Using a reprogrammed genetic code to modulate protein activity by novel post-translational control

Andrew M. Hartley

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

Cardiff University
School of Biosciences

September 2014

Submission of Thesis declaration and statements

DECLARATION

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

Signed …………………………………………. (candidate)   Date …………………………….

STATEMENT 1

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

Signed …………………………………………. (candidate)   Date …………………………….

STATEMENT 2

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

Signed …………………………………………. (candidate)   Date …………………………….

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …………………………………………. (candidate)   Date …………………………….

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …………………………………………. (candidate)   Date ……………….

Acknowledgements

I would first like to thank my supervisor, Dafydd Jones, for all the help and advice throughout the four years I have spent in the lab. Particular thanks for the feedback and advice I received during writing up, a process every PhD student is sad to end!

Thanks must go to all of the past and present members of the Jones lab, for all of their help, both in the lab and out of it. Thanks to Matt, Mike and Amy for helping me find my way around when I first arrived (and finding me a place to live), and thanks to Adam, Sarunas and all the newer members of the lab for their injections of enthusiasm and for making the coffee. A special thanks needs to be given to Sam and James, two people without whom my four years in Cardiff would have been a lot quieter and a lot more difficult.

There are many other people who have helped me with my work throughout my time in Cardiff. Firstly I must thank Pierre for all his help with crystallization, and for all his knowledge of space groups! Thanks must go to everyone in Physics who have been a pleasure to work with over the last couple of years. Athraa, Andy and Adam, thanks for of all your efforts. Thank you to everyone to whom I went to beg, borrow and steal chemicals and equipment, there are too many of you to mention individually! And lastly, thanks to Joan for all your help, you are irreplaceable.

Outside of the lab, great thanks must go to Sean, Basti and Claris, for the greatest suggestion at the worst possible timing. What a month! Thanks to all the members of the newly founded "Cardiff Running Club", best of luck next week! Thank you to all the members of the greatest 5-a-side team in Cardiff, bringing our own version of Total Football to pitch no. 5. Thanks to all my friends in Dentistry and in the Heath, much beer and much rugby has been enjoyed.

Finally, I must thank my family, without whom I would not have got to this point. Thanks to Mum and Dad for all your support and thanks to Jenn for the appreciated visits. Happy 30th by the way! Sorry I couldn't be there.

Abstract

Despite the diverse structures and functions sampled by the proteome, all proteins comprise 20 canonical amino acids that sample only a small percentage of available chemistry. This limitation is lifted somewhat through the use of post-translational modifications, however the limit imposed by the restricted number of amino acids inherently limits the variety of protein function and control that can be accessed. One powerful route to diversify the chemistry sampled by proteins is through genetically encoded unnatural amino acid (uAA) incorporation. The uAA *p*-azido-L-phenylalanine (AzPhe) can introduce two novel methods of control, photochemical covalent rearrangement and Click chemistry. AzPhe incorporation combined with these two methods of novel post-translational control were used to modulate the function of two distinct proteins; TEM β-lactamase and sfGFP.

This thesis introduces the use of uAAs and the technical modifications required to enable uAA incorporation *in vivo*. It describes the *in silico* approach taken to evaluate potential mutations based on the likelihood of them imparting novel changes to protein function. Nine positions in TEM β-lactamase were chosen for uAA incorporation and the effect on activity was then determined using kinetic analyses. AzPhe incorporation alone resulted in a variety of effects on enzyme activity, ranging from small increases to complete loss of activity. Subsequent post-translational modification using UV light resulted in only slight changes in activity. Modification via Click chemistry using dibenzyl cyclo-octyne (DBCO) derivatives resulted in either inhibition or increased catalytic activity, depending on the position of AzPhe incorporation and the type of adduct used.

Click chemistry was then used to modify TEM β-lactamase with other chemical modifications that enable the immobilization of proteins onto two different surfaces. The π-π stacking interaction between a DBCO-pyrene moiety and graphene was exploited to attach TEM β-lactamase to graphene in a defined and controlled manner, placing the active site in close proximity to the electron cloud of the sp^2-bonded material. TEM β-lactamase was then modified using two DNA oligonucleotides that define assembly of a DNA origami "tile". DNA origami can be used to immobilize

multiple proteins at several defined positions, enabling the re-creation of enzyme pathways or signalling cascades *in vitro*.

Finally, AzPhe was incorporated into sfGFP and the effects of its incorporation and subsequent modification on fluorescence were explored. The incorporation of AzPhe resulted in a blue shifted λ_{max}, a change that was reversed upon UV irradiation. X-ray crystallography suggested that a hydrogen-bonding network involving the chromophore and surrounding residues was disrupted upon AzPhe incorporation, but then reformed upon modification of the uAA. Click chemistry had a variable effect on fluorescence depending on the modification used. Modification of AzPhe with a large fluorescent dye had no effect on the sfGFP fluorescence spectrum, but enabled FRET between the two chromophores. Modification with a DBCO-amine had the same effect as UV irradiation.

Overall, this thesis has shown that the use of genetically encoded uAA incorporation coupled with novel post-translational modifications is a powerful approach for modifying protein function, and facilitating defined interfacing with new and useful materials.

Abbreviations

Amino acids are described using three-letter code (e.g. Tyr and Ser) throughout this thesis. Mutations are denoted with the protein first, the residue number, and then the mutant amino acid last in superscript (e.g. sfGFP^{His148AzPhe}).

Abbreviations used in this thesis;

2D	Two-dimensional
3D	Three-dimensional
λ_{em}	Emission wavelength
λ_{max}	Wavelength of highest absorbance
aaRS	Amino-acyl tRNA-synthetase
AzPhe	p-azido-L-phenyalanine
bp	Base pair
dNTP	deoxyribonucleotide phosphate
FRET	Forster resonance energy transfer
MMT	MES (2-(N-morpholino)ethanesulfonic acid), malic acid, Tris
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
SDS	Sodium dodesyl sulphate
sfGFP	Superfolder green fluorescent protein
SPAAC	Strain-promoted azide-alkyne cyclo-addition
TAE	Tris, acetate, EDTA
TEMED	Tetramethylethylenediamine
TLS	Translation – libration – screw
Tris	Tris-hydroxymethyl-aminomethane
tRNA	Transfer RNA
uAA	Unnatural amino acid
v/v	Volume per volume
wt (WT)	Wild type
w/v	Weight per volume

Contents

1. Introduction

1.1. Protein function and the genetic code

The central dogma of molecular biology describes the relationship and flow of information between DNA, RNA and proteins [1]. It outlines how DNA is used as a storage system, encoding all the critical information required by the cell until it is needed. RNA then acts as an interface, used to access the correct information at the correct time. The final and most critical part of the dogma outlines how this information is then translated into function by the proteome, performing the tasks required by the cell for survival. There are a vast array of processes needed to ensure survival, from metabolism and energy production, to defense and immunity, and this huge variety in cellular function requires a large and diverse repertoire of proteins.

Critical is the control of protein activity. The ability to regulate protein activity is pivotal to the cell's ability to respond to the changes presented to it. The system used to access protein sequence, and hence function, using both DNA and RNA offers a level of coarse control over temporal and spatial action (e.g. regulation of gene transcription). However this can be relatively slow to implement the effects of the loss or gain of a particular protein. A more rapid, direct and tunable response can be achieved by modulating activity directly at the protein level. Many proteins can be regulated through post-translational modification, which ranges from proteolytic cleavage (e.g. TEM β-lactamase signal sequence processing – [2]) to adduct addition (e.g. phosphorylation – [3]) to covalent rearrangement (e.g. GFP chromophore formation – [4]).

Despite the incredible diversity in protein structure and function, all proteins are synthesized using the 20 canonical amino acids, with the rare exceptions of pyrrolysine [5] and selenocysteine [6]. This limited diversity is despite the availability of 61 amino acid encoding trinucleotide codons (three codons are designated as stop codons). However, the genetic code is degenerate, with some trinucleotide codons coding for the same amino acid [7]. Arginine for example uses six different codons, isoleucine uses three, and phenylalanine uses two. This system of redundancy, whilst ensuring maximum fidelity (single base changes at the DNA level often result in

silent mutations – particularly at the third base of the codon), reduces the amino acid repertoire available to the cell.

This limitation on the number of amino acids that can be incorporated into proteins *in vivo* limits the chemistry that can be sampled. Given the diversity of organic chemistry, the 20 canonical amino acids only contain very rudimentary chemical groups and reactivity. Amine, methyl and hydroxyl groups offer some basic chemical functionality, and more complex chemical groups such as thiol, indole and carboxyl groups are present in some amino acid side chains, however these still represent only a tiny proportion of the physiologically stable chemical moieties that could be accessed within proteins. Alkenyl, carbonyl and imine groups are just three examples of basic chemical structures not present in any of the canonical amino acids. The opportunity to add further reactivity via other elements is also missed, as amino acids are largely composed of only carbon, nitrogen, oxygen and hydrogen (with the addition of sulphur in cysteine and methionine). These factors combined mean the chemical diversity of proteins, despite their functional and structural diversity, is actually quite limited. Thus, proteins often have to recruit new chemistry from other sources via post-translational modifications.

1.2. Post-translational modifications
The restricted chemical diversity sampled by the proteome is often expanded to an extent by the addition of specific chemical groups to particular amino acids, or amino acid sequences, in the form of post-translational modifications [8]. Phosphorylation is a common post-translational modification that adds a phosphate group to specific residues such as serine and threonine, resulting in increases or decreases in protein activity, depending on the target protein and the position of the modification. Phosphorylation provides a method of fine-tuning protein activity without large-scale protein production/degradation; the cells can respond quickly to changing conditions using a common metabolite (normally ATP). As well as phosphorylation, proteins can be acetylated, methylated and glycosylated, all of which have different effects on different proteins. Lysine acetylation of histones results in chromatin loosening and is associated with DNA transcription [9]. Methylation, although more commonly associated with DNA, can also occur to proteins including histones [10]. Glycosylation can be used to aid protein folding, increase protein stability, or is used
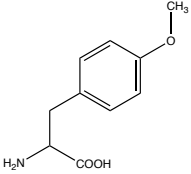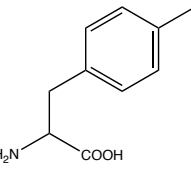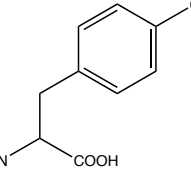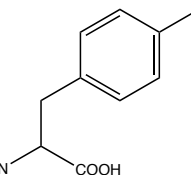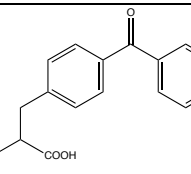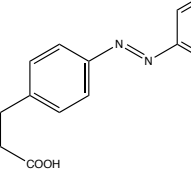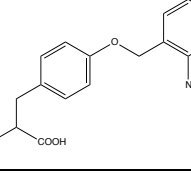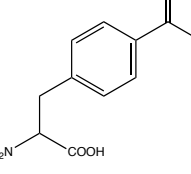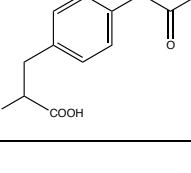
in cell-cell interactions [11]. Other mechanisms of control are available such as co-factor binding, protein-protein interactions, and pro-protein processing to modulate protein expression and activity.

Post-translational modifications have evolved to be highly accurate and reliable to help proteins function properly when and where they are required. However the fine-tuning of protein activity is very complex, often needing associated protein machinery to complete the modification [8]. Post-translational modifications can also be used in conjunction with other methods of protein control, adding a further level of control and complexity. This means that although post-translational mechanisms of protein control are used extensively in nature to modulate the activity of specific proteins in real time, natural post-translational modifications cannot be harnessed to modulate the activity of a single target protein in the laboratory, due to its diverse effects on the entire proteome.

1.3. Unnatural amino acids

One way to overcome the limited range of chemical properties sampled by proteins, and introduce novel post-translational modifications is to expand the amino acid repertoire available to the cell. Over recent years many unnatural amino acids (uAAs) have been synthesized that have a variety of chemical properties (some of which are shown in Table 1-1).

Table 1-1. Some of the uAAs that have been incorporated into proteins

Name	Structure	Feature	Ref.
o-methyltyrosine		First uAA genetically encoded	[12]
p-iodophenylalanine		Used in X-ray crystallography	[13]
p-cyanophenylalanine		Probe for IR imaging	[14]
p-azidophenylalanine		Photo-reactive and Click chemistry	[15]
p-benzoyl phenylalanine		Photo-reactive	[16]
p-azophenyl phenylalanine		Photo-isomerisable	[17]
o-nitrobenzyltyrosine		Photo-caged	[18]
p-acetylphenylalanine		Chemical modifications and labelling	[19]
p-carboxymethyl phenylalanine		Analog of phospho-tyrosine	[20]

For example, uAAs can be designed to mimic some of the specific post-translational modifications that occur naturally, such as phosphorylation and methylation [12]. Incorporation of a phosphotyrosine mimetic into a transcriptional activator was used to determine the effect that tyrosine phosphorylation had on protein activity and DNA binding [20]. Similar experiments have been performed using a phosphoserine uAA to analyze the effect of serine phosphorylation on protein activity [21]. uAAs can also be used to incorporate some of those chemical groups that are missing from the 20 canonical amino acids. The keto chemical group can be incorporated into proteins using uAAs, and then used to post-translationally modify the protein with several biomolecules [19].

The useful reactivity of some uAAs not present in the native amino acid set but compatible with biological systems can be utilized to introduce novel reactivity into proteins. One example is the addition of photo-reactive uAAs into normally non-photo-reactive proteins. This provides the opportunity to modulate protein activity using light. Light is an ideal method to control protein activity, as it is a gentle, non-invasive and highly compatible effector but can provide high spatial and temporal resolution so that specific areas of the cell can be targeted [22]. uAAs that can alter their structure upon illumination with light include photo-caged residues that can release part of their side chain upon irradiation [18], photo crosslinkers [16], and photo-isomerisable residues that can switch between *cis* and *trans* bond formations [17]. A photo-caged tyrosine was incorporated into an enzyme with a tyrosine in its active site to allow light-sensitive activation of the enzyme upon release of the photo-cage due to irradiation at a defined wavelength [18]. Photo-crosslinking using uAAs is a powerful technique to study many facets of protein behaviour. The uAA *p*-azido-L-phenylalanine, containing the photo-reactive azide chemical group, has been incorporated into several proteins to study protein-protein interactions [15], as well as the intrinsic properties of proteins [23]. Other uAAs utilize the light-sensitive benzoyl group and have been incorporated into multiple proteins allowing crosslinking upon photo-activation with both other proteins and DNA. It was used to probe the specific interactions between the subunits of the RNA polymerase II transcription initiation complex [24] and to probe the DNA binding of the catabolite repressor protein [25]. Isomerisable chemical structures have historically been used in membrane biology to introduce light-mediated control of membrane proteins [26, 27], however the

incorporation of the azobenzene group into a uAA has increased the versatility of this method, and has introduced light sensitive, and reversible, modulation of cellular proteins *in vivo* [17].

1.4. Incorporating uAAs

Whilst the use of uAAs as a means to introduce novel functionality into a target protein is attractive, this is only feasible if incorporation of the uAA is efficient and reliable. Originally, uAAs could only be incorporated into target proteins by chemical synthesis [28], however this is a slow and expensive process. Over recent years, many advances have been made that allow the incorporation of uAAs co-translationally, utilizing the natural machinery for protein synthesis [29]. To do this, several additions to the cellular machinery are required, and some modifications to the genetic code are needed to ensure uAA incorporation is effective. The successful translation of mRNA into a peptide sequence requires the accurate and reliable interactions between an amino-acyl tRNA-synthetase (aaRS) and its cognate transfer RNA (tRNA), and between the tRNA and the codon (Figure 1-1). For each of the 20 canonical amino acids there are specific tRNAs that work with a cognate aaRS to ensure maximum fidelity during translation [30]. The sequences of each tRNA differ from each other in several ways, including their anticodon that is used to align with the trinucleotide codons of the mRNA, and their 5' and 3' termini. In addition, prokaryote tRNAs have a long variable arm that aids recognition between the tRNA and the aaRS [31]. The aaRS enzymes differ in their amino acid substrate-binding site, all of which have evolved to bind only a single amino acid. Incorporating uAAs *in vivo* requires several adaptations to be made to this endogenous system so that the uAA can be reliably incorporated at defined positions, without perturbing the existing machinery. Firstly, the uAA needs to be assigned a codon that would result in minimal disruption to the existing genetic code. Of the 64 possible trinucleotide codons, 61 already code for an endogenous amino acid. This leaves only the three stop codons; TGA (opal), TAA (ochre) and TAG (amber). In *Escherichia coli*, the amber stop codon is used in only 7% of genes [32], which makes reprogramming the TAG codon the most desirable option. Very recently, whole genome approaches such as the removal of all TAG codons from the *E. coli* genome [33], or the recoding of TAG into a sense codon by removal of release factor-1 [34] have been used to completely eliminate the amber stop codon and increase the efficiency of uAA incorporation in response to TAG.
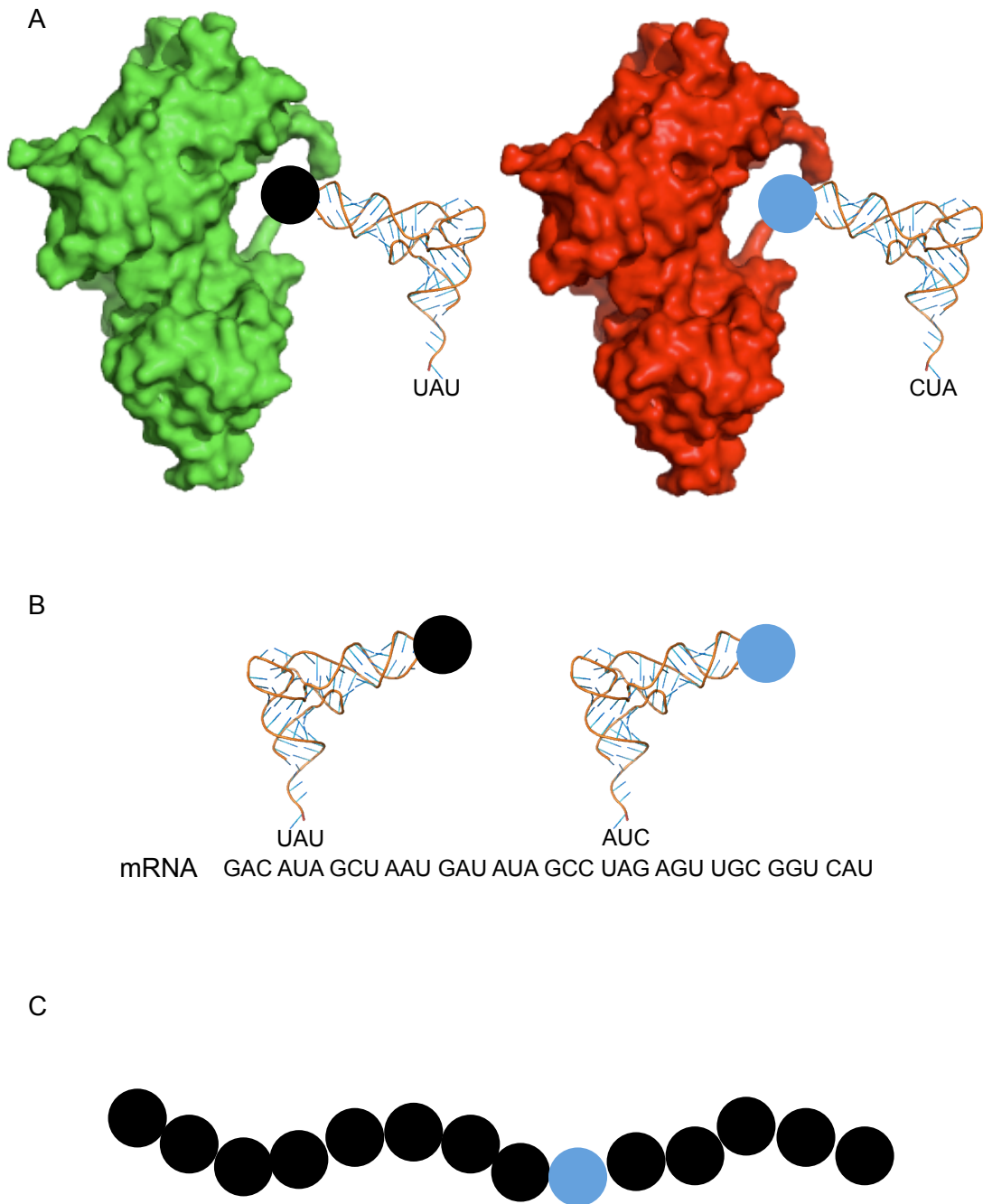
Figure 1-1. Schematic of the incorporation of unnatural amino acids into proteins. (A) The orthogonal amino-acyl tRNA-synthetase (red) charges its cognate tRNA with the unnatural amino acid (blue) during translation whilst the endogenous synthetases (green) charge their cognate tRNAs with the appropriate amino acid (black). (B) The orthogonal tRNA lines up with the amber stop codon in the mRNA. (C) The unnatural amino acid is incorporated into the linear sequence of amino acids.

