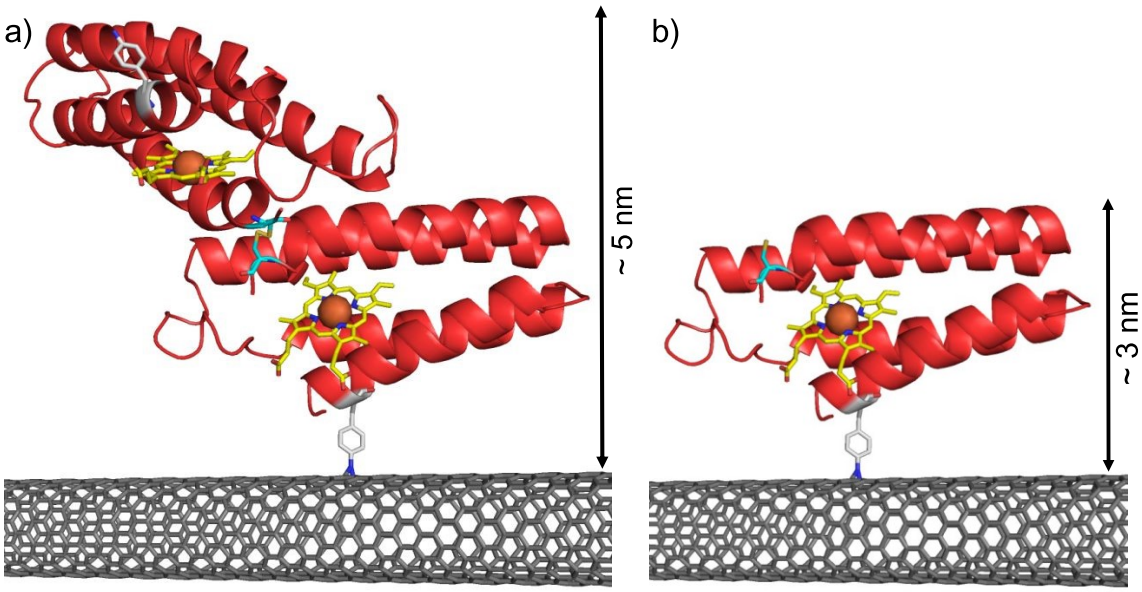


Figure 4.38 Approximate heights of cyt b_{562}^{SA} species on CNT. a) Cyt b_{562}^{SA} disulphide dimer conjugated to CNT surface via AzF-aziridine. Height ca. 5 nm. b) Cyt b_{562}^{SA} monomer conjugated to CNT surface via AzF-aziridine after reduction and separation by DTT. Height ca. 3 nm. Light grey: AzF5. Cyan: Cys104. Yellow: haem. PDB 256B.

In silico modelling of cyt b_{562}^{LA} dimensions suggests the slightly increased frequency of taller objects in the cyt b_{562}^{LA} distribution after reduction with DTT may be due to protein dimers conjugating to the CNT at both AzF reactive groups, then being split into monomers when reduced (modelled in Figure 4.39). The separation and inherent flexibility of the AzF and cysteine reactive handles in cyt b_{562}^{LA} appear to make this conformation feasible. They are located outside major secondary structures on end-loops, and are disordered in comparison to the reactive handles of cyt b_{562}^{SA} (Figure 4.40). A theoretical cyt b_{562}^{LA} disulphide dimer was created using ClusPro to dock two cyt b_{562} monomer structures at residue 21. This dimer model was manually edited using PyMOL, to join the cysteine thiols into a disulphide and conjugate both AzF reactive groups to a CNT via aziridine bonds, allowing estimation of the height of the proteins in dimer (< 4 nm) and monomer (> 5 nm) form. This arrangement may account for the increase in observed frequency of taller proteins above 1.5 nm following reduction with DTT, considering the typical underestimation of protein height using AFM. These findings demonstrate the bottom-up assembly and chemical modification of some interesting CNT-protein-protein complexes. We also show that designed placement of the AzF and

cysteine reactive groups can be used to construct different nanoscale structures with detectable structural differences.

Designed self-assembly of metalloprotein complexes could be used to construct artificial protein-based electron transport systems over relatively long distances, mimicking the electron transport chain in respiration. Della Pia *et al.*[169] previously demonstrated long-range electron transport over 2 nm from the haem iron centre of an engineered cyt b_{562} to a gold surface. The distance between haem cofactors could be varied by design, by linking proteins together and to CNTs via reactive handles placed at specified residues.

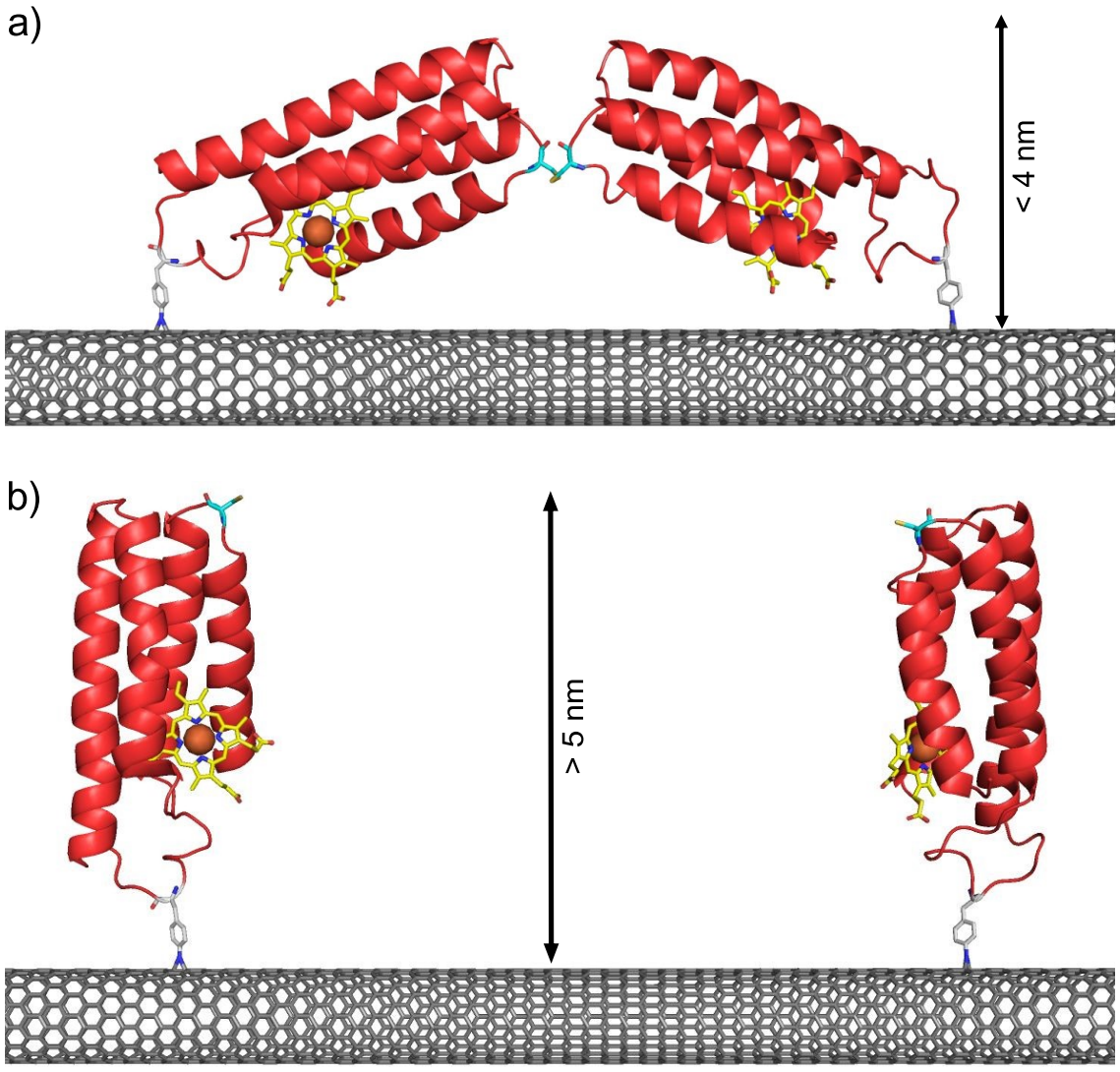


Figure 4.39 Approximate heights of cyt b_{562}^{LA} species on CNT. Possible conformations. a) Cyt b_{562}^{LA} disulphide dimer conjugated to CNT surface via two AzF-aziridines. Height > 5 nm. b) Cyt b_{562}^{LA} monomers conjugated to CNT surface via AzF-aziridine after reduction of the inter-protein disulphide by DTT. Height < 4 nm. Light grey: AzF50. Cyan: Cys21. Yellow: haem. PDB 256B.

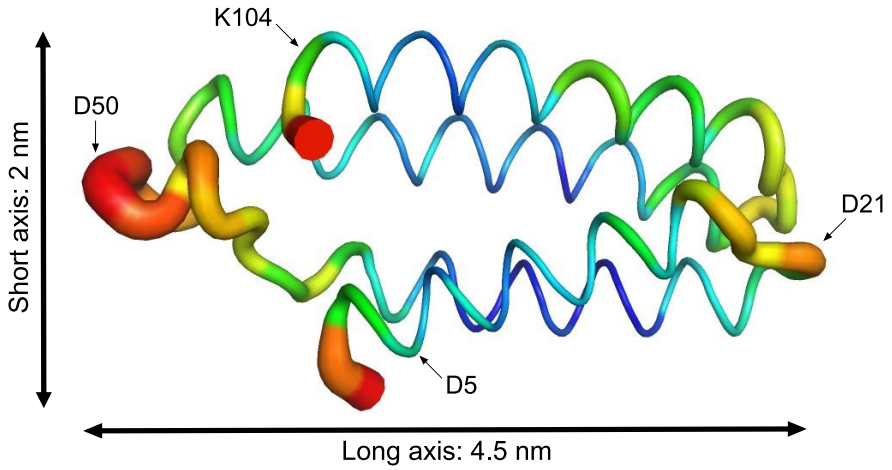


Figure 4.40 B-factor "sausage" or "putty" structure of cyt b_{562}. Greater backbone disorder is represented by a thicker line (orange and red). The cyt b_{562}^{LA} mutations (D21, D50) are in much more disordered parts of the protein than the cyt b_{562}^{SA} mutations (D5, K104).

4.3 Conclusions

This chapter sought to examine the general applicability of the genetically-encodable non-natural amino acid *p*-azido-L-phenylalanine to covalently functionalise S-SWCNTs with engineered proteins via a specifically-located UV-nitrene photoattachment. This approach can intimately conjugate populations of proteins to CNTs with defined orientation using just a short burst of UV-B irradiation, while avoiding the need to covalently modify the protein with a bulky linker molecule or functionalise CNTs in advance. Direct covalent photoattachment to CNT sidewalls via an AzF-aziridine bond was successfully applied to three unrelated proteins. This resulted in a similar level of CNT coverage (dispersion ca. 20-30 proteins per µm CNT) as was observed previously, when CNTs were functionalised with BLIP-II by adsorption via covalently-attached pyrene. Direct attachment to CNTs via aziridine may offer a more intimate form of protein conjugation than was achieved using the DBCO-pyrene linker (chapter 3), due to the small size of the AzF-aziridine bond (ca. 8 nm). The approach appears applicable to any recombinant protein of interest, for covalent integration with graphitic carbon surfaces in a user-defined orientation. In the case of proteins that are unsuitable for recombinant expression in bacteria, protocols now exist to genetically incorporate AzF in proteins for expression in a range of yeast[259] and mammalian[260] cell expression systems.

This work demonstrates that rational protein engineering and the new bioorthogonal chemistries offered by genetically incorporated non-natural amino acids can be used to design novel self-assembling complexes at the nanoscale. These systems can be designed *in silico* to have relatively predictable target binding behaviour, and in the case of the phenyl azide reactive group of AzF, can be activated on demand by UV-B light. AzF is an especially versatile tool for protein engineering, with reactivity that can be

easily directed in space using bioorthogonal SPAAC “click” chemistry to define a highly specific reaction target (as in chapter 3) or in time, using UV-B light to activate AzF photochemical crosslinking reactions when required. Importantly, these conjugation methods do not require harsh chemical conditions, high temperatures or nonpolar solvents to function, all conditions commonly used in CNT functionalisation but not tolerable for proteins. Engineered proteins containing AzF in this work retained their native fold and function throughout production and presumably also on the CNT surface, as little non-specific binding or surface contamination from denatured protein was evident from the AFM data. Apparent protein-protein interactions occurred at the CNT surface, with TEM^{WT} binding to CNT-immobilised $BLIP\text{-}II^{41AzF}$ detected as an increase in protein height, further indicating that CNT-immobilised proteins remain functional.

In future, it would be worthwhile to compare this covalent method of functionalising CNTs with the non-covalent adsorption method used in chapter 3. While both methods have been shown to reliably conjugate proteins to CNT surfaces, the photochemical method was not used to functionalise NTFET devices, so electrical effects of aziridine bonds at the CNT surface have not yet been tested. Carboxyl defects introduced by acid oxidation to provide CNT surface reactive groups are known to create defects and vacancies randomly across the graphitic network that can have detrimental effects on CNT conductivity[233]. Nitrene cycloaddition does not create vacancies, and only modifies the CNT graphitic network at the sites of protein functionalisation, rather than randomly across the surface, without breaking the σ-bonds between carbons. This represents an improvement in specificity of CNT modification over acid oxidation. The effects of the aziridine modification on analyte sensitivity in our BLIP-II-functionalised NTFET β-lactamase sensor is as yet unknown but the extent of modification can be controlled by using different concentrations of protein. Functionalisation of CNTs with pyrene has been previously shown to have little or no effect on CNT conductivity[209]. Further electrical analyses comparing NTFETs functionalised with BLIP-II by both approaches could show if either is preferable for this purpose.

The successful integration of TEM^{87AzF} with CNT sidewalls is a first step towards construction of an enzyme-functionalised NTFET sensor for detecting the catalysis of small molecule substrates. Sensing devices based on enzyme- or receptor-coated S-SWCNTs have been created several times since 2003[119,209,261]. However, these devices were all functionalised via intrinsic protein chemistries, most commonly by pre-coating CNTs with pyrene-NHS-ester to react with any random surface amine of the protein forming a covalent amide link. This results in CNTs functionalised with a population of proteins immobilised in a range of random orientations. Our genetically-incorporated AzF photoattachment system functionalises CNTs with engineered proteins, all via a selected

residue, allowing control over the spatial orientation of the whole final population. This specificity of conjugation enables all proteins interfaced on CNT surface to present a consistent electrostatic surface to the CNT and prevents binding/catalytic site obstruction. If individual proteins all influence the electrical properties of the CNT in approximately the same way, this should offer a superior signal-to-noise ratio as each protein reinforces a signal response, rather than averaging or cancelling out.

The AFM height findings from conjugation of cyt b_{562} dimers to CNTs suggested that post-assembly structural changes were made to the CNT-protein-protein complexes by chemical modification using DTT. This is interesting as it demonstrates the ability to selectively modify part of a covalently self-assembled biomolecular nano-complex; these could be engineered to self-assemble using several discrete chemistries, allowing later targeted modifications to be made. This kind of targeted reactivity could prove useful in the construction of protein-based nanodevices.

5 Switching protein metalloporphyrin binding specificity by design from iron to fluorogenic zinc

Work in this chapter contributed to a published paper where I am 1st author:

Bowen, B. J., McGarrity, A. R. *et al*. Switching protein metalloporphyrin binding specificity by design from iron to fluorogenic zinc. *Chemical Communications* **56**, 4308–4311 (2020).

5.1 Introduction

5.1.1 Porphyrin cofactors

Many natural proteins require cofactors to assist their function – these are non-protein small molecules or ions that bind proteins to form a functional holo-protein complex. An important class of cofactors is organometallic tetrapyrroles such as the metalloporphyrins that exist in all domains of life[262]. Metalloporphyrin cofactors can add new redox properties, enabling enzyme metal centres to switch between oxidation states to act as electron donors, acceptors and transport systems. These metalloporphyrin redox reactions form the basis of many biochemical systems, such as the reversible switching of iron protoporphyrin IX (haem) iron between Fe^{3+} and Fe^{4+} during peroxide catabolism (peroxidases)[263] or between Fe^{2+} and Fe^{3+} during oxidative phosphorylation (cytochrome *c*). Metalloporphyrin cofactors enable the free energy change associated with electron transport between metal centres of different redox potentials to be coupled to covalent bond formation, as during ATP synthesis in oxidative phosphorylation, making these some of the most important molecules in complex life[1]. Natural tetrapyrroles, including porphyrins, exist in complex with many different transition metals, but only four metals are known as biologically-active tetrapyrrole centres: Mg^{2+} (chlorophylls), Fe^{2+} (globins and cytochromes), Co^{2+} (vitamin B_{12}) and Ni^{2+} (coenzyme F_{430})[264].

Metalloporphyrins are of technological and biomedical interest due to their unusual photophysical and photochemical properties. These include light-induced electron transfer, whereby photoexcitation of a porphyrin can induce a singlet-excited electron to be donated to reduce a nearby molecule, instead of its relaxing to ground state with fluorescence emission. Photoinduced electron transfer has been observed from zinc porphyrins to semiconducting carbon nanotubes[265] and sodium titanate nanosheets[266], and zinc porphyrins have already been used as the basis for photovoltaic devices, converting solar energy into electrical energy with efficiencies up to 12%[174].

Some porphyrins and similar tetrapyrroles have been approved as photosensitisers for photodynamic therapy (PDT) for some cancers, such as porfimer sodium and protoporphyrin IX, while others containing metal centres are undergoing clinical trials, including tin ethyl etiopurpurin and motexafin lutetium. Photosensitisers can induce localised cell death with targeted excitation at specific wavelengths, by electron donation to target molecules, or the production of singlet oxygen or other reactive oxygen species[267,268]. Zinc protoporphyrin IX (ZnPP) has been successful as a light induced generator of singlet oxygen capable of killing T-cell leukaemia cells *in vitro*[269]. Metalloporphyrins containing different metal centres have also been used as reduced-toxicity contrast agents for several imaging techniques, including magnetic resonance imaging (MRI), using a paramagnetic manganese porphyrin[270], and positron emission/X-ray computed tomography (PET-CT), using radioactive copper-64 porphysome nanoparticles[271].

The construction of non-natural protein-metalloporphyrin complexes is the subject of research interest, with some such enzyme constructs showing catalytic properties not found in naturally-occurring biosystems. One example of this was a complex of a monoclonal antibody with a manganese TCP-porphine that was constructed and found to catalyse the peroxidation of pyrogallol[272]. A more recent paper reported the construction of an artificially evolved thermophile cytochrome P450 enzyme reconstituted with an abiological iridium methyl protoporphyrin IX (Ir(Me)-PIX), which was found to catalyse intramolecular carbene insertions not catalysed by natural enzymes, at a rate comparable to natural enzymatic functions[273,274].

5.1.2 ZnPP fluorescence

Fluorescence is a phenomenon caused by the absorption of energy from electromagnetic radiation (photon) by an electron in a fluorophore, e.g. from exposure to light. Fluorophores are typically organic molecules with conjugated π-bonds causing delocalisation across multiple p-orbitals. The electron is excited from ground state (S_0) to a higher energy singlet excited state (S_1), it can relax to the ground state via vibrational relaxation to a lower S_1 state followed by fluorescent emission of a lower-energy photon. Another route to relaxation of the electron to S_0 is phosphorescence, where the electron undergoes intersystem crossing to a triplet excited state (T_1) then relaxation to S_0 over a much longer timescale then fluorescence, with emission of a still lower energy photon. Emitted photons have a longer wavelength (redshift) due to their lower energy and this difference between the excitation and emission wavelengths is known as the Stokes shift. A wide Stokes shift is desirable in fluorescence spectroscopy and microscopy fluorophores as it minimises spectral overlap between the excitation and emission

wavelengths. The efficiency of fluorescence is measured by the ratio of photons emitted to photons absorbed: the fluorescence quantum yield (Φ_F). The fluorescence process is illustrated with a simplified Jablonski diagram in Figure 5.1[275].

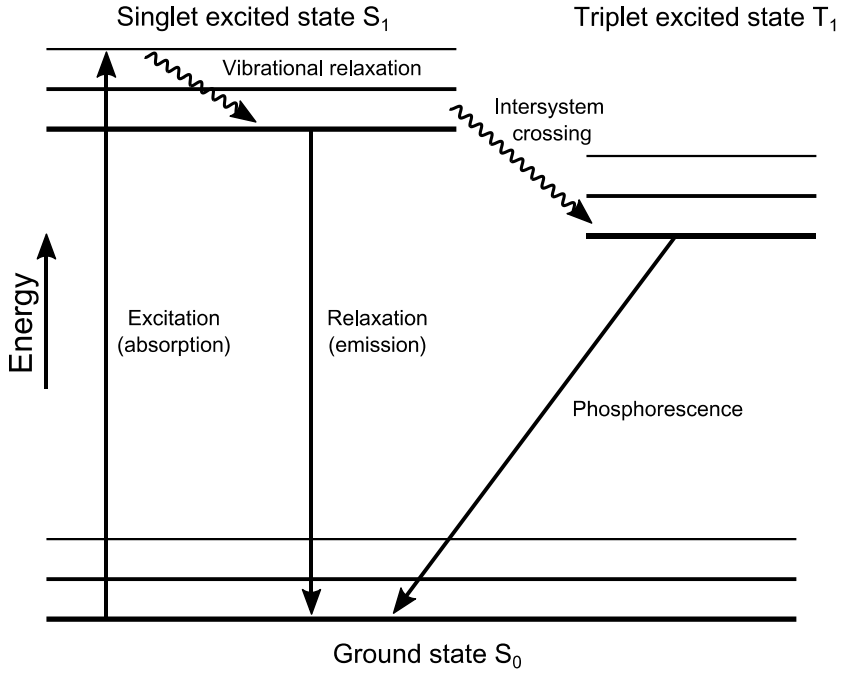


Figure 5.1 Simplified Jablonski diagram of electron excitation and relaxation. Absorption of a photon can excite an electron in a fluorophore from the ground state (S_0) to a high-energy excited singlet state (S_1). The electron can relax to S_0 via vibrational relaxation and subsequent fluorescent emission of a lower-energy photon, or via intersystem crossing to a triplet excited state (T_1) and subsequent phosphorescent emission of a photon. Wavy lines indicate non-radiative transitions.

ZnPP is a naturally-occurring metalloporphyrin that exists in red blood cells in trace amounts as a byproduct of defective haem biosynthesis due to low iron availability. Abnormally high levels of erythrocyte ZnPP are a biomarker of iron deficiency anaemia, especially within bone marrow, and is associated with chronic lead poisoning[276]. ZnPP shows several interesting properties including photo-induced electron transfer, fluorescence and as an analogue of haem, the ability to replace haem in many proteins. ZnPP is inherently fluorescent in organic solvents such as dimethyl sulfoxide (DMSO) and pyridine, with a quantum yield of 0.05 at the Q band maximum (λ_{ex} = 545 nm)[277] and can also be excited at the Soret band (max λ_{ex} = 410 nm). When excited at 410 nm ZnPP shows unusually large Stokes shifts of ca. 170 and 230 nm to its two major emission peaks (λ_{em} ca. 590 and 640 nm) (Figure 5.2).

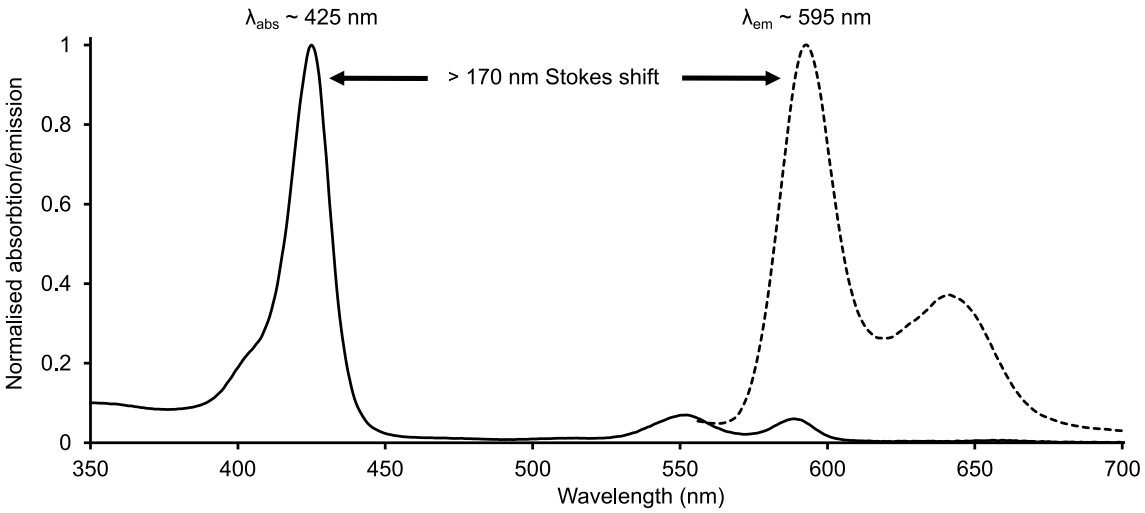


Figure 5.2 ZnPP absorbance and emission spectra in organic solvent. The overlap between emission (dashed line) and absorbance (solid line) bands in the 550-600 nm Q region causes re-absorption and the characteristic double emission peak of ZnPP.