Once a codon has been assigned for uAA incorporation, a corresponding tRNA that can recognize it is required. Initial attempts to incorporate uAAs into proteins *in vivo* were based around modifying endogenous tRNA/aaRS pairs so that they would no longer recognize their canonical amino acid, but would incorporate the uAA in response to the TAG codon instead [35]. This involved mutating specific bases of the glutaminyl tRNA (along with its anticodon) so that it had reduced affinity for its cognate aaRS. A library of mutant aaRSs were then tested to generate an enzyme that efficiently charged the mutant tRNA. This was completed successfully, however there remained a high level of endogenous amino acids incorporated in response to the TAG codon. Failure of this approach prompted the use of tRNA/aaRS pairs from other organisms as a method to incorporate uAAs. The first pair to be tested was the glutaminyl pair from *Saccharomyces serevisiae* (*Sc*). The *Sc* tRNA/*Sc* glutaminyl aaRS pair was found to work efficiently during translation in *E. coli* [36], however attempts to incorporate a uAA in place of glutamine once again proved impossible. The third strategy was to use a tRNA/aaRS pair from an archaeal species of bacteria, and the shift in focus from eukaryotic organisms to prokaryotic was accompanied by a shift in focus from the glutaminyl tRNA/aaRS pair to the tyrosyl tRNA/aaRS pair. The tyrosyl tRNA/aaRS pair from *Methanococcus jannaschii* (*Mj*) was found to efficiently incorporate tyrosine in response to the TAG codon (the anticodon was mutated to CUA) in *E. coli* [37]. The *Mj* aaRS was also found to be unable to amino-acylate any *E. coli* tRNAs, due to a difference in the 5' and 3' base pairing that is one of the elements crucial in tRNA/aaRS recognition [38], however the *E. coli* aaRS was found to amino-acylate the *Mj* tRNA. This work provided the basic elements that were required to enable uAA incorporation *in vivo*.

Before this tRNA/aaRS pair could be used effectively, several changes to the tRNA and the aaRS had to be made. The anticodon of the *Mj* tyrosyl tRNA was mutated to CUA to correspond to the amber stop codon, and specific nucleotides in the tRNA anticodon loop and other non-conserved nucleotides were randomized to reduce the recognition between the *Mj* tRNA and *E. coli* aaRSs, but preserve the recognition with other factors required for *in vivo* translation such as the ribosome and elongation factor Tu (EF-Tu) [12]. The mutants were then screened using a "negative and positive" selection process to identify a tRNA variant with reduced affinity for *E. coli* aaRSs. The negative selection was performed first without a cognate aaRS from *Mj*,

and involved the suppression of amber stop codons introduced into barnase, a lethal RNase. This selection step ensured any tRNA variants that were charged by *E. coli* aaRSs were removed from the library, because amino-acylation of the tRNA would result in the production of barnase leading to cell death. Any surviving cells contained tRNA variants that were either orthogonal to *E. coli* aaRSs or were non-functional. Non-functional tRNA sequences were then removed from the library in the positive selection, performed along with a cognate aaRS from *Mj*. This step required the suppression of an amber stop codon in β-lactamase, required for resistance against ampicillin. Surviving cells after the second step contained tRNA variants that were functional, but only functional when paired with their cognate aaRS.

A similar strategy was then used to engineer the *Mj* tyrosyl aaRS to switch its affinity from tyrosine to the chosen uAA (in the first instance this was *o*-methyl-tyrosine – [12]). Using available structures for aaRSs from other organisms as models [39], selected residues responsible for defining amino acid specificity were randomized (Figure 1-2). The resulting mutant library was then screened using an alternating positive and negative selection strategy. The positive selection (performed first) involved incorporating uAAs in response to amber stop codons inserted into chloramphenicol acetyl transferase, which confers bacterial resistance to chloramphenicol. This selected for aaRS variants that were able to charge its cognate tRNA with an amino acid, whether it was unnatural or not. The negative selection involved barnase and was performed in the absence of the uAA. This selected against aaRS variants that incorporate any of the 20 canonical amino acids in response to the amber stop codon, as incorporation of a canonical amino acid resulted in barnase production and cell death. The result of multiple rounds of positive and negative selection were aaRSs that charged the mutated tRNA with the desired uAA in *E. coli,* and none of the canonical amino acids. This process has been repeated for many different synthetases to optimize its recognition of a single uAA [40]. This has resulted in the generation of a library of synthetase sequences (Figure 1-2), which efficiently bind a single uAA at the expense of all 20 canonical amino acids, as well as other unnatural ones. Recent work however has highlighted the promiscuity of some of these evolved aaRSs. One study found aaRSs that still had a high affinity for tyrosine, despite mutations to change its substrate specificity [41], while another found an aaRS that could recognize and incorporate up to 18 different uAAs [42].
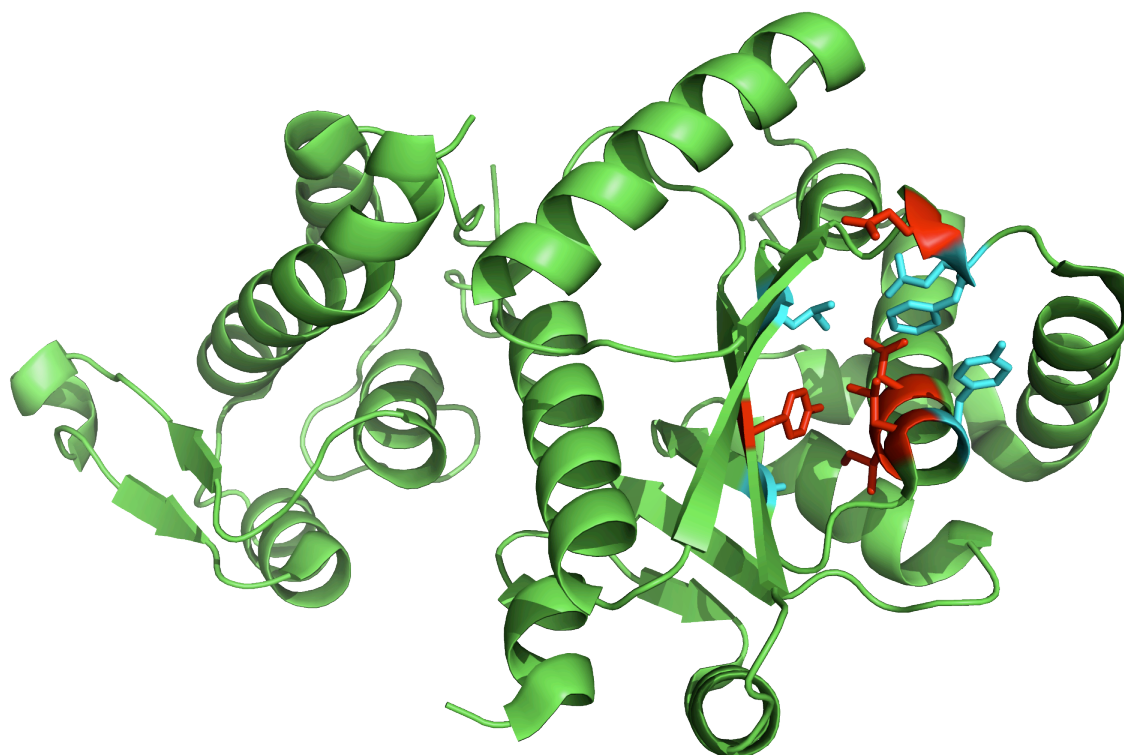
```
Tyrosyl   MDEFEMIKRNTSEIISEEELREVLKKDEKSAYIGFEPSGKIHLGHYLQIKKMIDLQNAGF 60
Napthyl   MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSGKIHLGHYLQIKKMIDLQNAGF 60
Acetyl    MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSGKIHLGHYLQIKKMIDLQNAGF 60
Benzo     -DEFEMIKRNTSEIISEEELREVLKKDEKSAGIGFEPSGKIHLGHYLQIKKMIDLQNAGF 60
Cyano     MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSGKIHLGHYLQIKKMIDLQNAGF 60
           **************************** ***************************

Tyrosyl   DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYR 120
Napthyl   DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSSFQLDKDYTLNVYR 120
Acetyl    DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYR 120
Benzo     DIIILLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSPFQLDKDYTLNVYR 120
Cyano     DIIIVLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEWMLDKDYTLNVYR 120
           ****:****************************************** : **********

Tyrosyl   LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGMEQRKIHMLA 180
Napthyl   LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNPAHYQGVDVVVGGMEQRKIHMLA 180
Acetyl    LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNGCHYRGVDVAVGGMEQRKIHMLA 180
Benzo     LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNTSHRLGVDVAVGGMEQRKIHMLA 180
Cyano     LALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNGAHYLGVDVAVGGMEQRKIHMLA 180
           ************************************ *  ****.************
```

Figure 1-2. Structure of the amino-acyl tRNA-synthetase (aaRS) amino acid binding pocket [39] with the initially randomized residues highlighted in red and further randomized residues highlighted in blue. The sequences of the first 180 residues of five evolved aaRSs are aligned and the same residues are highlighted.

After the initial attempts at evolving the glutaminyl tRNA and AARS to incorporate uAAs, all of the subsequent work was done using the tyrosyl tRNA/aaRS pair. As the variety of uAA structures increased, the need for more relevant starting structures increased as well. This has resulted in the evolution of orthogonal tRNA/aaRS pairs that formerly incorporated different canonical amino acids. After the *Mj* tyrosyl tRNA/aaRS pair, the next tRNA/aaRS pair to be exploited was a glutamyl pair [43]. The tRNA was developed using sequence alignments of multiple archaeal tRNA sequences that use glutamic acid as their substrate, and the aaRS was from *Pyrococcus horikoshii*. Evolution of a leucyl tRNA/aaRS pair enabled uAA incorporation in response to the opal stop codon (TGA) and quadruplet codons, as well as the amber stop codon [44]. This required extensive mutation of the anticodon loop to permit four-base recognition. The desire to use quadruplet codons led to the evolution of a lysyl tRNA/aaRS pair that successfully decoded the quadruplet codon, AGGA [45]. This system was shown to work in conjunction with the existing tyrosyl tRNA/aaRS pair to incorporate two different uAAs into the same protein at the same time. The discovery of pyrrolysine and its incorporation in response to the TAG/UAG codon in *Methanosarcina barkeri* [46] sparked interest in the use of this machinery to incorporate uAAs. Since this discovery, the pyrrolysine tRNA/aaRS pair from *M. barkeri* has been used to incorporate large lysine analogues into proteins co-translationally [47]. The evolution of a prolyl tRNA/aaRS pair [48] that successfully incorporated uAAs in response to the amber stop codon, as well as quadruplet codons, further increased the versatility of uAA design, as the structure of proline represents a large change from the structures of tyrosine, glutamic acid and pyrrolysine, whose translational machinery had been previously used. As the largest canonical amino acid, the tryptophanyl machinery is another appealing target for uAA incorporation. Evolution of a trytophanyl tRNA/aaRS pair from *Saccharomyces cerevisiae* was shown to be orthogonal in *E. coli*, and was used to incorporate multiple large uAAs [49]. The development of so many orthogonal tRNA/aaRS pairs gives chemists and biologists a huge library of potential structures to design evermore intricate and elaborate uAAs, enabling the incorporation of a vast array of useful functionalities into proteins.

The directed evolution approach has been applied to other components of the translation machinery as well in order to increase the efficiency of uAA incorporation.

Elongation factor-Tu (EF-Tu) is responsible for delivering the charged tRNA to the ribosome and this can hamper the incorporation of uAAs, especially ones with large aromatic groups, such as those based on the structure of phenylalanine, tyrosine or tryptophan [50]. The binding pocket can be mutated, in a similar manner to the binding pocket of the aaRSs, to increase the ability of EF-Tu to associate with large, aromatic uAAs [51]. One histidine residue was identified as a key component in the recognition and binding of phenylalanine and mutation affected the binding of several large amino acids, but had no effect on the binding of alanine and glycine [52]. These studies show that as well as a tailored aaRS for each uAA, a specific variant of EF-Tu could be used to increase the efficiency and selectivity of uAA incorporation. In addition to modification of EF-Tu, the ribosome has also undergone changes to enable greater incorporation of uAAs by increasing the number of codons available for uAA incorporation. A designer ribosome was created to work alongside the ever-increasing number of tRNA/aaRS pairs that are able to recognize quadruplet codons [53]. This technology allowed the incorporation of multiple uAAs simultaneously in response to several different quadruplet codons and greatly increases the efficiency of incorporating multiple uAAs into one protein.

Most of the work described above was completed in *E. coli*, however uAA incorporation has been developed for use in more complex model organisms as well. The incorporation of uAAs into yeast utilized an orthogonal tRNA/aaRS pair from *E. coli* [54]. This was found to function similarly to the *Mj* tRNA/aaRS pair in bacteria. Using this method, five disparate amino acids were incorporated in response to the amber stop codon. Unnatural amino acid technology was successfully transferred into mammalian cells, when an evolved aaRS from *E. coli* was coupled to a tRNA from *Bacillus stearothermophilus* [13]. This enabled incorporation of a uAA in response to the amber stop codon in mammalian cells, with a fidelity and incorporation efficiency of over 95%. More recently, uAAs have been incorporated co-translationally in *Drosophila melanogaster* [55]; uAA incorporation was targeted to different tissues, different subsets of cells within a single tissue, and to different stages of development. One final hurdle was passed when genetic code expansion and uAA incorporation was introduced into a whole organism. The genetic code of *Caenorhabditis elegans* was expanded to incorporate uAAs in response to the amber stop codon using three

orthogonal tRNA/aaRS pairs, two from *E. coli*, and the pyrrolysyl pair from *Methanosarcina* [56].

Recent advances have also been made that enable the incorporation of uAAs via cell-free protein synthesis, which allows greater yields of proteins but can be expensive. *E. coli* based cell-free expression systems have been utilized to incorporate uAAs using the *Mj* tRNA/aaRS, similar to the *in vivo* system of protein expression. One system used a tRNA derived from a linearized plasmid or PCR product to increase the concentration of orthogonal tRNA available [57], whereas another increased the concentration of orthogonal aaRS available [58]. These methods produced up to 600 μg/ml protein, compared with 5-20 μg/ml produced using traditional cellular based methods [59]. This increase in efficiency has also allowed the incorporation of multiple uAAs into a single protein using cell-free synthesis [60]. Cell-free systems of expression also represent an ideal target for quick screening of proteins containing uAAs due to their ability to quickly generate large quantities of the desired protein [61]. Additionally, due to the nature of using linearized plasmids as expression templates, this technique may be suitable for the further directed evolution of tRNAs and aaRSs for the incorporation of current, and future uAAs, as the products of error-prone PCR can be directly supplied to the expression system.

1.5. p-Azido-L-phenylalanine

One uAA that has been successfully incorporated into a variety of proteins [15, 23], and in a variety of organisms [15, 62], is *p*-azido-L-phenylalanine (AzPhe – Table 1-1). It is a popular choice for incorporation as it has several interesting properties that make it an ideal candidate for controlling proteins post-translationally. It is an analogue of phenylalanine, with an azide functional group (N_3) at the *para* position of the benzene group (a phenyl azide). It is the azide, or more accurately the phenyl azide that makes this uAA particularly useful as it introduces a new set of non-native chemistry into proteins.

Phenyl azides are particularly sensitive to UV light, a feature that can be used to modulate a target protein. Upon radiation with light below 310 nm [15], phenyl azides lose N_2 and convert to a reactive nitrene species with two lone pairs of electrons. This nitrene radical can follow various routes depending on its immediate environment

(Figure 1-3). One route is ring expansion to form a dehydroazepine that can be attacked by a nucleophile (such as an amine group). Alternatively, the nitrene can be protonated to form a nitrenium ion or can insert into carbon-hydrogen bonds [63]. This particular reaction mechanism can result in photo-crosslinking to nearby amino acids in a protein's folded state which can cause local structural changes, leading to changes in protein stability or activity. This feature of AzPhe has been utilized in the past for studying protein-protein and protein-ligand interactions [64]. Similarly, the cross-linking feature of AzPhe has been utilized to aid the functional analysis of native proteins by immobilizing proteins on surfaces for surface plasmon resonance analysis [65]. The crosslinking reaction has been recently characterized by X-ray crystallography [23].
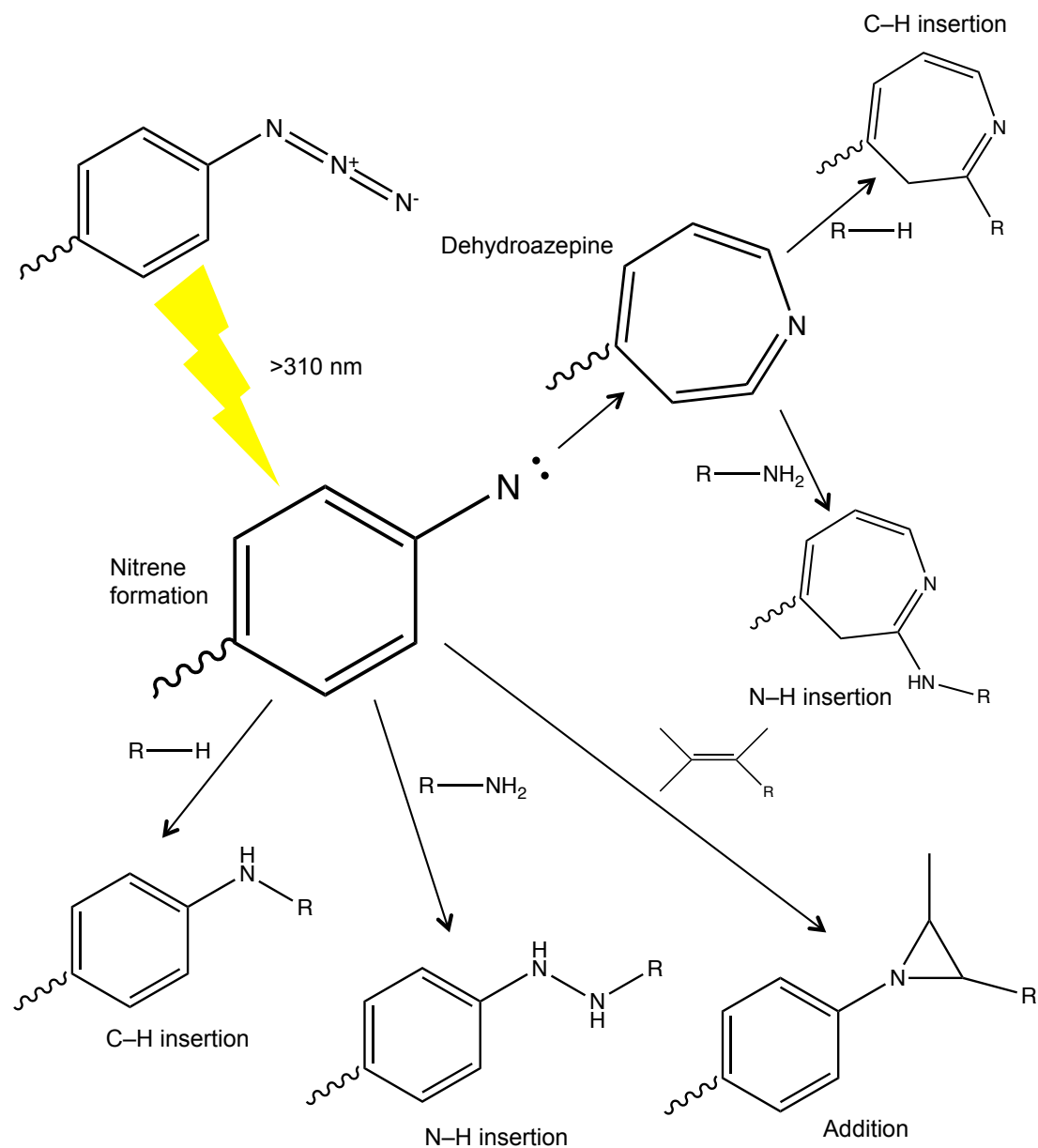
Figure 1-3. Photolysis of *p*-azido-L-phenylalanine (AzPhe). Irradiation using light below 310 nm results in the formation of the nitrene radical caused by the loss of molecular nitrogen. This species can then go down several reaction pathways such as insertion reactions with both N-H and C-H bonds and addition reactions with C=C double bonds. Photolysis at room temperature in physiological conditions promotes the formation of the dehydroazepine, which can subsequently react with nucleophiles.

A second useful feature of AzPhe is Click chemistry. Click chemistry is a term used to encompass a wide range of chemical reactions that generate high yields of non-toxic products and byproducts [66]. In addition, Click reactions are compatible with physiological conditions, are generally quick and end with a single, stereospecific, and physiologically stable product. Therefore, many Click reactions can occur *in vivo*. One such "Click" reaction is strain-promoted azide-alkyne cycloaddition (SPAAC). The phenyl azide chemical handle found in AzPhe is able to undergo highly specific reactions with alkyne groups under ring strain without the need for any catalyst to form a triazole link (Figure 1-4). The original azide-alkyne cycloaddition reactions required cytotoxic levels of copper as a catalyst, but putting the alkyne under ring strain alleviates the need for this catalyst, allowing the reaction to occur under physiological conditions [67]. A huge variety of chemical modifications is available that contain ring strained alkyne groups [68]. This means a single residue in a single protein can be targeted with a wide variety of post-translational modifications that can add a diverse and complex variety of chemical functionality, adding new protein function or modulating existing protein activity. For example, proteins can be modified with two different fluorescent probes in a site-specific manner to enable single molecule FRET (Forster resonance energy transfer), a technique used to interrogate protein structure and dynamics [69].
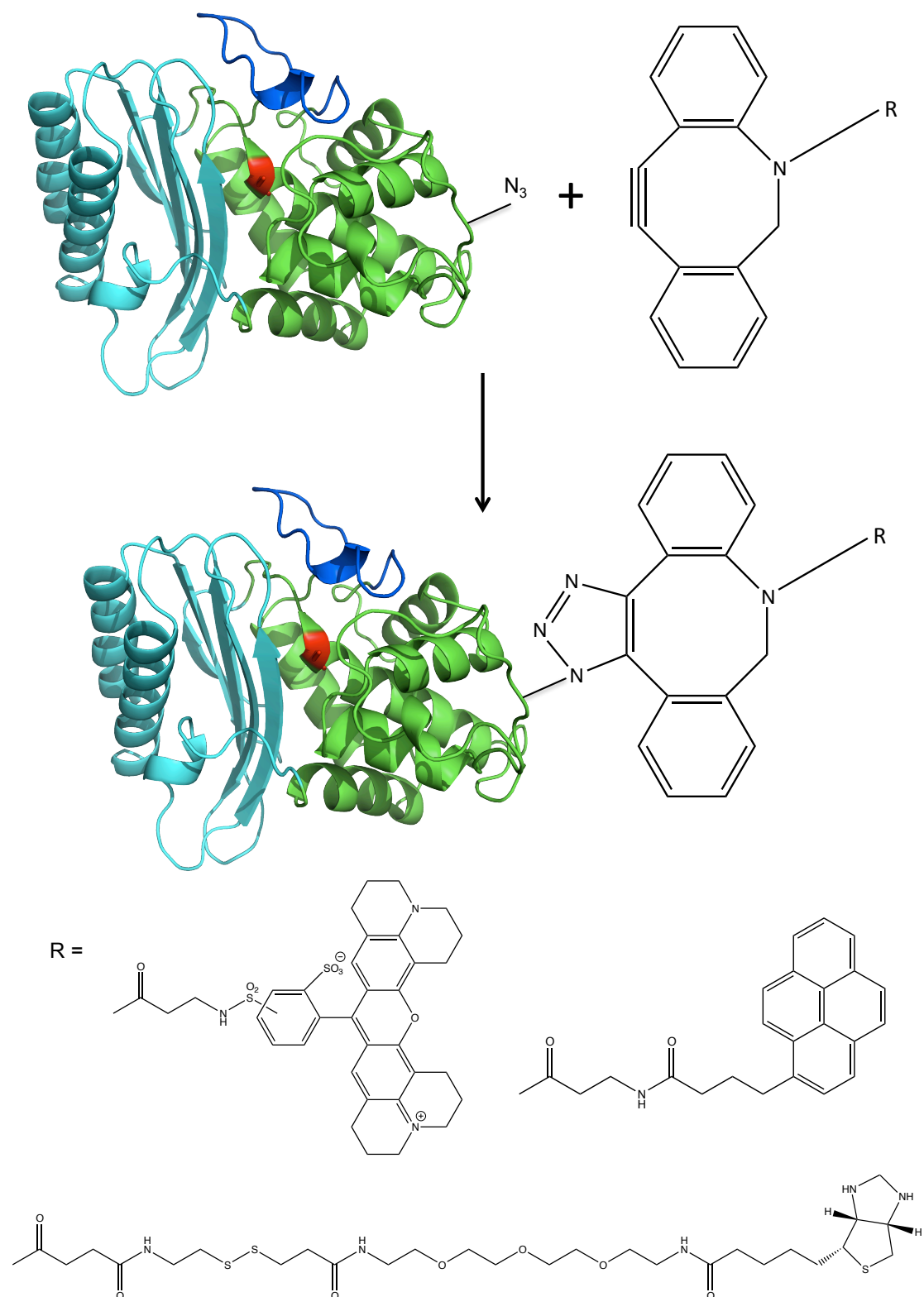
Figure 1-4. Schematic of strain promoted azide-alkyne cyclo-addition (SPAAC) forming a triazole. The dibenzyl cyclo-octyne (DBCO) group can be functionalized with many different chemical groups at the R position such as fluorescent dyes, pyrene and PEG (polyethylene glycol) chains.

1.6. TEM β-lactamase

The β-lactamases are a family of enzymes responsible for the breakdown of many commonly used β-lactam antibiotics such as the penicillins through hydrolysis of the amide bond of the β-lactam ring [70]. β-lactamase enzymes are split into four classes: A, B, C and D. Classes A, C and D hydrolyze the amide ring of the β-lactam ring via a catalytic serine residue whereas class B β-lactamases are metallo-enzymes and require zinc to function [71]. TEM (named after the patient who provided the first sample) is a class A β-lactamase and is the most widely prevalent enzyme responsible for bacterial antibiotic resistance [72]. It is highly efficient at hydrolyzing penicillin-type β-lactam antibiotics such as ampicillin, but is less active against others groups of β-lactam antibiotics such as the cephalosporin family of antibiotics. The structure of TEM is made up of two distinct domains, an α domain (consisting of six α-helices) and an α/β domain that is formed of both the N and C-termini regions (Figure 1-5A). The catalytic residue (Ser70) is located in the α domain on helix 2, with the catalytic pocket formed at the sub-domain interface. The structure of the protein and the position of the active site within the cleft formed by the two domains means there are a number of residues that either directly influence the catalytic process, or are in such close proximity that they contribute indirectly to the hydrolysis reaction and/or substrate recognition and binding (Figure 1-5B). One key feature of the enzyme structure is the omega loop (Ω-loop – residues 161-179), which sits above the active site formed by the two domains (Figure 1-5C). This structure has a major role in substrate specificity and catalysis [73].
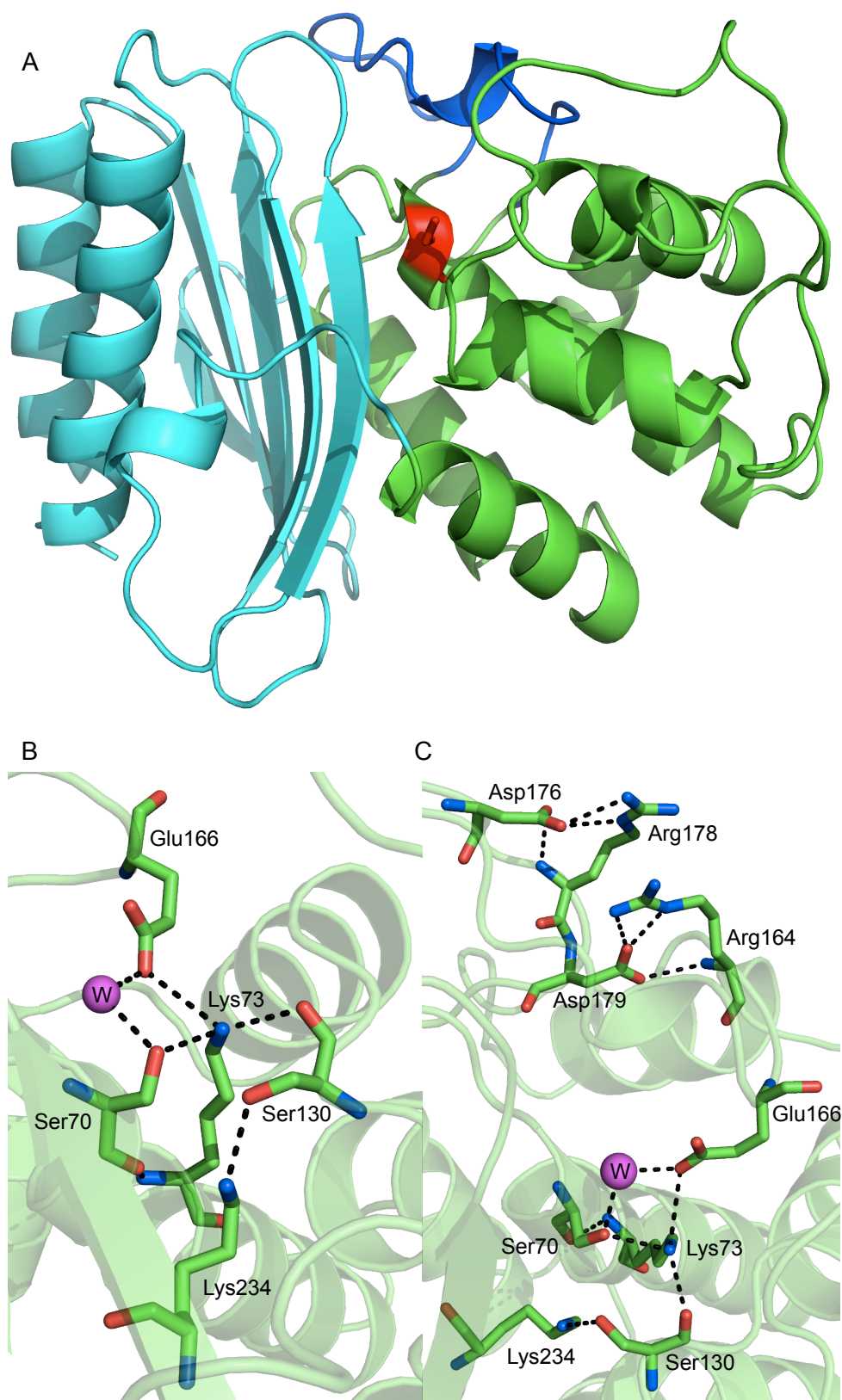
Figure 1-5. (A) Structure of TEM β-lactamase (PDB code 1BTL – [74]) indicating the location of the active serine (red) and the Ω-loop (blue). (B) Residues and interactions involved in catalysis. (C) Residues involved in formation of salt bridges on the Ω-loop

The first step of the catalytic mechanism of TEM β-lactamase is the attack on the β-lactam ring of the antibiotic by the hydroxyl group of the active serine residue (Figure 1-6). This results in the formation of a high-energy intermediate, quickly proceeding to the formation of the briefly stable acyl enzyme as the serine forms an ester with the antibiotic. A hydrolytic water molecule subsequently attacks the acyl-enzyme, resulting in another high-energy intermediate that decays to form the hydrolyzed antibiotic (Figure 1-6). The hydroxyl group on the serine is regenerated and the enzyme is able to hydrolyze more substrate upon product release. Although the general mechanism of action is widely accepted, there remains debate over the precise residue that serves as the general base in the acylation and deacylation steps. It is generally considered to be Glu166, present on the Ω-loop, that functions as the base in the deacylation step of the reaction mechanism, however there is still controversy over that role in the initial acylation step of the mechanism [75, 76]. It is thought the carboxylate group of Glu166 activates the hydrolytic water that attacks the acyl-enzyme, thereby serving as the catalytic base in the deacylation step. This theory has been supported by numerous experimental techniques including site-directed mutagenesis [77]. There have been two mechanisms of acylation proposed however. One suggests the deprotonated form of Lys73 acts as the general base [78], whereas another suggests it is Glu166, acting through a bound water molecule [79]. Apart from Ser70, Lys73 and Glu166, other residues are involved either directly or indirectly in the catalytic mechanism. Ser130 is a part of the "SDN" loop that is found in a loop between helices 3 and 4. These three residues are highly conserved in β-lactamases and are associated with having roles in active site formation and transition state stabilization [80]. Lys234 is part of another conserved trio of residues and is the fifth member of the active site to have a role in catalysis. It has been suggested Lys234 also has a role in transition state stabilization [81].
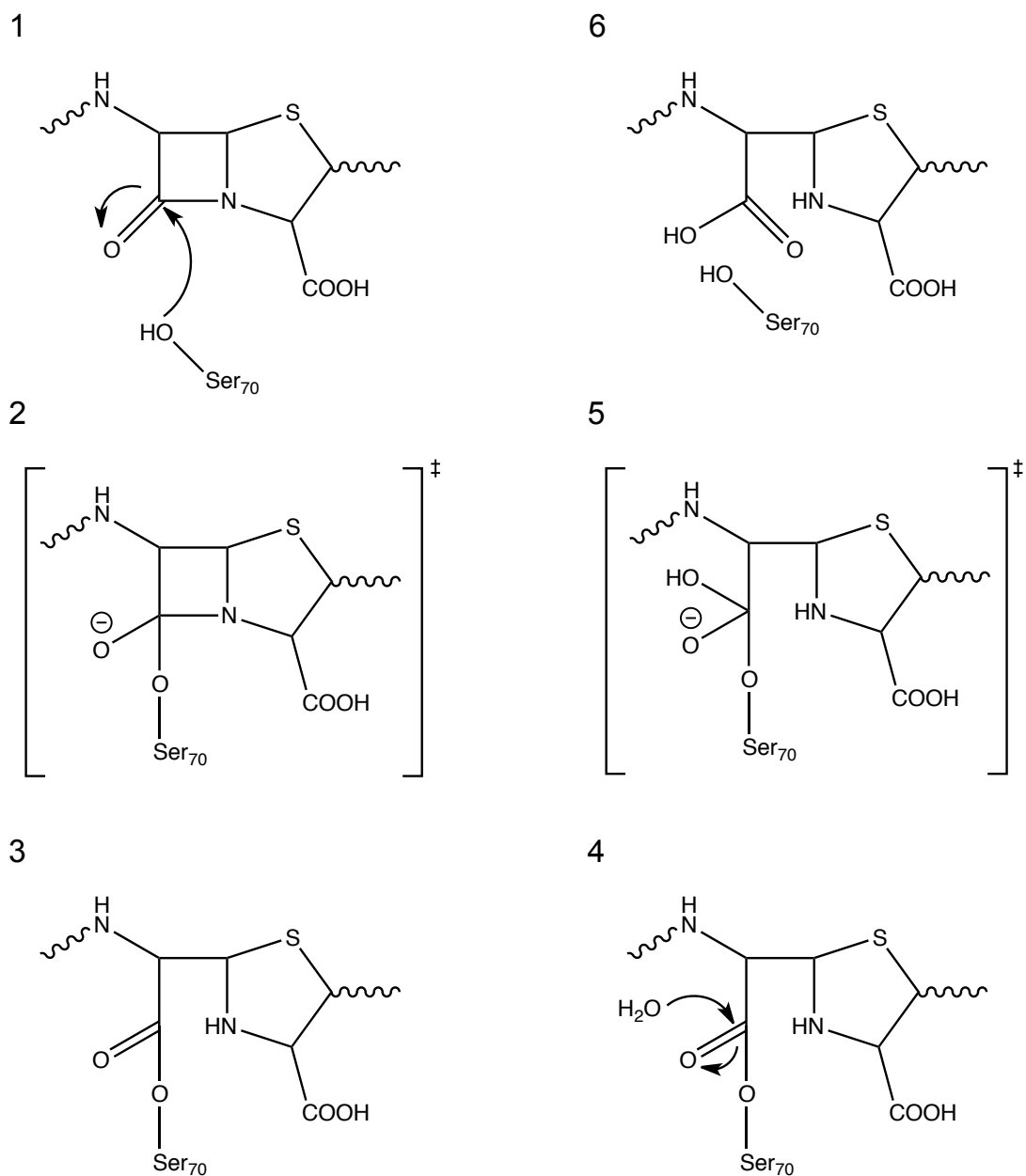
Figure 1-6. Catalytic mechanism of TEM β-lactamase. (1) The enzyme attacks the β-lactam ring of the antibiotic via the hydroxyl group of a serine. (2) This results in the formation of a high-energy intermediate, (3) and then a briefly stable acyl enzyme. (4) A hydrolytic water molecule attacks the acyl-enzyme. (5) This forms a second high-energy intermediate that decays to form the hydrolyzed antibiotic (6).

Due to its role in conferring resistance to some of the most common clinically used antibiotics, β-lactamase structure and function has been investigated extensively. These studies have led to the discovery of "extended spectrum" variants, which are active towards synthetic β-lactams developed to overcome original resistance. The increased ability of TEM β-lactamase variants to hydrolyze synthetic β-lactams such as cephalosporins is permitted largely by the flexibility of the Ω-loop. The Ω-loop comprises residues 161 to 179 [82] and sits over the gap between the two domains of the protein. Four residues, Arg164, Asp176, Arg178 and Asp179, are involved in forming salt bridges that are integral to maintaining the structure of the Ω-loop (Figure 1-5). Several mutations in the Ω-loop are associated with increased spectrum variants of TEM β-lactamase and these include Trp165Ser [83] and Arg164Ser [84]. These mutations, and other mutations [85], not only show that the Ω-loop has a key role in determining extended spectrum activity, possibly by increasing the flexibility of the Ω-loop, enlarging the size of the substrate-binding site, but also that the Ω-loop is tolerant to mutations at key sites in its structure.

1.7. Green Fluorescent Protein

Green Fluorescent Protein (GFP) from *Aequorea victoria* [86] is another protein that has been shown to tolerate the incorporation of uAAs [87], and that novel functional aspects can be accessed with uAAs [23]. GFP is formed by 11 β-sheets that wrap around each other to form a β-barrel, with a single α-helix running through the centre of the barrel (Figure 1-7A). The chromophore is a *p*-hydroxybenzylideneimidazolinone formed by the rearrangement of Ser65, Tyr66 and Gly67, which are found in the middle of the barrel [88]. The rearrangement of the chromophore occurs spontaneously in the presence of O_2 and without any co-factors or other proteins (Figure 1-7B). Since its discovery, GFP has been extensively studied and engineered for new passive and active uses. Its most widely used role is as a passive, genetically encoded probe for monitoring many cellular functions such as protein trafficking and cell migration [89]. As a result of this, the original GFP sequence has been modified and mutated to increase both its fluorescence and its stability under cellular conditions. This includes the generation of two variants in particular that have been widely used due to their improved properties. Firstly, the chromophore was mutated from the original serine to a threonine (Ser65Thr) to

produce a simplified excitation spectrum [90]. This resulted in the generation of EGFP (enhanced GFP), which included one other mutation (Phe64Leu) designed to improve folding at 37°C [91]. EGFP was then the template for directed evolution which resulted in the generation of superfolder GFP (sfGFP), a variant that folds efficiently when fused to other proteins, including those with poor overall stability and solubility [92]. Like EGFP, sfGFP has a major excitation peak at 475 nm, corresponding to the phenolate form of the Tyr66 moiety in the chromophore, and an emission peak at 508 nm.

As previously said, the chromophore of GFP is able to rearrange itself into a fluorescent structure without the need for co-factors or other proteins (Figure 1-7B). The first step is a series of torsional rotations that results in Thr65 being brought into close proximity to Glu67. Secondly, nucleophilic attack on Thr65 by the amide nitrogen of Glu67 results in cyclisation, releasing water in the process. Fluorescence is gained after oxidation of the α-β carbon bond in Tyr66 by molecular oxygen, forming an extended delocalized π-electron system. Critically, only the folded protein is fluorescent; unfolded GFP is no longer fluorescent, despite the presence of the chromophore. A network of hydrogen bonds formed with residues and integral water molecules found within the barrel of the protein supports this structure and helps maintain and define fluorescence. Key residues include Thr203 (whose mutation to tyrosine results in YFP – yellow fluorescent protein [93]), His148 and Glu222. His148, along with Thr203, forms a hydrogen bond with Tyr66 and plays a key role in determining the protonation state of its hydroxyl group [94]. Glu222 fulfills a variety of roles in defining fluorescence by affecting the protonation state of the chromophore and stabilizing intermediates during chromophore maturation [95].
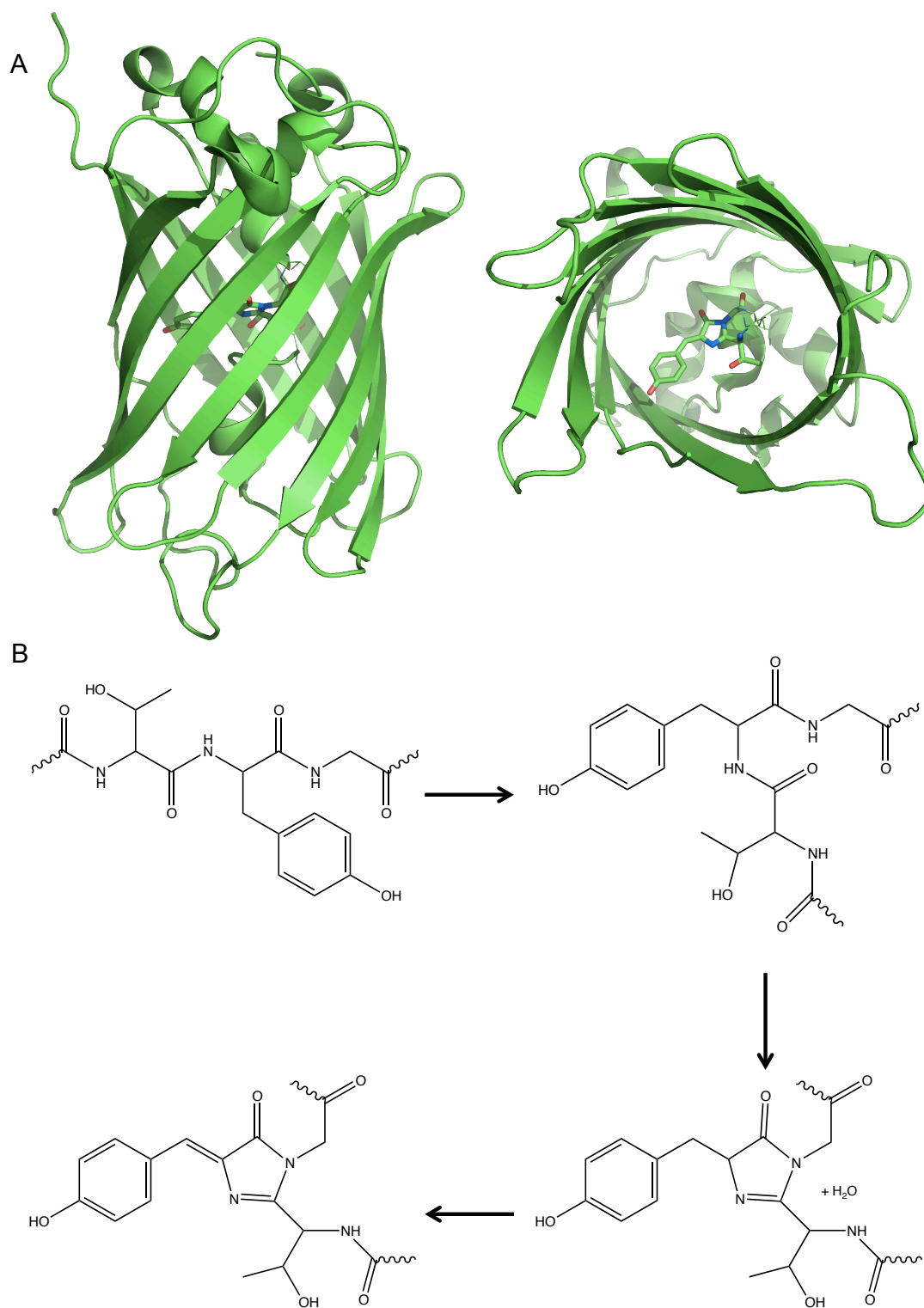
Figure 1-7. (A) Structure of sfGFP (PDB code 2B3P – [92]) consisting of a β-barrel with a single helix in the middle. (B) Maturation of the GFP chromophore. A series of torsional rotations results in Thr65 being brought into close proximity to Gly67. Nucleophilic attack on Thr65 by the amide nitrogen of Gly67 results in cyclisation, releasing water in the process. The α-β carbon bond in Tyr66 is oxidized by molecular oxygen, forming a delocalized π-electron system and fluorescence.