ZnPP fluorescence is largely quenched in aqueous solution[277] making it of little use as an imaging molecule in biological systems, however it has been found that binding to a haemprotein (by replacing the haem and/or iron metal centre) can shield the porphyrin from the water environment, enhancing its emission at least tenfold[278]. These properties show cyt b_{562} has promise as a potential basis for genetically-encodable, chemically switchable fluorogenic protein labels for cell imaging. In this work, we sought to switch the binding affinity of a haem-binding protein to preferentially bind zinc protoporphyrin IX (ZnPP) as a cofactor.

The protein chosen for this work, cytochrome $b_{562,}$ is a small bacterial electron transport protein that binds its haem cofactor with high affinity, with a K_D around 9 nM[279]. Cyt b_{562} is a suitable model system for protein engineering due to its temperature and solvent stability, small size and high expression. Of special interest is the ability of cyt b_{562} to bind a range of porphyrins with reasonable affinity, including ZnPP (ca. 500 nM)[171,280]. For these reasons, this laboratory has extensive previous experience with the protein. Previous work has involved reconstitution of cyt b_{562} with various exogenous protoporphyrins containing other metal centres, including ZnPP and copper protoporphyrin IX, and characterisation of the altered redox and spectral properties that resulted[171]. Other publications saw cyt b_{562} conjugated intact to a gold (III) surface via engineered cysteine residues, confirmed by atomic force microscopy (AFM). This enabled the study and control of haem redox states and electron transfer to the Au (III) surface for individual cyt b_{562} molecules using scanning tunnelling microscopy (STM) techniques[170,281]. Cyt b_{562} was also incorporated into an engineered fusion construct with the unrelated green fluorescent protein, superfolder GFP (sfGFP). This construct was

found to exhibit close to 100% resonance energy transfer between the sfGFP chromophore and the haem cofactor of cyt b_{562}. When excited at the sfGFP excitation maximum of 488 nm, protein emission was fully quenched on binding to haem, indicating complete resonance energy transfer between the sfGFP and cyt b_{562} domains[168]. To combine the established photoinduced electron transport properties of ZnPP with the ZnPP-binding and electron transport ability of cyt b_{562} would be of interest for construction of nanoscale photoelectronic devices. This aim ties in with the other chapters of this thesis, the construction of carbon nanotube (CNT)-protein based constructs for possible bionanoelectronic device construction.

This chapter sought to improve the ZnPP binding affinity of cyt b_{562} while reducing its affinity for haem, by designing a zinc-optimised cyt b_{562} mutant using computational tools.

5.2 Results and discussion

5.2.1 Computational design of zinc-optimised cytochrome b_{562} (performed with Adam McGarrity)

A range of cyt b_{562} mutants were previously designed and optimised *in silico* by former lab member Adam McGarrity in an iterative computational process. Initially, the cyt b_{562} residues forming the haem iron axial ligands (H102 and M7; see Figure 5.4a) that coordinate the porphyrin metal centre were examined to optimise Zn coordination. This was performed using the ZincExplorer server web app[282] (http://systbio.cau.edu.cn/ZincExplorer), Figure 5.3 shows the output of this analysis. ZincExplorer scores each residue of a protein sequence with a prediction score (max. 1) representing its probability of acting as a zinc-coordinating ligand, with a score of 0.4 being the cutoff value for a reasonable zinc ligand, and a score over 0.5 giving high confidence in the ligand being optimal for zinc coordination.

The imidazole side chain of residue H102 was considered an optimal Zn-coordinating ligand, but the *S*-methyl thioether group of residue M7 that coordinated haem in cyt b_{562}^{WT} was not scored as a Zn ligand at all. This suggests a neighbouring ligand may coordinate zinc in cyt b_{562}^{WT}, or coordination of the Zn ion may be pentadentate. Thus, M7 was repeatedly changed in the cyt b_{562}^{WT} sequence and the sequences submitted to the ZincExplorer server for relative scoring as Zn ligands. The mutation M7C was identified as likely to be an improved Zn ligand, though at 0.33 it still scored below the cutoff value of 0.4. Cysteine is a ubiquitous ligand in Zn-coordinating residues, as demonstrated by their utilisation in zinc finger domains[283]. It is to be noted that ZincExplorer is designed for optimising the coordination of lone Zn^{2+} ions as in ZF domains, typically by four

ligands, rather than as part of a zinc porphyrin in a hexadentate system such as this, but the web app was still considered useful to inform residue choice.

The second round of optimisation focused on changing several neighbouring residues close to the porphyrin in helix 1 of cyt b_{562}^{M7C} (L3, E4, T9 and L10) and repeating the scoring process using ZincExplorer. The best result was obtained with the five mutations L3K, E4D, M7C, T9C and L10R, with this arrangement M7C scored 0.33 compared to 0.61 for H102, still suboptimal but the best available choice to improve Zn binding (**Figure 5.3**).

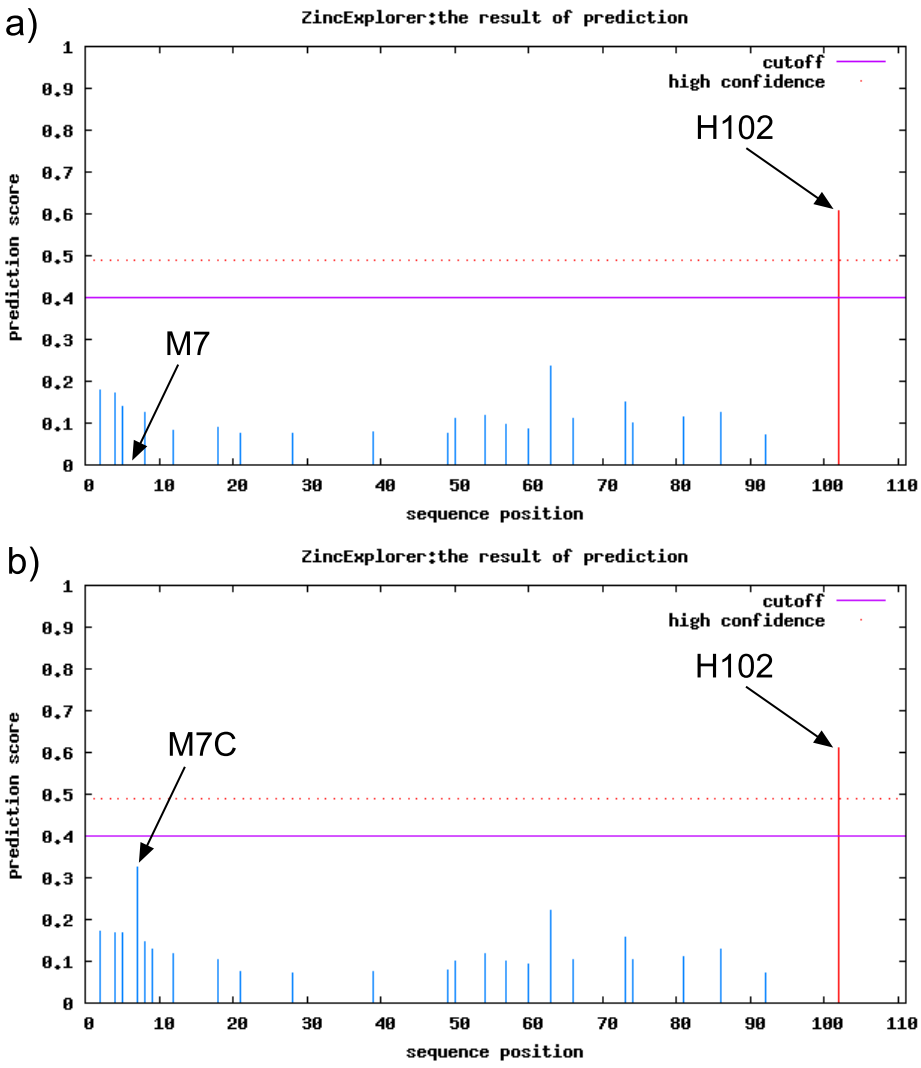


Figure 5.3 ZincExplorer scores for ZnPP axial ligand potential among cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} residues. Mutation of M7 to cysteine along with L3K, E4D, T9C and L10R provided the best improvement in potential zinc coordination (M7C = 0.33), though still below the cutoff threshold of 0.4. Ligand residue H102 was considered optimal already at 0.61.

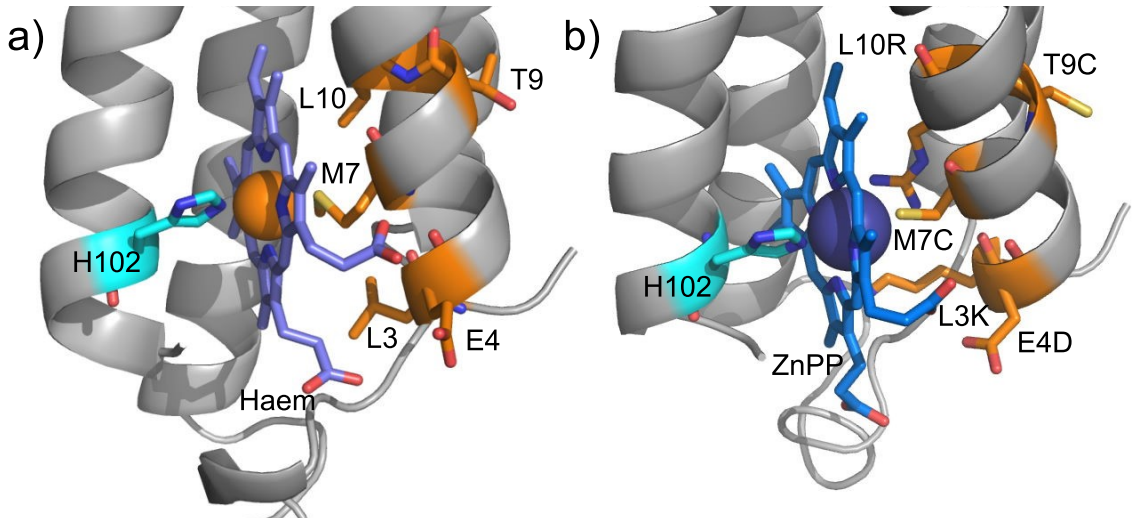


Figure 5.4 Cyt b_{562} porphyrin binding sites. a) Structure of haem binding pocket of cyt b_{562}^{WT}. The iron centre of haem (orange ion) is coordinated by two axial ligands: the side chains of H102 (guanidinium) and M7 (thioether). From PDB: 256B[38] b) Predicted ZnPP binding site of cyt b_{562}^{ZnPP} showing geometry-optimised structure for ZnPP, original axial ligand H102, and M7C and the other mutations predicted to improve ZnPP binding (orange).

McGarrity used the cyt b_{562}^{WT} crystal structure PDB:256B[38] as a basis for creating a library of cyt b_{562} mutant structure files containing every permutation of amino acid mutation for residues 1-20, the entire first helix, using automated scripts in PyMOL[11]. These files were used as starting points for energy minimisation using molecular dynamics simulations in GROMACS software[284]. The lowest energy state conformation for cyt b_{562} was used as a starting point for molecular docking simulations. A geometry optimised structure file of ZnPP was made using Avogadro[285] and GAMESS-US[286] software, this was used as the ligand file for ligand docking simulations using AutoDock4 software[15]. The optimised ZnPP model shows a “puckered” macrocycle conformation with the Zn^{2+} ion in an out-of-plane position, probably due to the increased ionic radius of Zn^{2+} (74 pm) compared to the Fe^{3+} (60 pm) in haem, a planar porphyrin (**Figure 5.5**). The mutants with the lowest free binding energies from this process were further analysed using RosettaLigand[287] software; these results were compared with those from AutoDock 4 to inform the choice of cyt b_{562} variant candidates that would be produced for further analysis and use.

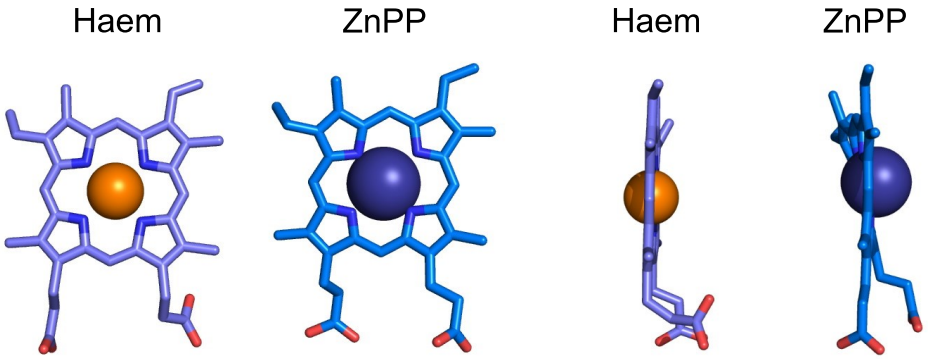


Figure 5.5 Conformation of porphyrins used in this work. The macrocycle of haem is planar, and the $Fe^{2+/3+}$ metal centre is an ionic radius of 60 pm. ZnPP has a "puckered" macrocycle with a non-coplanar Zn^{2+} metal centre with a larger ionic radius of 74 pm.

Overall, the *in silico* design process identified five mutations in helix 1 of cyt b_{562} with a predicted improvement to Zn coordination: L3K, E4D, M7C, T9C and L10R (see Figure 5.4b). The only mutation here with an obvious effect is M7C, with methionine being a poor ligand for Zn^{2+}, scored zero by ZincExplorer, and cysteine being a common Zn-coordinating residue. While it is not known exactly why the other four mutations should improve ZnPP binding and reduce haem binding, the mutations radically change the electrostatic character of the cyt b_{562} porphyrin binding pocket from nonpolar to positively charged (Figure 5.6, surface electrostatics by APBS[211]). In its typical oxidised form, haem iron carries a 3+ charge while ZnPP zinc carries a 2+ charge. It is possible that the positively charged porphyrin binding pocket of cyt b_{562}^{ZnPP} exerts greater repulsion on haem due to its more positive charge. The combination of a better ligand plus increased repulsion may explain a switch in specificity.

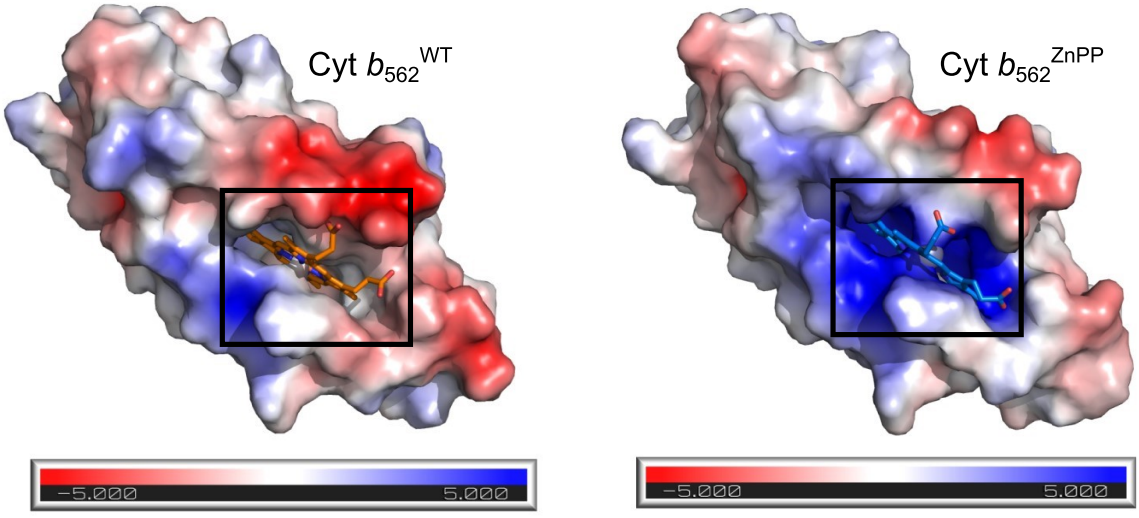


Figure 5.6 Surface electrostatic potential of cyt b_{562}^{WT} and predicted cyt b_{562}^{ZnPP} (blue = positive, red = negative, grey = neutral. Units -5 to +5 $k_b \cdot T \cdot e_c^{-1}$). The electrostatic character of the porphyrin binding pocket (boxed) in cyt b_{562}^{WT} is nonpolar. Comparison of the modelled cyt b_{562}^{ZnPP} structure shows the mutant binding pocket to be overwhelmingly positively charged. PDB: 256B. Surface electrostatics by APBS[211].

In this work, the variant with all five mutations, named **cyt b_{562}^{ZnPP}**, is produced and characterised, its predicted binding affinity is shown in Table 5.1. The new cyt b_{562}^{ZnPP} mutant was predicted to reduce haem binding affinity by 3.3 kcal/mol, while improving ZnPP binding affinity by -0.8 kcal/mol. These predictions were tested by producing the designed protein and quantifying its biophysical properties in the following sections.

Table 5.1 Predicted affinity of new cyt b_{562}^{ZnPP} mutant cf. wild type.

	Predicted affinity (kcal/mol)	
Variant	ZnPP	Haem
Cyt b_{562}^{WT}	-8.7	-12.3
Cyt b_{562}^{ZnPP}	-9.5	-9

5.2.2 Design of cytochrome b_{562}^{ZnPP} construct

A mutant gene, cyt b_{562}^{ZnPP}, was produced containing the five amino acid mutations previously designed to enhance cyt b_{562} zinc binding, as discussed above. This was produced using the existing wild type construct, pBAD-cyt b_{562} as a template. The mutations made to the cyt b_{562} gene and protein sequence are shown below, with mutations in red:

```
N-term - M  A  D  L  E  D  N  M  E  T  L  N  D  N  L  K
5′ -     ATGGCAGATCTTGAAGACAATATGGAAACCCTCAACGACAATTTAA
         |||||||||   || ||||||   |||  ||  ||||||||||||||
5′ -     ATGGCAGATAAGGACGACAATTGTGAATGCCGCAACGACAATTTAA
N-term - M  A  D  K  D  D  N  C  E  C  R  N  D  N  L  K
```

The five mutations were introduced in a single round of site-directed, ligase independent mutagenesis (SLIM)[188] by PCR (method: chapter 2.2.2.6). As the mutations are all in the first α-helix of cyt b_{562} within a seven amino acid stretch, all mutations could be introduced simultaneously using two pairs of SLIM primers (listed in 2.2.2.3).

5.2.3 Cytochrome b_{562}^{ZnPP} production

5.2.3.1 Cytochrome b_{562}^{ZnPP} molecular biology

To introduce the designed mutations by SLIM PCR mutagenesis, the existing pBAD-cyt b_{562} plasmid was used as a template. The pBAD-cyt b_{562} construct contains the gene encoding cyt b_{562}^{WT} with its native signal sequence removed, cloned into the pBAD plasmid, and was originally made by Sam Reddington[253]. On PCR amplification a band

of ca. 4.3 kbp was expected. Analysis of SLIM PCR by agarose gel electrophoresis revealed a single band between 4-5 kbp (Figure 5.7) indicating that the reaction had worked. The product was then digested with DpnI to remove any residual parental template, ligated and transformed into *E. coli*. More than 20 colonies were present the following day on an LB agar plate, six were cultured, plasmid DNA isolated and the insert sequenced to confirm the presence of the mutations. All of the sequenced clones contained the mutations and were thus taken forward for production of the protein termed cyt b_{562}^{ZnPP}.

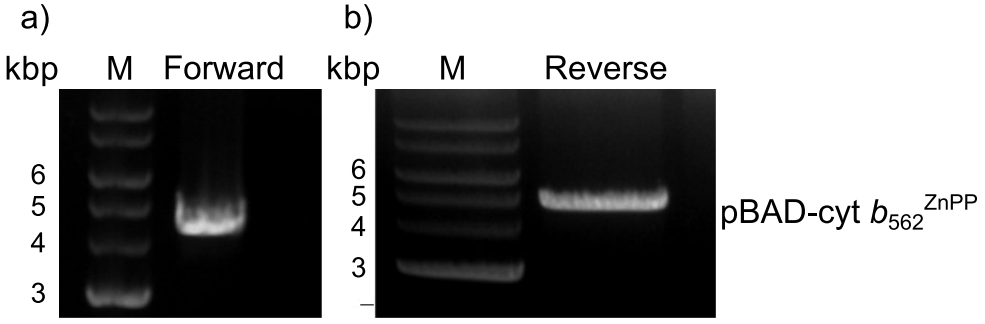


Figure 5.7 Agarose gel analysis of the a) forward and b) reverse DNA strands of the 4.3 kbp pBAD-Cyt b_{562}^{ZnPP} plasmid, after PCR mutagenesis and amplification using the SLIM method. One band was present at 4-5 kpb for each strand. M: molecular weight marker ladder.

5.2.3.2 Cytochrome b_{562}^{WT} and cyt b_{562}^{ZnPP} expression and purification

Cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} were expressed in *E. coli* TOP10 cells as described previously (chapter 4). After centrifuging cell cultures, the cyt b_{562}^{WT} cell pellet was red, characteristic of cyt b_{562}^{WT} expression. Cyt b_{562}^{ZnPP} cells were colourless (Figure 5.8a), this could indicate a lack of expression, or expression of cyt b_{562}^{ZnPP} but no haem binding. Lysate from the colourless cells was analysed by SDS-PAGE along with non-transformed *E. coli* lysate, showing overexpression of a small protein (< 17 kDa) only in pET-cyt b_{562}^{ZnPP} transformed cells (Figure 5.8b). This band on SDS-PAGE matches the previously observed band size for cyt b_{562} variants in chapter 4.

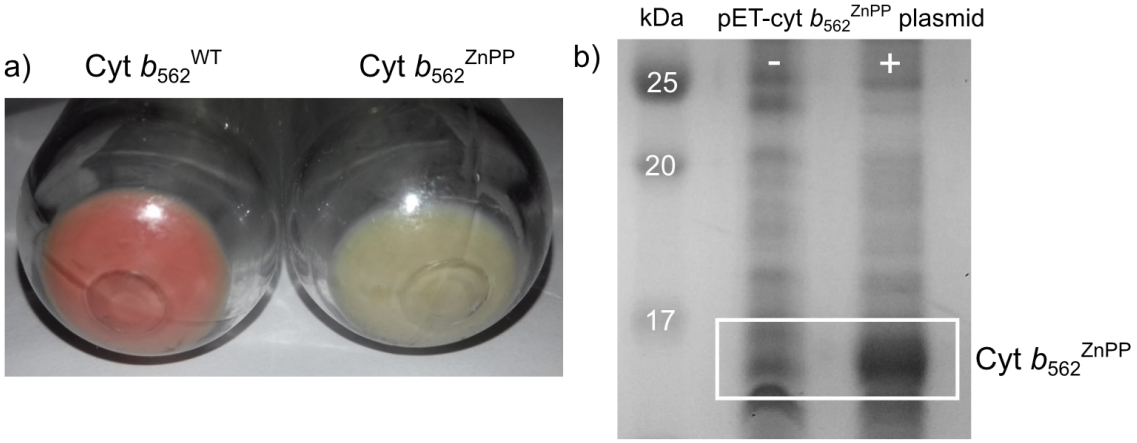


Figure 5.8 Expression of cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} in *E. coli*. a) Cell pellets expressing cyt b_{562}^{WT} and cyt b_{562}^{ZnPP}. The characteristic red colour of protein-bound haem in cyt b_{562} is absent in the cyt b_{562}^{ZnPP} mutant. b) SDS PAGE analysis of cell lysate from cells with (+) and without (-) the pBAD-cyt b_{562}^{ZnPP} plasmid, showing expression of cyt b_{562}^{ZnPP} only in the transformed cells.

Cells containing cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} were lysed and the target proteins part-purified using ammonium sulphate precipitation as described previously. Precipitated proteins were re-dissolved in buffer and the sample containing cyt b_{562}^{WT} was purified by size exclusion chromatography (SEC) using a Superdex 200 26/600 column (Cytiva) (Figure 5.9). This yielded impure cyt b_{562}^{WT}, so fractions containing cyt b_{562}^{WT} were concentrated to 2 ml using a 3 kDa Vivaspin concentrator (Cytiva) and SEC was repeated using a Superdex 75 16/600 column (Cytiva). Cyt b_{562}^{WT} eluted at V_e 65-75 mL and was ≥ 99% pure, as estimated using ImageJ (Figure 5.10).

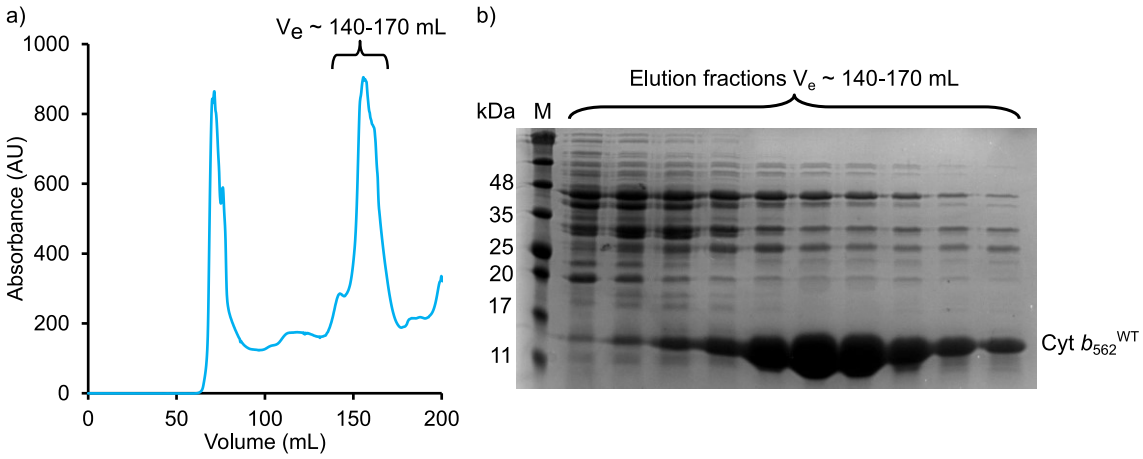


Figure 5.9 Purification of cyt b_{562}^{WT} by size exclusion chromatography (Superdex 200 26/600). a) Elution of cyt b_{562}^{WT} from column at V_e 140-170 mL: only 280 nm absorbance was available (blue line), cyt b_{562}^{WT} elution was tracked by viewing red band on the column. b) SDS-PAGE analysis of red fractions: no pure protein at this stage.