1.8. Aims & Objectives

The limit placed on proteins by using only 20 naturally occurring amino acids restricts the chemical diversity that can be sampled by proteins and ultimately their functionality. The incorporation of uAAs into proteins can be used to lift this restriction. The overall aim of the project is to demonstrate that incorporation of uAAs via a reprogrammed genetic code can introduce novel and useful properties into proteins, which can be subsequently used to modulate protein activity at two levels; firstly, co-translational control through the presence and absence of the uAA, and secondly, post-translational control through the location and modification of the uAA. The latter will be the focus of this thesis.

The first objective is to use TEM β-lactamase as a model protein to show that AzPhe can be incorporated into enzymes and new chemical reactivity can be used to modulate protein function. TAG mutations will be incorporated into the *bla* gene using site-directed mutagenesis and protein containing a uAA at a single, defined position will be isolated. Activity assays will then be used to determine the effect of uAA incorporation on enzyme kinetics. UV irradiation will then be used to activate the azide group of AzPhe and the effect of this process on enzyme activity will also be investigated. Click chemistry will be used to attach a range of chemical groups to the protein at key residues and the impact these have on enzyme activity will be investigated. Click chemistry will also be used to enable defined attachment onto surfaces for use in single molecule studies.

AzPhe incorporation and Click chemistry will also be used to modulate the function of sfGFP. The effect of AzPhe incorporation near the chromophore will be assessed, as well as the effect of UV irradiation and Click chemistry on sfGFP fluorescence. X-ray crystallography will be used to determine the precise molecular events that give rise to any changes in fluorescence caused by AzPhe incorporation and subsequent modification via UV irradiation.

2. Materials & Methods

2.1. Materials
2.1.1. Specialized chemicals

p-Azido-L-phenylalanine (AzPhe) was purchased from Bachem (Bubendorf, Switzerland). It was prepared by suspending the appropriate weight required for a 1 mM concentration in MilliQ water. The amino acid was then dissolved by titrating 1 M NaOH until fully dissolved. This solution was then added to the growth medium to give the final concentration of 1 mM. Nitrocefin (BD, Maryland, USA) for enzyme assays was prepared by dissolving 1 mg in 100 µL DMSO (Sigma-Aldrich). 1.9mL reaction buffer (see Section 2.5.1.1.) was added to generate a working concentration of 500 µg/mL. All buffers/solutions were dissolved in MilliQ water and autoclaved or filter sterilized using a 0.2 µm filter prior to use. For antibiotics used in enzyme assays see Section 2.1.4.

2.1.2. Cells

Escherichia coli cells were used throughout the work for this thesis. Plasmids for use in cloning and mutagenesis were grown in either *E. coli* DH5α cells (Invitrogen, Paisley, UK) or *E. coli* NEB5α cells (New England Biolabs – NEB, Hitchin, UK). Protein production was performed in the *araD* (arabinose catabolism) deficient *E. coli* Top10 cells (Invitrogen).

2.1.3. Media

LB (Lysogeny broth) agar (Melford, Ipswich, UK) was prepared to a concentration of 37 g/L in MilliQ water and autoclaving. LB broth (Melford) was prepared to a concentration of 20 g/L in MilliQ water and autoclaving. Auto-induction medium was prepared in a similar fashion to that described by Studier [96]. Briefly, tryptone and yeast extract (both Melford) were prepared to a concentration of 10 g/L and 5 g/L respectively in MilliQ water and autoclaved. Buffers and chemicals were then added according to Table 2-1. SOC (Super optimal broth with catabolite repression) medium (Melford) was prepared to a concentration of 30.5 g/L SOB (Super optimal broth) in MilliQ water and autoclaving, followed by the addition of 0.4% (w/v) glucose.

Table 2-1. Components used in the auto-induction media

Component	Concentration
Tryptone	1% w/v
Yeast extract	0.5% w/v
Glycerol	0.5% v/v
Glucose	0.025% w/v
Lactose	0.2% w/v
Na_2HPO_4	25 mM
KH_2PO_4	25 mM
NH_4Cl	50 mM
Na_2SO_4	5 mM
$MgSO_4$	2 mM
Trace metals[a]	1x
L-arabinose	0.05% w/v

[a] 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, 0.4 μM H_3BO_3 and 10 μM $FeCl_3$

2.1.4. Antibiotics

Antibiotics for use in cell culture were prepared as follows. Ampicillin and kanamycin (both Melford) were prepared by dissolving 100 mg/mL or 25 mg/mL respectively in MilliQ water and then filter sterilizing using a 0.2 μm filter. Tetracycline (Fluka) was prepared by dissolving 5 mg/mL in 100% ethanol and filter sterilizing using a 0.2 μm filter. Stocks were stored at -20°C. The final working concentrations of ampicillin, kanamycin and tetracycline were 100 μg/mL, 25 μg/mL and 25 μg/mL respectively.

Antibiotics for use in β-lactamase activity assays (see Section 2.5.1.2) were prepared as follows. Ampicillin was dissolved at a stock concentration of 100 mM in MilliQ water, filter sterilized using a 0.2 μm filter and stored at -20°C. Cephalothin (Sigma-Aldrich) was prepared at a stock concentration of 10 mM by dissolving in MilliQ water, filter sterilizing using a 0.2 μm filter and storage at -20°C. The ampicillin and cephalothin stocks were diluted to 10 mM and 1 mM respectively for use in the β-lactamase activity assays.

2.2. Molecular Biology

2.2.1. Oligonucleotides

The sequences of the synthetic oligonucleotides (Integrated DNA Technologies – IDT) used as primers for PCR are found in Table 2-2. The melting temperatures of designed primers were calculated using an on-line Tm calculator, which determines the melting temperature of primers in the particular buffer used (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator).

Table 2-2. Sequences of primers used. Restriction sites are underlined, TAG mutations are capitalized, and His-tags are in bold.

Name	Sequence	Tm (°C)	Comments
Remove ampicillin resistance from pBAD			
AHpBADf	gcggccgcatttatcagggttattgtctc	61	Remove ampR from pBAD
AHpBADr	cagctggtaactgtcagaccaagtttactc	61	
Insert amber stop codons			
AHtem105f	ggttgagtaGtcaccagtcacag	62	Insert TAG at Tyr105
AHtem105r	aagtcattctgagaatagtgtatgcg	64	
AHtem164f	ccttgatTAGtgggaaccgg	65	Insert TAG at Arg164
AHtem164r	cgagttacatgatcccccatg	67	
AHtem237f	aatctggaTAGggtgagcgtg	62	Insert TAG at Ala237
AHtem237r	tatcagcaataaaccagccag	62	
AHtem72f	gatgagcactTAGaaagttctgctatg	62	Insert TAG at Phe72
AHtem72r	attggaaaacgttcttcgg	62	
AHtem179f	cgagcgtTAGaccacgatg	62	Insert TAG at Asp179
AHtem179r	tcgtttggtatggcttcattc	64	
AHtem235f	ctgataaaTAGggagccggtg	64	Insert TAG at Ser235
AHtem235r	caataaaccagccagccg	65	
AHtem65f	ccgaagaaTAGtttccaatgatg	62	Insert TAG at Arg65
AHtem65r	ggcgaaaactctcaaggatc	63	
AHtem174f	gaagccataTAGaacgacgagc	63	Insert TAG at Asp174
AHtem174r	attcagctccggttccc	63	
AHtem165f	gatcgtTAGgaaccggagc	66	Insert TAG at Trp165
AHtem165r	aaggcgagttacatgatcccc	66	
AHtem79f	gcaatgaaaTAGggtgagctgg	65	Insert TAG at Ala79
AHtem79r	catacggaattccggatgag	65	
Remove *bla* from another plasmid			
AH*bla*f	ccatggatgagtattcaacatttccgtgtcg	67	Add *Nco*I site
AH*bla*r	ggtacctta**atggtgatgatgatggtggctg cc**ccaatgcttaatcagtgaggcac	67	Add *Kpn*I site/His tag

29

2.2.2. Polymerase chain reaction

Polymerase chain reaction (PCR) was performed using the Phusion Site-Directed Mutagenesis kit (NEB). All components were mixed according to manufacturer's instructions. Briefly, dNTPs (0.4 mM each), oligonucleotide primers (1 μM each), template DNA (10 pg) and polymerase (1 unit) were mixed in reaction buffer (10 mM Tris-HCl, 0.1% (v/v) Triton X-100, pH 8.8). Thermocycling was performed using a Techne TC-3000 thermocycler, using a 3-step program as recommended by the Phusion DNA polymerase suppliers for performing whole-plasmid amplification (Table 2-3). The annealing temperature was dependent on the specific primers used (listed in Table 2-2), and the extension time was dependent on the length of fragment generated by the reaction.

Table 2-3. Thermocycling procedure for PCR.

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	30 seconds	1
Denaturation	98°C	5-10 seconds	25-35
Annealing[a]	64-72°C	10-30 seconds	
Extension[b]	72°C	15-30 seconds/kb	
Final extension	72°C	5 minutes	1

[a] Annealing temperature was based on the primer used.

[b] Extension time was changed according to the length of the template DNA.

2.2.3. Restriction digestion

Restriction digests of DNA were performed using restriction endonucleases from NEB and completed using the appropriate buffer supplied with the enzyme: NEBuffer 1 (10 mM Bis-Tris Propane-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol, pH 7.0), NEBuffer 2 (10 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM NaCl, 1 mM dithiothreitol, pH 7.9), NEBuffer 3 (50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM dithiothreitol, pH 7.9) and NEBuffer 4 (20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol, pH7.9). The majority of reactions were completed using NEBuffer 4. The restriction digestions were incubated at 37°C for 1 hour using a MJ Research PTC-100 heat block. The amounts of DNA used per reaction ranged from 1 to 50 ng.

2.2.4. Phosphorylation

PCR products were phosphorylated using T4 polynucleotide kinase (NEB) in 66 mM Tris-HCl, 10 mM MgCl$_2$, 1 mM dithiothreitol, 1 mM ATP, 7.5% (w/v) polyethyleneglycol (PEG6000), pH 7.6. The reaction was incubated at 37°C for 30 minutes.

2.2.5. Ligation

DNA fragments and linearized plasmids were ligated using Quick Ligase (NEB) in the same buffer as the phosphorylation reactions (Section 2.2.4). Reactions were incubated at room temperature for 10 minutes. For the insertion of a DNA fragment into a plasmid, a three-fold molar excess of insert was used.

2.2.6. Agarose gel electrophoresis

Agarose gels were made by mixing the appropriate weight of agarose (Melford) to a volume of TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.8) to make the required percentage agarose (0.8 – 1% (w/v)). Ethidium bromide (Melford) was added to a final concentration of 40 ng/mL to visualize the DNA. Prior to loading the gel, loading buffer (2.5% Ficoll-400, 11 mM EDTA, 3.3 Tris-HCl, 0.017% SDS (sodium dodecyl sulphate) and 0.015% bromophenol blue) was added to the DNA sample to be analyzed. Electrophoresis was performed at 120 V until sufficient separation of bands was achieved. DNA size was analyzed by comparison to either the 1 kb DNA ladder (NEB) or the 100 bp DNA ladder (NEB) standards loaded on the same gel. Bands were visualized using a GelDoc-It UV-transilluminator (Ultra-Violet Products, Cambridge, UK).

2.2.7. DNA isolation and purification
2.2.7.1. Plasmid DNA isolation

Isolation and purification of plasmid DNA from cell cultures was performed using the QIAprep Miniprep kit (Qiagen). Cell cultures (5 mL) were centrifuged at 4000 rpm (2800 x g) to create a cell pellet. The pellet was then resuspended and lysed via alkaline lysis using the buffers supplied with the kit. The lysed cells were then spun again at 13000 rpm (15600 x g) to remove cell debris and genomic DNA. The plasmid DNA was extracted from the supernatant by binding to a silica membrane under low salt conditions, and then eluted under high salt conditions.

2.2.7.2. DNA extraction from agarose gel

The extraction of DNA from an agarose gel matrix was performed using the QIAquick Gel Extraction Kit and protocol (Qiagen). The required DNA band was excised from the gel using a sterile scalpel and then extracted and purified according to the manufacturer's instructions.

2.2.7.3. DNA purification from enzymatic reactions

The isolation and purification of DNA from enzymatic reactions such as phosphorylation and ligation was performed using the QIAquick MinElute kit and protocol (Qiagen).

2.2.8. Transformation

Electro-competent *E. coli* cells were transformed with plasmid DNA using a Flowgen Cellject electroporator. A 50 µL aliquot of electro-competent cells were thawed on ice and mixed with the desired plasmid DNA. The quantity of DNA used ranged from 10 pg to 5 ng depending on the source of the DNA (e.g. plasmid DNA or ligation reactions). Cells were pulsed with 2400 V before recovery in 350 µL SOC medium at 37°C for 1 hour. Cells were then plated onto LB Agar plates supplemented with the appropriate antibiotic and incubated at 37°C overnight.

2.2.9. Sequencing

All sequencing was done via the Cardiff University Molecular Biology Support Unit Sequencing Core.

2.3. Mutagenesis modeling

Putative mutations in TEM β-lactamase were modeled *in silico* using a combination of Swiss-Model (http://swissmodel.expasy.org - Swiss Institute of Bioinformatics) and Swiss PDB Viewer (http://spdbv.vital-it.ch) [97, 98]. Residues were mutated *in silico* to tyrosine using the mutation tool in Swiss PDB Viewer. The new sequence was then uploaded using the online server into the alignment mode in Swiss-Model along with the wild type sequence (PDB code 1BTL). The program aligns sequences using iterative least squares algorithms, scoring structures using force field energy and

steric hindrances, and then performs energy minimisation using the GROMOS96 force field [99]. The results were returned in PDB file format and analyzed using Swiss PDB Viewer and PyMol. All images were generated using PyMol.

2.4. Protein production

2.4.1. Recombinant protein production

E. coli Top10 cells containing the two required plasmids (pBAD/pBAK (see Section 3.2.2) containing the gene of interest and pDULE containing the tRNA and the amino-acyl tRNA-synthetase) for protein production were grown in a 5 mL LB starter culture overnight. This starter culture was diluted 1/200 to inoculate 50 mL auto induction medium (Table 2-1) supplemented with 100 µg/mL ampicillin or 10 µg/ml kanamycin (depending on the protein produced), and 10 µg/ml tetracycline. For protein variants containing a TAG mutation, 1 mM AzPhe was added to the culture before incubation. Cultures were incubated at 37°C in a shaking incubator (200 rpm) for 24 hours. Cultures were grown in the dark to prevent photolysis of the azide group.

For the preparation of large cultures (>50 mL), the same method as above was used except for the following. A 10 mL starter culture (per 1 L main culture volume) was used to inoculate auto-induction medium.

2.4.2. Cell lysis

2.4.2.1. Cell lysis using BugBuster

Small cultures were lysed chemically using BugBuster (Novagen) for quick analysis of protein expression. Cells were pelleted by centrifugation at 2800 x *g* in a microfuge. Cells were resuspended in 1 mL BugBuster and incubated at room temperature for 30 minutes in a shaking incubator. The mixture was pelleted by centrifugation at 17000 x *g* in a bench top centrifuge. The supernatant containing the cell lysate was recovered and the pellet was discarded.

2.4.2.2. Cell lysis using sonication

Sonication was used to lyse medium size cultures (50-200 mL). Cells were pelleted as above and then resuspended in phosphate buffered saline (PBS – 100 mM sodium phosphate, 300 mM NaCl, pH 8). Resuspended cells were then exposed to 8 cycles of

10 seconds high frequency sound waves (20 kHz, 40% amplitude) followed by 30 seconds cooling. Cells were kept on ice at all times during sonication.

2.4.2.3. Cell lysis using French press

Lysis by pressure was used for all large cultures (>200 mL) using an American Instrument Company French press. Cells were pelleted and resuspended in PBS (as in Section 2.4.2.2) in an equivalent of a 50-fold dilution of the original culture volume. The cells were then passed through the French press two to three times at 1260 psi to ensure full lysis. Cell lysates were spun at 20000 x *g* and the pellet was discarded.

2.4.3. SDS-PAGE electrophoresis

SDS – polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze crude cell extracts and purified protein as described by Laemmli [100]. Electrophoresis was performed using a mini-PROTEAN 3 electrophoresis system (Bio-Rad). A 15% gel (to resolve proteins between 15 kDa and 60 kDa) was used throughout the work, and was made using the recipe shown in Table 2-4.

Table 2-4. Components in SDS-PAGE stacking and resolving (15%) gels

	Stacking	**Resolving**
Acrylamide/bis-Acrylamide[a]	5% (w/v)	15% (w/v)
Tris-HCl[b]	65 mM	0.375 mM
SDS	0.2% (w/v)	0.1% (w/v)
Ammonium persulphate (APS)	0.1% (w/v)	0.05% (w/v)
TEMED	0.02% (w/v)	0.02% (w/v)

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

[b] pH 6.8 and 8.8 Tris-HCl buffers were used in stacking and resolving gels respectively

Prior to loading onto the gel, the samples were mixed with 5 x loading dye (final concentrations; 2% (w/v) SDS, 0.2 M Tris-HCl pH 6.8, 0.04% (w/v) bromophenol blue, 8% (w/v) glycerol, 10% (v/v) β-mercaptoethanol) and heated at 95°C for 5 minutes. When loading whole cell extracts for analysis, cells were pelleted and then resuspended in 1 x loading buffer (same concentrations as above) to an optical density

of 10 (600 nm), before incubating at 95°C for 10 minutes. Samples were run with a broad-range, pre-stained protein marker (NEB) for estimation of band migration. Gels were stained using Coomassie blue stain (40% (v/v) methanol, 10% (v/v) acetic acid and 0.1% (w/v) Coomassie blue R250 (Fisher)) and destained using 40% (v/v) methanol, 10% (v/v) acetic acid.

2.4.4. Protein purification

2.4.4.1. Nickel affinity chromatography

Nickel affinity chromatography was performed using a HisTrap column (GE Healthcare) connected to an AKTAPurifier (GE Healthcare). Cell lysates were spun at 17000 x g to ensure no cell debris was loaded onto the column. The soluble fraction of lysed cells was loaded onto the column equilibrated with PBS. An imidazole (Fisher) gradient from 0 mM to 250 mM (in PBS) was then applied to the column over 20 column volumes to elute the protein. Eluted fractions were measured for absorbance at 280 nm (and 485 nm for sfGFP) and presence of protein was confirmed using SDS-PAGE and nitrocefin activity assays for β-lactamase activity (see Section 2.5.1.1.).

For small-scale protein purification, His SpinTrap columns (GE Healthcare) were used to purify the target protein. Purification was completed in a benchtop microcentrifuge at 100 x g, as per the supplier's recommended protocol. Briefly, samples were bound to a nickel resin in PBS, washed in a low concentration of imidazole to remove weakly bound proteins from the resin, and then eluted at a higher concentration (500 mM) of imidazole. Fractions were analyzed using SDS-PAGE.

2.4.4.2. Ion exchange chromatography

To further purify sfGFP variants for crystallography, ion exchange chromatography was used after nickel affinity chromatography. Imidazole and NaCl were removed via buffer exchange using 3000 MWCO centrifugal concentrator units (Merck Millipore) prior to purification. The protein was then loaded onto a MonoQ anion exchange column (GE). The protein was then eluted using a concentration gradient of NaCl, from 0 mM to 500 mM over 20 column volumes.

2.4.5. Dialysis

Dialysis was used to remove any unwanted solutes from fractions containing the pure protein. Fractions were pooled and applied to a SpectraPor dialysis membrane (Spectrum Labs). The dialysis tubing used for all pure proteins had a molecular weight cut off of 3500 Da. The tubing was closed at both ends using plastic clips and the membranes were left in a single rinse of 5 litres PBS overnight in a cold room. The buffer was kept stirring throughout.

2.5. β-lactamase activity analysis

All enzymatic activity analyses were performed using a Hewlett Packard 8452A Diode Array Spectrophotometer. Measurements were recorded using a 1 cm pathway in a QS quartz cuvette (Hellma).

2.5.1. Nitrocefin hydrolysis assay

TEM β-lactamase activity in cell lysates was initially analyzed by nitrocefin hydrolysis assay in reaction buffer (10 mM sodium phosphate, 100 mM NaCl, pH 7) at room temperature. Cell lysates (50 μL) were analyzed in 1 mL reaction volume. Reactions were initiated by addition of nitrocefin to a final concentration of between 50-100 μM and monitored by an increase in absorbance at 485 nm. Concentration of hydrolysed nitrocefin was determined using an extinction coefficient of 14060 M^{-1} cm^{-1} [101].

The activity of TEM β-lactamase variants after Click chemistry modification (see Section 2.8.1) was also analyzed using the nitrocefin hydrolysis assay. Click-modified TEM β-lactamase was diluted to a final concentration of 250 ng/μL in PBS. Reactions were initiated by addition of nitrocefin to a final concentration of between 50-100 μM and monitored by an increase in absorbance at 485 nm.

2.5.2. Antibiotic hydrolysis

The kinetics of β-lactamase-dependent β-lactam hydrolysis were determined using ampicillin (235 nm, $\Delta = 1500$ M^{-1} cm^{-1}) and cephalothin (262 nm, $\Delta = 7660$ M^{-1} cm^{-1}) as substrates. Hydrolysis assays were carried out in a 1 mL reaction volume. Purified enzyme was diluted to a final concentration of 250 ng/μL in 50 mM sodium phosphate buffer, pH 8 at room temperature. Reactions were started by addition of

either ampicillin or cephalothin and hydrolysis was measured by the decrease in absorption at the extinction wavelength of each antibiotic. Substrate concentrations ranged from 50 µM to 800 µM for ampicillin and from 10 µM to 150 µM for cephalothin.

2.5.3. Kinetics analysis

Kinetic parameters were calculated using initial rate of hydrolysis at each substrate concentration (n=3) and then fitting to the Michaelis-Menten equation (Equation 1) using GraphPad Prism.

$$v = \frac{V_{max}[S]}{K_M + [S]}$$
<div align="right">Equation 1</div>

This is a non-linear fitting equation that presents values for V_{max} (maximum velocity) and K_M (substrate concentration at which enzyme velocity is half V_{max} – a constant that does not describe strength of binding, but rather a constant that addresses both substrate affinity and substrate hydrolysis). This equation assumes a linear production of product by a single enzyme and that substrate concentration sufficiently exceeds enzyme concentration so that the concentration of free substrate remains effectively the same throughout the assay. It also assumes there is no co-operativity of substrate binding. Catalytic turnover (k_{cat}) was then determined using the calculated V_{max} and the known enzyme concentration (Section 2.5.2.).

$$CE = \frac{k_{cat}}{K_M}$$
<div align="right">Equation 2</div>

Catalytic efficiency (CE) was then determined using Equation 2.

$$\frac{e_{CE}}{CE} = \sqrt{\left(\frac{e_a}{a}\right)^2 + \left(\frac{e_b}{b}\right)^2}$$
<div align="right">Equation 3</div>

Relative errors for catalytic efficiency were determined using Equation 3 where a represents k_{cat}, e_a represent k_{cat} error, b represents K_M and e_b represents K_M error.

2.5.4. Photolysis of proteins

TEM β-lactamase and the AzPhe-containing variants were irradiated with UV light using a UVM-57 handheld UV lamp (UVP) in 50 mM sodium phosphate buffer. Enzymes were irradiated for two minutes at a distance of 1 cm from the sample, unless otherwise stated, sampling a range of UV wavelengths (275 – 380 nm) at 4 watts, as defined by the light source. Enzymes were irradiated at a concentration of 25 μg/mL, prior to dilution for antibiotic hydrolysis experiments.

2.6. Superfolder GFP (sfGFP) fluorescence spectroscopy

Fluorescence spectroscopy was performed using a Varian Cary Eclipse spectrophotometer. Spectra were recorded at a scan rate of 600 nm/min with a slit width of 5 nm. Emission spectra were recorded up to 650 nm from a single excitation wavelength (485 nm). Excitation spectra were recorded by measuring emission at 511 nm over a range of excitation wavelengths from 350 to 511 nm. For whole cell fluorescence, cells were pelleted and resuspended in TNG buffer (50 mM Tris-HCl, 150 mM NaCl, 10% (v/v) glycerol, pH 8) to an OD of 0.1. Pure proteins were analysed at a concentration of 1 μM.

2.6.1. Quantum yield determination

Quantum yields of sfGFP^{His148AzPhe} were determined by comparison with fluorescein, a reference with a known quantum yield at a similar wavelength (496 nm). sfGFP^{His148AzPhe} and fluorescein were diluted to an A_{485} of 0.05 and emission spectra were recorded at the excitation maxima of sfGFP^{His148AzPhe} and integrated between 5 nm above the excitation wavelength and 650 nm, and then inserted into Equation 4.

$$\phi_x = \phi_R \cdot \left(\frac{Int_x}{Int_R}\right) \cdot \left(\frac{\eta_x^2}{\eta_R^2}\right) \qquad\qquad \text{Equation 4}$$

ϕ is the quantum yield, *Int* is the integrated emission spectrum and η is the refractive index of the solvent. The subscripts *x* and *R* refer to the sample and reference respectively. Brightness was determined as a product of molar extinction coefficient and quantum yield.

2.7. X-ray crystallography

Purified sfGFP containing AzPhe at residue 148 was dialyzed into 50 mM Tris, 150 mM NaCl, pH 8, and concentrated to 10 mg/ml. Crystal formation was screened using the sitting drop vapour diffusion method across a variety of pH conditions (4-9 in steps of 0.5) and a variety of $(NH_4)_2SO_4$ concentrations (50-90% saturation in steps of 10%). Screens were set up in duplicate, at 4°C and at 25°C. Drops were set up with equal volumes of protein and precipitant solution (0.2 µl). The only condition to produce crystals was 50 mM MMT (Malic acid, MES, Tris), 60% saturation with $(NH_4)_2SO_4$ (2.52 M), pH 8.5, which yielded two crystals. 1 mM ethylene glycol was added as a cryo-protectant before harvesting one of the crystals. The second crystal was then irradiated for 60 minutes using a UVM-57 handheld UV lamp (UVP). This second crystal was then harvested in the same manner.

Data were collected by Dr. Pierre Rizkallah (Cardiff Medical School, Cardiff University) at the Diamond Light Source, Harwell, UK. Data were reduced using the XIA2 package [102], assigned a space group using POINTLESS [103], scaled using SCALA [103] and merged using TRUNCATE [104]. Structures were solved by molecular replacement with PHASER, using a previously determined sfGFP structure (PDB code 2B3P). Structures were then adjusted manually using COOT and refined by TLS restrained refinement using RefMac. All the above programs were accessed via the CCP4 package (http://www.ccp4.ac.uk/) [104].
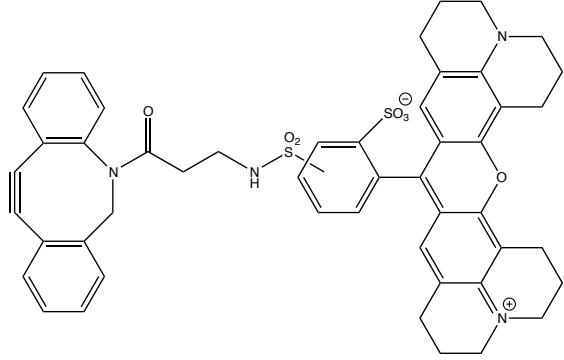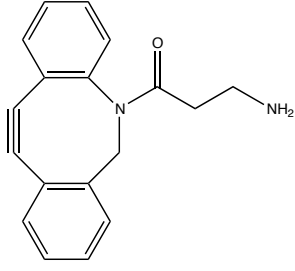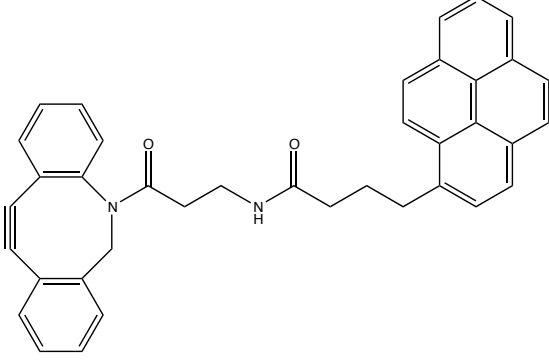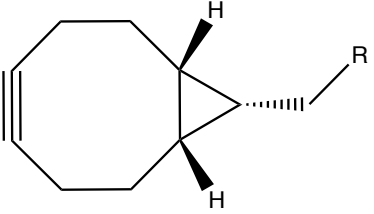
2.8. Click Chemistry

2.8.1. Strain-promoted azide-alkyne cyclo-addition (SPAAC)

A full list of the strained ring alkyne reagents used is shown in Table 2-5. SPAAC reactions were performed on pure protein (5 µM – 15 µM) using a five-fold molar excess of Click modification to protein in PBS. Reactions were left overnight at room temperature (reactions involving DNA strands were left for 48 hours). DBCO-585 and DBCO-amine were obtained from Click Chemistry Tools, while DBCO-pyrene was a gift from S. Reddington. It was synthesized from DBCO-amine and 1-pyrenebutanoic acid, succinimidyl ester (Anaspec) via nucleophilic substitution in an argon atmosphere with TEA. The reaction was ended by addition of aqueous HCl (0.1 M) and extracted using EtOAc. DNA strands were sourced from ATDBio (University of Southampton).

Proteins modified with DNA oligonucleotides were separated from unmodified proteins using ion exchange chromatography. Proteins were applied to a MonoQ anion exchange column (GE). Successfully modified proteins bound to the column via the DNA and were eluted using a concentration gradient of NaCl from 0 mM to 500 mM over 20 column volumes.

Table 2-5. Chemical structures of the strained ring alkynes used for post-translational modification of proteins

Name	Structure
DBCO-585	
DBCO-amine	
DBCO-pyrene	
DNA-oligonucleotides	R1 – GGTGGAGTGATGGATGGGAG R2 – TTTTTTACTCACATTAAT TGCGACCTGTCGTGCCAGCT

2.8.2. Fluorescent imaging of clicked protein

SPAAC reactions with DBCO-585 were analyzed via SDS-PAGE and subsequent imaging of the fluorescent dye. Proteins were separated via SDS-PAGE as per the standard procedure (see Section 2.4.3). Prior to staining, protein gels were visualized using a GelDoc-It (Bio-Imaging Systems). Band intensities were then quantified using ImageJ software (NIH). Gels were then stained with Coomassie blue as per the standard procedure to confirm the same quantity of protein was loaded into each well.

2.8.3. Relative FRET efficiency calculation

The relative FRET efficiency between the chromophore of sfGFP^{His148AzPhe} (donor) modified with DBCO-585 and the fluorophore of DBCO-585 (acceptor) was calculated experimentally using Equation 5, where I_A and I_D are the fluorescence intensities of the acceptor and donor respectively when excited at 400 nm.

$$E_{Rel} = \frac{I_A}{I_A + I_D} \hspace{5cm} \text{Equation 5}$$

This equation determines the ratio between the emission intensities of the acceptor and the donor fluorophores.

2.8.4. AFM imaging of proteins on carbon allotropes
2.8.4.1. AFM imaging of proteins on HOPG

HOPG (highly ordered pyrolytic graphite) was labeled with DBCO-pyrene prior to the addition of protein. Flakes of HOPG were immersed in 0.2 mM DBCO-pyrene in 100% DMSO before washing with 100% DMSO and MilliQ water. TEM β-lactamase was then incubated with the HOPG for 10 minutes before washing as above. Additionally, HOPG was also incubated with pre-modified protein and washed as above. AFM imaging was performed by Ms. Athraa Zaki (School of Physics & Astronomy, Cardiff University). Measurements were taken in air using a Multimode microscope with a Nanoscope III controller (DI Veeco), using the tapping mode with polysilicon probes.

2.8.4.2. AFM imaging of proteins on graphene

Graphene (on a copper base) was labeled with TEM β-lactamase already clicked to DBCO-pyrene (see Section 2.8.1 and Table 2-5). Flakes of graphene were immersed in clicked protein for 15 minutes before two washing steps. Firstly, the flakes were washed in two changes of 100% DMSO for a total of 10 minutes. Secondly, the flakes were rinsed with MilliQ water prior to imaging. Imaging was undertaken as above.

3. *In silico* modeling of TEM β-lactamase mutations and DNA cloning

3.1. Introduction

The incorporation of unnatural amino acids (uAAs) has great potential for introducing new and useful properties into proteins through expansion of the chemistry available. Protein engineering and more recently, synthetic biology utilizes two general strategies to generate protein mutants; directed evolution through the use of libraries comprised of randomly generated mutations [105], or the rational design of site-specific mutations. Rational design requires the use of structural information (either at the linear amino acid sequence level or the 3-dimensional protein level) to aid the selection of the specific targets, because mutating every residue in a protein with every possible combination of mutations is, in practical terms, an impossible task. Additionally, the design of mutants increases the chance of introducing specific properties to proteins. In this chapter, the selection of potential variants was assessed using computer modeling to hypothesize, firstly, the effect of uAA incorporation on the target protein, and secondly to hypothesize the effect of subsequent novel post-translational modifications on TEM β-lactamase activity. This was used in conjunction with existing literature to select a range of residues that when mutated and modified, may impact protein activity, either directly through the active site, or indirectly through interactions with the Ω-loop and the structure of the enzyme.

Once suitable residues have been selected, the successful incorporation of uAAs requires the mutation of endogenous codons to the amber stop codon, TAG. The TAG codon is then reprogrammed through the use of engineered orthogonal tRNA and amino acyl-tRNA synthetase pairs, which incorporate uAAs in response to the TAG trinucleotide sequence (see Chapter 1.4). This enables the uAA to be incorporated into the protein during cellular protein synthesis in a variety of different organisms [54-56]. This chapter will also detail the steps taken to modify existing vectors for use with an auto-inducible system of expression, used for high-yield production of TEM β-lactamase and sfGFP in *E. coli*, the two proteins studied in this thesis, and then the steps taken to introduce the amber stop codon into the target genes.

3.2. Results and discussion

3.2.1. Modeling TEM β-lactamase mutations in silico

3.2.1.1. Using Swiss-PDB Viewer/Swiss Model

A rational design approach was chosen so that specific interactions caused by the incorporation of *p*-azido-L-phenylalanine (AzPhe) – and the subsequent post-translational modification of the novel side chain – could be targeted. To facilitate this design process and residue selection, structural modeling of potential mutations were undertaken using a combination of Swiss PDB Viewer [97], to generate the mutations and visualize their effects, and the Swiss-Model online service [98] to hypothesize the effects that may occur by the substitution of the existing amino acid to a uAA. Although at the time this modeling was initiated it was not possible to use the structure of a uAA such as AzPhe, the structure of tyrosine is very similar, and so makes it a suitable replacement for this purpose. Both AzPhe and tyrosine are large, aromatic residues, with the only difference at the *para* position of the phenyl ring. Tyrosine has a hydroxyl group at this position whereas AzPhe has an azide group (N_3). Individual residues were mutated to tyrosine *in silico* and energy minimization was performed to determine a structure that represents the most likely orientation of the mutated aromatic side chain. The modeling process not only looks at the effect placed on the selected residue alone, but also the local microenvironment surrounding the particular side chain. Mutation to a large, aromatic amino acid such as tyrosine, especially when the canonical amino acid is a small amino acid (e.g. glycine or alanine) can have a large impact on the surrounding residues, and the precise orientation of their side chains.

Residues targeted for mutation were chosen based on three criteria; proximity to residues involved in catalysis, location within the Ω-loop, and/or positions that may have effects on the tertiary structure of the protein such as affecting Ω-loop flexibility, or affecting the formation of the domain interface where the active site lies. A full list of the residues chosen can be found in Table 3-1.

Table 3-1. List of residues chosen for mutation based on the three criteria; proximity to the active site, proximity to the Ω-loop and/or potential effects on tertiary structure.

Residue	Potential interactions	Rationale
Tyr105	Val216	Tyr105 forms a large part of the domain interface near the active site, mutation could reduce the space available for substrates
Ala237	Tyr105	Ala237 is found in the active site and may have a role in substrate orientation during catalysis, mutation could disturb substrate binding
Phe72	Glu166, water	Found close to Ser70, could disrupt Glu166 or a conserved water involved in the catalytic process reducing catalytic turnover
Ser235	Steric clashes	Ser235 is found in the active site near Lys234 which has a role in catalysis, mutation could cause steric clashes due to the larger side chain
Arg164	Salt bridge network	One of the residues that form salt bridges to maintain the structure of the Ω-loop, mutation could disrupt the structure of the Ω-loop
Asp179	Salt bridge network	One of the residues that form salt bridges to maintain the structure of the Ω-loop, mutation could disrupt the structure of the Ω-loop
Trp165	Gly143	Found in the Ω-loop close to Glu166, mutation could induce an interaction with Gly143, reducing the flexibility of the Ω-loop
Pro174	Gly267	Pro174 is found on the Ω-loop and may regulate flexibility by restricting an unstructured loop region, mutation could increase this flexibility
Arg65	Asn175	Found behind the Ω-loop, mutation may result in crosslinking with the Ω-loop, reducing the flexibility of the loop
Ala79	Glu145	Found buried in the core of the enzyme, mutation could result in crosslinking with Glu145 and disturb the positioning of Ser70

The residues directly involved in the catalytic mechanism of the TEM β-lactamase are Ser70, Glu166, Lys73, Lys234 and Ser130 [70]. Any positions that were selected for mutation due to their proximity to these catalytic residues are highlighted in yellow in Figure 3-1. A key feature of TEM β-lactamase is the Ω-loop, formed by residues 161 to 179 [106]. This loop plays a major role in substrate recognition and defines access to the active site, and many TEM β-lactamase variants that hydrolyze synthetic antibiotics (increased spectrum mutants) have residue substitutions in this structure [82]. Residues chosen that are present on the Ω-loop are shown in pink on Figure 3-1. Many mutations that affect enzyme activity do not directly influence the active site, but help by increasing protein folding or stability [107, 108]. This is due to the complex network of interactions present in the tertiary structure of an enzyme that help to position and orientate the catalytic residues. Any disruption in these interactions can have a subsequent affect on activity. Any positions that are not in close proximity to the active site, or are not present on the Ω-loop, but fulfill this criterion are shown in orange on Figure 3-1.
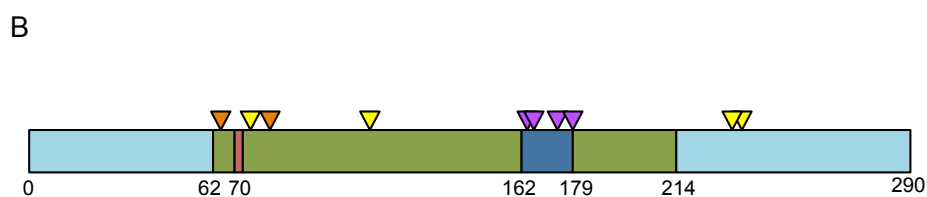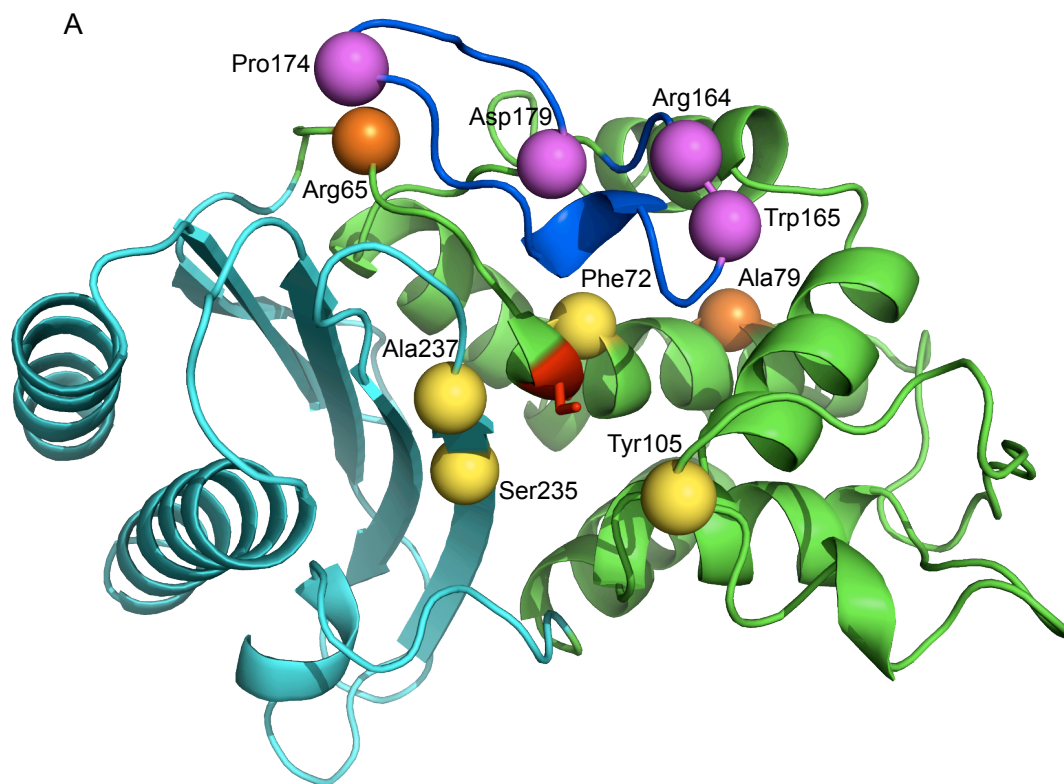
Figure 3-1. Residues of TEM β-lactamase targeted for uAA incorporation. **A**. Structure of TEM β-lactamase highlighting the two-domain structure, with the active Ser70 residue (red) in the domain interface. The α-domain is shown in green and the α-β domain shown in cyan. The Ω-loop is shown in blue. Spheres represent residues chosen for site-directed mutagenesis; (i) due to their proximity to the active site (yellow), (ii) due to their location in the Ω-loop (pink), (iii) due to their effect on the tertiary structure (orange). **B**. Schematic of the structure of TEM β-lactamase showing the two-domain structure, Ser70 and the Ω-loop. This highlights the location of the mutations in the linear amino acid structure, surrounding both the active site and the Ω-loop primarily.