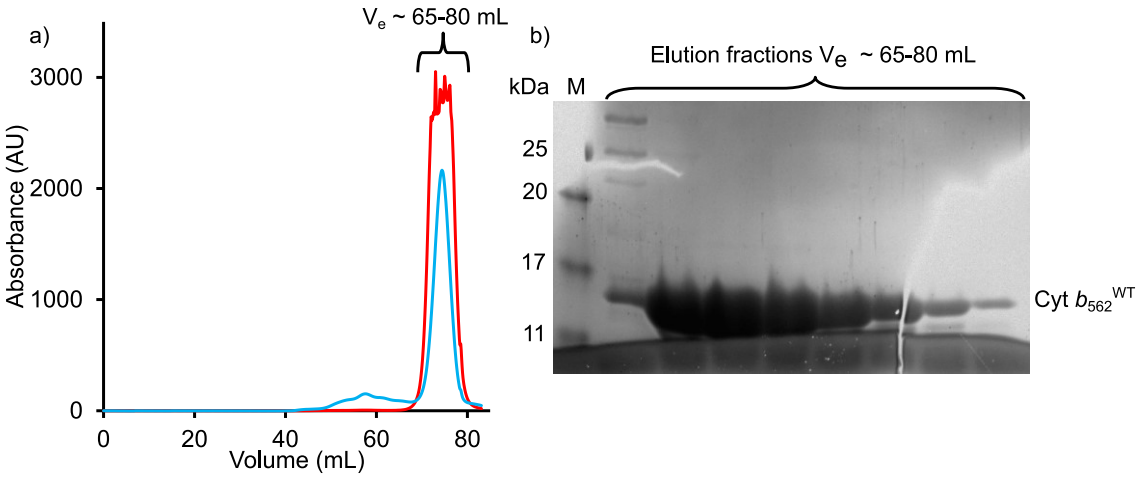


Figure 5.10 Purification of cyt b_{562}^{WT} by size exclusion chromatography (Superdex 75 16/600). a) Elution of cyt b_{562}^{WT} from column at V_e 65-80 mL: 280 nm absorbance (blue line), 418 nm absorbance (red). b) SDS-PAGE analysis of fractions from major A418 peak at V_e 65-80 mL. Latter fractions ≥ 99% pure.

The cyt b_{562}^{ZnPP} protein sample was desalted using a PD10 gel filtration column (Cytiva) and applied to a MonoQ anion exchange column, then eluted using a 0-500 mM NaCl gradient. SDS-PAGE analysis showed that cyt b_{562}^{ZnPP} did not bind the MonoQ column and was all present in the flow-through fraction. This was concentrated to 2 ml using a 3 kDa Vivaspin concentrator. The sample was further purified by SEC using a Superdex 75 column, with cyt b_{562}^{ZnPP} eluting at V_e 65-75 mL (Figure 5.12a). The elution peak on the chromatogram for cyt b_{562}^{ZnPP} showed a sixfold higher peak absorbance at 280 nm than at the haem Soret maximum of 418 nm. The molar extinction coefficients for cyt b_{562}^{WT} are 3 $mM^{-1}{\cdot}cm^{-1}$ at 280 nm and 117.4 $mM^{-1}{\cdot}cm^{-1}$ at 418 nm for holo-protein. This implies that less than 0.5% of cyt b_{562}^{ZnPP} was in the haem-bound state, reflecting the lack of red colouration of the protein. Purity of cyt b_{562}^{ZnPP} elution fractions was assessed using SDS-PAGE (Figure 5.12b) and ImageJ, latter fractions were ≥ 99% pure.

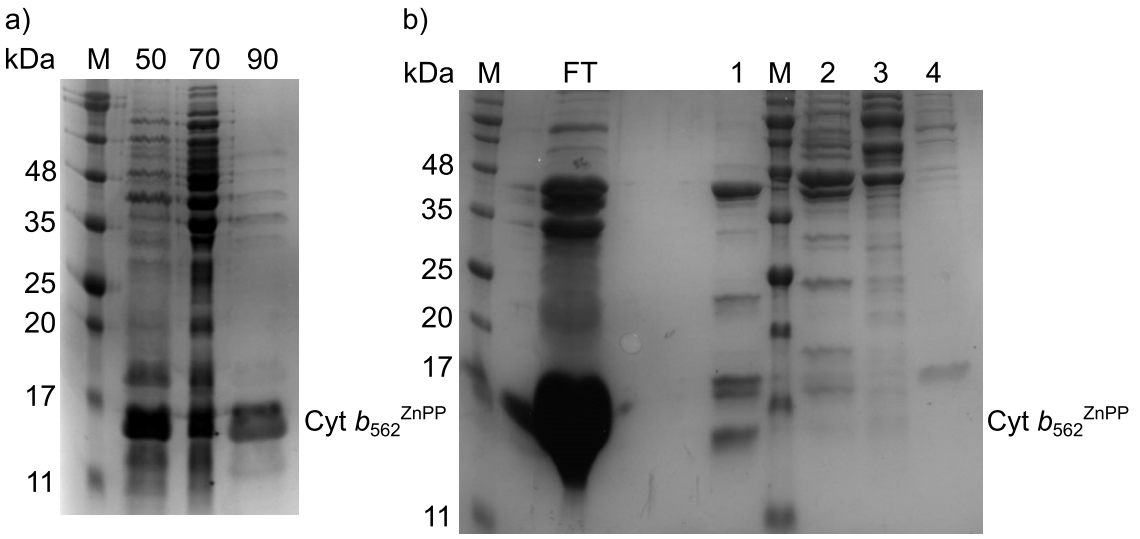


Figure 5.11 SDS-PAGE analyses of cyt b_{562}^{ZnPP} purification. a) Proteins precipitated from *E. coli* lysate at 50%, 70% and 90% (w/v) ammonium sulphate saturation. A band < 17 kDa indicates cyt b_{562}^{ZnPP} in all. b) Fractions eluted from anion exchange column with NaCl gradient. Cyt b_{562}^{ZnPP} did not bind and remained in the flow-through (FT), but many contaminants bound the column (fractions 1-4), partly purifying the sample. M: molecular weight marker ladder.

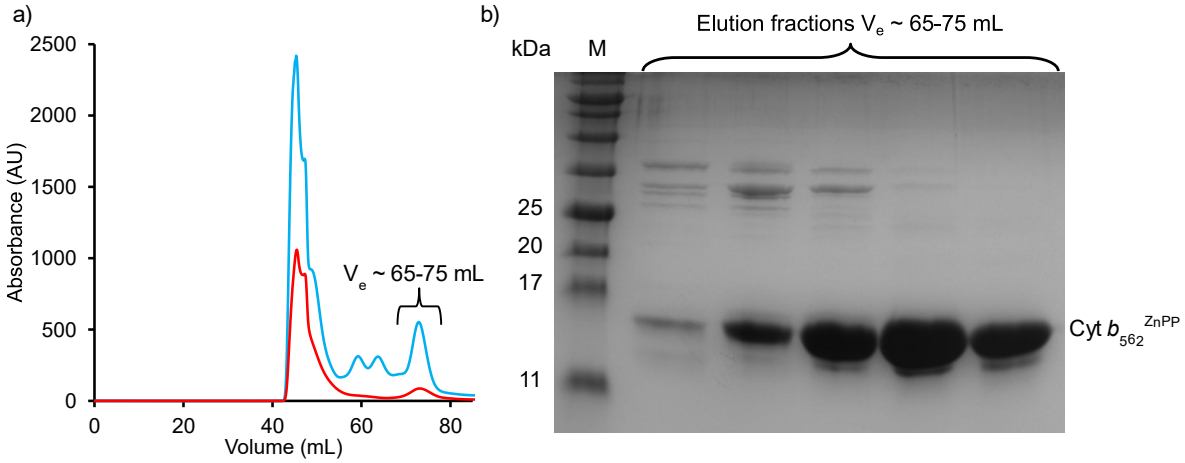


Figure 5.12 Purification of cyt b_{562}^{ZnPP} by size exclusion chromatography. a) Chromatogram of cyt b_{562}^{ZnPP} elution at V_e 70 mL. 280 nm absorbance (blue line) and 418 nm absorbance (red line). b) SDS-PAGE analysis of the fractions from cyt b_{562}^{ZnPP} elution at V_e 65-75 mL. The latter two fractions were ≥99% pure cyt b_{562}^{ZnPP} (gel analysis with ImageJ). M: molecular weight marker ladder.

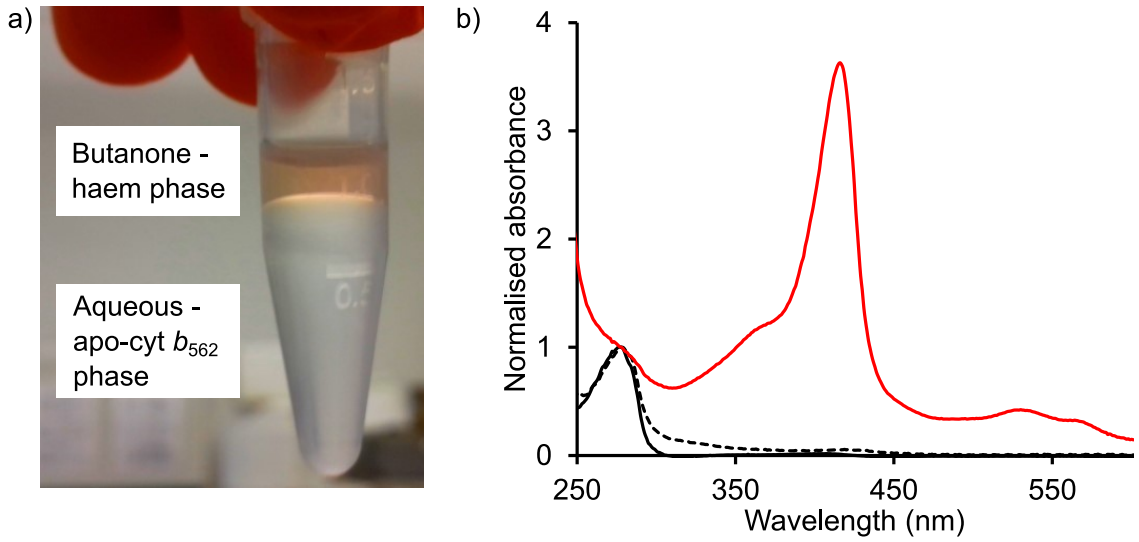


Figure 5.13 Production of apo-cyt b_{562}^{WT}. a) Organic extraction of haem from cyt b_{562}^{WT}, with the haem dissolved into the upper butanone phase for removal and apo-cyt b_{562} remaining in the lower aqueous phase. b) Absorbance spectra for holo-cyt b_{562}^{WT} (red line), apo-cyt b_{562} (black) and cyt b_{562}^{ZnPP} (dashed line). The absence of the major haem Soret peak at 418 nm following haem extraction shows it was successful. Cyt b_{562}^{ZnPP} showed no apparent haem peak.

To investigate porphyrin binding parameters, any bound haem cofactor needs to be removed by organic extraction to yield apo-protein (method: 2.2.4.6) (Figure 5.13a). Protein haem content was then assessed by UV-visible spectrophotometry: holo-cyt b_{562}^{WT}, apo-cyt b_{562}^{WT} (after haem extraction), and cyt b_{562}^{ZnPP} absorbance spectra were compared (Figure 5.13b). Holo-cyt b_{562}^{WT} showed its characteristic intense haem Soret band peak at 418 nm and smaller Q band peaks at 500-575 nm, while these features were absent in apo-cyt b_{562}^{WT} indicating removal of the haem cofactor. The cyt b_{562}^{ZnPP} sample did not undergo organic extraction yet showed negligible absorbance at typical haem wavelengths, implying a near complete lack of haem binding. This supports the observation from the colourless cyt b_{562}^{ZnPP} cells that this mutant protein does not bind haem to any significant extent, in accordance with the design aims.

5.2.4 Characterisation of cytochrome b_{562}^{ZnPP}

5.2.4.1 Cytochrome b_{562}^{ZnPP} porphyrin binding

The mutations introduced in the basic cytochrome b_{562} were designed *in silico* to make the protein preferentially bind the fluorogenic zinc protoporphyrin IX over its normal haem cofactor. The physiochemical properties of metalloporphyrin, such as the spectral properties of ZnPP are sensitive to factors affected by protein binding such as metal oxidation state, coordination by axial ligands, and solvent environment. In the case of ZnPP, it is hypothesised that binding to protein will sequester it from the aqueous environment so promoting fluorescence. Thus, protein-bound porphyrins are distinguishable from porphyrins free in solution using absorbance and fluorescence

spectrophotometry. The effects of the binding mutations were assessed by comparing the spectral properties of the cyt b_{562}^{ZnPP} mutant with wild type cyt b_{562}.

5.2.4.2 Binding of metalloporphyrin to cyt b_{562}^{ZnPP}: absorbance spectroscopy.

The features of metalloporphyrin absorbance spectra are affected by the coordination geometry of the metal centre, polarity of the solvent environment, metal centre oxidation state and porphyrin aggregation[288]. These factors affect the wavelength, intensity, relative peak width and intensity ratios of observed Soret and Q band absorbance features, providing information about the porphyrin environment such as solvation and protein binding. The absorbance spectra of cyt b_{562} and cyt b_{562}^{ZnPP} were analysed using UV-visible spectrophotometry to determine porphyrin binding characteristics. Major features of the cyt b_{562}^{WT} absorbance spectrum in complex with both haem and ZnPP have previously been reported[171,289] and are shown here as a reference (**Figure 5.14**). Haem used here was in its oxidised (Fe^{3+}) state.

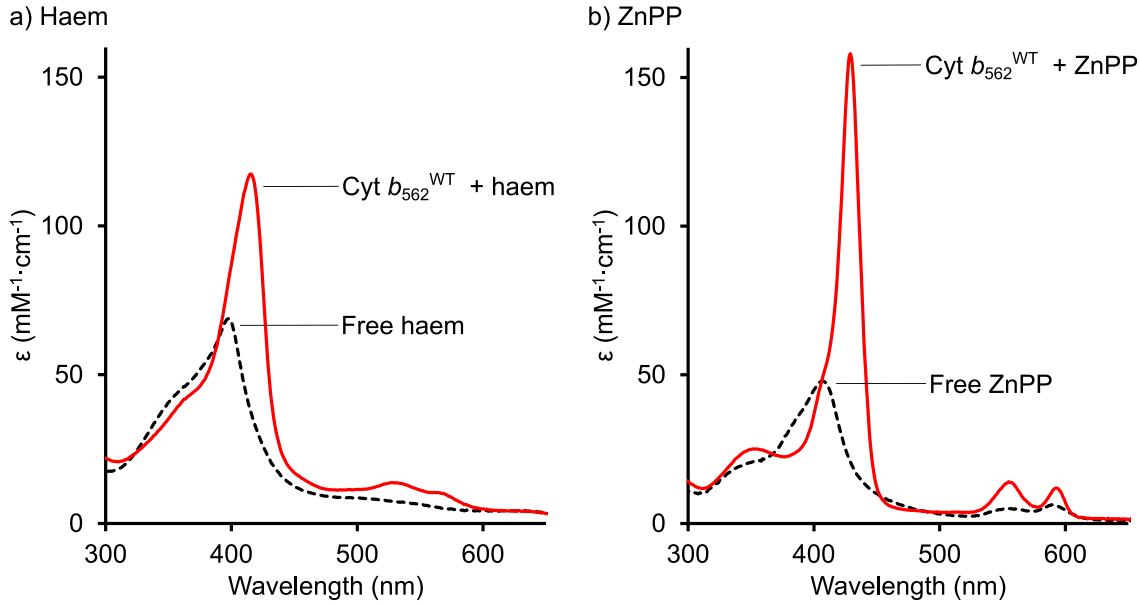


Figure 5.14 Comparison of cyt b_{562}^{WT} absorbance spectra in sodium phosphate buffer, pH 6.2. a) Spectra of free haem (oxidised, Fe^{3+}), and cyt b_{562}^{WT} reconstituted with haem. b) Spectra of free ZnPP, and cyt b_{562}^{WT} reconstituted with ZnPP.

Haem is poorly soluble in aqueous solution and readily dimerises and aggregates by π-π stacking. Its absorbance spectrum in aqueous buffer shows multiple overlapping Soret band absorbance peaks between 350-400 nm due to a mixed population of haem monomer and dimer[290] with a maximum molar extinction coefficient (ε_{max}) of ca. 60 $mM^{-1}\cdot cm^{-1}$ at 396 nm. The spectral features and molar extinction coefficients of haem in aqueous buffer vary depending on its proportion of monomeric, dimeric or higher aggregate species, which is affected by buffer composition and time spent in solution[291,292]. On binding to apo-cyt b_{562}^{WT}, haem Soret peaks consolidated into a single

sharper, redshifted peak at 418 nm with ε = 117.4 $mM^{-1}·cm^{-1}$ and discrete Q band peaks became discernible at ε_β ca. 530 nm, ε_α ca. 570 nm (Figure 5.14a), consistent with reconstitution of cyt b_{562}^{WT} holo-protein (cf. holo-cyt b_{562}^{WT} spectrum in Figure 5.13b). ZnPP is also poorly soluble and subject to aggregation in aqueous solution[293], tending toward a broad Soret peak with ε_{max} ca. 50 $mM^{-1}·cm^{-1}$ at ca. 406 nm with Q band peaks at ε_β ca. 555 nm, ε_α ca. 590 nm. On binding to apo-cyt b_{562}^{WT}, the ZnPP Soret peak redshifted to 429 nm, ε = 158 $mM^{-1}·cm^{-1}$ and the intensity ratio of the ε_β:ε_α peaks was reversed (Figure 5.14b), this is consistent with previously published observations of ZnPP-cyt b_{562}^{WT} binding[171] and ZnPP-haemprotein binding generally[294].

To assess the basic porphyrin binding ability of cyt b_{562}^{ZnPP}, a 5 μM sample of the protein was incubated with haem or ZnPP (6 μM) to allow formation of protein-porphyrin complexes. Absorbance spectra were recorded and compared to those of of free porphyrins in aqueous solution, and porphyrins bound to cyt b_{562}^{WT}. The spectrum of haem incubated with cyt b_{562}^{ZnPP} showed few of the signs of protein-haem binding observed with cyt b_{562}^{WT}. The haem-cyt b_{562}^{ZnPP} spectrum closely resembled that of free haem, except for an 8 nm Soret peak redshift (396-404 nm) and small increase in intensity. These slight changes after incubation with a molar excess of haem imply a very small proportion of haem binding to cyt b_{562}^{ZnPP}, with the majority of cyt b_{562}^{ZnPP} existing as apo-protein. This finding further supports the previous observations that *E. coli* cells expressing cyt b_{562}^{ZnPP} lack the red colour characteristic of haem-bound cyt b_{562}^{WT} (Figure 5.8a) and that purified cyt b_{562}^{ZnPP} lacks haem spectral features (Figure 5.13b). When incubated with cyt b_{562}^{ZnPP}, the Soret band λ_{max} of ZnPP redshifted from 406 to 426 nm, the peak sharpened and increased its intensity from ca. 50 to 158 $mM^{-1}·cm^{-1}$. This indicates ZnPP binding protein binding as previously published[171] (Figure 5.15b).

Binding to both cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} caused the intensity ratio of the ZnPP β and α peaks to swap around, suggesting altered coordination geometry of the metal centre due to protein-porphyrin complex formation. The spectral features of protein-bound ZnPP more closely resemble ZnPP in a nonpolar solution with a sharp Soret peak around 430 nm, compared to the broad ca. 406 nm peak of ZnPP in aqueous solution due to ZnPP aggregation. (**Figure 5.16**)[293] This suggests that cyt b_{562} surrounds ZnPP with its hydrophobic core, shielding it from water.

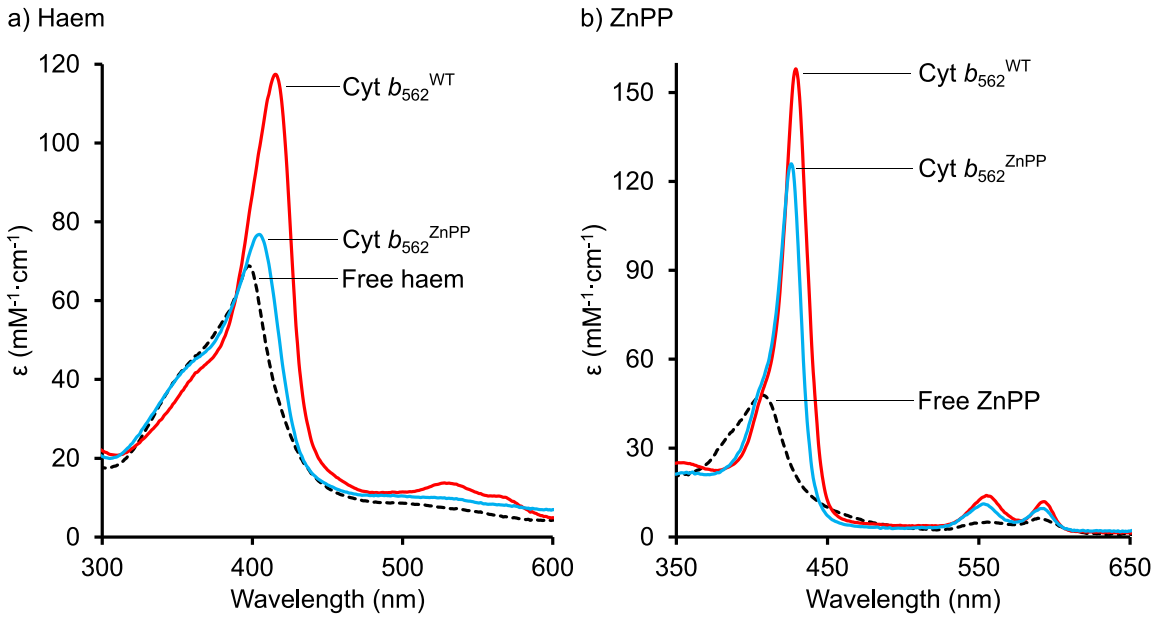


Figure 5.15 Absorbance spectra of free porphyrin, and porphyrin incubated with cyt b_{562}^{WT} or cyt b_{562}^{ZnPP} in aqueous buffer. a) Free haem (dashed line), cyt b_{562}^{WT} (red line) and cyt b_{562}^{ZnPP} (blue). Cyt b_{562}^{WT} Soret peak shows characteristic redshift from 395 to 418 nm and increase in Q band α/β peaks at 500-600 nm on haem binding. Cyt b_{562}^{ZnPP} lacks these features implying a lack of haem binding. b) Spectra of free ZnPP (dashed line), cyt b_{562}^{WT} (red line) and cyt b_{562}^{ZnPP} (blue). Cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} show redshift of the ZnPP Soret peak from 406 nm to 426 and 429 nm respectively, with reversal of Q band α/β peak ratio that is characteristic of ZnPP-protein binding.

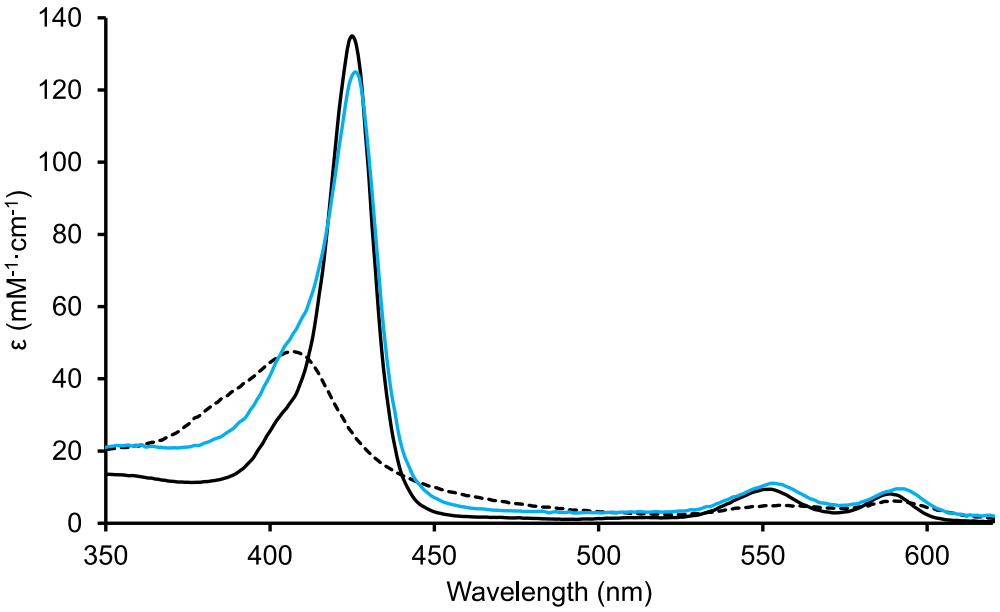


Figure 5.16 Comparison of ZnPP absorbance spectra. Free ZnPP in aqueous solution (dashed line) and in nonpolar pyridine (black line) and ZnPP bound to cyt b_{562}^{ZnPP} (blue). The spectral features of protein-bound ZnPP strongly resemble those of ZnPP in nonpolar solution.

Metalloporphyrin titration binding assays were performed to determine the dissociation constants and of cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} for haem and ZnPP, using the absorbance values at the redshifted protein-porphyrin peaks. The titration was performed using cyt b_{562}^{WT} or cyt b_{562}^{ZnPP} at 5 μM and porphyrin solutions at 50 μM, with the porphyrin titrated

into the protein. The absorbance values for cyt b_{562}^{WT} with haem at the complex Soret ε_{max} of 418 nm were fitted to Equation 2.4 (chapter 2.2.8.7) using Prism software, this equation calculates dissociation constant while accounting for dilution by the porphyrin solution. Background free haem absorbance was subtracted from the values. When fitted to the published K_D of this interaction (9 nM[279]), the data fit with a pseudo-R^2 of 99.6% (Figure 5.17). No discrete redshifted absorbance peak was observed during the titration of haem into cyt b_{562}^{ZnPP}, at any concentration of haem, implying negligible binding affinity (cf. Figure 5.15a). This finding supports the earlier observations that cyt b_{562}^{ZnPP} expressing cells lacked the characteristic red colour of protein-bound haem (Figure 5.8), and that the absorbance spectrum of purified cyt b_{562}^{ZnPP} lacked the characteristic 418 nm Soret peak of protein-bound haem (Figure 5.13).

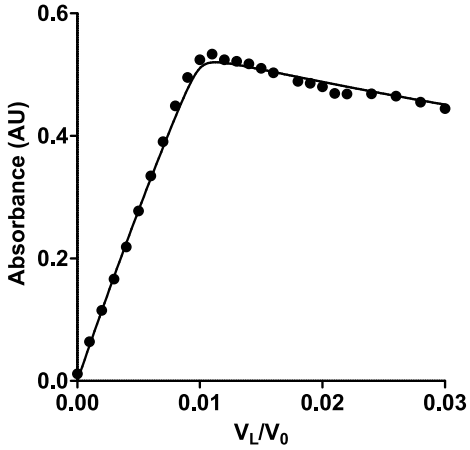


Figure 5.17 Verification of K_D for cyt b_{562}^{WT} with haem. The data matched the published figure of 9 nM.

The cyt b_{562}-ZnPP titration absorbance spectra recorded were not useful for calculating dissociation constants because the absorbance peak for free ZnPP (ca. 406-416 nm) overlaps considerably with the absorbance peak of the protein-porphyrin complex at ca. 427 nm. This effect means that as ZnPP concentration is increased during titration, the absorbance contribution from the free ZnPP in the system also raises the ca. 427 nm absorbance value, rendering it inaccurate for fitting to the binding equation. After the protein becomes saturated with ZnPP, the free ZnPP overtakes and obscures the protein-porphyrin peak altogether (Figure 5.18). This masking effect was greater for the cyt b_{562}^{ZnPP}-ZnPP titration due to the lower molar absorption coefficient of the cyt b_{562}^{ZnPP}-ZnPP complex. Wavelengths from the red edge of the protein-ZnPP peak were analysed but were still not sufficiently far from the free ZnPP to avoid interference. Attempts to subtract the free ZnPP spectra and deconvolute the overlapping peaks were made using a|e software[295] in Matlab, but results from these were unsatisfactory. While it was clear from the discrete peak at ca. 427 nm that both cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} were binding to

ZnPP, the affinity could not be quantified by this method. It was decided to try to estimate the K_D of these complexes by fluorescence instead.

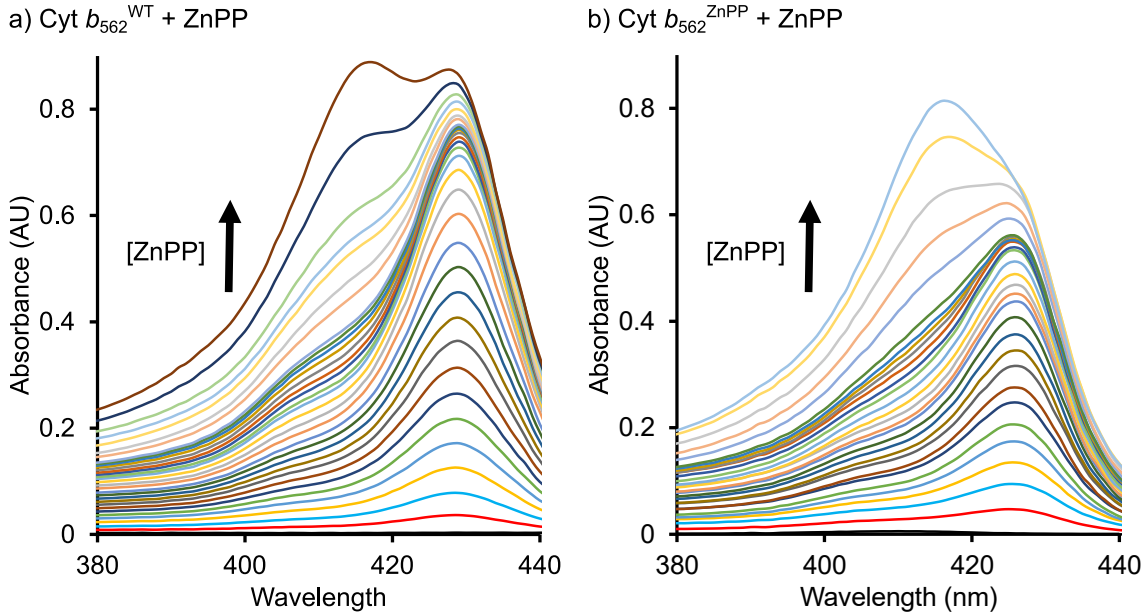


Figure 5.18 Titration of ZnPP into cyt b_{562}^{WT} and cyt b_{562}^{ZnPP}. The free ZnPP absorbance peak (ca. 410 nm) overlaps and merges with the protein-ZnPP complex absorbance peak (ca. 427 nm), eventually obscuring it altogether. This effect makes the absorbance values at ca. 427 nm for the protein-ZnPP complex inaccurate and unusable for binding calculations.

5.2.4.3 ZnPP binding properties probed using fluorescence.

During cyt b_{562}-ZnPP titrations, the free porphyrin Soret peaks overlapped with and masked the redshifted protein-porphyrin peaks on the absorbance spectra, making the results unsuitable for fitting to binding equations. What the absorbance data did indicate was that cyt b_{562}^{ZnPP} had a much lower affinity for haem backing up the observation of cell colouration differences on recombinant protein production. The second aspect of the design was to increase ZnPP binding affinity above that of cyt b_{562}^{WT}. While the absorbance data indicated that cyt b_{562}^{ZnPP} binds ZnPP in the same way as cyt b_{562}^{WT}, the improvement in binding affinity of cyt b_{562}^{ZnPP} was quantified using fluorescence.

The same titration experiment as previously was repeated using the inherent fluorescent characteristics of ZnPP, exploiting the difference in fluorescence characteristics of ZnPP between the aqueous and protein-bound environments. In aqueous solution the excitation maxima of ZnPP are at 410 nm for free porphyrin and 425 nm complexed with cyt b_{562}^{ZnPP}, for this experiment, samples were excited at 430 nm to minimise emission from free ZnPP.

It was found that when ZnPP is solubilised using NaOH, its emission (λ_{em} 597 nm) on binding to cyt b_{562}^{ZnPP} is enhanced ca. 60-fold. However, when ZnPP is solubilised using DMSO, its emission on binding to cyt b_{562}^{ZnPP} is enhanced approx. tenfold (Figure 5.19).

This unexpected difference in free ZnPP emission may be due to residual DMSO solvating ZnPP molecules and shielding them from the quenching effects of water, this increasing emission.

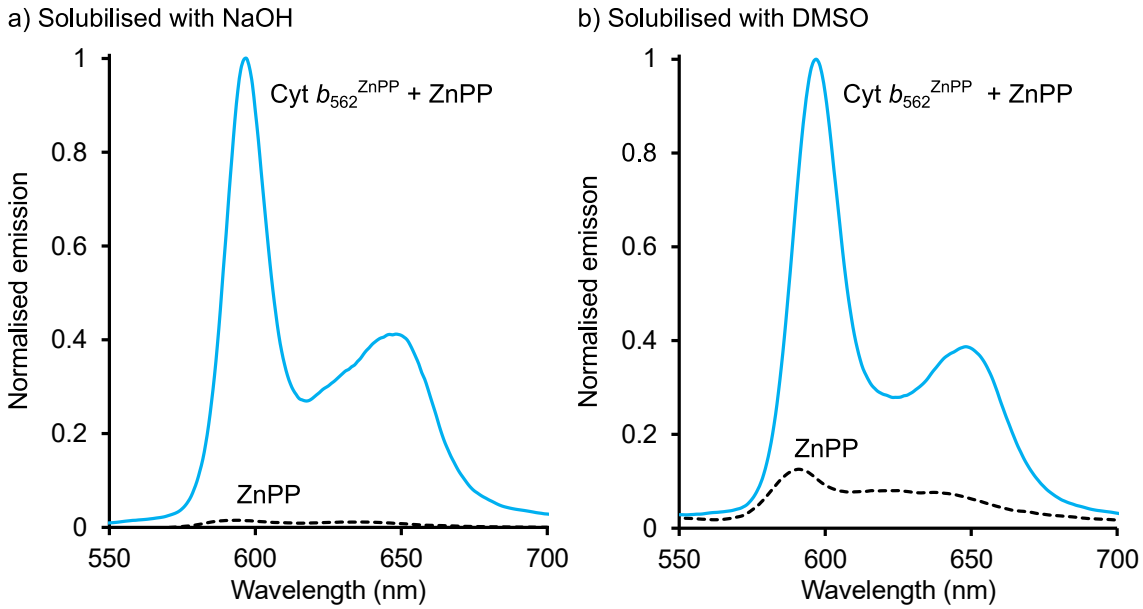


Figure 5.19 Emission spectra for free ZnPP and cyt b_{562}^{ZnPP}+ZnPP in sodium phosphate buffer, λ_{ex} 430 nm. Normalised to a maximum of 1 to compare relative fluorescence of unbound ZnPP. a) Peak ZnPP emission is enhanced ca. 60-fold on binding to cyt b_{562}^{ZnPP} when ZnPP was solubilised with NaOH. b) Peak ZnPP emission is enhanced approx. tenfold on binding to cyt b_{562}^{ZnPP} when ZnPP was solubilised with DMSO.

Table 5.2 Porphyrin dissociation constants (K_D) of cyt b_{562} variants, as calculated by fitting peak fluorescence data to Equation 2.5 (chapter 2.2.8.8). Errors are SE of fitting to Equation 2.5.

Variant	ZnPP K_d
Wild type cyt b_{562}	626 ± 28 nM
Cyt b_{562}^{ZnPP}	34 ± 10 nM
Pseudo-R^2 values > 0.98 for both fits.	

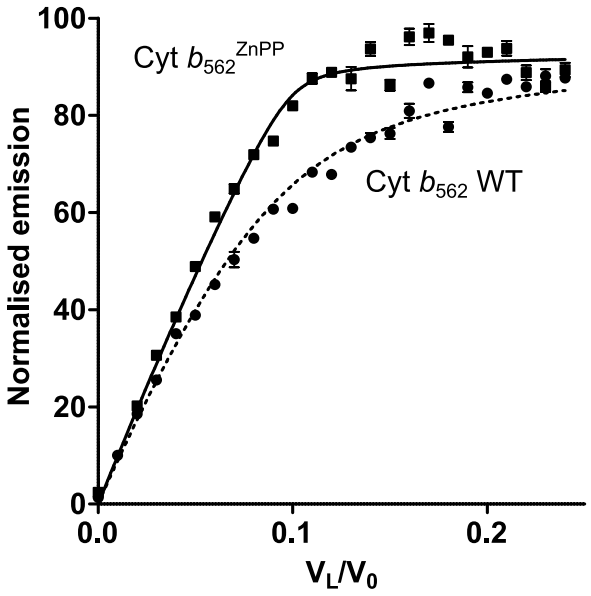


Figure 5.20 Fluorescence emission for the cyt b_{562}-ZnPP complex (λ_{ex} = 430 nm, λ_{em} = 597 nm) for a 0-12 µM titration of ZnPP into 5 µM WT cyt b_{562} (circles, dashed line) or cyt b_{562}^{ZnPP} (squares, solid line). Cyt b_{562}^{ZnPP} reaches binding saturation at a lower ZnPP concentration, indicating higher binding affinity than WT. Error bars are SEM of three biological replicates.

The minimum tenfold enhancement of emission seen on binding to cyt b_{562} allowed the peak emission values (λ_{em} 597 nm) for ZnPP titrations into cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} to be fitted to Equation 2.5 (chapter 2.2.8.8) using Prism to estimate K_D values for cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} (Figure 5.20, Table 5.2). This equation calculates dissociation constant while accounting for dilution and the lower fluorescence contribution from free ZnPP in the system. Using this fluorescence approach, protein-ZnPP dissociation constants were calculated to be 626 ± 28 nM for cyt b_{562}^{WT} and 34 ± 10 nM for cyt b_{562}^{ZnPP}, with this result for cyt b_{562}^{WT} being consistent with the published value of 500 nM[171].

The data from the fluorescence titration showed that the cyt b_{562}^{ZnPP} mutations caused an 18-fold improvement in ZnPP affinity compared to cyt b_{562}^{WT}. The new cyt b_{562}^{ZnPP} mutant was found to have a ZnPP affinity close to that of cyt b_{562}^{WT} for haem (34 ± 10 nM vs. 9 nM[279]). When combined with the near-total abolition of haem binding affinity in the cyt b_{562}^{ZnPP} mutant, established by absorbance spectrophotometry in section 5.2.4.2, this represents an almost complete switch of cofactor specificity for cyt b_{562}^{ZnPP}. This outcome fulfils the original aim of the design process and shows the power of computational design methods in making rational changes to protein structure towards a practical outcome. The single mutation variant cyt b_{562}^{M7C} was produced during preliminary work by Adam McGarrity, and did not display the drastic shift in affinity observed for cyt b_{562}^{ZnPP}, with haem affinity of 206 nM and ZnPP affinity of 181 nM, similar to cyt b_{562}^{WT}. This demonstrates the contribution of multiple residues to the affinity shift of cyt b_{562}^{ZnPP}.

The titration data collected was not ideal for the purpose, with the concentration of both cyt b_{562}^{ZnPP} and ZnPP used being far above the K_D value. The concentration of cyt b_{562}^{ZnPP} should be below the anticipated K_D to derive an accurate value of K_D. Using these concentrations practically all ZnPP at each titration step below protein concentration is bound, rendering the point at which ZnPP saturates half of cyt b_{562}^{ZnPP} (K_D) impossible to discern[296]. However, the equipment available did not allow such low concentrations and the values estimated were still useful for comparing cyt b_{562}^{ZnPP} to cyt b_{562}^{WT}.

5.2.4.4 Analysis of holo-cyt b_{562}^{ZnPP} using red-edge excitation shift (REES) and circular dichroism (CD) spectroscopy

Fluorescence spectra were recorded for both ZnPP and for the holo-cyt b_{562}^{ZnPP} complex, with excitation wavelengths from 411-441 nm; the data was analysed with the help of Dr. Chris Pudney at the University of Bath, using red-edge excitation shift (REES) and circular dichroism (CD) spectroscopy. REES is an optical phenomenon whereby low energy excitation of a fluorophore (i.e. at longer wavelength) can cause a redshift in the centre of spectral mass (CSM) of the emission peaks; this effect is influenced by the

solvation environment of the fluorophore and can be used to probe fluorophore binding[297]. Both major emission maxima of ZnPP (λ_{em}(1) ca. 590 nm; λ_{em}(2) ca. 630 nm) were analysed in free and protein-bound form. Free ZnPP showed a REES effect of 0.56 ± 0.02 nm^{-1} at λ_{em}(1), and 0.21 ± 0.02 nm^{-1} at λ_{em}(2); when bound to cyt b_{562}^{ZnPP}. This effect was reduced to 0.12 ± 0.01 nm^{-1} at λ_{em}(1) and -0.04 ± 0.03 nm^{-1} at λ_{em}(2) (**Figure 5.21**a). These results were interpreted to mean that free ZnPP inhabits a range of different solvation environments in solution, while protein-bound ZnPP exists in a single, or very limited range of solvation environments, or is shielded from water altogether by cyt b_{562}^{ZnPP}. This suggests tight ZnPP binding by cyt b_{562}^{ZnPP} in a discrete coordination state.

To assess the conformational states of cyt b_{562}^{ZnPP} in apo-form and bound to ZnPP, far-UV CD spectroscopy was carried out before and after ZnPP binding. Apo-cyt b_{562}^{WT} is relatively unstructured compared to its four α-helix, haem-bound holo-form, the binding of haem can be observed using CD spectroscopy. On binding its haem cofactor, the CD spectral features associated α-helix content (troughs at 208 and 222 nm) become more prominent as cyt b_{562}^{WT} folds around its cofactor, becoming more structured[298]. The CD spectra of cyt b_{562}^{ZnPP} were compared to see whether binding to ZnPP is associated with a similar increase in organisation, to eliminate the possibility of the previously observed binding effects being non-specific. The CD spectra show that apo-cyt b_{562}^{ZnPP} displays substantially less helical character than holo-cyt b_{562}^{ZnPP}, evidenced by the shallower troughs at 208 and 222 nm (Figure 5.21b)[299], consistent with a partially unfolded apo-protein. The cyt b_{562}^{ZnPP}-ZnPP holo-complex had a CD spectrum similar to holo-cyt b_{562}^{WT} (with haem)[300], implying that cyt b_{562}^{ZnPP} folds arounds a ZnPP core in the same manner as the cyt b_{562}^{WT}-haem complex.

These results further support the evidence from the previous experiments that the ZnPP cofactor specifically binds within the haem pocket of cyt b_{562}^{ZnPP}. The single coordination/solvation state implied by REES and the gain in α-helical character on ZnPP binding from CD spectrophotometry reinforce the changes observed in absorbance and fluorescence spectral features, that cyt b_{562}^{ZnPP} binds ZnPP similarly to the natural cyt b_{562}^{WT}-haem complex, in accordance with the original design aim.

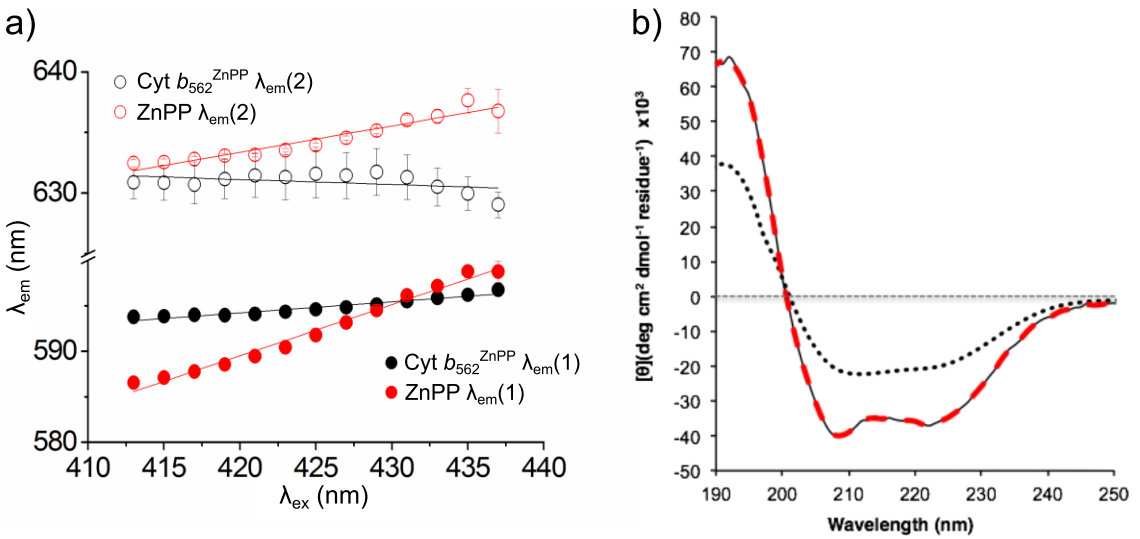


Figure 5.21 ZnPP binding to cyt b_{562}^{ZnPP}. a) REES profile of free ZnPP (red) and protein-bound ZnPP (black) for the two major ZnPP emission peaks at around 590 and 630 nm. b) Far-UV CD spectra of apo-cyt b_{562}^{ZnPP} (black, dotted line) and cyt b_{562}^{ZnPP} bound to ZnPP (red, dashed line), showing partially folded apo-form folds into native four-α-helix bundle on binding ZnPP. REES analysis and figure: Dr. Chris Pudney. CD analysis and figure: Dr. D. Dafydd Jones.

5.2.4.5 Cytochrome b_{562}^{ZnPP} fluorescence in whole cells

For cyt b_{562}^{ZnPP} to be useful as a genetically-encodable, switchable fluorescent marker for optical microscopy, the system needs to work *in vivo*, the protein must express within the target cell and interact and fluoresce with the ZnPP cofactor (either supplied or endogenous) as seen *in vitro*. In addition, ZnPP needs to cross the plasma membrane of the target cells to bind to and "switch on" the fluorescent cyt b_{562}^{ZnPP}. This was tested by incubating *E. coli* cells for six hours with 50 μM ZnPP in 0-5 % concentrations of the cell membrane permeabilising agent dimethyl sulphoxide (DMSO)[301,302]. Cells transformed with pBAD-Cyt b_{562}^{ZnPP} plasmid were compared to negative control *E. coli* cells to ensure fluorescent effects could be attributed to intracellular cyt b_{562}^{ZnPP}, whole cell contents were analysed by SDS-PAGE to confirm protein expression (Figure 5.22a). All cells were washed several times with clean buffer to remove extracellular ZnPP and DMSO, and fluorescence spectra taken. Cells containing cyt b_{562}^{ZnPP} and negative control cells both emitted at 587 nm when excited at 400 nm, indicating intracellular free ZnPP. All cyt b_{562}^{ZnPP}-expressing cells incubated with both ZnPP and DMSO showed prominent fluorescence at a peak of 593 nm when excited at 430 nm, characteristic of cyt b_{562}-bound ZnPP. The cells lacking cyt b_{562}^{ZnPP} showed no significant fluorescence when excited at 430 nm (Figure 5.22b).

These results show that expressed intracellular cyt b_{562}^{ZnPP} is functional in a ZnPP-binding role, and that it modulates the fluorescent properties of ZnPP in the same way as was demonstrated *in vitro*: enhancing ZnPP emission on excitation at the optimal wavelength of 430 nm. It was also demonstrated that percentages of DMSO as low as

1% permeabilise *E. coli* membranes sufficiently to allow ZnPP to enter the cytosol; cells incubated with ZnPP but without DMSO showed no significant intracellular fluorescence from either free or protein-bound ZnPP.

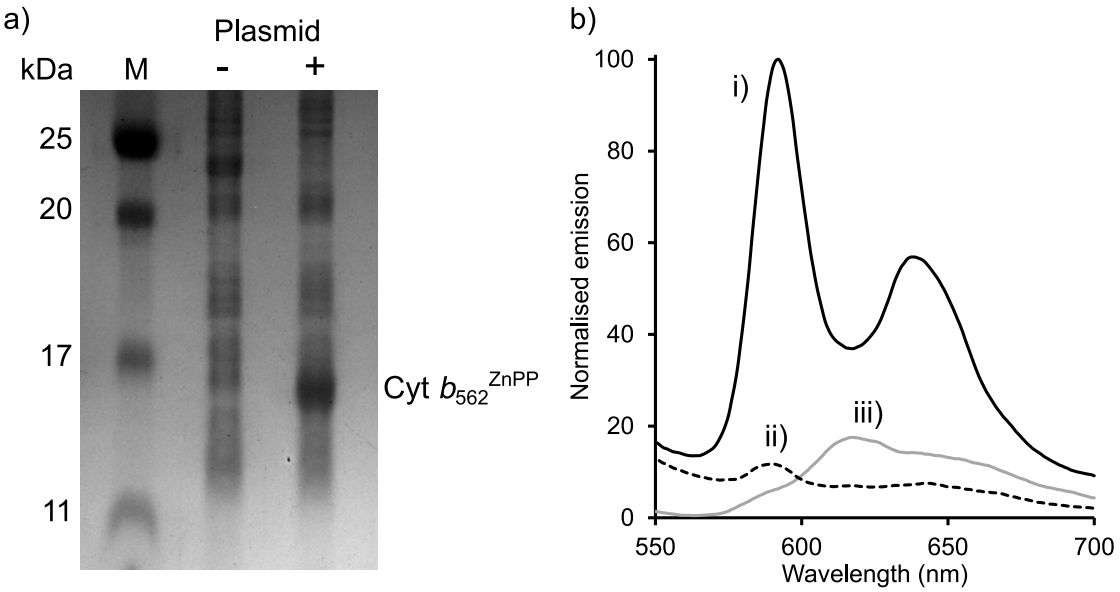


Figure 5.22 Cyt b_{562}^{ZnPP} expression and *in vivo* fluorescence. a) SDS-PAGE gel of whole-cell lysate showing expression of cytochrome b_{562}^{ZnPP} in *E. coli* transformed with the pBAD-Cyt b_{562}^{ZnPP} plasmid. b) Emission spectra from 430 nm excitation of: i) *E. coli* cells expressing cyt b_{562}^{ZnPP} (black line); ii) *E. coli* cells without plasmid transformation (dashed black line); and iii) No cells, just 50 μM ZnPP (grey). Cells were incubated with 50 μM ZnPP and 1% DMSO. Cell samples showed free ZnPP emission peaks at 587 nm when excited at 400 nm indicating intracellular ZnPP; when excited at 430 nm, only ZnPP-Cyt b562^{ZnPP} complexes show emission.