3.2.1.2. Residues directly impacting the active site

The introduction of AzPhe into TEM β-lactamase accesses two potential methods of post-translational control of activity, photo-reactivity and Click chemistry. During the selection of potential mutations, particularly those residues that would directly influence the active site, these two routes of post-translational control were considered when selecting which residues to mutate. Four positions were selected that if AzPhe were to be incorporated, there could be interesting effects on the enzyme activity caused by photo-reactivity and/or Click modification. The first, Tyr105, forms one side of the active site cleft where the catalytic residues sit (Figure 3-2A). Any modification of the structure of the protein here is likely to have significant effects on substrate binding and access to the active site. Modeling was not performed (the residue is already a tyrosine) but the position of the tyrosine and the orientation of its side chain in the structure suggests a potential interaction upon photo-activation with Val216, located on the α-β domain on the other side of the cleft (Figure 3-2A – inset). Any crosslinking here (valine has no nucleophilic properties but insertion into carbon-hydrogen bonds are possible) would affect the size and shape of the cleft, limiting the space for substrates to occupy. Additionally, modification of this residue via Click chemistry could significantly affect substrate binding as well. The addition of any chemical groups (such as those used in Cu-free Click chemistry) to the active site would leave little room for substrates to gain access to the active site, and could interfere with the orientation of residues responsible for hydrolysis and/or substrate binding.
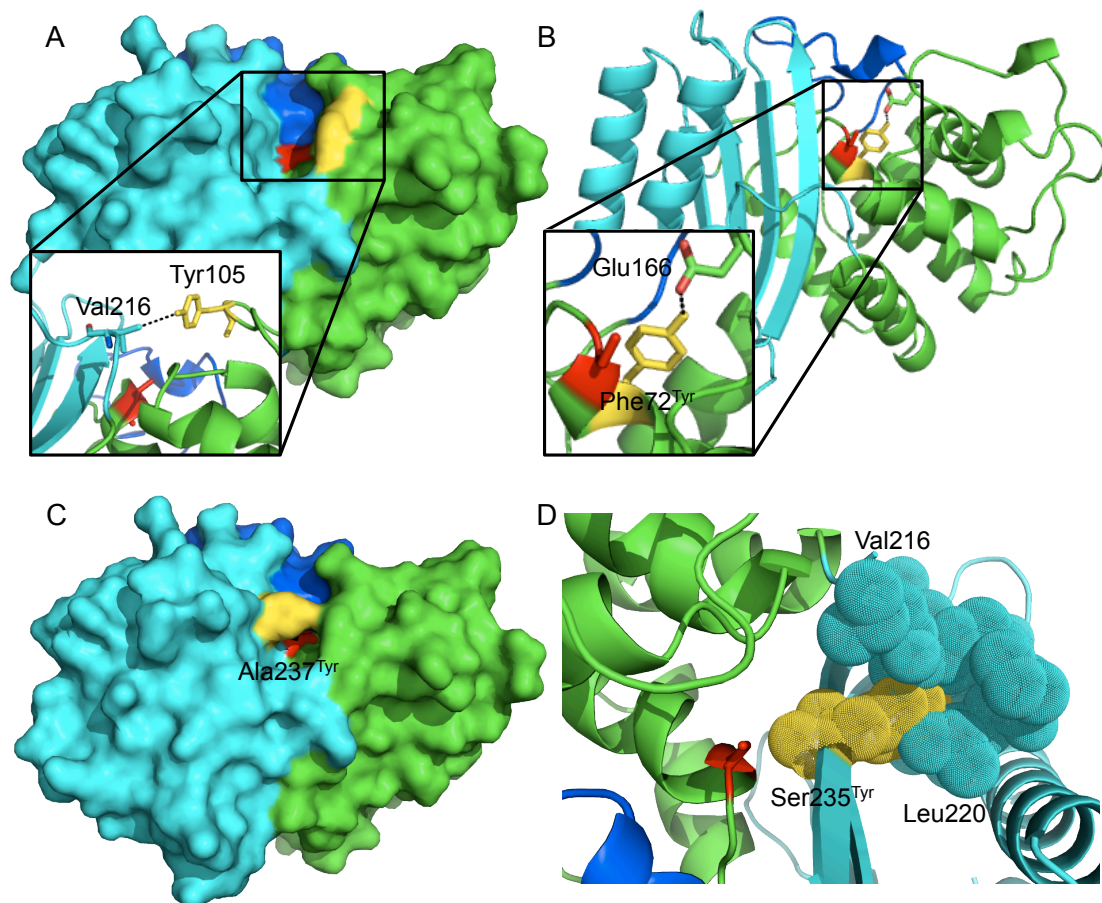
Figure 3-2. **A**. Position of Tyr105 (yellow) in the folded protein. It sits along the cavity formed by the two domains near the active site. The active Ser70 residue is shown in red, the α-domain is shown in green, the α-β domain is shown in cyan and the Ω-loop is shown in blue. Val216 is present on the opposite side of the active site (inset). **B-D**. Modeling mutations at residues (B) Phe72, (C) Ala237 and (D) Ser235. **B**. Modeling a mutation at Phe72 (yellow) revealed potential interactions with Glu166 (inset), which has a key role in the catalytic mechanism. **C**. Mutation to tyrosine at Ala237 highlights the impact a larger side chain may have in this position of the enzyme. **D**. Modeling the mutation at Ser235 reveals potential clashes with surrounding residues. The incorporation of a larger side chain could result in structural changes to accommodate the new side chain.

Phe72 represents a second potentially useful site due to its location two residues away from the catalytic Ser70 residue within the core α-helix. Mutation to tyrosine during modeling revealed a potential interaction with Glu166 upon photo-activation, a key residue in the proton shuttle cascade critical to catalysis (Figure 3-2B). The incorporation of the larger azide group may alter the position of Glu166, affecting this proton cascade. Additionally, upon photo-activation of the azide group, amine formation (see Figure 1-3) could cause an interaction with Glu166, in the same manner as the OH group of the mutated tyrosine (Figure 3-2B – inset). This interaction could also affect the proton cascade and thus activity. Thirdly, the hydroxyl group on the side chain of the glutamate could provide the nucleophile for attack on the activated nitrene species outlined in Chapter 1.5, resulting in crosslinking between the two residues. Alternatively, the conserved water molecule that is present between Glu166 and Ser70, thought to act as a shuttle for the proton during catalysis, could provide the nucleophile required for addition reactions. Interactions with this crucial water molecule could rob the enzyme of this relay system. The location of this residue near the core of the protein would make modification via Click chemistry difficult. Using NetSurfP [109] to estimate surface accessibility, Phe72 has a surface exposed area of 10 $Å^2$ (relative surface area (RSA) of 0.05), compared to 113 $Å^2$ (RSA = 0.53) for Tyr105. This highlights the buried nature of the residue and would suggest post-translational modification of this residue would be negligible. However, it has been shown in previous studies that surface accessibility is not the only factor in determining the likelihood of a successful Click reaction [110]. Other factors that were shown to impact Click modification efficiency were residue flexibility and the microenvironment surrounding a side chain.

The third residue selected was Ala237. This residue sits in the catalytic pocket but has no direct role in the catalytic mechanism, however it has been shown to help orient the β-lactam ring of the antibiotic in the active site [111]. Alanine is one of the smallest amino acids available, so mutation to an aromatic uAA such as AzPhe is likely to have a large impact on the catalytic pocket and the ability for substrates to gain access to Ser70. The estimated orientation of the mutated tyrosine residue in the model suggests the catalytic pocket would be almost entirely closed by this mutation (Figure 3-2C). The side chain points towards Tyr105 on the α-domain and any interaction with this residue caused by photo-activation of the azide group would

completely block off the active site. Any post-translational modifications via Click chemistry would also limit substrate access to the active site and would likely have a negative impact on activity. The addition of chemical groups at this position could act as a novel inhibitor, as any modification may occupy the same space substrates occupy during catalysis. A fourth residue targeted to impact the catalytic mechanism is Ser235. Ser235 is part of a conserved triad of residues including Lys234 (which has a direct role in catalysis – [81]) that have been shown to be involved in the catalytic process [112]. The hydroxyl group of Ser235 has been shown to be important in both penicillinase activity and cephalosporinase activity. The side chain of the modeled tyrosine residue points away from the active site (even though the hydroxyl group of Ser235 points towards the active site), so mutation to a uAA here is likely to greatly impact catalytic activity. The orientation of the modeled tyrosine at Ser235 suggests potential steric clashes with nearby residues that may cause an effect on the positioning of other side chains in the local environment (Figure 3-2D). This could lead to subtle changes in the overall structure of the enzyme. The orientation of the predicted side chain suggests post-translational modification via Click chemistry is unlikely to occur, as the residue is buried in the enzyme with very limited accessibility to solvent (8 Å^2).

3.2.1.3. Residues on the Ω-loop

A further four residues were selected that are present on the Ω-loop of TEM β-lactamase. The first of these is Trp165, one residue N-terminal to the catalytically important Glu166 and one of the residues mutated in increased spectrum variants of TEM β-lactamase [83]. Substitution of this residue with tyrosine during modeling identified possible interactions with the carbonyl oxygen of Gly143 (Figure 3-3A – inset). Any interactions here caused by the photo-activation of the azide group would alter the flexibility of the Ω-loop, a characteristic that is integral to the function of this protein [113]. Glu166 plays a major role in the catalytic mechanism of the enzyme and is present on the Ω-loop, so any changes to the precise orientation of this residue may have a significant effect on activity. The orientation of the side chain in the model suggests post-translational modification would be possible, as the side chain of the modeled tyrosine points away from the core of the protein and is surface accessible, suggesting a high modification efficiency is possible. The addition of large chemical groups via Click chemistry so close to Glu166 and the other catalytic

residues is likely to result in decreases in activity caused by changes in the precise orientation of these residues.

A second residue selected on the Ω-loop is Pro174. Although Pro174 is located on the Ω-loop, it is on the opposite side to Glu166 and the rest of the active site (Figure 3-3B). Proline is a unique amino acid as it is the only secondary amine in the 20 canonical amino acids. Proline is a highly rigid amino acid due to its unique structure, a feature that is likely to be important here. The Ω-loop is largely unstructured, save for a small α-helix, so a proline here is likely to have a role in regulating the flexibility of the loop. Mutation to tyrosine could potentially release some of this rigidity and increase the flexibility of the Ω-loop. Modeling revealed a potential interaction with Gly267 (Figure 3-3B – inset), which could have a further impact on the flexibility of the Ω-loop, and the orientation of Glu166. Due to the location of this residue on the Ω-loop, AzPhe-specific post-translational modifications are unlikely to have an impact due to the distance between the modified residue and the active site.

A key feature of the Ω-loop is a network of salt bridges that are formed between four residues present in the structure [74]. Arg164, Asp 176, Arg178 and Asp179 all interact to stabilize the Ω-loop (Figure 3-3C). These salt bridges stabilize the structure of the Ω-loop and any disruption in these interactions destabilizes the enzyme, leading to a loss of activity against penicillins [114]. Two of these residues were targeted for mutation, Arg164 and Asp179. The initial incorporation of AzPhe will likely disrupt the interactions that maintain the structure of the Ω-loop, but these interactions may be reformed upon photolysis of the azide group and subsequent cross-linking. This could lead to an activity switch built into the enzyme accessible by light. Modeling the substitution of tyrosine for all four residues individually suggested the side chain orientations of Arg164^{Tyr} and Asp179^{Tyr} only would have potential for cross-linking, as modeling suggested the orientation of the phenol side chains at positions 176 and 178 would render interactions unlikely. The models suggest that the mutated phenol side chain at position 164 would point in a similar direction to the side chain of the native arginine residue, whereas modeling the mutation at 179 suggested a slightly different side chain orientation, one that may yield interactions across the loop, with the backbone of Ala171 or Ile172. Modification of these residues via Click chemistry would remove the potential for restoring the interactions, as any suitable chemistry

introduced (such as amine groups that could form interactions with other residues) would be too far from the other residues to form useful interactions due to the size of the modification.
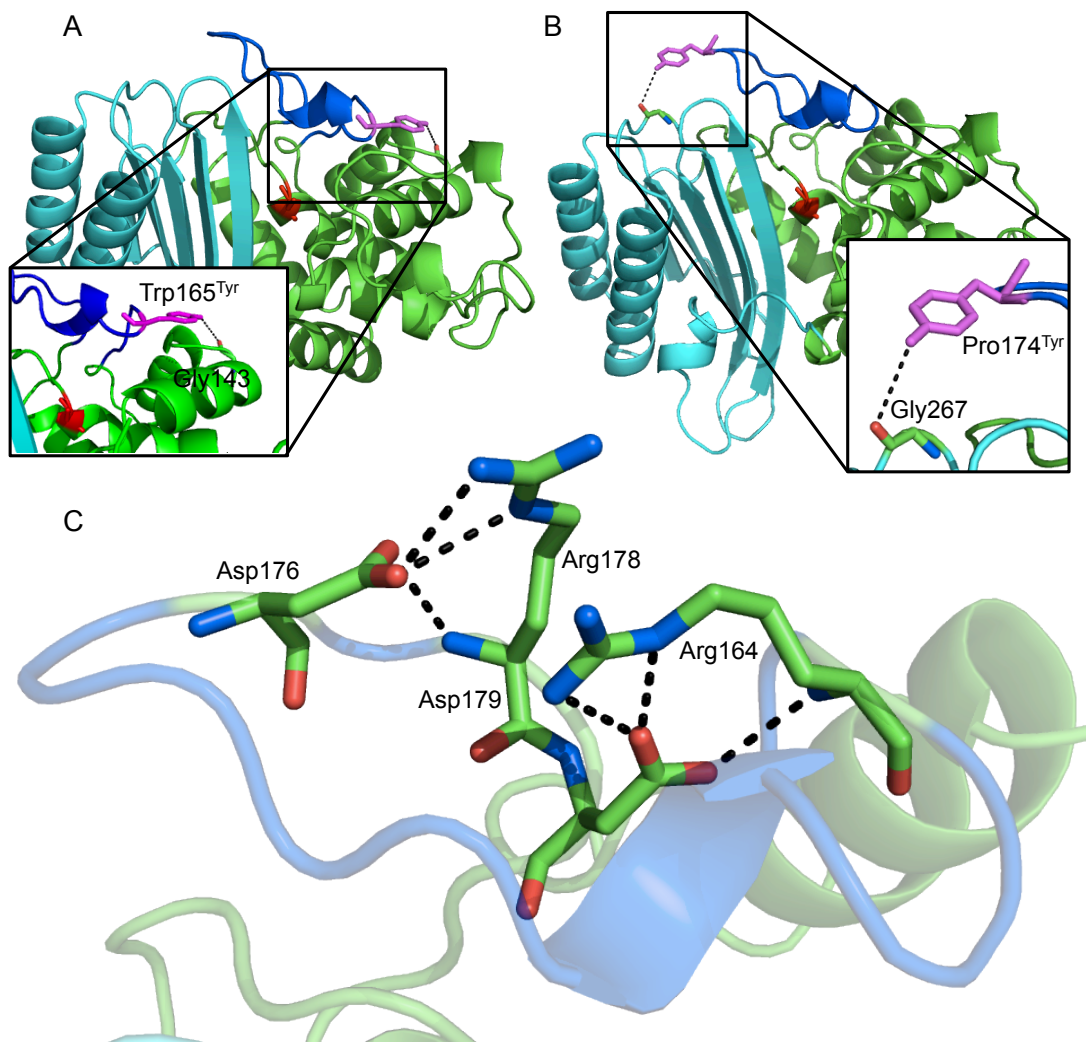
Figure 3-3. Modeling mutations at residues (A) Trp165 and (B) Pro174, and (C) the network of interactions present on the Ω-loop. The active Ser70 residue is shown in red, the α-domain is shown in green, the α-β domain is shown in cyan and the Ω-loop is shown in blue. **A**. The position of Trp165 (pink) in the Ω-loop. **B**. Modeling the mutation at Pro174 (pink) highlighted a potential interaction with Gly267 (inset). **C**. The structure of the four residues that interact to form salt bridges that maintain the structure of the Ω-loop.

3.2.1.4. Other modeled residues

Two more residues were selected in positions that may influence the activity of TEM β-lactamase through indirect interactions. Arg65 is located close to Ser70 in the linear structure but in the 3D structure of the protein is nearer to the Ω-loop (Figure 3-4A). Modeling the substitution of tyrosine for arginine highlighted a potential interaction with Asn175, which is located on the Ω-loop, but on the opposite side to Glu166 and the active site (Figure 3-4A – inset). Crosslinking between these two residues would have an impact on the flexibility of the Ω-loop, a key factor in determining accessibility to the active site. This could have an impact, not only on overall enzyme activity, but also on substrate specificity. A second residue selected for mutation was Ala79 (Figure 3-4B). As alanine is one of the smallest amino acids, the substitution to tyrosine represents a large change in side chain size. Ala79 is in the core of the enzyme and is not very surface accessible (8 $Å^2$) so this change occurs near the hydrophobic core of the protein where there is limited space for the addition of a substantially larger amino acid side chain. These factors suggest this mutation would have an impact on the precise positioning of Ser70, however modeling suggested no change in the orientation of the α-helix, with Ser70 remaining in the same place. This is an ideal outcome as any change in the position of Ser70 would likely result in an inactive enzyme. Modeling also suggests a potential interaction with Glu145 (Figure 3-4B). Glutamate presents an ideal partner for photo-crosslinking with AzPhe and any interactions between these two residues would have an impact on the flexibility of the Ω-loop. The buried nature of this residue means, novel post-translational modifications via Click chemistry are unlikely to occur at a high level of efficiency, making control of protein function in this manner unlikely.
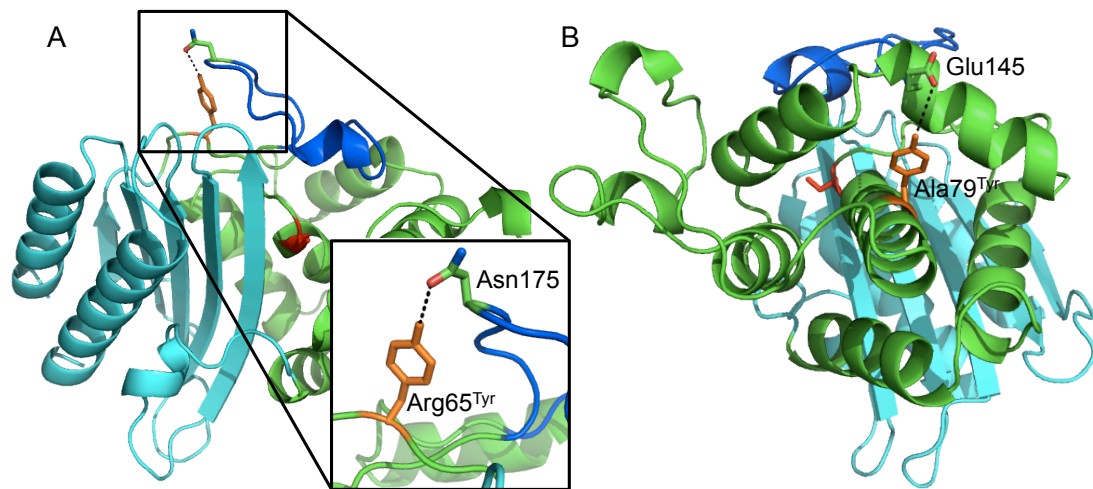
Figure 3-4. Modeling mutations at residues (A) Arg65 and (B) Ala79. The active Ser70 residue is shown in red, the α-domain is shown in green, the α-β domain is shown in cyan and the Ω-loop is shown in blue. **A**. Mutating Arg65 (orange) with tyrosine highlighted a potential interaction with Asn175 (inset) on the Ω-loop. **B**. The incorporation of tyrosine at position 79 (orange) identified a possible interaction with Glu145 on the Ω-loop.

3.2.2. Construction of pBAK vector

Despite the technological advances in recent years that have allowed the incorporation of uAAs, the machinery used to do this, especially the engineered aminoacyl-tRNA synthetases (aaRSs), are not as efficient as the endogenous machinery and can often struggle to maintain reliable incorporation, especially when there is a sudden demand for increased protein production, such as that found with traditional IPTG-induced overexpression of target proteins (e.g. pET vectors). One way around this problem is to use a gentler, auto-induction approach, which enables a higher cell density and a slower, more gradual increase in protein production. This gives more time for the orthogonal aaRSs to incorporate the uAAs, resulting in an increased yield of target proteins that contain the uAA.

To enable this proposed auto-induction approach, the *bla* gene encoding TEM β-lactamase was cloned into the pBAD plasmid system (Invitrogen). This is a tightly regulated, auto-inducible system based around the P_{BAD} promoter. Expression is controlled by the concentration of L-arabinose, which binds to AraC, a repressor protein of the P_{BAD} promoter, resulting in a change in structure and subsequent activation of the promoter [115]. A secondary system of control mediated by cAMP allows a high cell density to be reached before protein production is switched on (for more information see Section 4.2.1.). These two aspects of arabinose-inducible expression are beneficial when using an expanded genetic code, as the orthogonal aaRS used to incorporate the uAA is not as efficient as the endogenous aaRSs that are responsible for incorporating the 20 canonical amino acids. Reducing the rate at which the target protein is produced increases the reliability of the uAA incorporation. Expression can also be tailored by optimizing the concentration of L-arabinose in the media.