5.2.4.6 Cytochrome b_{562}^{ZnPP} crystallisation

An attempt was made to obtain X-ray crystallographic data for the holo-cyt b_{562}^{ZnPP} complex, to determine the three-dimensional crystal structure so elucidate the specifics of ZnPP binding to the mutant protein. A crystal trial was set up to grow cyt b_{562}^{ZnPP} crystals; the trial was set up in a 96-well vapour diffusion sitting-drop type plate with 1 μL protein and 1 μL buffer per well, with 100 μL buffer in each reservoir; the buffers used were from a JCSG-plus HT96 kit (Molecular Dimensions). The plates were incubated at 20 °C for several weeks and were examined by microscope with UV illumination. Some fluorescent objects in one well appeared to be possible holo-cyt b_{562}^{ZnPP} crystals, these were taken to Diamond Light source by Dr. Pierre Rizkallah and but did not diffract when analysed using the Diamond X-ray beamline (Figure 5.23). No time was left to repeat crystal trials.

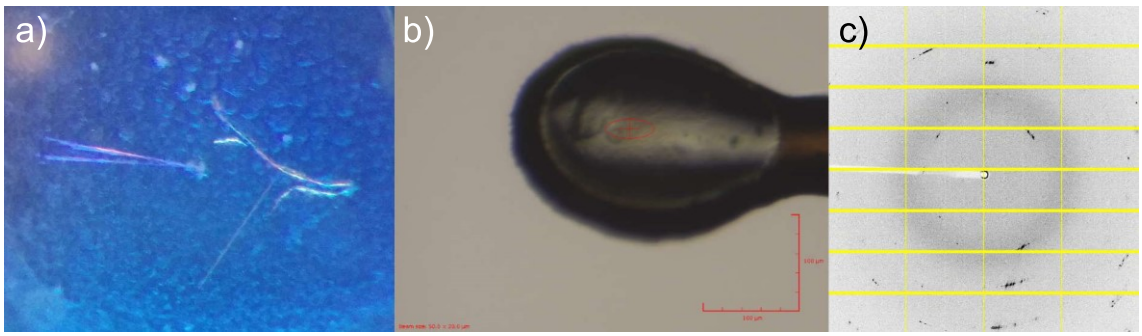


Figure 5.23 Cyt b_{562}^{ZnPP} crystallography. a) Possible crystals of cyt b_{562}^{ZnPP} from crystal screen under UV illumination (left). b) Object in X-ray beamline at Diamond Light Source. c) Lack of diffraction pattern from X-ray beam.

5.3 Conclusions

This work demonstrates that using a computationally guided approach to protein engineering, it was possible to switch cyt b_{562} metalloporphyrin cofactor specificity from the native haem cofactor to an abiotic zinc protoporphyrin. The data recorded for cyt b_{562}^{WT} and cyt b_{562}^{ZnPP} all support the conclusion that the mutations designed *in silico* had the intended effect: improving cyt b_{562} affinity for the fluorescent metalloporphyrin ZnPP while reducing affinity for the native haem cofactor. Binding affinity predictions from the design stage proved to be underestimates, cyt b_{562}^{ZnPP} was predicted to show slightly improved ZnPP binding and reduced haem binding (Table 5.3). The measured values showed that affinity for ZnPP was increased 18-fold, while haem binding affinity was reduced such as to be considered abolished, this constitutes a near-total switch in specificity from a haem cofactor to ZnPP.

Table 5.3 Predicted and measured binding affinity for cyt b_{562}^{WT} and cyt b_{562}^{ZnPP}.

	Predicted affinity (K_D) [a]		**Measured affinity (K_D, ±SE)**	
Variant	ZnPP	Haem	ZnPP	Haem
Cyt b_{562}^{WT}	425 nM	1 nM	626 ± 28 nM	~ 9 nM [b]
Cyt b_{562}^{ZnPP}	110 nM	256 nM	34 ± 10 nM	ND [c]

[a] Converted from AutoDock binding energy (kcal/mol) using $\Delta G/(R \times T) = \ln K_D$

[b] Fitted to published figure with pseudo-R^2 value of 99.6%.

[c] Not determined; no evidence of binding observed.

The change in appearance of absorbance spectral features and dissociation constant were further supported by evidence of a single ZnPP coordination/solvation state from

REES analysis and of a typical cyt b_{562} apo- to holo-protein folding transition from CD analysis.

During preliminary binding assays by Adam McGarrity, prior to this work, a cyt b_{562}^{M7C} single mutation variant was found to bind ZnPP better than cyt b_{562}^{WT}, but worse than cyt b_{562}^{ZnPP} with an estimated K_D of ca. 180 nM. While cysteine is considered a far superior ligand for zinc than methionine[303], the single M7C mutation did not drastically improve ZnPP binding by itself. But when combined with the other four mutations into the cyt b_{562}^{ZnPP} variant, binding affinity was improved more than tenfold over cyt b_{562}^{WT}. Metalloporphyrin coordinating affinity is influenced by both the choice of residue for axial ligands and the immediate environment of the ligand and metal centre. Examination of a predicted structure of cyt b_{562}^{ZnPP} (Figure 5.24) suggests that the mutated L10R residue on the neighbouring turn of α-helix 1 has its side chain guanidine group within approximately 3 Å of the axial ligand M7C thiol. It is possible that the proximity of R10 may contribute to the improved dissociation coefficient of cyt b_{562}^{ZnPP} by abstraction of a proton from C7 to become protonated (NH_2^+), deprotonating the C7 thiolate (S^-) group. Thiolate is a better Zn-coordinating ligand than reduced thiol[304] and many Zn-bound cysteines in natural metalloproteins, such as in zinc finger domains, are stabilised in the thiolate form due to the proximity of neighbouring charged groups[305].

A contribution to the switch in specificity may also be made by the change in electrostatic character of the porphyrin binding pocket. As discussed in section 5.2.1 and **Figure 5.6**, the mutations to cyt b_{562} helix 1 switch the porphyrin binding site from overwhelmingly nonpolar to mostly positive. This may be partly responsible for the drastic loss of haem affinity: haem iron (charge typically 3+) has a more positive charge than ZnPP zinc (2+) and the more positive binding pocket may exert increased repulsion on haem.

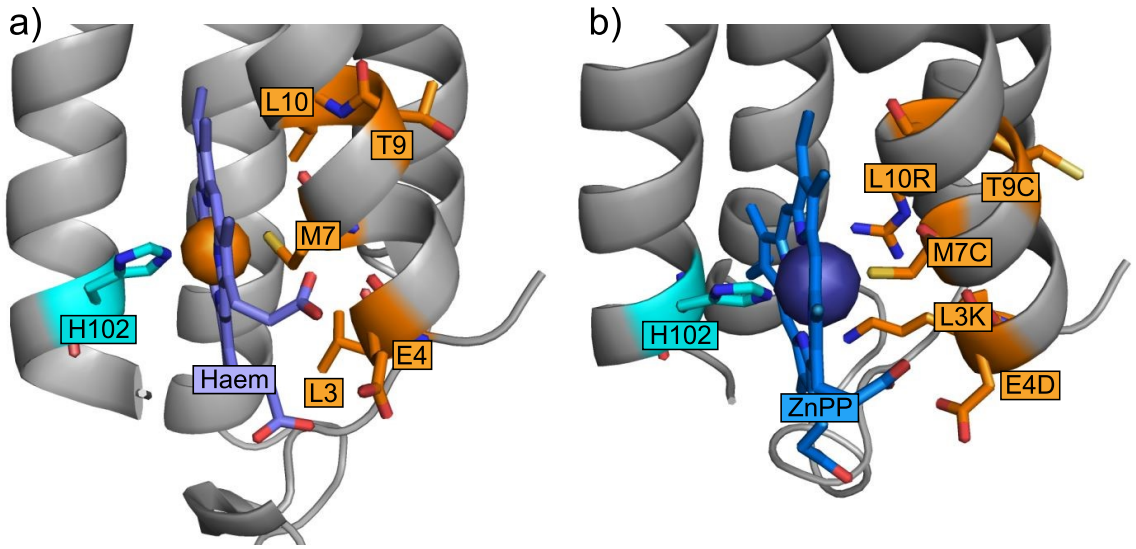


Figure 5.24 a) Haem binding site of cyt b_{562}^{WT}. Axial ligands are H102 and M7, residues mutated for cyt b_{562}^{ZnPP} are shown in orange. b) A predicted structure of ZnPP bound to cyt b_{562}^{ZnPP} showing the proximity (< 3 Å) of R10 to axial ligand C7, possibly influencing the ionisation state of the C7 thiol. Mutated residues shown in orange.

The spectral features of holo-cyt b_{562}^{ZnPP} were characterised and found to be similar to those of free ZnPP and the WT cyt b_{562}-ZnPP complex, but at shifted wavelengths. On binding cyt b_{562}^{ZnPP}, the ZnPP Soret absorbance peak narrowed, redshifted from 406 nm to 426 nm and the intensity ratio of the β and α Q band peaks swapped around. WT cyt b_{562} showed the same effect, but all peaks were 2 nm redshifted relative to cyt b_{562}^{ZnPP}. This small spectral shift between protein variants likely reflects electrostatic changes to the ZnPP coordination environment from the mutations to the axial ligand and its surrounding residues. The fluorescence excitation maximum of free ZnPP was at 410 nm and its two emission maxima were at 588 and 638 nm with an approx. 2:1 intensity ratio. Cyt b_{562}^{ZnPP} bound to ZnPP showed the same features but redshifted to λ_{ex} 425 nm and λ_{em} 597 and 648 nm, with significantly increased emission. The redshift and emission enhancement on protein binding means that, on excitation at 430 nm, binding to cyt b_{562}^{ZnPP} increased ZnPP emission intensity at ca. 590 nm more than tenfold; this is potentially a genetically encodable, switchable fluorescent protein. Features that make cyt b_{562}^{ZnPP} attractive as an expressed fluorescent label are its unusually large Stokes shifts, from its major excitation peak to its two emission peaks, shifts of 172 nm and 223 nm. This wide separation eliminates the possibility of crosstalk between the excitation and emission wavelengths, preventing self-quenching and improving signal-to-noise ratios for fluorescence microscopy[306]. It is also possible to excite holo-cyt b_{562}^{ZnPP} at its β and α Q band maxima at 554 and 593 nm respectively, with lower emission; the ability to excite at multiple, longer wavelengths would be a useful property for fluorescent imaging.

It was demonstrated that intracellularly expressed cyt b_{562}^{ZnPP} modulates the fluorescence of ZnPP *in vivo* in the same way as *in vitro.* It was also found that ZnPP added to an *E. coli* cell sample was able to cross a DMSO-permeabilised plasma membrane and bind to intracellular cyt b_{562}^{ZnPP}. This makes cyt b_{562}^{ZnPP} a promising candidate for use as a genetically encoded fluorescent fusion tag that can be activated as required by the addition of ZnPP to the system; the protein remains in its non-fluorescent apo form until ZnPP is added. As a potential fluorescent labelling protein, cyt b_{562}^{ZnPP} has other useful physical properties: it is extremely robust and resistant to extremes of pH, temperature and high solute concentrations, readily refolding if denatured; it is a very small protein at 12 kDa, half the mass of GFP, which may be useful to minimise its steric impact as a fusion tag; the protein is also extremely soluble, potentially aiding fusion partner proteins to resist aggregation.

For future use of a cyt b_{562}^{ZnPP}-ZnPP system as a fluorescent label, the possibility of off-target binding of ZnPP *in vivo* by other endogenous haemproteins was a concern, since cyt b_{562}^{WT} binds ZnPP with a relatively high affinity so other haemproteins also might. However, the lack of significant fluorescence enhancement observed in the negative controls during the intracellular fluorescence experiment suggests that in *E. coli* at least, this appears not to occur. Further intracellular experiments in mammalian cells are planned to assess this approach.

6 General discussion

6.1 Overview

Protein engineering is still only a few decades old, and the *de novo* construction of molecular systems matching the complexity of naturally-selected protein systems is not currently feasible. The functions of natural systems derive from protein fold, chemistry and dynamics, which are dictated by their primary sequence and perfected by millions of years of evolution. The number of potential backbone conformations in a small protein such as cytochrome b_{562} are on the scale of 10^{100}, which indicates the process of folding cannot be random but must follow a pathway or energy landscape to the native state[307,308]. The prediction of folding and stabilising interactions is too complex for all but the simplest proteins to be designed from scratch. While progress in computational modelling and prediction constantly improves, currently the only practical approach to protein engineering is to use natural proteins as a basis.

This thesis set out to harness the unique properties of four different natural proteins while expanding their utility by interfacing with non-natural materials. The aims were to use a rational design process to effect predictable changes to our proteins of interest, producing new properties. The work had two main threads: 1) Functionalisation of carbon nanotubes with proteins in defined orientations though artificial post-translational modification of a non-natural amino acid, for biosensing purposes. 2) Switching the cofactor specificity of cyt b_{562} from haem to fluorogenic zinc protoporphyrin IX (ZnPP) using computational design and mutagenesis, as a potential new genetically-encodable fluorescent label.

The non-natural amino acid *p*-azido-L-phenylalanine (AzF) was incorporated into three different proteins: BLIP-II, TEM β-lactamase and cytochrome b_{562}. These AzF-containing protein variants were purified to ≥ 99% homogeneity and characterised, then subject to artificial post-translational modification specifically at the mutated residue, using the bioorthogonal chemistry of the phenyl azide side chain. This enabled the proteins to be 1) labelled with a fluorescent dye at the targeted residue through SPAAC click reaction, 2) conjugated to DBCO-pyrene via SPAAC for specifically-oriented binding to a CNT surface, and 3) conjugated directly to the π-bond network of a CNT in defined orientation by photoactivated UV-nitrene cycloaddition.

Using the DBCO-pyrene attachment method, β-lactamase sensing devices were constructed, based on a CNT field-effect transistor (NTFET) design, and functionalised with oriented BLIP-II molecules as recognition elements. These prototype devices were able to detect the antibiotic resistance enzyme TEM β-lactamase at concentrations as

low as 10 nM (30 ng/mL) in buffer of physiological ionic strength (DPBS), and as low as 50 nM (150 ng/mL) in a complex protein mixture (10% serum in DPBS). This represents sensitivity approaching that of traditional western blotting[309] or enzyme-linked immunosorbent assay (ELISA)[310], two of the most common lab-based assays for specific detection of unlabelled proteins. An electronic device based on our NTFET design could form the basis of a fast point-of-care test for β-lactamases in patient samples with minimal on-site equipment requirements. This system could be functionalised with other binding proteins as a platform for detection of analytes from proteins to small molecules.

The porphyrin cofactor specificity of the small haem-binding metalloprotein cyt b_{562} was effectively switched from haem to the zinc-containing fluorogenic haem analogue ZnPP. This was achieved through computational design, which suggested multiple amino acid mutations to improve cyt b_{562} ZnPP-binding function. A cyt b_{562} variant with these mutations was made and characterised, revealing almost complete abolition of haem binding with an 18-fold improvement in ZnPP affinity. The ZnPP fluorescence-enhancing properties of cyt b_{562}^{ZnPP} were demonstrated both in solution and in live *E. coli* cells, indicating that cyt b_{562}^{ZnPP} variant may fulfil a useful role as a genetically-encoded fluorescence labelling tag that can be activated at will by addition of the ZnPP cofactor. Cyt b_{562}^{ZnPP} is around half the size of the common fluorescent protein GFP, and extremely soluble and robust, all properties desirable in a fluorescent fusion tag for fluorescence microscopy.

6.2 Computational design of engineered proteins

Computational tools are essential to almost every aspect of protein engineering, from visualising the 3D protein structure to designing mutagenic PCR primers. In this thesis, 3D structural models of protein complexes were examined using PyMOL to rationally select residues for mutation to AzF for CNT conjugation (chapters 3 and 4). Residues were chosen as conjugation points based on their surface accessibility and location relative to binding and catalytic sites. This allowed CNTs to be functionalised with populations of proteins all in approximately the same orientation, which is a key requirement for measuring a consistent signal for the whole population. The online molecular simulation tool Phyre2 (www.sbg.bio.ic.ac.uk/~phyre2) proved useful for estimating the effect of AzF mutations on protein fold, allowing protein variants to be screened for major structural disruptions before production. ClusPro docking simulation (https://cluspro.bu.edu) was used to predict and rank probable geometries for formation of cyt b_{562} dimers via disulphide bridges and to predict TEM binding orientation with BLIP-II AzF variants.

The mutations that caused a shift in cyt b_{562} porphyrin cofactor specificity from haem to ZnPP (chapter 5) were designed by Adam McGarrity using several *in silico* tools. These included the ZincExplorer web app (http://systbio.cau.edu.cn/ZincExplorer), which was used to suggest improved axial ligands for Zn. To optimise surrounding residues to the axial ligand M7, GROMACS was used for molecular dynamics simulation, GAMESS-US for geometry optimisation and AutoDock and LigandDock for porphyrin docking simulation. The cyt b_{562}^{ZnPP} variant was designed entirely using advanced computational tools and this would not have been feasible using traditional mutagenesis approaches without considerable amounts of trial and error.

6.3 Genetically-incorporated non-natural amino acids: expanding the chemical repertoire

The inclusion of nnAAs in engineered proteins by genetic code expansion allows proteins to be expressed by cells *in situ*, carrying chemical groups that do not occur in the standard proteinogenic amino acids. The limited chemical variety available using the 20 canonical amino acids can be an obstacle to achieving specific conjugation or labelling. The most common chemistries used for artificial modification of canonical amino acid side chains are succinimide-amine reactions to Lys residues and maleimide-thiol reactions to Cys residues. However, the vast majority of natural proteins have multiple native Lys residues, and at least one native Cys[311]. This prevents introduction of these residues by mutagenesis as specific reactive handles, except in rare cases such as cyt b_{562} that naturally have no Cys residues. While natural systems are highly specific and controlled through enzyme-mediated signal pathways, this level of subtle control is far beyond current capabilities. To enable specific targeting or modification of individual residues in recombinant proteins, genetically encodable nnAA systems can be used.

More than 200 nnAAs have been developed to date for genetic incorporation using the amber suppression co-translation approach described in chapter 1.1.3, in eukaryotic, prokaryotic and cell-free expression systems[62,312]. This range of available nnAAs greatly expands the chemical repertoire available for protein engineering and can be incorporated into potentially any location of any protein. Some purposes for engineering nnAAs into a protein include: enabling bioorthogonal reactions for labelling or conjugating via a specific targeted residue (e.g. with AzF in this thesis); heavy element labelling at a defined location to aid structure determination by x-ray crystallography, (e.g. with *p*-iodo-L-phenylalanine[63]); or site-specific labelling with a fluorescent amino acid (e.g. L-(7-hydroxycoumarin-4-yl)ethylglycine[313]) for spectroscopy.

The current approaches to incorporating nnAAs has some shortcomings, notably variable efficiency of nnAA incorporation in different proteins, and at different locations in the same protein. The protein release factor 1 (RF1) competes with the engineered $tRNA_{CUA}$ to terminate peptide elongation instead of incorporating the nnAA in response to a TAG/UAG codon, resulting in reduced yield relative to wild-type proteins or mutants without nnAAs, down to as little as 10% yield[73]. *E. coli* expression strains have been developed in which all known TAG amber stop codons in the genome have been replaced with TAA ochre stop codons, allowing deletion of RF1 for improved incorporation of non-natural amino acids[77]. Transfer RNAs featuring four-base anticodons have also been produced to avoid competition with RF1 and allow multiple different nnAAs to be incorporated in the same protein[314].

Other approaches exist for inclusion of nnAAs in recombinant proteins, including the original strategy of expressing proteins in an *E. coli* strain auxotrophic for a specific amino acid, in media deficient in that essential amino acid but supplemented with a non-natural analogue, to force translation of the analogue by native cell machinery[315]. This approach suffers from lower efficiency than amber suppression and is now rarely used. Another method is expressed protein ligation (EPL) where a synthesised peptide containing the nnAA is ligated to an expressed protein after production[316,317]. This has the disadvantage of the expense and time delay associated with solid-phase peptide synthesis and limits the position of the nnAA to within 50-70 residues of the N- or C-termini, the maximum current length of peptide synthesis[318]. The best currently available method of nnAA incorporation is using amber suppression.

6.4 Non-natural post-translational modification with AzF

AzF is one of the most widely-used genetically incorporated nnAAs due to its range of physical and chemical properties; it has been added to the genetic codes of bacteria[79], yeasts[259,319], mammalian cell lines[260] and animals[320,321]. The phenyl azide reactive side chain is infrared-active has proven useful as a probe for IR spectroscopy, absorbing at wavenumber 2100 cm^{-1} [322]. The phenyl azide of AzF can also undergo several specific bioorthogonal reaction routes, including cycloaddition to an alkyne moiety, as used in chapter 3 where a reaction efficiency of ca. 50-60% was achieved using SPAAC click chemistry; and a range of UV light-mediated phenylnitrene radical reactions, as detailed in chapter 4.1. These specific reactions make AzF a good candidate for targeted non-natural post-translational modifications in an engineered protein.

In this thesis, SPAAC was used to modify proteins post-translationally, including labelling with fluorescent dyes and conjugation of a DBCO-pyrene linker for interfacing with CNTs. Using rationally-designed AzF substitutions and SPAAC click chemistry, engineered

proteins can be easily conjugated to practically any surface or structure, in a predetermined orientation. A recent project here in the Jones lab saw engineered green fluorescent proteins combined into designed artificial dimers, with one GFP in each dimer having an AzF reactive group and the other GFP having a cyclooctyne-lysine (SCO) nnAA. These proteins were joined via SPAAC to form a covalently-linked dimer that demonstrated functional synergy between the two GFP chromophores[69]. Many azide-carrying molecules are available for conjugation to AzF-bearing proteins from companies such as Click Chemistry Tools. These include fluorophores, biotinylation/PEGylation reagents or adapters such as DBCO-amine and DBCO-sulfhydryl that can be used to make customised molecules for specific conjugation to proteins, as with the DBCO-pyrene CNT linker we made in chapter 3 of this thesis. The photochemical properties of AzF were used in chapter 4 of this thesis to functionalise CNTs directly with populations of several different proteins. Other previous uses of AzF phenylnitrene photochemistry were demonstrated by the Jones lab in a paper detailing the creation of two photo-controllable GFP variants containing engineered AzF residues close to the GFP chromophore. The fluorescent function of one non-fluorescent variant was able to be switched on by UV irradiation, while the other variant could be inactivated by UV irradiation. These are just a few examples of the wide possibilities of genetically-encoded AzF residues for manipulating and studying protein structure and function by artificial post-translational modification.

Drawbacks to AzF incorporation in protein engineering include reduced protein yield compared to WT, as mentioned in 6.3, and the necessity to keep AzF and any AzF-containing proteins in dark conditions at all times to protect the phenylnitrene group from photolysis to phenyl amine. Working with AzF requires samples and equipment to be kept in a darkroom, with samples and chromatography columns foil-wrapped as far as possible to minimise light exposure. AzF is also an expensive reagent and must be dissolved at relatively high concentration (1 mM, 206 mg/L) in growth media to enable co-translation. AzF was impossible to source for much of 2019 after publication of a report by Richardson *et al*. (2018)[231] on its previously undocumented explosive decomposition. The reaction rate of the SPAAC reaction is relatively low compared to other labelling chemistries such maleimide-thiol reactions, which contributes to the relatively low reaction efficiency but is made up for by the bioorthogonal specificity of SPAAC.

6.5 Integration of proteins with carbon nanotubes for biosensing

Biosensing field-effect transistors (Bio-FETs) have been traditionally based on ion sensitive FET (ISFET) designs with a semiconducting silicon transistor channel in which conductance is influenced by the analyte, while later designs used a silicon nanowire (SiNW) channel. S-SWCNTs have superior properties for this application with smaller diameters than SiNWs, and all charge carriers at the surface of the hollow CNT contributing to the extreme sensitivity of CNTs to their chemical environment. CNTs also show high mechanical strength, thermal and chemical stability, and ease of functionalisation with receptors, binding proteins or other recognition elements.

This thesis used two distinct methods to interface proteins with CNT surfaces, both using genetically incorporated AzF residues: adsorption via a DBCO-pyrene linker molecule and direct UV-nitrene photoattachment. Both methods yielded CNTs functionalised with a similar dispersion of engineered proteins, ca. 20 to 30 proteins per µm of CNT, and both methods showed no evidence of non-specific binding of proteins to the CNT. It is not clear at this point which functionalisation strategy is superior as only the DBCO-pyrene adsorption method was used for construction of β-lactamase biosensors. The UV-nitrene photoattachment method shows potential to be superior as less pre-treatment of CNT devices is required, and the aziridine bond formed is extremely short, possibly improving the intimacy of β-lactamase analyte interaction with the CNT.