The selectable marker present in the pBAD vector confers ampicillin resistance so has similar activity to the target protein TEM β-lactamase. Therefore, to ensure all observed activity is due to uAA-containing variants of TEM β-lactamase alone, a new vector was constructed in which ampicillin resistance was replaced with kanamycin resistance (pBAK). Primers were designed so that ampicillin resistance could be removed from the pBAD vector, and simultaneously add specific restriction sites (*Not*I and *Sal*I) to the PCR product (Figure 3-5A). Once a single PCR product was

generated, the resulting linear DNA was digested with *Not*I and *Sal*I to produce sticky ends. A DNA cassette containing kanamycin resistance was digested with the same two enzymes (Figure 3-5B) and the two linear DNA fragments were ligated to produce a new kanamycin resistant vector termed pBAK (Figure 3-5C). PCR was then used to amplify the *bla* gene from another vector (Figure 3-5D). As before, restriction sites (*Nco*I and *Kpn*I) were incorporated into the primers so that the linear DNA could be digested to produce sticky ends compatible with the multiple cloning site of the new vector. A C-terminal hexahistidine tag was also incorporated into the gene at this point to simplify protein purification. The vector DNA was also treated with *Nco*I and *Kpn*I (both found in the multiple cloning site) to produce complimentary sticky ends (Figure 3-5E). These two linear fragments were then ligated together to complete the new, kanamycin resistant vector. The resulting vector (Figure 3-5F) contained the *bla* gene in the multiple cloning site and had removed any other source of ampicillin resistance from the vector. This enabled the vector to be used in the overexpression of TEM β-lactamase and the uAA-containing variants, and allow kanamycin to act as an independent selectable marker. A list of all the primers used for these steps can be found in Table 2-2.
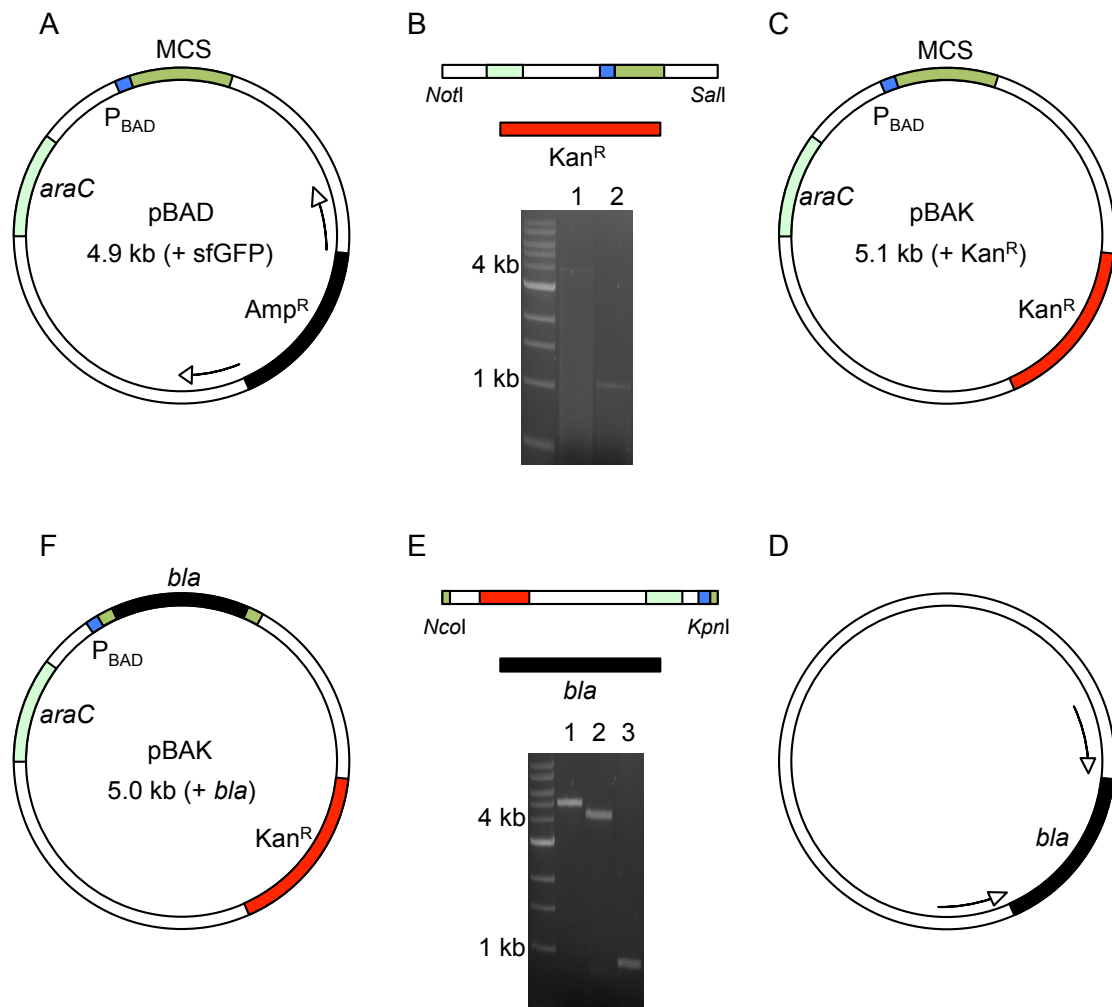
Figure 3-5. Schematic of the steps taken to remove ampicillin resistance from pBAD and replace it with kanamycin resistance to form pBAK, and subsequently insert the *bla* gene into the multiple cloning site. **A**. With the pBAD vector as a template, primers annealed to either side of the Amp^R gene and PCR was used to amplify the vector without the selectable marker. **B**. Digestion of pBAD with *Not*I and *Sal*I (Lane 1) created complimentary sticky ends with the Kan^R insert (Lane 2) that was digested with the same enzymes. **C**. Kan^R was ligated into the vector to form pBAK, a kanamycin resistant auto-induction compatible vector. **D**. PCR was used to amplify the *bla* gene from another vector (pNOM – [116]). **E**. *Nco*I and *Kpn*I form complimentary sticky ends in both the pBAK (Lane 2 – Lane 1 is a single digest with *Nco*I) vector and the *bla* gene (Lane 3). F. The *bla* gene was ligated into the MCS of the pBAK vector.

3.2.3. Site-directed mutagenesis

Now that a vector had been constructed for overexpressing TEM β-lactamase, and the desired mutations had been modeled and selected, the necessary amber stop codon (TAG) mutations were incorporated into the *bla* gene at the respective positions via site-directed mutagenesis. To do this, a whole-plasmid PCR approach was taken. This involves the use of two primers, one mutagenic and one normal, which are designed to amplify the vector back to back (no overlapping). This results in a linear fragment of DNA that can be blunt-end ligated to form the circularized vector complete with the mutated gene. This process eliminates the need for treatment of the template vector, which is required in other site-directed mutagenesis protocols, or any steps involved in re-cloning genes back into vectors. A list of all positions mutated, and all the primers used for this can be found in Table 2-2.

3.3. Conclusions and summary

Modeling potential mutations *in silico* is a great resource when mutating target proteins by rational design. Several factors such as time and cost limit the number of mutations that can be introduced into a target protein so to ensure that the best mutation candidates are selected in the early stages of project design is of great importance. Many resources are available to model proteins in a variety of ways, such as online mutation tools (such as Rosetta) or molecular dynamics simulators (such as CHARMM [117]. In this chapter, an online server designed to hypothesize the likely side chain orientation of simple point mutations was used to aid the design of mutations to incorporate AzPhe, an unnatural amino acid with novel post-translational modification opportunities. This system was used to mutate a single amino acid to tyrosine, the closest available analogue of AzPhe, and hypothesize the effect of this mutation on the positioning of the mutated side chain. This is an important factor to appreciate, as any post-translational events that occur to AzPhe would occur at the *para* position of the aromatic ring, far away from the backbone of the same residue.

Some potential mutations were not used in the subsequent analysis because the *in silico* models suggested no interactions with other residues. An example of this is Glu104. Next to Tyr105 in the protein, this was thought to present a similar opportunity to interact with the α-β domain on the other side of the catalytic cleft. However when tyrosine was incorporated in place of the native glutamate, the side chain was flipped out, pointing away from the rest of the molecule, suggesting no interactions with other residues would be likely.

4. Assessing the effect of AzPhe incorporation and novel post-translational modifications on TEM β-lactamase activity

4.1. Introduction

Although uAA incorporation is a relatively new technique in biochemistry, much progress has been made in both the chemical synthesis of new and useful amino acids, and the methods used to incorporate them into proteins. This in turn has enabled many uAAs with chemically diverse and reactive side chains to be incorporated into a plethora of proteins, introducing new functionality. The orthogonal nature of uAA incorporation is an important feature that enables uAAs to be incorporated precisely by design. In this chapter, uAA incorporation into TEM β-lactamase will be described, including a demonstration of the reliability of uAA incorporation, and the purification of uAA-containing proteins.

Due to its historic relevance to antibiotic resistance, TEM β-lactamase is a well-studied enzyme in terms of its structure and function, including enzyme kinetic analyses to understand the mechanism of hydrolysis. Site-directed mutagenesis experiments have been performed historically to piece together the precise role that each residue in the active site has in the catalytic mechanism. In addition, mutagenesis has led to the generation of many variants that are active against a wide variety of substrates including naturally occurring antibiotics, such as penicillin, and many synthetic compounds, such as cephalosporin antibiotics, developed to counteract emerging resistance. In this chapter, I present the results of uAA-incorporation into TEM β-lactamase, including detailed kinetic analyses on uAA-containing variants to understand their influence on activity and on the catalytic mechanism. These detailed kinetic analyses will include the measurement of three kinetic parameters; k_{cat} (a measure of catalytic turnover – how fast the enzyme hydrolyzes the substrate), K_M (a measure of substrate affinity – how well the substrate binds) and catalytic efficiency (k_{cat}/K_M – a measure of total enzyme efficiency).

One particular benefit of using uAAs is the novel mechanisms of protein control that become available, most notably through the use of non-natural post-translational modifications. By using *p*-azido-L-phenylalanine (AzPhe) two methods of post-

translational control are available, control via phenyl azide photochemistry and control by adduct addition via Click chemistry. UV irradiation of the azide group results in loss of molecular nitrogen and the formation of a reactive nitrene species, which can then react with nucleophiles, carbon-hydrogen bonds and carbon-nitrogen bonds. The azide present in AzPhe can also undergo highly specific reactions with alkyne groups to form a stable triazole link. As neither the azide or alkyne groups are naturally present in proteins, this offers a very specific and highly selective method of post-translational control. In this chapter, both of these methods of post-translational control are explored, and their effect on TEM β-lactamase activity is determined.

4.2. Results and discussion

4.2.1. Production of AzPhe-containing TEM β-lactamase

The theory behind the incorporation of uAAs into proteins is discussed in detail in Section 1.4. Briefly, to incorporate uAAs into proteins co-translationally, the gene of interest (under the control of an arabinose-inducible P_{BAD} promoter) is mutated to place a TAG trinucleotide codon at the desired position, and the cell is provided with additional translational machinery on a second plasmid that allows incorporation of uAAs in response to the TAG codon (Figure 4-1). Description of the methods taken to modify the pBAD plasmid, and then insert the gene of interest (particularly the *bla* gene) can be found in Section 3.2.2. The pDULE plasmid, which contains the engineered translational machinery, is a constitutive system, with the genes for the amino-acyl tRNA-synthetase (aaRS) required to incorporate the uAA, and the tRNA engineered to work alongside this aaRS, under the control of a constitutive promoter.
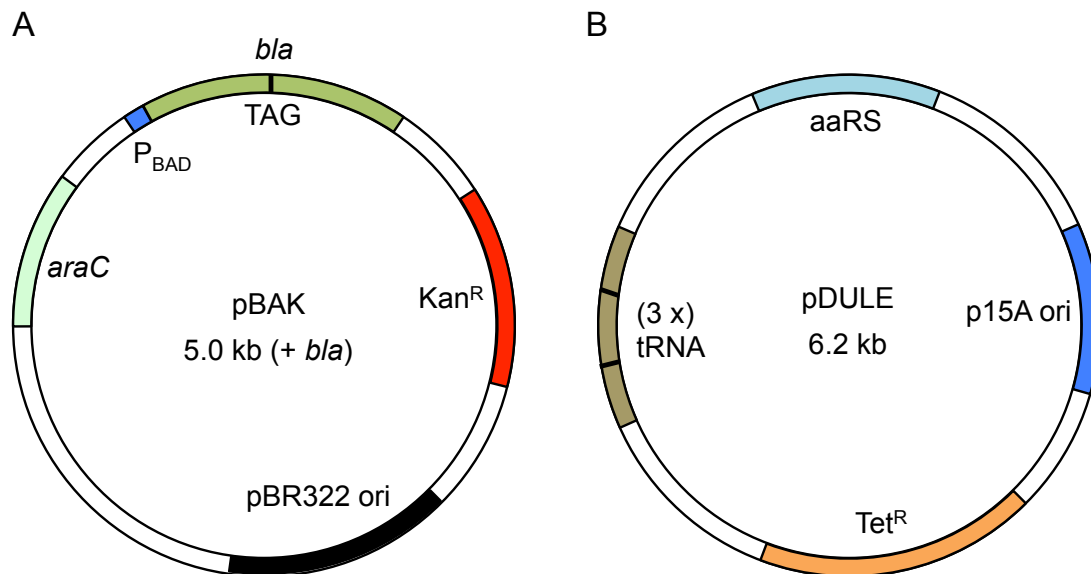
Figure 4-1. Plasmid maps of the two plasmids used to incorporate uAAs into TEM β-lactamase in *E. coli*. **A**. pBAK delivers the *bla* gene with a TAG codon into the cell. The P_{BAD} promoter is an auto-inducible promoter in part controlled by *AraC*. **B**. pDULE delivers the engineered orthogonal tRNA/aaRS pair to the cell. Three copies of the tRNA are present on pDULE to increase incorporation efficiency.

The P_{BAD} promoter can be used very effectively in conjunction with an auto-inducible growth medium system. The auto-induction system used here relies on two things, the cell's natural preference for different sugars, in this case glucose and arabinose, and the activation of the P_{BAD} promoter by both AraC and cAMP. Glucose is the preferred sugar source for the cell, so in the early stages of cell growth arabinose is not used as an energy source and is not transported into the cell. Whilst glucose is used as the primary energy source, cAMP levels are kept low, reducing its ability to induce target protein expression through its interactions with cAMP receptor protein (CRP). As the cells use up glucose as their primary energy source, the switch in sugar usage from glucose to arabinose switches on production of the target protein. The influx of arabinose into the cell and its use as an energy source, also results in arabinose binding to the AraC repressor protein, which exists as a dimer and binds to the P_{BAD} promoter preventing gene expression. Arabinose binding to AraC causes dissociation of the two subunits forming a monomer, which acts as an activator of expression [115]. At the same time, cAMP-CRP levels increase and help activate expression of the target protein [118]. These control mechanisms combine to tightly regulate the expression of the target protein. Slower and more continuous expression of the target

protein reduces the energetic impact protein production has on the cell, allowing the less-efficient orthogonal aaRSs more time to incorporate the uAA, increasing the yield of uAA-containing proteins. Due to the importance of the concentration of these two sugars, a specialized auto-induction medium is used to optimize the concentrations of both glucose and arabinose [96]. *E. coli* Top10 cells are used along with this auto-induction medium as they have the arabinose catabolism gene (*araD*) knocked out, allowing protein expression to be directly proportional to the arabinose concentration in the medium.

To ensure that uAA (AzPhe) incorporation was reliable and with high fidelity, proteins were produced in the presence and the absence of AzPhe in the growth medium. In the presence of AzPhe, it is incorporated into the target protein by the orthogonal tRNA/aaRS pair supplied to the cell. If there is no AzPhe in the growth media, the TAG codon reverts to its natural function as a stop codon, and no functional protein should be produced. To determine the effect of the presence and absence of AzPhe in the growth medium had on functional target protein production, nitrocefin activity assays were performed on cell lysates. Cells were grown in small culture volumes, in the presence and absence of AzPhe, standardized to the same optical density at 600 nm, and then lysed using BugBuster. Activity in cell lysates was measured by changes in absorbance at 475 nm (Figure 4-2).
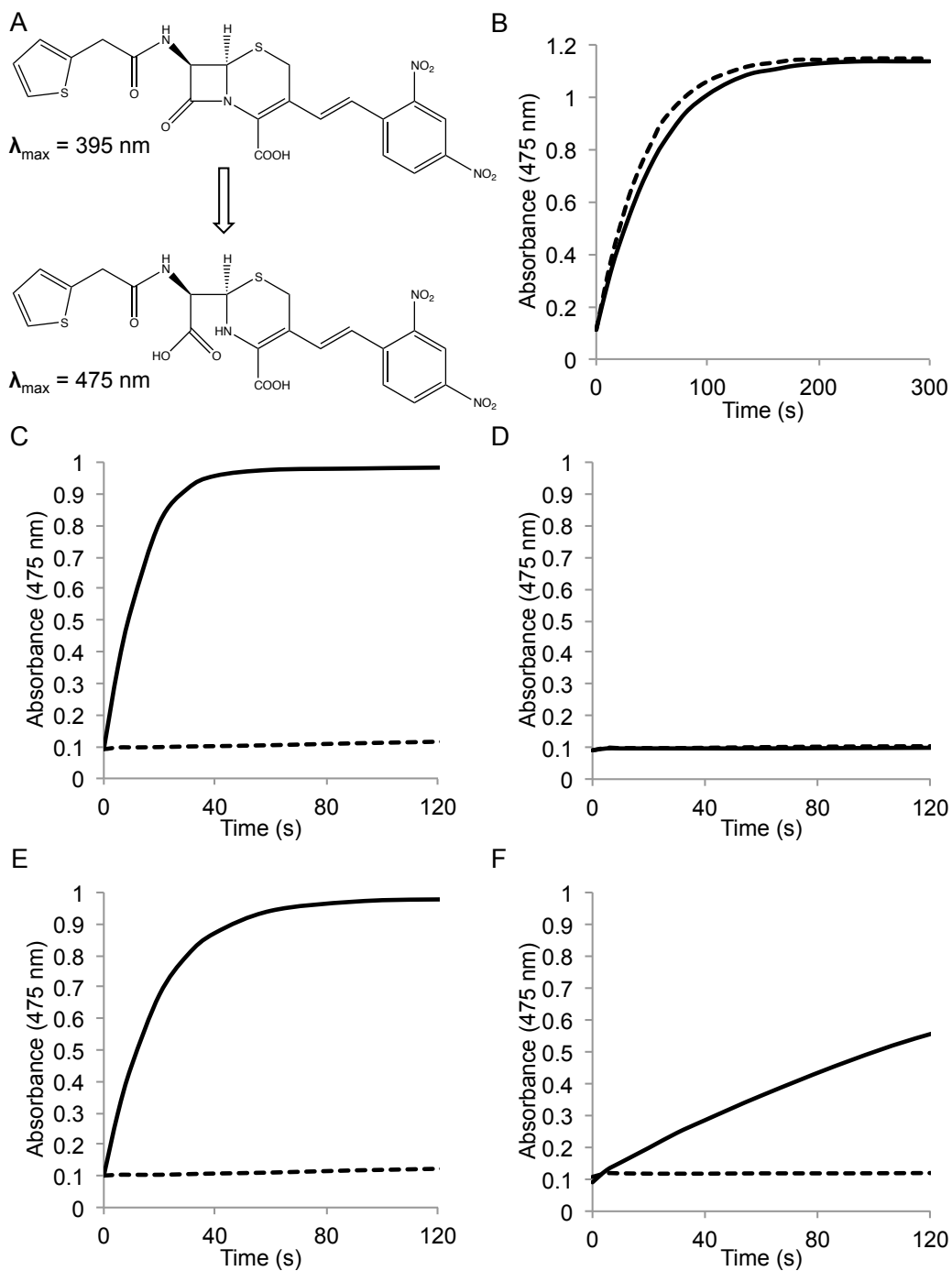
Figure 4-2. **A**. Structure of nitrocefin. Nitrocefin is a β-lactam that changes its absorbance from 395 nm to 475 nm upon hydrolysis. This change can be measured using spectrophotometry. **B-F**. Activity assays of (B) wild type TEM β-lactamase and four AzPhe-containing variants, (C) TEM[Tyr105AzPhe], (D) TEM[Arg164AzPhe], (E) TEM[Phe72AzPhe] and (F) TEM[Ala237AzPhe] in the presence (solid line) and absence (dashed line) of AzPhe in the growth medium. Absorbance was measured at 475 nm over a time period of 5 minutes (wild type was diluted to enable visualization of the initial rate) or 2 minutes.

As expected, the activity of wild type TEM β-lactamase in both the presence and the absence of AzPhe were the same (Figure 4-2B), demonstrating that the presence of AzPhe, in the growth medium is not negatively affecting target protein production. The orthogonality of AzPhe incorporation was confirmed for all TAG-containing TEM β-lactamase variants (Figure 4-2 shows four examples to demonstrate). No nitrocefin hydrolysis activity was seen in the absence of AzPhe for any variant, proving that expression of full-length, active TEM β-lactamase is dependent on the presence of AzPhe in the medium. It also proves the engineered tRNA/aaRS pair are orthogonal, and do not incorporate any of the 20 canonical amino acids in response to the TAG codon.

Several variants showed no activity, even in the presence of AzPhe in the growth medium (Figure 4-2D). TEM^{Arg65AzPhe}, TEM^{Arg164AzPhe}, TEM^{Asp179AzPhe} and TEM^{Ser235AzPhe} all showed no activity in the nitrocefin hydrolysis assay. The side chains of Arg164 and Asp179 are key contributors to the formation of a hydrogen-bonding network on the Ω-loop, a key structural feature of TEM β-lactamase that is involved in substrate specificity [73]. Therefore removal of the hydrogen bonding capabilities of the side chains at these positions has a detrimental effect on enzyme activity. Arg65, although not itself in the Ω-loop, is another residue that forms interactions with other side chains that help to stabilize the Ω-loop, in this case Asn175 [119]. Thus, the mutation of residues not on the Ω-loop can be equally as disruptive. Ser235 is located in the active site and plays a role in positioning substrates for hydrolysis [120]. Mutation of this serine to AzPhe could disrupt the catalytic mechanism.

4.2.2. Impact of UV irradiation and Click chemistry on enzyme activity in cell lysates

One of the benefits of incorporating AzPhe into proteins is the light-sensitive reactivity of the phenyl azide. When irradiated with UV light below 310 nm, molecular nitrogen is lost forming a reactive nitrene species (see Section 1.5). This can then undergo several reaction pathways including reduction, insertion into carbon-hydrogen bonds and addition reactions (Figure 1-3). After the effect of AzPhe-incorporation on TEM β-lactamase was determined, further analysis was undertaken to determine the effect of UV irradiation on the activity of the AzPhe-containing

variants. Cell lysates were subjected to two and ten minutes of UV irradiation using a handheld lamp that emits a range of wavelengths from 275 nm to 380 nm. Cell lysates were irradiated in quartz cuvettes, with a constant distance between the lamp and the sample to standardize the intensity of UV irradiation. Upon irradiation with UV light, three variants showed changes in activity, TEM^{Tyr105AzPhe}, TEM^{Phe72AzPhe} and TEM^{Pro174AzPhe} (Figure 4-3). Both TEM^{Tyr105AzPhe} and TEM^{Phe72AzPhe} show decreases in activity after just 2 minutes of UV exposure (TEM^{Pro174AzPhe} has no reduction in activity after 2 minutes), and activity is reduced further for all three variants after extended times of exposure. Incorporation of AzPhe at Tyr105 and subsequent irradiation resulted in almost a 25% decrease in activity (Table 4-1; Figure 4-3A). TEM^{Phe72AzPhe} suffered a 40% decrease in activity (Table 4-1; Figure 4-3B), and the activity of TEM^{Pro174AzPhe} decreased by almost 25%, but only after 10 minutes irradiation (Table 4-1; Figure 4-3C). One other variant, TEM^{Ala79AzPhe} also showed a decrease in activity, however this decrease was smaller (Figure 4-3D). TEM^{Ala237AzPhe} and TEM^{Trp165AzPhe} showed very little decrease in activity (Figure 4-3E-F), proving loss of activity was not universal to all AzPhe variants. Of the variants that showed no activity after AzPhe incorporation (TEM^{Arg65AzPhe}, TEM^{Arg164AzPhe}, TEM^{Asp179AzPhe} and TEM^{Ser235AzPhe}), none showed increases in activity upon UV irradiation.