Other means of functionalising CNTs include acid oxidation, which provides carboxylic acid groups at the CNT surface that could easily be linked to AzF-bearing proteins via pre-functionalisation with DBCO-amine. Acid oxidation is known to introduce vacancy defects in the CNT sidewall, altering conduction properties[233], but whether these defects would be detrimental to the function of our protein-functionalised NTFET is not known. DNA wrapping is a common means of solubilising CNTs in aqueous solution, as the DNA bases adsorb to CNT sidewalls by π-π stacking. Engineered proteins containing AzF could be conjugated to commercially sold DBCO-linked DNA oligonucleotides (IDT) before wrapping. This has been done previously by our collaborator at Queen Mary, University of London, Dr. Mark Freeley, with cyt b_{562}-DNA constructs self-assembled on CNTs[323]. However, this results in a low coverage of proteins (< 5 proteins per µm) and the DNA wrapping is likely to shield the CNT from the electrostatic effects of bound proteins.

6.6 Future work

Further experiments are planned for BLIP-II-functionalised NTFET devices, to optimise their signal response and analyse the lowest detectable concentration of analyte (TEM β-lactamase). It is also intended to characterise the signal response of the devices to other class A serine β-lactamases, including CTX-M, KPC and Bla1. It is thought that the different surface charges of these other β-lactamases may elicit different responses, as was observed when different surfaces of TEM were presented to the NTFET device (chapter 3.3.5.2). The aim is to be able discern the β-lactamase analyte type from its signal response, which would be of major interest as an analytical/diagnostic tool, and possibly form the basis of a more general technology for sensing protein-protein interactions. During the writing of this thesis, the global pandemic of SARS-CoV-2 virus occurred, suggesting an ideal application for this kind of rapid protein recognition device. The Jones lab is currently sourcing the plasmids necessary to integrate human ACE2 receptor protein into an NTFET sensing device in a range of orientations, via genetically encoded AzF residues. ACE2 is the receptor target for the SARS-CoV-2 virus spike protein antigen and binds the virus with high affinity (K_D: 44 nM)[324]. ACE2 is also a relatively small protein, at ca. 5.5 × 5.5 × 7.5 nm, and is likely to bind the spike protein within range of the NTFET to electrostatically influence conductance, i.e. within the Debye length.

No protein crystals were yielded in the crystal screen that was performed with cyt b_{562}^{ZnPP} and due to time constraints and protein shortage, the screen could not be repeated with different conditions. To resolve a crystal structure of cyt b_{562}^{ZnPP} would be extremely useful, enabling comparison of the predicted results from the computational modelling with the actual crystal structure of the holo-protein. Future directions for cyt b_{562}^{ZnPP} should include its expression as a fusion with a target protein, to assess its effectiveness as a small fluorescence tag that can be selectively activated by addition of ZnPP. *In vivo* fluorescence microscopy imaging was attempted in HEK293 human cells, but conclusive evidence of expression was not shown before the end of this PhD.

Several opportunities and ideas arose during this work that could not be realised due to time constraints, and would make interesting future research:

The photochemical crosslinking properties of AzF were to be exploited by attempting to trap proteins associated with cyt b_{562} *in vivo*. Cyt b_{562} is a native *E. coli* protein and despite its being widely-used in protein engineering, little is known about its natural function and interactions. It was hoped to overexpress cyt b_{562}^{SA} or cyt b_{562}^{LA} in *E. coli* then irradiate the cells with UV-B light to activate the azide groups to phenylnitrene. As the nitrene can insert into adjacent N-H and C-H bonds to form permanent covalent links, it was hoped

to trap cyt b_{562} associated proteins for purification, analysis and possible identification and characterisation. This can hopefully be realised.

A research thread that could not be fully followed up was the end-functionalisation of CNTs with DBCO reactive groups, for targeted conjugation of one protein per CNT. Preliminary AFM data suggested successful 1:1 end-linking of cyt b_{562}^{SA} dimers to CNT ends by SPAAC click chemistry (Figure 6.1a). Interestingly, the AFM data also showed evidence of cyt b_{562} junctions formed between two CNTs, apparently formed when CNTs become conjugated to both of a dimer's AzF groups (Figure 6.1b). Protein-based CNT junctions have potential for integrating individual proteins as highly sensitive components in electronic circuits, such as sensors or transistors. Single-molecule protein transistors have been fabricated from individual antibodies bridging a gap between gold nanoparticles[325], but these rely on traditional antibody-antigen affinity to self-assemble. A predefined self-assembling CNT-protein-CNT junction would be useful for molecular electronics, allowing practically any protein to form a CNT junction by engineering two AzF residues into the protein at the desired sites. A protein-CNT junction incorporating cyt b_{562}^{ZnPP} would be an interesting next step, providing a direct and self-assembling way to integrate the photoelectronic properties of ZnPP into electronic circuits as a single-molecule bioelectronic component. ZnPP derivatives have served as photosensitisers in solar cells with efficiencies of up to 12%[326] and as photodetectors with wide spectral range (UV-A to near-IR) and high sensitivity[327,328]. Cyt b_{562} could be engineered to act as a scaffold and organising element for the bottom-up assembly of CNT-protein-porphyrin arrays for molecular-scale solar power harvesting or photodetection.

Finally, it was also intended to end-functionalise CNTs with BLIP-II and TEM AzF variants, to try to create self-assembling CNT junctions held together by the high-affinity BLIP-II – TEM binding interaction only. This was tried several times, but no junctions were observed similar to those formed by the cyt b_{562}^{SA} dimers. This remains an interesting possibility and should be repeated with varying conditions to see if non-covalent junctions can be achieved.

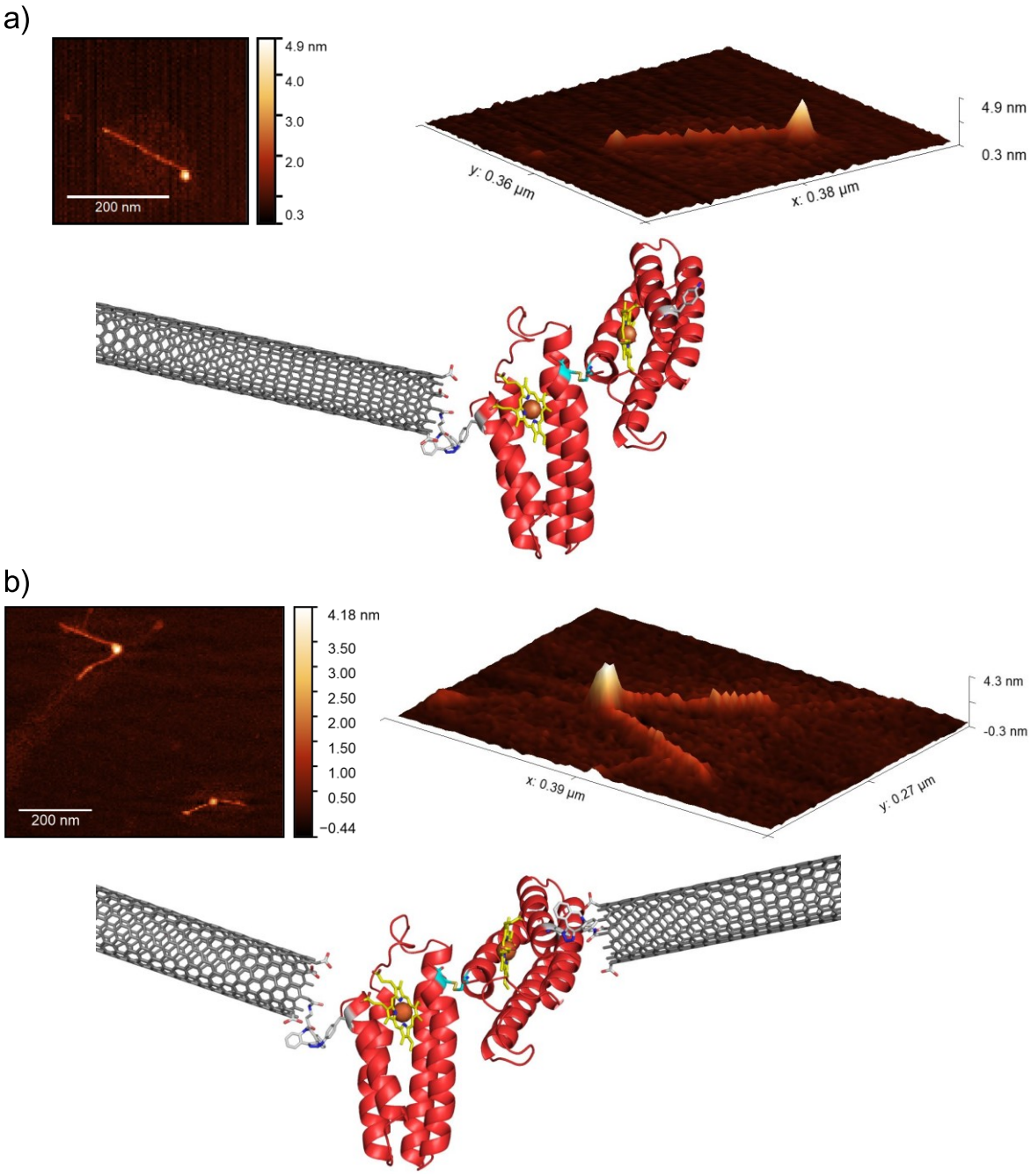


Figure 6.1 CNTs end-functionalised with cyt b_{562}^{SA} dimers, preliminary findings. AFM data in 2D and 3D and renders from a hypothetical structure. a) Single CNT decorated with one cyt b_{562}^{SA} dimer conjugated at its end. b) Two CNTs joined by a cyt b_{562}^{SA} dimer acting as a protein junction.

7 Bibliography

1. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Biochemistry, 7th Edition*. *W H Freeman* (2012).
2. Rosenthaler, L. Durch Enzyme bewirkte asymmetrische Synthesen. *Biochemische Zeitschrift* **14**, 238–253 (1908).
3. Lutz, S. & Iamurri, S. M. Protein engineering: Past, present, and future. in *Methods in Molecular Biology* vol. 1685 1–12 (Humana Press Inc., 2018).
4. Estell, D. A., Graycar, T. P. & Wells, J. A. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *Journal of Biological Chemistry* **260**, 6518–6521 (1985).
5. Saunders, C. T. & Baker, D. Evaluation of structural and evolutionary contributions to deleterious mutation prediction. *Journal of Molecular Biology* **322**, 891–901 (2002).
6. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Research* **2**, 28–33 (1992).
7. Wong, T. S., Tee, K. L., Hauer, B. & Schwanenberg, U. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Research* **32**, (2004).
8. Moore, J. C. & Arnold, F. H. Directed Evolution of a Para-Nitrobenzyl Esterase for Aqueous-Organic Solvents. *Nature Biotechnology* **14**, 458–467 (1996).
9. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994).
10. Ness, J. E., Kim, S., Gottman, A., Pak, R., Krebber, A., Borchert, T. V., Govindarajan, S., Mundorff, E. C. & Minshull, J. Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently. *Nature Biotechnology* **20**, 1251–1255 (2002).
11. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version 1.3*. (2010).
12. Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **60**, 2288–2294 (2004).
13. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
14. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D. & Vajda, S. The ClusPro web server for protein-protein docking. *Nature Protocols* **12**, 255–278 (2017).
15. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* **30**, 2785–2791 (2009).
16. Cardiff University. Advanced research computing. https://www.cardiff.ac.uk/advanced-research-computing/about-us/our-supercomputers (2020).
17. Korendovych, I. V. & DeGrado, W. F. De novo protein design, a retrospective. *Quarterly Reviews of Biophysics* **53**, (2020).
18. Kranaster, R. & Marx, A. Engineered DNA Polymerases in Biotechnology. *ChemBioChem*

11, 2077–2084 (2010).

19. Frenzel, A., Hust, M. & Schirrmann, T. Expression of recombinant antibodies. *Frontiers in Immunology* **4**, 217 (2013).
20. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. & Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology* **21**, 86–89 (2003).
21. Los, G. V. *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology* **3**, 373–382 (2008).
22. Sauna, Z. E., Lagassé, H. A. D., Alexaki, A., Simhadri, V. L., Katagiri, N. H., Jankowski, W. & Kimchi-Sarfaty, C. Recent advances in (therapeutic protein) drug development. *F1000Research* **6**, (2017).
23. King, D. J. *Applications And Engineering Of Monoclonal Antibodies*. (CRC Press, 2014).
24. Zhang, Y. & Gladyshev, V. N. General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni, and Se. *Journal of Biological Chemistry* **285**, 3393–3405 (2010).
25. Mousa, R., Notis Dardashti, R. & Metanis, N. Selenium and Selenocysteine in Protein Chemistry. *Angewandte Chemie - International Edition* **56**, 15818–15827 (2017).
26. Steinbrenner, H., Speckmann, B. & Klotz, L. O. Selenoproteins: Antioxidant selenoenzymes and beyond. *Archives of Biochemistry and Biophysics* **595**, 113–119 (2016).
27. Schmidt, R. L. & Simonović, M. Synthesis and decoding of selenocysteine and human health. *Croatian Medical Journal* **53**, 535 (2012).
28. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296**, 1459–62 (2002).
29. Brugère, J.-F., Atkins, J. F., O'Toole, P. W. & Borrel, G. Pyrrolysine in archaea: a 22nd amino acid encoded through a genetic code expansion. *Emerging Topics in Life Sciences* **2**, 607–618 (2018).
30. Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Breitwieser, F. P., Chang, Y. C., Madugundu, A. K., Pandey, A. & Salzberg, S. L. CHESS: A new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biology* **19**, 208 (2018).
31. Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M. C. & Caracausi, M. Human protein-coding genes and gene feature statistics in 2019. *BMC Research Notes* **12**, 315 (2019).
32. Smith, L. M., Kelleher, N. L. & Proteomics, T. C. for T. D. Proteoform: a single term describing protein complexity. *Nature methods* **10**, 186 (2013).
33. Ponomarenko, E. A., Poverennaya, E. V, Ilgisonis, E. V, Pyatnitskiy, M. A., Kopylov, A. T., Zgoda, V. G., Lisitsa, A. V & Archakov, A. I. The Size of the Human Proteome: The Width and Depth. *International Journal of Analytical Chemistry* **2016**, 7436849 (2016).
34. Aebersold, R. *et al.* How many human proteoforms are there? *Nature chemical biology* **14**, 206–214 (2018).

35. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **24**, 79–88 (2006).
36. Lim, D., Park, H. U., De Castro, L., Kang, S. G., Lee, H. S., Jensen, S., Lee, K. J. & Strynadka, N. C. J. Crystal structure and kinetic analysis of β-lactamase inhibitor protein-II in complex with TEM-1 β-lactamase. *Nature Structural Biology* **8**, 848–852 (2001).
37. Chen, Y., Delmas, J., Sirot, J., Shoichet, B. & Bonnet, R. Atomic resolution structures of CTX-M β-lactamases: Extended spectrum activities from increased mobility and decreased stability. *Journal of Molecular Biology* **348**, 349–362 (2005).
38. Lederer, F., Glatigny, A., Bethge, P. H., Bellamy, H. D. & Mathews, F. S. Improvement of the 2.5 Å resolution model of cytochrome b562 by redetermining the primary structure and using molecular graphics. *Journal of Molecular Biology* **148**, 427–448 (1981).
39. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: Nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **4**, 565–583 (2012).
40. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
41. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Computational Biology* **11**, e1004049 (2015).
42. Loeb, K. R. & Haas, A. L. The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins. *Journal of Biological Chemistry* **267**, 7806–7813 (1992).
43. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. Structure of ubiquitin refined at 1.8Åresolution. *Journal of Molecular Biology* **194**, 531–544 (1987).
44. Komander, D. & Rape, M. The Ubiquitin Code. *Annual Review of Biochemistry* **81**, 203–229 (2012).
45. Mann, M. & Jensen, O. N. Proteomic analysis of post-translational modifications. *Nature Biotechnology* **21**, 255–261 (2003).
46. Peisley, A., Wu, B., Xu, H., Chen, Z. J. & Hur, S. Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I. *Nature* **508**, 110–114 (2014).
47. Black, C. B., Huang, H. W. & Cowan, J. A. Biological coordination chemistry of magnesium, sodium, and potassium ions. Protein and nucleotide binding sites. *Coordination Chemistry Reviews* **135**–**136**, 165–202 (1994).
48. Stanyon, H. F., Cong, X., Chen, Y., Shahidullah, N., Rossetti, G., Dreyer, J., Papamokos, G., Carloni, P. & Viles, J. H. Developing predictive rules for coordination geometry from visible circular dichroism of copper(II) and nickel(II) ions in histidine and amide main-chain complexes. *FEBS Journal* **281**, 3945–3954 (2014).
49. Shimazaki, Y., Takani, M. & Yamauchi, O. Metal complexes of amino acids and amino acid side chain groups. Structures and properties. *Dalton Transactions* 7854–7869 (2009) doi:10.1039/b905871k.
50. Harding, M. M. The architecture of metal coordination groups in proteins. *Acta*

Crystallographica Section D: Biological Crystallography **60**, 849–859 (2004).

51. Harding, M. M. Geometry of metal-ligand interactions in proteins. *Acta Crystallographica Section D: Biological Crystallography* **57**, 401–411 (2001).
52. Weller, M., Overton, T., Rourke, J. & Armstrong, F. *Inorganic Chemistry*. (Oxford University Press, 2014).
53. Toseland, C. P. Fluorescent labeling and modification of proteins. *Journal of chemical biology* **6**, 85–95 (2013).
54. Qi, D., Tann, C. M., Haring, D. & Distefano, M. D. Generation of new enzymes via covalent modification of existing proteins. *Chemical Reviews* **101**, 3081–3111 (2001).
55. Abuchowski, A., McCoy, J. R., Palczuk, N. C., van Es, T. & Davis, F. F. Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *Journal of Biological Chemistry* **252**, 3582–3586 (1977).
56. Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational modification of histones. *Current Opinion in Chemical Biology* **45**, 35–47 (2018).
57. Oliveira, S. F., Bisker, G., Bakh, N. A., Gibbs, S. L., Landry, M. P. & Strano, M. S. Protein functionalized carbon nanomaterials for biomedical applications. *Carbon* **95**, 767–779 (2015).
58. UniProt. UniProtKB/TrEMBL protein database release 2019_09 statistics. https://www.ebi.ac.uk/uniprot/TrEMBLstats (2019).
59. Kim, Y., Ho, S. O., Gassman, N. R., Korlann, Y., Landorf, E. V., Collart, F. R. & Weiss, S. Efficient site-specific labeling of proteins via cysteines. *Bioconjugate Chemistry* **19**, 786–791 (2008).
60. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
61. Reynolds, N. M., Vargas-Rodriguez, O., Söll, D. & Crnković, A. The central role of tRNA in genetic code expansion. *Biochimica et Biophysica Acta - General Subjects* **1861**, 3001–3008 (2017).
62. Vargas-Rodriguez, O., Sevostyanova, A., Söll, D. & Crnković, A. Upgrading aminoacyl-tRNA synthetases for genetic code expansion. *Current Opinion in Chemical Biology* **46**, 115–122 (2018).
63. Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G. & Schultz, P. G. The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nature Biotechnology* **22**, 1297–1301 (2004).
64. Seitchik, J. L., Peeler, J. C., Taylor, M. T., Blackman, M. L., Rhoads, T. W., Cooley, R. B., Refakis, C., Fox, J. M. & Mehl, R. A. Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes. *Journal of the American Chemical Society* **134**, 2898–2901 (2012).
65. Plass, T., Milles, S., Koehler, C., Schultz, C. & Lemke, E. A. Genetically encoded copper-free click chemistry. *Angew Chem Int Ed Engl* **50**, 3878–3881 (2011).
66. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a

photocrosslinking amino acid to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 11020–11024 (2002).

67. Tang, Y. & Tirrell, D. A. Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host [13]. *Journal of the American Chemical Society* **123**, 11089–11090 (2001).
68. Seo, M. H., Han, J., Jin, Z., Lee, D. W., Park, H. S. & Kim, H. S. Controlled and oriented immobilization of protein by site-specific incorporation of unnatural amino acid. *Analytical Chemistry* **83**, 2841–2845 (2011).
69. Worthy, H. L., Auhim, H. S., Jamieson, W. D., Pope, J. R., Wall, A., Batchelor, R., Johnson, R. L., Watkins, D. W., Rizkallah, P., Castell, O. K. & Jones, D. D. Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry. *Communications Chemistry* **2**, 83 (2019).
70. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences of the United States of America* **103**, 9482–9487 (2006).
71. Smolskaya, S. & Andreev, Y. A. Site-specific incorporation of unnatural amino acids into escherichia coli recombinant protein: Methodology development and recent achievement. *Biomolecules* **9**, 255 (2019).
72. Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins [16]. *Journal of the American Chemical Society* **122**, 5010–5011 (2000).
73. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
74. Anderson, J. C. & Schultz, P. G. Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression. *Biochemistry* **42**, 9598–9608 (2003).
75. Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B. L., Mori, H. & Horiuchi, T. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Molecular Systems Biology* **2**, (2006).
76. Loscha, K. V, Herlt, A. J., Qi, R. H., Huber, T., Ozawa, K. & Otting, G. Multiple-Site Labeling of Proteins with Unnatural Amino Acids. *Angewandte Chemie-International Edition* **51**, 2243–2246 (2012).
77. Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M. & Isaacs, F. J. Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013).
78. Kobayashi, T., Nurekil, 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. *Nature Structural Biology* **10**, 425–432 (2003).
79. Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *Journal of the American*

Chemical Society **124**, 9026–9027 (2002).

80. Ye, S., Zaitseva, E., Caltabiano, G., Schertler, G. F. X., Sakmar, T. P., Deupi, X. & Vogel, R. Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* **464**, 1386–1389 (2010).
81. Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie International Edition* **40**, 2004–2021 (2001).
82. Gaetke, L. M., Chow-Johnson, H. S. & Chow, C. K. Copper: toxicological relevance and mechanisms. *Archives of Toxicology* **88**, 1929–1938 (2014).
83. Gritsan, N. & Platz, M. Photochemistry of Azides: The Azide/Nitrene Interface. in *Organic Azides: Syntheses and Applications* (ed. Banert, K.) 311–372 (John Wiley & Sons, Ltd., 2010).
84. Marth, G., Hartley, A. M., Reddington, S. C., Sargisson, L. L., Parcollet, M., Dunn, K. E., Jones, D. D. & Stulz, E. Precision Templated Bottom-Up Multiprotein Nanoassembly through Defined Click Chemistry Linkage to DNA. *ACS Nano* **11**, 5003–5010 (2017).
85. Freeley, M., Worthy, H. L., Ahmed, R., Bowen, B., Watkins, D., Macdonald, J. E., Zheng, M., Jones, D. D. & Palma, M. Site-Specific One-to-One Click Coupling of Single Proteins to Individual Carbon Nanotubes: A Single-Molecule Approach. *Journal of the American Chemical Society* **139**, 17834–17840 (2017).
86. Morris, J. L., Reddington, S. C., Murphy, D. M., Jones, D. D., Platts, J. A. & Tippmann, E. M. Aryl azide photochemistry in defined protein environments. *Org Lett* **15**, 728–731 (2013).
87. Reddington, S. C., Rizkallah, P. J., Watson, P. D., Pearson, R., Tippmann, E. M. & Jones, D. D. Different photochemical events of a genetically encoded phenyl azide define and modulate GFP fluorescence. *Angew Chem Int Ed Engl* **52**, 5974–5977 (2013).
88. Reddington, S. C., Driezis, S., Hartley, A. M., Watson, P. D., Rizkallah, P. J. & Jones, D. D. Genetically encoded phenyl azide photochemistry drives positive and negative functional modulation of a red fluorescent protein. *RSC Advances* **5**, 77734–77738 (2015).
89. Reddington, S. C., Baldwin, A. J., Thompson, R., Brancale, A., Tippmann, E. M. & Jones, D. D. Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chemical Science* **6**, 1159–1166 (2015).
90. Zaki, A. J., Hartley, A. M., Reddington, S. C., Thomas, S. K., Watson, P., Hayes, A., Moskalenko, A. V., Craciun, M. F., Macdonald, J. E., Jones, D. D. & Elliott, M. Defined covalent assembly of protein molecules on graphene using a genetically encoded photochemical reaction handle. *RSC Advances* **8**, 5768–5775 (2018).
91. Nature. nan'o·tech·nol'o·gy n. *Nature Nanotechnology* **1**, 10 (2006).
92. Knowles, T. P. J., Oppenheim, T. W., Buell, A. K., Chirgadze, D. Y. & Welland, M. E. Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nature nanotechnology* **5**, 204–7 (2010).
93. Graugnard, E., Hughes, W. L., Jungmann, R., Kostiainen, M. A. & Linko, V. Nanometrology and super-resolution imaging with DNA. *MRS Bulletin* **42**, 951–958

(2017).

94. Karajanagi, S. S., Yang, H., Asuri, P., Sellitto, E., Dordick, J. S. & Kane, R. S. Protein-assisted solubilization of single-walled carbon nanotubes. *Langmuir* **22**, 1392–1395 (2006).

95. Kam, N. W. S., Jan, E. & Kotov, N. A. Electrical stimulation of neural stem cells mediated by humanized carbon nanotube composite made with extracellular matrix protein. *Nano Letters* **9**, 273–278 (2009).

96. Oliverio, M., Perotto, S., Messina, G. C., Lovato, L. & De Angelis, F. Chemical Functionalization of Plasmonic Surface Biosensors: A Tutorial Review on Issues, Strategies, and Costs. *ACS Applied Materials and Interfaces* **9**, 29394–29411 (2017).

97. Birkert, O., Haake, H. M., Schütz, A., Mack, J., Brecht, A., Jung, G. & Gauglitz, G. A streptavidin surface on planar glass substrates for the detection of biomolecular interaction. *Analytical Biochemistry* **282**, 200–208 (2000).

98. Kumeria, T., Kurkuri, M. D., Diener, K. R., Parkinson, L. & Losic, D. Label-free reflectometric interference microchip biosensor based on nanoporous alumina for detection of circulating tumour cells. *Biosensors and Bioelectronics* **35**, 167–173 (2012).

99. Khan, N. I. & Song, E. Lab-on-a-chip systems for aptamer-based biosensing. *Micromachines* **11**, 1–30 (2020).

100. Xu, J. J., Luo, X. L. & Chen, H. Y. Analytical aspects of FET-based biosensors. *Frontiers in Bioscience* **10**, 420–430 (2005).

101. Syu, Y.-C., Hsu, W.-E. & Lin, C.-T. Review—Field-Effect Transistor Biosensing: Devices and Clinical Applications. *ECS Journal of Solid State Science and Technology* **7**, Q3196–Q3207 (2018).

102. Justino, C. I. L., Freitas, A. C., Amaral, J. P., Rocha-Santos, T. A. P., Cardoso, S. & Duarte, A. C. Disposable immunosensors for C-reactive protein based on carbon nanotubes field effect transistors. *Talanta* **108**, 165–170 (2013).

103. Pimková, K., Bocková, M., Hegnerová, K., Suttnar, J., Čermák, J., Homola, J. & Dyr, J. E. Surface plasmon resonance biosensor for the detection of VEGFR-1-a protein marker of myelodysplastic syndromes. *Analytical and Bioanalytical Chemistry* **402**, 381–387 (2012).

104. Damborský, P., Švitel, J. & Katrlík, J. Optical biosensors. *Essays in Biochemistry* **60**, 91–100 (2016).

105. Liu, S., Shen, Q., Cao, Y., Gan, L., Wang, Z., Steigerwald, M. L. & Guo, X. Chemical functionalization of single-walled carbon nanotube field-effect transistors as switches and sensors. *Coordination Chemistry Reviews* **254**, 1101–1116 (2010).

106. Zheng, H. Y., Alsager, O. A., Zhu, B., Travas-Sejdic, J., Hodgkiss, J. M. & Plank, N. O. V. Electrostatic gating in carbon nanotube aptasensors. *Nanoscale* **8**, 13659–13668 (2016).

107. Sorgenfrei, S., Chiu, C.-Y., Johnston, M., Nuckolls, C. & Shepard, K. L. Debye screening in single-molecule carbon nanotube field-effect sensors. *Nano letters* **11**, 3739–43 (2011).

108. Chu, C.-H., Sarangadharan, I., Regmi, A., Chen, Y.-W., Hsu, C.-P., Chang, W.-H., Lee, G.-Y., Chyi, J.-I., Chen, C.-C., Shiesh, S.-C., Lee, G.-B. & Wang, Y.-L. Beyond the Debye length in high ionic strength solution: direct protein detection with field-effect transistors

(FETs) in human serum. *Scientific Reports* **7**, 5256 (2017).

109. Chen, R. J., Bangsaruntip, S., Drouvalakis, K. A., Wong Shi Kam, N., Shim, M., Li, Y., Kim, W., Utz, P. J. & Dai, H. Noncovalent functionalization of carbon nanotubes for highly specific electronic biosensors. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 4984–4989 (2003).
110. Mubeen, S., Zhang, T., Yoo, B., Deshusses, M. A. & Myung, N. V. Palladium nanoparticles decorated single-walled carbon nanotube hydrogen sensor. *Journal of Physical Chemistry C* **111**, 6321–6327 (2007).
111. Meyyappan, M. Carbon Nanotube-Based Chemical Sensors. *Small* **12**, 2118–2129 (2016).
112. Xu, X., Clément, P., Eklöf-Österberg, J., Kelley-Loughnane, N., Moth-Poulsen, K., Chávez, J. L. & Palma, M. Reconfigurable Carbon Nanotube Multiplexed Sensing Devices. *Nano Letters* **18**, 4130–4135 (2018).
113. Wang, T., Chen, C., Larcher, L. M., Barrero, R. A. & Veedu, R. N. Three decades of nucleic acid aptamer technologies: Lessons learned, progress and opportunities on aptamer development. *Biotechnology Advances* **37**, 28–50 (2019).
114. Lakhin, A. V., Tarantul, V. Z. & Gening, L. V. Aptamers: Problems, solutions and prospects. *Acta Naturae* **5**, 34–43 (2013).
115. Kim, J. P., Lee, B. Y., Hong, S. & Sim, S. J. Ultrasensitive carbon nanotube-based biosensors using antibody-binding fragments. *Analytical Biochemistry* **381**, 193–198 (2008).
116. Carrington, G., Tomlinson, D. & Peckham, M. Exploiting nanobodies and Affimers for superresolution imaging in light microscopy. *Molecular Biology of the Cell* **30**, 2737–2740 (2019).
117. Choi, Y., Weiss, G. A. & Collins, P. G. Single molecule recordings of lysozyme activity. *Physical Chemistry Chemical Physics* **15**, 14879–14895 (2013).
118. Akhterov, M. V., Choi, Y., Olsen, T. J., Sims, P. C., Iftikhar, M., Gul, O. T., Corso, B. L., Weiss, G. A. & Collins, P. G. Observing Lysozymes Closing and Opening Motions by High-Resolution Single-Molecule Enzymology. *ACS Chemical Biology* **10**, 1495–1501 (2015).
119. Hatada, M., Tran, T. T., Tsugawa, W., Sode, K. & Mulchandani, A. Affinity sensor for haemoglobin A1c based on single-walled carbon nanotube field-effect transistor and fructosyl amino acid binding protein. *Biosensors and Bioelectronics* **129**, 254–259 (2019).
120. Gritsan, N. P., Yuzawa, T. & Platz, M. S. Direct Observation of Singlet Phenylnitrene and Measurement of Its Rate of Rearrangement. *Journal of the American Chemical Society* **119**, 5059–5060 (1997).
121. Holzinger, M., Vostrowsky, O., Hirsch, A., Hennrich, F., Kappes, M., Weiss, R. & Jellen, F. Sidewall functionalization of carbon nanotubes. *Angewandte Chemie - International Edition* **40**, 4002–4005 (2001).
122. Holzinger, M., Steinmetz, J., Samaille, D., Glerup, M., Paillet, M., Bernier, P., Ley, L. & Graupner, R. [2+1] Cycloaddition for cross-linking SWCNTs. *Carbon* **42**, 941–947 (2004).
123. Tuci, G., Luconi, L., Rossin, A., Berretti, E., Ba, H., Innocenti, M., Yakhvarov, D., Caporali,

S., Pham-Huu, C. & Giambastiani, G. Aziridine-Functionalized Multiwalled Carbon Nanotubes: Robust and Versatile Catalysts for the Oxygen Reduction Reaction and Knoevenagel Condensation. *ACS Applied Materials and Interfaces* **8**, 30099–30106 (2016).

124. Zheng, M., Jagota, A., Semke, E. D., Diner, B. A., McLean, R. S., Lustig, S. R., Richardson, R. E. & Tassi, N. G. DNA-assisted dispersion and separation of carbon nanotubes. *Nature Materials* **2**, 338–342 (2003).
125. Chen, R. J., Zhang, Y., Wang, D. & Dai, H. Noncovalent Sidewall Functionalization of Single-Walled Carbon Nanotubes for Protein Immobilization. *J Am Chem Soc* **123**, 3838–3839 (2001).
126. Choi, Y., Moody, I. S., Sims, P. C., Hunt, S. R., Corso, B. L., Perez, I., Weiss, G. A. & Collins, P. G. Single-molecule lysozyme dynamics monitored by an electronic circuit. *Science* **335**, 319–324 (2012).
127. Aminov, R. I. A brief history of the antibiotic era: Lessons learned and challenges for the future. *Frontiers in Microbiology* **1**, 134 (2010).
128. Barlow, M. & Hall, B. G. Phylogenetic analysis shows that the OXA β-lactamase genes have been on plasmids for millions of years. *Journal of Molecular Evolution* **55**, 314–321 (2002).
129. Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews : MMBR* **74**, 417–33 (2010).
130. Zaman, S. Bin, Hussain, M. A., Nye, R., Mehta, V., Mamun, K. T. & Hossain, N. A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* **9**, e1403 (2017).
131. Hall, B. G. & Barlow, M. Evolution of the serine β-lactamases: Past, present and future. *Drug Resistance Updates* **7**, 111–123 (2004).
132. Aslam, B., Wang, W., Arshad, M. I., Khurshid, M., Muzammil, S., Rasool, M. H., Nisar, M. A., Alvi, R. F., Aslam, M. A., Qamar, M. U., Salamat, M. K. F. & Baloch, Z. Antibiotic resistance: a rundown of a global crisis. *Infection and drug resistance* **11**, 1645–1658 (2018).
133. Abeylath, S. C. & Turos, E. Drug delivery approaches to overcome bacterial resistance to β-lactam antibiotics. *Expert Opinion on Drug Delivery* **5**, 931–949 (2008).
134. Yocum, R. R., Waxman, D. J., Rasmussen, J. R. & Strominger, J. L. Mechanism of penicillin action: Penicillin and substrate bind covalently to the same active site serine in two bacterial D-alanine carboxypeptidases. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 2730–2734 (1979).
135. Massova, I. & Mobashery, S. Kinship and Diversification of Bacterial Penicillin-Binding Proteins and β-Lactamases. *Antimicrobial Agents and Chemotherapy* **42**, 1 (1998).
136. Meini, M.-R., Llarrull, L. I. & Vila, A. J. Overcoming differences: The catalytic mechanism of metallo-β-lactamases. *FEBS Letters* **589**, 3419–3432 (2015).
137. Shaikh, S., Fatima, J., Shakil, S., Rizvi, S. M. D. & Kamal, M. A. Antibiotic resistance and extended spectrum beta-lactamases: Types, epidemiology and treatment. *Saudi Journal of Biological Sciences* **22**, 90–101 (2015).

138. Jacoby, G. A. β-lactamase nomenclature. *Antimicrobial Agents and Chemotherapy* vol. 50 1123–1129 (2006).
139. Pitkälä, A., Salmikivi, L., Bredbacka, P., Myllyniemi, A.-L. & Koskinen, M. T. Comparison of Tests for Detection of β-Lactamase-Producing Staphylococci. *Journal of Clinical Microbiology* **45**, 2031–2033 (2007).
140. Thai, Q. K., Bös, F. & Pleiss, J. The lactamase engineering database: A critical survey of TEM sequences in public databases. *BMC Genomics* **10**, 390 (2009).
141. Peimbert, M. & Segovia, L. Evolutionary engineering of a β-Lactamase activity on a D-Ala D-Ala transpeptidase fold. *Protein Engineering* **16**, 27–35 (2003).
142. Hartley, A. M., Zaki, A. J., McGarrity, A. R., Robert-Ansart, C., Moskalenko, A. V., Jones, G. F., Craciun, M. F., Russo, S., Elliott, M., MacDonald, J. E. & Jones, D. D. Functional modulation and directed assembly of an enzyme through designed non-natural post-translation modification. *Chemical Science* **6**, 3712–3717 (2015).
143. Rice, P., Longden, L. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277 (2000).
144. Drawz, S. M. & Bonomo, R. A. Three Decades of β-Lactamase Inhibitors. *Clinical Microbiology Reviews* **23**, 160 (2010).
145. Petrosino, J., Rudgers, G., Gilbert, H. & Palzkill, T. Contributions of aspartate 49 and phenylalanine 142 residues of a tight binding inhibitory protein of beta-lactamases. *The Journal of biological chemistry* **274**, 2394–400 (1999).
146. Kang, S. G., Park, H. U., Lee, H. S., Kim, H. T. & Lee, K. J. New beta-lactamase inhibitory protein (BLIP-I) from Streptomyces exfoliatus SMF19 and its roles on the morphological differentiation. *The Journal of biological chemistry* **275**, 16851–6 (2000).
147. Brown, N. G., Chow, D. C., Ruprecht, K. E. & Palzkill, T. Identification of the β-Lactamase Inhibitor Protein-II (BLIP-II) interface residues essential for binding affinity and specificity for class A β-lactamases. *Journal of Biological Chemistry* **288**, 17156–17166 (2013).
148. Brown, N. G., Chow, D. C. & Palzkill, T. BLIP-II is a highly potent inhibitor of Klebsiella pneumoniae carbapenemase (KPC-2). *Antimicrobial Agents and Chemotherapy* **57**, 3398–3401 (2013).
149. Doi, N. & Yanagawa, H. Design of generic biosensors based on green fluorescent proteins with allosteric sites by directed evolution. *FEBS Letters* **453**, 305–307 (1999).
150. Hu, R., Yap, H. K., Fung, Y. H., Wang, Y., Cheong, W. L., So, L. Y., Tsang, C. S., Lee, L. Y. S., Lo, W. K. C., Yuan, J., Sun, N., Leung, Y. C., Yang, G. & Wong, K. Y. Light up protein-protein interaction through bioorthogonal incorporation of a turn-on fluorescent probe into β-lactamase. *Molecular BioSystems* **12**, 3544–3549 (2016).
151. Kim, H., Yoon, H. K. & Yoo, T. H. Engineering β-lactamase zymogens for use in protease activity assays. *Chemical Communications* **50**, 10155–10157 (2014).
152. Chow, D. C., Rice, K., Huang, W., Atmar, R. L. & Palzkill, T. Engineering Specificity from Broad to Narrow: Design of a β-Lactamase Inhibitory Protein (BLIP) Variant That Exclusively Binds and Detects KPC β-Lactamase. *ACS Infectious Diseases* **2**, 969–979 (2016).

153. Adamski, C. J. & Palzkill, T. Systematic substitutions at BLIP position 50 result in changes in binding specificity for class A β-lactamases. *BMC Biochemistry* **18**, 2 (2017).
154. Brown, N. G. & Palzkill, T. Identification and characterization of beta-lactamase inhibitor protein-II (BLIP-II) interactions with beta-lactamases using phage display. *Protein engineering, design & selection : PEDS* **23**, 469–78 (2010).
155. Brown, N. G., Chow, D.-C., Sankaran, B., Zwart, P., Prasad, B. V. V. & Palzkill, T. Analysis of the binding forces driving the tight interactions between beta-lactamase inhibitory protein-II (BLIP-II) and class A beta-lactamases. *The Journal of biological chemistry* **286**, 32723–35 (2011).
156. Adamski, C. J. & Palzkill, T. BLIP-II Employs Differential Hotspot Residues To Bind Structurally Similar Staphylococcus aureus PBP2a and Class A β-Lactamases. *Biochemistry* **56**, 1075–1084 (2017).
157. Doran, J. L., Leskiw, B. K., Aippersbach, S. & Jensen, S. E. Isolation and Characterization of a Beta-Lactamase-Inhibitory Protein From Streptomyces Clavuligerus and Cloning and Analysis of the Corresponding Gene. *Journal of Bacteriology* **172**, 4909–4918 (1990).
158. Park, H. U. & Lee, K. J. Cloning and heterologous expression of the gene for BLIP-II, a β-lactamase-inhibitory protein from Streptomyces exfoliatus SMF19. *Microbiology* **144**, 2161–2167 (1998).
159. Lim, D., Park, H. U., De Castro, L., Kang, S. G., Lee, H. S., Jensen, S., Lee, K. J. & Strynadka, N. C. J. Crystal structure and kinetic analysis of β-lactamase inhibitor protein-II in complex with TEM-1 β-lactamase. *Nature Structural Biology* **8**, 848–852 (2001).
160. Arnesano, F., Banci, L., Bertini, I., Faraone-Mennella, J., Rosato, A., Barker, P. D. & Fersht, A. R. The solution structure of oxidized Escherichia coli cytochrome b562. *Biochemistry* **38**, 8657–8670 (1999).
161. Barker, P. D., Nerou, E. P., Freund, S. M. & Fearnley, I. M. Conversion of cytochrome b562 to c-type cytochromes. *Biochemistry* **34**, 15191–15203 (1995).
162. Wittung-Stafshede, P., Lee, J. C., Winkler, J. R. & Gray, H. B. Cytochrome b562 folding triggered by electron transfer: Approaching the speed limit for formation of a four-helix-bundle protein. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6587–6590 (1999).
163. Thomas, S. K., Jamieson, W. D., Gwyther, R. E. A., Bowen, B. J., Beachey, A., Worthy, H. L., Macdonald, J. E., Elliott, M., Castell, O. K. & Jones, D. D. Site-Specific Protein Photochemical Covalent Attachment to Carbon Nanotube Side Walls and Its Electronic Impact on Single Molecule Function. *Bioconjugate Chemistry* **31**, 584–594 (2020).
164. Solomon, L. A., Kodali, G., Moser, C. C. & Dutton, P. L. Engineering the assembly of heme cofactors in man-made proteins. *Journal of the American Chemical Society* **136**, 3192–3199 (2014).
165. Feng, Y., Sligar, S. G. & Wand, A. J. Solution structure of apocytochrome b562. *Nat Struct Biol* **1**, 30–35 (1994).
166. Oohora, K., Onoda, A. & Hayashi, T. Supramolecular assembling systems formed by heme–heme pocket interactions in hemoproteins. *Chemical Communications* **48**, 11714–

11726 (2012).

167. Edwards, W. R., Williams, A. J., Morris, J. L., Baldwin, A. J., Allemann, R. K. & Jones, D. D. Regulation of beta-lactamase activity by remote binding of haem: Functional coupling of unrelated proteins through domain insertion. *Biochemistry* (2010) doi:10.1021/bi100793y.
168. Arpino, J. A., Czapinska, H., Piasecka, A., Edwards, W. R., Barker, P., Gajda, M. J., Bochtler, M. & Jones, D. D. Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. *J Am Chem Soc* **134**, 13632–13640 (2012).
169. Della Pia, E. A., Chi, Q., Jones, D. D., Macdonald, J. E., Ulstrup, J. & Elliott, M. Single-molecule mapping of long-range electron transport for a cytochrome b(562) variant. *Nano Letters* **11**, 176–182 (2011).
170. Della Pia, E. A., Macdonald, J. E., Elliott, M. & Jones, D. D. Direct binding of a redox protein for single-molecule electron transfer measurements. *Small* **8**, 2341–2344 (2012).
171. Della Pia, E. A., Chi, Q., Elliott, M., Emyr MacDonald, J., Ulstrup, J. & Dafydd Jones, D. Redox tuning of cytochrome b562 through facile metal porphyrin substitution. *Chemical Communications* **48**, 10624–10626 (2012).
172. Sommer, D. J., Vaughn, M. D., Clark, B. C., Tomlin, J., Roy, A. & Ghirlanda, G. Reengineering cyt b562 for hydrogen production: A facile route to artificial hydrogenases. *Biochimica et Biophysica Acta - Bioenergetics* **1857**, 598–603 (2016).
173. Ishida, Y., Konishi, K., Nagamune, T. & Aida, T. Template catalysis: Catalytic metalation of porphyrins by apocytochrome b562 [9]. *Journal of the American Chemical Society* vol. 121 7947–7948 (1999).
174. Li, L. L. & Diau, E. W. Porphyrin-sensitized solar cells. *Chem Soc Rev* **42**, 291–304 (2013).
175. Ishihara, S., Labuta, J., Van Rossom, W., Ishikawa, D., Minami, K., Hill, J. P. & Ariga, K. Porphyrin-based sensor nanoarchitectonics in diverse physical detection modes. *Physical Chemistry Chemical Physics* vol. 16 9713–9746 (2014).
176. Mol, J. A., Lau, C. S., Lewis, W. J. M., Sadeghi, H., Roche, C., Cnossen, A., Warner, J. H., Lambert, C. J., Anderson, H. L. & Briggs, G. A. D. Graphene-porphyrin single-molecule transistors. *Nanoscale* **7**, 13181–13185 (2015).
177. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207–234 (2005).
178. Swiss Institute of Bioinformatics. SwissSidechain - PyMOL Plugin (https://www.swisssidechain.ch/visualization/pymol.php). (2015).
179. JCrystalSoft. Nanotube Modeler v1.8.0 (www.jcrystal.com/products/wincnt). (2018).
180. Ji Ram, V., Sethi, A., Nath, M. & Pratap, R. Three-Membered Ring Heterocycles. in *The Chemistry of Heterocycles* 19–92 (Elsevier, 2019). doi:10.1016/b978-0-08-101033-4.00003-6.
181. Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A. & Brewer, S. H. Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry* **48**, 5953–5962 (2009).

182. Dippel, A. B., Olenginski, G. M., Maurici, N., Liskov, M. T., Brewer, S. H. & Phillips-Piro, C. M. Probing the effectiveness of spectroscopic reporter unnatural amino acids: a structural study. *Acta Crystallographica Section D* **72**, 121–130 (2016).
183. GSL Biotech. SnapGene® software (www.snapgene.com). (2020).
184. Patel, M. P., Fryszczyn, B. G. & Palzkill, T. Characterization of the global stabilizing substitution A77V and its role in the evolution of CTX-M β-Lactamases. *Antimicrobial Agents and Chemotherapy* **59**, 6741–6748 (2015).
185. Sosa-Peinado, A., Mustafi, D. & Makinen, M. W. Overexpression and biosynthetic deuterium enrichment of TEM-1 β-lactamase for structural characterization by magnetic resonance methods. *Protein Expression and Purification* **19**, 235–245 (2000).
186. Serial Basics. Serial Cloner software (http://serialbasics.free.fr). (2013).
187. Nicolas Jullien. AmplifX 1.7.0 software. (2013).
188. Chiu, J., March, P. E., Lee, R. & Tillett, D. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic acids research* **32**, e174 (2004).
189. Heitman, J., Ivanenko, T. & Kiss, A. DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in Escherichia coli. *Molecular Microbiology* **33**, 1141–1151 (2002).
190. Teale, F. W. J. Cleavage of the haem-protein link by acid methylethylketone. *BBA - General Subjects* **35**, 543 (1959).
191. Neu, H. C. & Heppel, L. A. The release of enzymes from Escherichia coli by osmotic shock and during the formation of spheroplasts. *The Journal of biological chemistry* **240**, 3685–92 (1965).
192. Nossal, N. G. & Heppel, L. A. The release of enzymes by osmotic shock from Escherichia coli in exponential phase. *The Journal of biological chemistry* **241**, 3055–62 (1966).
193. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
194. Ball, D. W. *Field Guide to Spectroscopy*. *Field Guide to Spectroscopy* (2006). doi:10.1117/3.682726.
195. Johnson, K. A. & Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper. *Biochemistry* **50**, 8264–8269 (2011).
196. GraphPad Software. GraphPad Prism version 5.03 for Windows. (2009).
197. Morrison, J. F. *Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors*. *Biochimica et Biophysica Acta (BBA) - Enzymology* vol. 185 (Elsevier, 1969).
198. Murphy, D. J. Determination of accurate KI values for tight-binding enzyme inhibitors: An in silico study of experimental error and assay design. *Analytical Biochemistry* **327**, 61–67 (2004).
199. Wang, Z. X., Ravi Kumar, N. & Srivastava, D. K. A novel spectroscopic titration method for determining the dissociation constant and stoichiometry of protein-ligand complex. *Analytical Biochemistry* **206**, 376–381 (1992).

200. Nosjean, O., Souchaud, S., Deniau, C., Geneste, O., Cauquil, N. & Boutin, J. A. A simple theoretical model for fluorescence polarization binding assay development. *Journal of biomolecular screening* **11**, 949–58 (2006).
201. Motulsky, H. J. & Neubig, R. R. Analyzing Binding Data. *Current Protocols in Neuroscience* **52**, 7.5.1-7.5.65 (2010).
202. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development* **82**, 518–529 (2015).
203. Li, P., Martin, C. M., Yeung, K. K. & Xue, W. Dielectrophoresis aligned single-walled carbon nanotubes as ph ensors. *Biosensors* **1**, 23–35 (2011).
204. Nečas, D. & Klapetek, P. Gwyddion: An open-source software for SPM data analysis. *Central European Journal of Physics* vol. 10 181–188 (2012).
205. Saifuddin, N., Raziah, A. Z. & Junizah, A. R. Carbon Nanotubes: A Review on Structure and Their Interaction with Proteins. *Journal of Chemistry* **2013**, 676815 (2013).
206. Tung, N. T., Tue, P. T., Thi Ngoc Lien, T., Ohno, Y., Maehashi, K., Matsumoto, K., Nishigaki, K., Biyani, M. & Takamura, Y. Peptide aptamer-modified single-walled carbon nanotube-based transistors for high-performance biosensors. *Scientific Reports* **7**, 1–9 (2017).
207. Heller, I., Janssens, A. M., Männik, J., Minot, E. D., Lemay, S. G. & Dekker, C. Identifying the mechanism of biosensing with carbon nanotube transistors. *Nano Letters* **8**, 591–595 (2008).
208. Kase, H., Negishi, R., Arifuku, M., Kiyoyanagi, N. & Kobayashi, Y. Biosensor response from target molecules with inhomogeneous charge localization. *Journal of Applied Physics* **124**, 064502 (2018).
209. Besteman, K., Lee, J. O., Wiertz, F. G. M., Heering, H. A. & Dekker, C. Enzyme-coated carbon nanotubes as single-molecule biosensors. *Nano Letters* **3**, 727–730 (2003).
210. Sims, P. C., Moody, I. S., Choi, Y., Dong, C., Iftikhar, M., Corso, B. L., Gul, O. T., Collins, P. G. & Weiss, G. A. Electronic measurements of single-molecule catalysis by cAMP-dependent protein kinase A. *J Am Chem Soc* **135**, 7861–7868 (2013).
211. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Science* **27**, 112–128 (2018).
212. Benjamin, J., Macdonald, J. E., Elliott, M., Castell, O. K. & Dafydd, D. Site-specific protein covalent attachment to nanotube side walls and its electronic impact on single molecule function . *Bioconjugate Chemistry* 1–19 (2019) doi:10.1021/acs.bioconjchem.9b00719.
213. Tan, Y. H., Liu, M., Nolting, B., Go, J. G., Gervay-Hague, J. & Liu, G. Y. A nanoengineering approach for investigation and regulation of protein immobilization. *ACS Nano* **2**, 2374–2384 (2008).
214. Santos, S., Barcons, V., Christenson, H. K., Font, J. & Thomson, N. H. The Intrinsic Resolution Limit in the Atomic Force Microscope: Implications for Heights of Nano-Scale Features. *PLoS ONE* **6**, e23821 (2011).
215. Lyubchenko, Y., Shlyakhtenko, L., Harrington, R., Oden, P. & Lindsay, S. Atomic force

microscopy of long DNA: Imaging in air and under water. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 2137–2140 (1993).

216. Bustamante, C., Erie, D. A. & Keller, D. Biochemical and structural applications of scanning force microscopy. *Current Opinion in Structural Biology* **4**, 750–760 (1994).
217. Rao Movva, N., Nakamura, K. & Inouye, M. Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of Escherichia coli. *Journal of Biological Chemistry* **255**, 27–29 (1980).
218. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**, 3784–3788 (2003).
219. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
220. Stojanoski, V., Chow, D. C., Hu, L., Sankaran, B., Gilbert, H. F., Prasad, B. V. V. & Palzkill, T. A triple mutant in the Ω-loop of TEM-1 β-lactamase changes the substrate profile via a large conformational change and an altered general base for catalysis. *Journal of Biological Chemistry* **290**, 10382–10394 (2015).
221. Zafaralla, G., Manavathu, E. K., Lerner, S. A. & Mobashery, S. Elucidation of the Role of Arginine-224 in the Turnover Processes of Class A β-Lactamases. *Biochemistry* **31**, 3847–3852 (1992).
222. Fryszczyn, B. G., Adamski, C. J., Brown, N. G., Rice, K., Huang, W. & Palzkill, T. Role of β-lactamase residues in a common interface for binding the structurally unrelated inhibitory proteins BLIP and BLIP-II. *Protein Science* **23**, 1235–1246 (2014).
223. Kvålseth, T. O. Cautionary note about r2. *American Statistician* **39**, 279–285 (1985).
224. Khoury, G. A., Smadbeck, J., Kieslich, C. A. & Floudas, C. A. Protein folding and de novo protein design for biotechnological applications. *Trends in Biotechnology* **32**, 99–109 (2014).
225. Madeira, F., Park, Y. mi, Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. & Lopez, R. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**, W636–W641 (2019).
226. Bush, K. Past and present perspectives on β-lactamases. *Antimicrobial Agents and Chemotherapy* **62**, (2018).
227. Worthy, H. L. Novel routes to defined post translational modifications using non-canonical amino acids. PhD thesis available at http://orca.cf.ac.uk/119409. (Cardiff University, 2018).
228. Gruttner, C., Muller, K. & Teller, J. Comparison of strain-promoted alkyne-azide cycloaddition with established methods for conjugation of biomolecules to magnetic nanoparticles. *IEEE Transactions on Magnetics* **49**, 172–176 (2013).
229. Xu, D., Bum-Erdene, K., Si, Y., Zhou, D., Ghozayel, M. K. & Meroueh, S. O. Mimicking Intermolecular Interactions of Tight Protein–Protein Complexes for Small-Molecule Antagonists. *ChemMedChem* **12**, 1794–1809 (2017).
230. Carlson, H. A., Smith, R. D., Khazanov, N. A., Kirchhoff, P. D., Dunbar, J. B. & Benson,

M. L. Differences between high- and low-affinity complexes of enzymes and nonenzymes. *Journal of Medicinal Chemistry* **51**, 6432–6441 (2008).

231. Richardson, M. B., Brown, D. B., Vasquez, C. A., Ziller, J. W., Johnston, K. M. & Weiss, G. A. Synthesis and Explosion Hazards of 4-Azido- l -phenylalanine. *Journal of Organic Chemistry* **83**, 4525–4536 (2018).
232. Knight, K. L. & McEntee, K. Tyrosine 264 in the recA protein from Escherichia coli is the site of modification by the photoaffinity label 8-azidoadenosine 5'-triphosphate. *The Journal of biological chemistry* **260**, 10185–91 (1985).
233. Collins, P. G. Defects and disorder in carbon nanotubes. in *Oxford Handbook of Nanoscience and Technology: Volume 2: Materials: Structures, Properties and Characterization Techniques* (2010). doi:10.1093/OXFORDHB/9780199533053.013.2.
234. Pang, H. L., Liu, J., Hu, D., Zhang, X. H. & Chen, J. H. Immobilization of laccase onto 1-aminopyrene functionalized carbon nanotubes and their electrocatalytic activity for oxygen reduction. *Electrochimica Acta* **55**, 6611–6616 (2010).
235. Siu, H. & Duhamel, J. Molar absorption coefficient of pyrene aggregates in water. *Journal of Physical Chemistry B* **112**, 15301–15312 (2008).
236. Moniruzzaman, M. & Winey, K. I. Polymer nanocomposites containing carbon nanotubes. *Macromolecules* **39**, 5194–5205 (2006).
237. Cole, N. B. Site-specific protein labeling with SNAP-tags. *Current Protocols in Protein Science* **2013**, 30.1.1-30.1.16 (2013).
238. Steen Redeker, E., Ta, D. T., Cortens, D., Billen, B., Guedens, W. & Adriaensens, P. Protein engineering for directed immobilization. *Bioconjugate Chemistry* **24**, 1761–1777 (2013).
239. Lee, J. O., So, H. M., Jeon, E. K., Chang, H., Won, K. & Kim, Y. H. Aptamers as molecular recognition elements for electrical nanobiosensors. *Analytical and Bioanalytical Chemistry* **390**, 1023–1032 (2008).
240. Liu, W., Brock, A., Chen, S., Chen, S. & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nature Methods* **4**, 239–244 (2007).
241. Wentrup, C. Flash Vacuum Pyrolysis of Azides, Triazoles, and Tetrazoles. *Chemical Reviews* vol. 117 4562–4623 (2017).
242. Du Bois, J. Rhodium-catalyzed C-H amination. an enabling method for chemical synthesis. *Organic Process Research and Development* **15**, 758–762 (2011).
243. Holzinger, M., Abraham, J., Whelan, P., Graupner, R., Ley, L., Hennrich, F., Kappes, M. & Hirsch, A. Functionalization of single-walled carbon nanotubes with (R-)oxycarbonyl nitrenes. *Journal of the American Chemical Society* **125**, 8566–8580 (2003).
244. Castanheira, M., Farrell, S. E., Deshpande, L. M., Mendes, R. E. & Jones, R. N. Prevalence of β-lactamase-encoding genes among Enterobacteriaceae bacteremia isolates collected in 26 U.S. Hospitals: Report from the SENTRY antimicrobial surveillance program (2010). *Antimicrobial Agents and Chemotherapy* **57**, 3012–3020 (2013).
245. Cantón, R., González-Alba, J. M. & Galán, J. C. CTX-M enzymes: Origin and diffusion. *Frontiers in Microbiology* **3**, (2012).

246. Marchesan, S. & Prato, M. Under the lens: Carbon nanotube and protein interaction at the nanoscale. *Chemical Communications* **51**, 4347–4359 (2015).
247. Möller, C., Allen, M., Elings, V., Engel, A. & Müller, D. J. Tapping-mode atomic force microscopy produces faithful high-resolution images of protein surfaces. *Biophysical Journal* **77**, 1150–1158 (1999).
248. ChemAxon (http://www.chemaxon.com). MarvinSketch. (2020).
249. Palzkill, T. Structural and mechanistic basis for extended-spectrum drug-resistance mutations in altering the specificity of TEM, CTX-M, and KPC β-lactamases. *Frontiers in Molecular Biosciences* **5**, 16 (2018).
250. Salverda, M. L. M., de Visser, J. A. G. M. & Barlow, M. Natural evolution of TEM-1 β-lactamase: Experimental reconstruction and clinical relevance. *FEMS Microbiology Reviews* **34**, 1015–1036 (2010).
251. Davis, J. J., Green, M. L. H., Hill, H. A. O., Leung, Y. C., Sadler, P. J., Sloan, J., Xavier, A. V. & Tsang, S. C. The immobilisation of proteins in carbon nanotubes. *Inorganica Chimica Acta* **272**, 261–266 (1998).
252. Damblon, C., Raquet, X., Lian, L. Y., Lamotte-Brasseur, J., Fonze, E., Charlier, P., Roberts, G. C. K. & Frère, J. M. The catalytic mechanism of β-lactamases: NMR titration of an active-site lysine residue of the TEM-1 enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1747–1752 (1996).
253. Reddington, S. C. Introducing novel protein functionality using unnatural amino acids. PhD thesis available at http://orca.cf.ac.uk/56226. (Cardiff University, 2013).
254. Stevens, J. M., Daltrop, O., Higham, C. W. & Ferguson, S. J. Interaction of heme with variants of the heme chaperone CcmE carrying active site mutations and a cleavable N-terminal His tag. *The Journal of biological chemistry* **278**, 20500–6 (2003).
255. Salgado, E. N., Ambroggio, X. I., Brodin, J. D., Lewis, R. A., Kuhlman, B. & Tezcan, F. A. Metal templated design of protein interfaces. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 1827–1832 (2010).
256. Pattison, D. I., Rahmanto, A. S. & Davies, M. J. Photo-oxidation of proteins. *Photochemical and Photobiological Sciences* **11**, 38–53 (2012).
257. Durchschlag, H., Fochler, C., Feser, B., Hausmann, S., Seroneit, T., Swientek, M., Swoboda, E., Winklmair, A., Wlček, C. & Zipper, P. Effects of X-and UV-irradiation on proteins. *Radiation Physics and Chemistry* **47**, 501–505 (1996).
258. Tippmann, E. M., Culpepper, S., Bunnel, W. & Appel, N. New perspectives on aryl azide noncanonical amino acid use in yeast. *Photochemical and Photobiological Sciences* **18**, 253–258 (2019).
259. Palzer, S., Bantel, Y., Kazenwadel, F., Berg, M., Rupp, S. & Sohn, K. An expanded genetic code in Candida albicans to study protein-protein interactions in vivo. *Eukaryotic Cell* **12**, 816–827 (2013).
260. Jakob, L., Gust, A. & Grohmann, D. Evaluation and optimisation of unnatural amino acid incorporation and bioorthogonal bioconjugation for site-specific fluorescent labelling of proteins expressed in mammalian cells. *Biochemistry and Biophysics Reports* **17**, 1–9

(2019).

261. Zhu, Z., Garcia-Gancedo, L., Flewitt, A. J., Xie, H., Moussy, F. & Milne, W. I. A Critical Review of Glucose Biosensors Based on Carbon Nanomaterials: Carbon Nanotubes and Graphene. *Sensors* **12**, 5996–6022 (2012).
262. Hou, S., Larsen, R. W., Boudko, D., Riley, C. W., Karatan, E., Zimmer, M., Ordal, G. W. & Alam, M. Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature* **403**, 540–544 (2000).
263. Pandey, V. P., Awasthi, M., Singh, S., Tiwari, S. & Dwivedi, U. N. A Comprehensive Review on Function and Application of Plant Peroxidases. *Biochemistry & Analytical Biochemistry* **06**, (2017).
264. Crichton, R. Biological Ligands for Metal Ions. in *Biological Inorganic Chemistry* 81–118 (Elsevier, 2019). doi:10.1016/b978-0-12-811741-5.00004-7.
265. Chitta, R., Sandanayaka, A. S. D., Schumacher, A. L., D'Souza, L., Araki, Y., Ito, O. & D'Souza, F. Donor - Acceptor nanohybrids of zinc naphthalocyanine or zinc porphyrin noncovalently linked to single-wall carbon nanotubes for photoinduced electron transfer. *Journal of Physical Chemistry C* **111**, 6947–6955 (2007).
266. Biswas, S., Mukherjee, D., De, S. & Kathiravan, A. Probing the Highly Efficient Electron Transfer Dynamics between Zinc Protoporphyrin IX and Sodium Titanate Nanosheets. *The Journal of Physical Chemistry A* **120**, 7121–7129 (2016).
267. Imran, M., Ramzan, M., Qureshi, A. K., Azhar Khan, M. & Tariq, M. Emerging applications of porphyrins and metalloporphyrins in biomedicine and diagnostic magnetic resonance imaging. *Biosensors* **8**, (2018).
268. Huang, H., Song, W., Rieffel, J. & Lovell, J. F. Emerging applications of porphyrins in photomedicine. *Frontiers in Physics* **3**, 23 (2015).
269. Regehly, M., Greish, K., Rancan, F., Maeda, H., Böhm, F. & Röder, B. Water-Soluble Polymer Conjugates of ZnPP for Photodynamic Tumor Therapy. *Bioconjugate Chemistry* **18**, 494–499 (2007).
270. Jing, L., Liang, X., Li, X., Lin, L., Yang, Y., Yue, X. & Dai, Z. Mn-porphyrin conjugated Au nanoshells encapsulating doxorubicin for potential magnetic resonance imaging and light triggered synergistic therapy of cancer. *Theranostics* **4**, 858–871 (2014).
271. Liu, T. W., MacDonald, T. D., Shi, J., Wilson, B. C. & Zheng, G. Intrinsically Copper-64-Labeled Organic Nanoparticles as Radiotracers. *Angewandte Chemie International Edition* **51**, 13128–13131 (2012).
272. Harada, A., Fukushima, H., Shiotsuki, K., Yamaguchi, H., Oka, F. & Kamachi, M. Peroxidation of Pyrogallol by Antibody−Metalloporphyrin Complexes. *Inorganic Chemistry* **36**, 6099–6102 (1997).
273. Key, H. M., Dydio, P., Clark, D. S. & Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **534**, 534–537 (2016).
274. Dydio, P., Key, H. M., Nazarenko, A., Rha, J. Y. E., Seyedkazemi, V., Clark, D. S. & Hartwig, J. F. An artificial metalloenzyme with the kinetics of native enzymes. *Science* **354**, 102–106 (2016).

275. Atkins, P. & de Paula, J. Atkins' Physical Chemistry, Eighth Edition. in 492–494 (W. H. Freeman and Company, 2006).
276. Labbe, R. F., Vreman, H. J. & Stevenson, D. K. Zinc protoporphyrin: A metabolite with a mission. *Clin Chem* **45**, 2060–2072 (1999).
277. Hu, Y., Geissinger, P. & Woehl, J. C. Potential of protoporphyrin IX and metal derivatives for single molecule fluorescence studies. *Journal of Luminescence* **131**, 477–481 (2011).
278. REPESSÉ, Y., DIMITROV, J. D., PEYRON, I., MOSHAI, E. F., KIGER, L., DASGUPTA, S., DELIGNAT, S., MARDEN, M. C., KAVERI, S. V. & LACROIX-DESMAZES, S. Heme binds to factor VIII and inhibits its interaction with activated factor IX. *Journal of Thrombosis and Haemostasis* **10**, 1062–1071 (2012).
279. Robinson, C. R., Liu, Y., Thomson, J. A., Sturtevant, J. M. & Sligar, S. G. Energetics of heme binding to native and denatured states of cytochrome b562. *Biochemistry* **36**, 16141–16146 (1997).
280. Takeda, S., Kamiya, N. & Nagamune, T. Rational design of a protein-based molecular device consisting of blue fluorescent protein and zinc protoporphyrin IX incorporated into a cytochrome b $_{562}$ scaffold. *Biotechnology Letters* **26**, 121–125 (2004).
281. Della Pia, E. A., Chi, Q., Macdonald, J. E., Ulstrup, J., Jones, D. D. & Elliott, M. Fast electron transfer through a single molecule natively structured redox protein. *Nanoscale* **4**, 7106–7113 (2012).
282. Chen, Z., Wang, Y., Zhai, Y. F., Song, J. & Zhang, Z. ZincExplorer: An accurate hybrid method to improve the prediction of zinc-binding sites from protein sequences. *Molecular BioSystems* **9**, 2213–2222 (2013).
283. Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G. & Raschellà, G. Zinc-finger proteins in health and disease. *Cell Death Discovery* vol. 3 (2017).
284. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **91**, 43–56 (1995).
285. Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E. & Hutchison, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* **4**, 17 (2012).
286. Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M. & Montgomery, J. A. General atomic and molecular electronic structure system. *Journal of Computational Chemistry* **14**, 1347–1363 (1993).
287. Meiler, J. & Baker, D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **65**, 538–548 (2006).
288. Cowan, J. A. & Gray, H. B. Q-band splitting in the spectra of heme proteins. *Inorganic Chemistry* **28**, 4554–4556 (1989).
289. Itagaki, E., Palmer, G. & Hager, L. P. Studies on cytochrome b562 of Escherichia coli. II. Reconstitution of cytochrome b562 from apoprotein and hemin. *The Journal of biological*

chemistry **242**, 2272–7 (1967).

290. Brown, S. B., Dean, T. C. & Jones, P. Catalatic activity of iron(III)-centred catalysts. Role of dimerization in the catalytic action of ferrihaems. *The Biochemical journal* **117**, 741–744 (1970).
291. Kuželová, K., Mrhalová, M. & Hrkal, Z. Kinetics of heme interaction with heme-binding proteins: The effect of heme aggregation state. *Biochimica et Biophysica Acta - General Subjects* **1336**, 497–501 (1997).
292. Renney, C. M., Fukuhara, G., Inoue, Y. & Davis, A. P. Binding or aggregation? Hazards of interpretation in studies of molecular recognition by porphyrins in water. *Chemical Communications* **51**, 9551–9554 (2015).
293. Luo, L., Chang, C.-H., Chen, Y.-C., Wu, T.-K. & Diau, E. W.-G. Ultrafast Relaxation of Zinc Protoporphyrin Encapsulated within Apomyoglobin in Buffer Solutions. *The Journal of Physical Chemistry B* **111**, 7656–7664 (2007).
294. Leonard, J. J., Yonetani, T. & Callis, J. B. A fluorescence study of hybrid hemoglobins containing free base and zinc protoporphyrin IX. *Biochemistry* **13**, 1460–1464 (1974).
295. FluorTools. a|e - UV-Vis-IR Spectral Software 1.2. (2015).
296. Pollard, T. D. MBOC technical perspective: A guide to simple and informative binding assays. *Molecular Biology of the Cell* **21**, 4061–4067 (2010).
297. Catici, D. A. M., Amos, H. E., Yang, Y., van den Elsen, J. M. H. & Pudney, C. R. The red edge excitation shift phenomenon can be used to unmask protein structural ensembles: implications for NEMO-ubiquitin interactions. *The FEBS Journal* **283**, 2272–2284 (2016).
298. Robinson, C. R., Sligar, S. G., Liu, Y., O'Brien, R. & Sturtevant, J. M. A differential scanning calorimetric study of the thermal unfolding of apo- and holo-cytochrome b562. *Protein Science* **7**, 961–965 (2008).
299. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876–2890 (2007).
300. Feng, Y. & Sligar, S. G. Effect of heme binding on the structure and stability of Escherichia coli apocytochrome b562. *Biochemistry* **30**, 10150–10155 (1991).
301. Ansel, H. C., Norred, W. P. & Roth, I. L. Antimicrobial Activity of Dimethyl Sulfoxide Against Escherichia coli, Pseudomonas aeruginosa, and Bacillus megaterium. *Journal of Pharmaceutical Sciences* **58**, 836–839 (1969).
302. Notman, R., Noro, M., O'Malley, B. & Anwar, J. Molecular basis for dimethylsulfoxide (DMSO) action on lipid membranes. *Journal of the American Chemical Society* **128**, 13982–13983 (2006).
303. Laitaoja, M., Valjakka, J. & Janis, J. Zinc coordination spheres in protein structures. *Inorg Chem* **52**, 10983–10991 (2013).
304. Pace, N. J. & Weerapana, E. Zinc-binding cysteines: diverse functions and structural motifs. *Biomolecules* **4**, 419–434 (2014).
305. Simonson, T. & Calimet, N. CysxHisy-Zn2+ interactions: Thiol vs. thiolate coordination. *Proteins: Structure, Function, and Genetics* **49**, 37–48 (2002).
306. Ren, T. B., Xu, W., Zhang, W., Zhang, X. X., Wang, Z. Y., Xiang, Z., Yuan, L. & Zhang, X.

B. A General Method to Increase Stokes Shift by Introducing Alternating Vibronic Structures. *Journal of the American Chemical Society* **140**, 7716–7722 (2018).

307. Levinthal, C. How to fold graciously. *Mossbauer spectroscopy in biological systems* **67**, 22–24 (1969).
308. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Bioinformatics* **21**, 167–195 (1995).
309. Goldman, A., Harper, S. & Speicher, D. W. Detection of proteins on blot membranes. *Current Protocols in Protein Science* **2016**, 10.8.1-10.8.11 (2016).
310. Zhang, S., Garcia-D'Angeli, A., Brennan, J. P. & Huo, Q. Predicting detection limits of enzyme-linked immunosorbent assay (ELISA) and bioanalytical techniques in general. *Analyst* **139**, 439–445 (2013).
311. Miseta, A. & Csutora, P. Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms. *Molecular Biology and Evolution* **17**, 1232–1239 (2000).
312. Gao, W., Cho, E., Liu, Y. & Lu, Y. Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Frontiers in Pharmacology* **10**, 611 (2019).
313. Kuhn, S. M., Rubini, M., Müller, M. A. & Skerra, A. Biosynthesis of a fluorescent protein with extreme Pseudo-stokes shift by introducing a genetically encoded non-natural amino acid outside the fluorophore. *Journal of the American Chemical Society* **133**, 3708–3711 (2011).
314. Hohsaka, T., Ashizuka, Y., Taira, H., Murakami, H. & Sisido, M. Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* **40**, 11060–11064 (2001).
315. Budisa, N., Minks, C., Medrano, F. J., Lutz, J., Huber, R. & Moroder, L. Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: Structure and stability of the per-thiaproline mutant of annexin V. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 455–459 (1998).
316. Lockless, S. W. & Muir, T. W. Traceless protein splicing utilizing evolved split inteins. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 10999–11004 (2009).
317. Minnihan, E. C., Yokoyama, K. & Stubbe, J. Unnatural amino acids: better than the real things? *F1000 Biology Reports* **1**, (2009).
318. Behrendt, R., White, P. & Offer, J. Advances in Fmoc solid-phase peptide synthesis. *Journal of Peptide Science* **22**, 4–27 (2016).
319. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z. & Schultz, P. G. An expanded eukaryotic genetic code. *Science* **301**, 964–967 (2003).
320. Greiss, S. & Chin, J. W. Expanding the genetic code of an animal. *J Am Chem Soc* **133**, 14196–14199 (2011).
321. Bianco, A., Townsley, F. M., Greiss, S., Lang, K. & Chin, J. W. Expanding the genetic code of Drosophila melanogaster. *Nature Chemical Biology* **8**, 748–750 (2012).

322. Ye, S., Huber, T., Vogel, R. & Sakmar, T. P. FTIR analysis of GPCR activation using azido probes. *Nature Chemical Biology* **5**, 397–399 (2009).
323. Freeley, M. Nanoscale Platforms for Single-Molecule Investigations. PhD thesis. (Queen Mary, University of London, 2018).
324. Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A. & Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
325. Chen, Y. S., Hong, M. Y. & Huang, G. S. A protein transistor made of an antibody molecule and two gold nanoparticles. *Nature Nanotechnology* **7**, 197–203 (2012).
326. Yella, A., Lee, H. W., Tsao, H. N., Yi, C., Chandiran, A. K., Nazeeruddin, M. K., Diau, E. W. G., Yeh, C. Y., Zakeeruddin, S. M. & Grätzel, M. Porphyrin-sensitized solar cells with cobalt (II/III)-based redox electrolyte exceed 12 percent efficiency. *Science* **334**, 629–634 (2011).
327. Li, L., Huang, Y., Peng, J., Cao, Y. & Peng, X. Highly responsive organic near-infrared photodetectors based on a porphyrin small molecule. *Journal of Materials Chemistry C* **2**, 1372–1375 (2014).
328. Xiao, L., Chen, S., Chen, X., Peng, X., Cao, Y. & Zhu, X. High-detectivity panchromatic photodetectors for the near infrared region based on a dimeric porphyrin small molecule. *Journal of Materials Chemistry C* **6**, 3341–3345 (2018).