Table 4-1. Initial velocity of six AzPhe-containing variants of TEM β-lactamase after no irradiation, 2 minutes irradiation, and 10 minutes irradiation.

	Initial velocity ($\mu M^{-1}\ min^{-1}$)		
	0min UV	2min UV	10min UV
TEM^{Tyr105AzPhe}	238	180	76.4
TEM^{Phe72AzPhe}	186	115	87.3
TEM^{Trp165AzPhe}	313	314	256
TEM^{Pro174AzPhe}	8.77	8.88	5.45
TEM^{Ala237AzPhe}	18.3	21.3	17.0
TEM^{Ala79AzPhe}	7.54	3.93	3.42
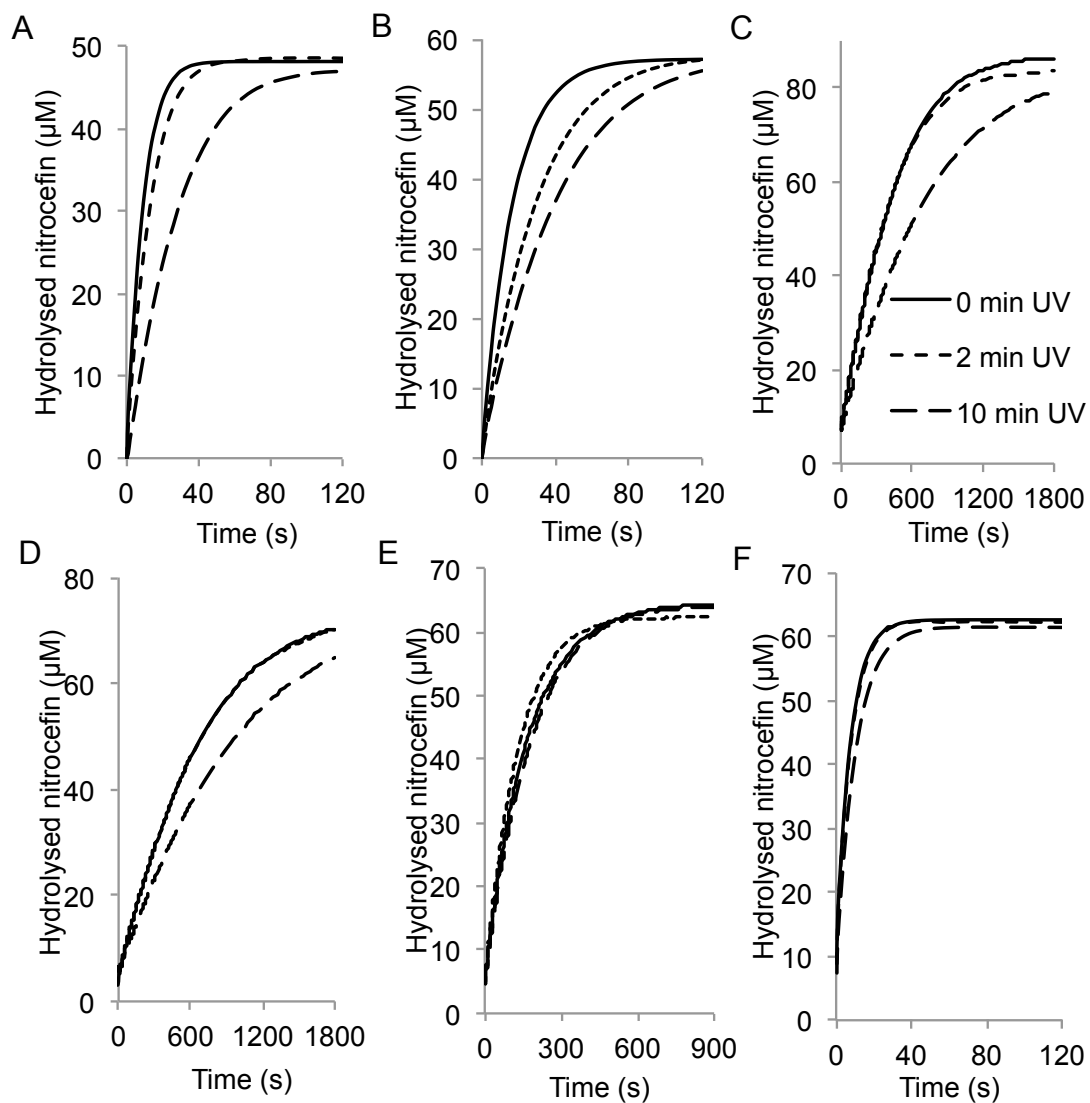
Figure 4-3. **A-F**. Effect of UV irradiation on (A) TEM^{Tyr105AzPhe}, (B) TEM^{Phe72AzPhe}. (C) TEM^{Pro174AzPhe}, (D) TEM^{Ala79AzPhe}, (E) TEM^{Ala237AzPhe} and (F) TEM^{Trp165AzPhe} in cell lysates.

In addition to its photochemical properties, the azide group of AzPhe can also undergo chemical modification via strain promoted azide-alkyne cycloaddition (SPAAC). This is a highly versatile technique as any number of novel chemical modifications can be added to the azide group, and as no other amino acids contain an azide group, the specificity offered by this type of modification is very high. Two modifications were used to determine the effect of SPAAC on TEM β-lactamase activity, one containing a large, hydrophobic dye (**1** – Figure 4-4) and a second that can form long-range hydrogen bonds (**2**).

TEM^{Tyr105AzPhe} and TEM^{Ala237AzPhe} both showed a large reduction in activity after modification with both adducts (Figure 4-4). After modification with **1**, the activity of TEM^{Tyr105AzPhe} decreased by over 90% (Table 4-2; Figure 4-4A), and the activity of TEM^{Ala237AzPhe} decreased by 80% (Table 4-2; Figure 4-4E). TEM^{Pro174AzPhe} also showed a decrease in activity after Click modification, losing 60% of activity after modification with **1** (Table 4-2; Figure 4-4C). In each case, modification with **2**, the smaller modification, resulted in a smaller decrease in activity. TEM^{Phe72AzPhe} (Figure 4-4B), TEM^{Ala79AzPhe} (Figure 4-4D) and TEM^{Trp165AzPhe} (Figure 4-4F) all showed very little effect on activity when modified with either chemical adduct.

Table 4-2. Initial velocity of six AzPhe-containing variants of TEM β-lactamase with no modification, and modified with **1** and **2** (Figure 4-4).

	Initial velocity (μM^{-1} min^{-1})		
	Unmodified	**1**	**2**
TEM^{Tyr105AzPhe}	169	15.6	66.0
TEM^{Phe72AzPhe}	175	178	173
TEM^{Trp165AzPhe}	385	285	313
TEM^{Pro174AzPhe}	8.77	3.39	5.94
TEM^{Ala237AzPhe}	6.53	1.39	5.06
TEM^{Ala79AzPhe}	7.54	1.91	4.80
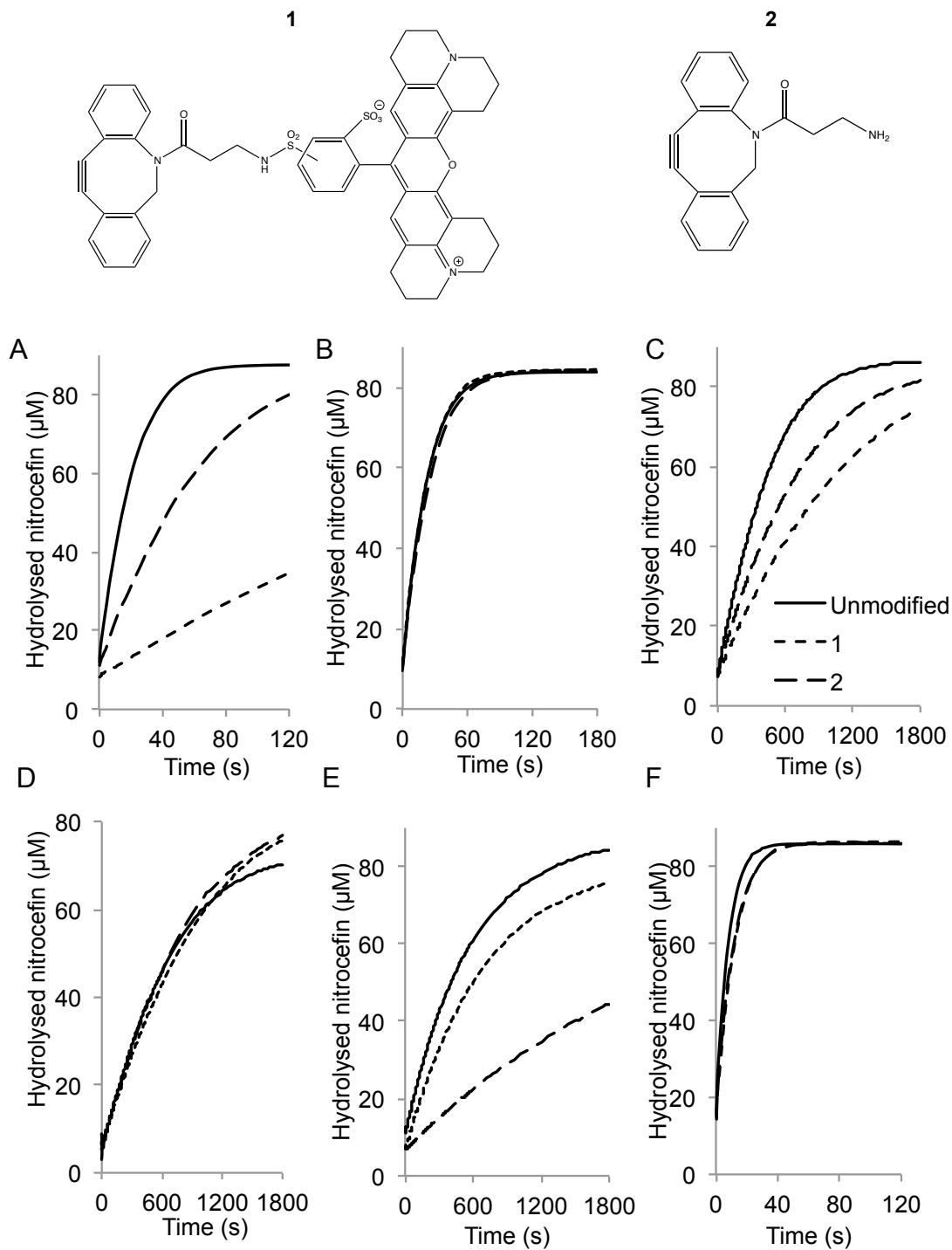
Figure 4-4. Chemical structures of the two DBCO- (dibenzyl cyclooctyne)-based modifications used in click chemistry, DBCO-sulforhodamine 101 (**1**) and DBCO-amine (**2**). **A-F.** Effect of Click modification on (A) TEM$^{\text{Tyr105AzPhe}}$, (B) TEM$^{\text{Phe72AzPhe}}$. (C) TEM$^{\text{Pro174AzPhe}}$, (D) TEM$^{\text{Ala79AzPhe}}$, (E) TEM$^{\text{Ala237AzPhe}}$ and (F) TEM$^{\text{Trp165AzPhe}}$ in cell lysates.

After the initial selection of mutations based on the *in silico* models, ten mutations were selected to be produced and tested using the nitrocefin activity assay. After these activity tests, six of the original ten mutations were then selected for further analysis via more detailed kinetic analyses; TEM^{Tyr105AzPhe}, TEM^{Ala237AzPhe}, TEM^{Pro174AzPhe}, TEM^{Trp165AzPhe}, TEM^{Ala79AzPhe} and TEM^{Phe72AzPhe}. The position of Tyr105 in the folded protein suggested that modification of this residue, especially using Click chemistry would result in a large decrease in activity. This hypothesis proved to be correct, as activity was largely reduced upon modification with both **1** and **2**, however the change in activity after UV irradiation was not as great. The effect of UV irradiation on all of the variants was difficult to hypothesize *in silico* due to the uncertainty around nitrene reactivity. In practice, UV irradiation had a limited effect on activity, regardless of the mutation position. This is likely due to the relatively small changes in the structure of the phenyl azide caused by UV irradiation. Although many of the changes suggested by modeling, the scale of the changes was somewhat smaller. Mutation at Phe72 suggested large changes in activity caused by UV irradiation in particular, however only a small decrease was observed, and no decrease when Click chemistry was used to modify the protein.

Generally, the hypotheses presented after *in silico* modeling of the individual mutations proved to be correct, even if the changes were smaller than expected, however there were some exceptions. Modeling tyrosine at Trp165 suggested a large change in activity caused by modification due to the proximity of the residue to active residues such as Ser70 and Glu166. This proved to not be the case however, as modification with both UV irradiation and Click chemistry resulted in no changes in activity. Modeling mutations at Arg164 and Asp179 suggested the destruction of the salt bridges that maintain the structure of the omega loop. This assumption proved to be correct, however the hypothesis that activity may be regained via UV irradiation of the phenyl azide proved to be false. For this reason, both of these mutations were not selected for further analysis. As well as Arg164 and Asp179, Arg65 and Ser235 were not selected for kinetic analysis, as these variants were also inactive, even after post-translational modification.

4.2.3. Purification of TEM β-lactamase

After the general effect of AzPhe incorporation on enzyme activity had been determined in cell lysates, the active variants were produced in larger volumes (1 L) and purified using Nickel-affinity chromatography. This process involves the use of immobilized Ni^{2+} to bind the imidazole side chain of the histidine amino acid. A hexa-histidine tag (His-tag) was present at the end of the protein to accelerate the process of protein purification. Clarified cell lysates were passed over a column of Ni^{2+} immobilized beads. The target protein was then eluted using an elution buffer containing imidazole, which competes for the Ni^{2+} with the histidine residues. The elution buffer is applied to the column via a concentration gradient, to obtain protein with a high level of purity (Figure 4-5A). During the purification process, absorbance at 280 nm was monitored to track wild type elution from the column. As monitoring at 280 nm would likely result in photolysis of the phenyl azide, this was turned off during the purification of AzPhe-containing variants. After the purification of these variants (as well as wild type), a combination of SDS-PAGE and nitrocefin hydrolysis assays were used to determine the fractions containing the purified TEM β-lactamase, and their purity (Figure 4-5B-C). The fractions containing only TEM β-lactamase were pooled and buffer exchanged into PBS for more detailed analysis. Buffer exchange was performed via dialysis using a 3500 MWCO dialysis membrane.
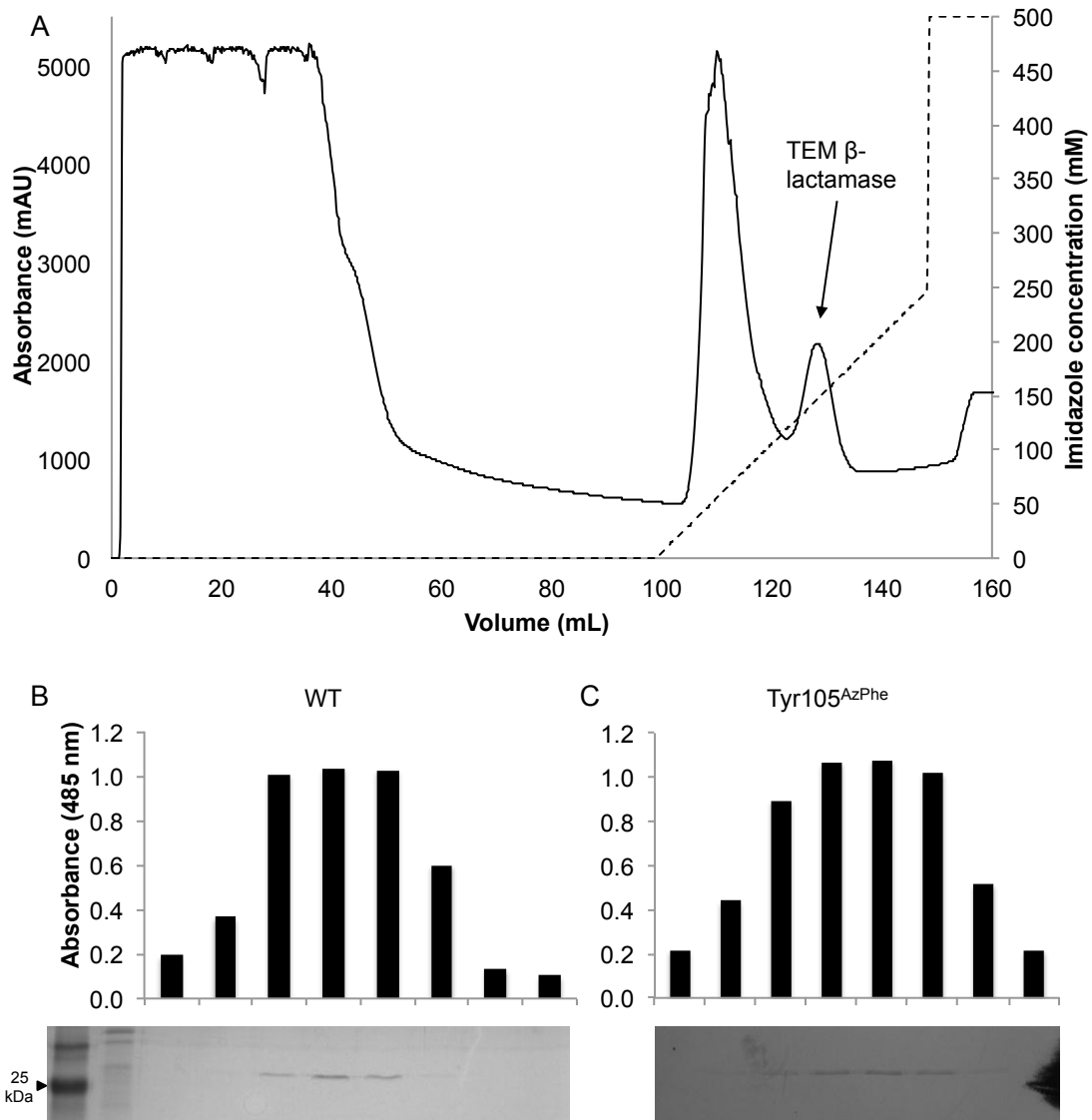
Figure 4-5. Nickel-affinity chromatography purification of TEM β-lactamase. **A**. Absorbance at 280 nm monitoring wild type elution (black line) on application of an imidazole gradient (dashed line). The peak to elute during the gradient corresponds to non-specific binding of proteins to the Ni^{2+} matrix. The peak corresponding to TEM β-lactamase is indicated. **B-C**. SDS-PAGE analysis and activity analysis of TEM β-lactamase-containing fractions from the purification of (B) wild type and (C) TEM^{Tyr105AzPhe}.

4.2.4. Kinetic analysis of AzPhe-containing TEM β-lactamase variants

4.2.4.1. Effect of AzPhe incorporation on TEM β-lactamase activity

After wild type TEM β-lactamase and the AzPhe-containing mutants were purified, detailed kinetic analysis was undertaken to determine the exact effect of AzPhe incorporation on the enzyme's ability to both bind and hydrolyze the substrate. Whilst the colourimetric substrate nitrocefin was used in the preliminary analysis of the AzPhe-containing variants, two other, more clinically relevant, substrates were chosen for kinetic analysis, ampicillin and cephalothin. Both of these are β-lactam antibiotics, however ampicillin is a naturally occurring penicillin-based antibiotic which is hydrolyzed by TEM β-lactamase with a much greater efficiency, whereas cephalothin is a synthetic cephalosporin originally developed to overcome the evolution of resistance against penicillin-based antibiotics. They have disparate structures and interact with the protein in slightly different ways [119, 121].

Incorporating AzPhe in place of Tyr105 had no negative impact on activity with regards to both ampicillin and cephalothin hydrolysis. Compared to wild type, TEM$^{\text{Tyr105AzPhe}}$ k_{cat} was slightly improved, indicating a small increase in catalytic turnover. However this was offset by a small increase in K_{M}, suggesting a slight impact on substrate binding. The overall result was an enzyme with a very similar overall catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) for both substrates to that of wild type TEM β-lactamase (Table 4-3; Figure 4-6). The largely similar k_{cat} and K_{M} suggest the addition of an azide group at this position does not have a significant impact on the ability of the enzyme to bind and then hydrolyze either substrate. Tyr105 is an aromatic amino acid that contributes to the domain interface and is highly surface exposed (Figure 4-7C). Despite this, it is not directly involved in the catalytic mechanism, however it does lie in close proximity to many of the residues that do have direct or indirect roles in substrate hydrolysis. As predicted by the original hypothesis (see Section 3.2.1.), the addition of a single atom at the *para* position of the phenyl ring (OH to N_3), coupled with the surface exposed nature of the residue, suggests the incorporation of AzPhe here would not have a detrimental effect on enzyme activity, which is the case. It is possible the increased bulk caused by the azide group at the *para* position decreases binding efficiency by slightly reducing the space available for the substrate in the catalytic pocket.

Table 4-3. The effect of AzPhe-incorporation on TEM β-lactamase activity.

	Ampicillin					
	Wild type	**TEM[Tyr105AzPhe]**	**TEM[Ala237AzPhe]**	**TEM[Pro174AzPhe]**	**TEM[Trp165AzPhe]**	**TEM[Leu201AzPhe]**
k_{cat} (s^{-1})	550 ± 21	753 ± 17	98 ± 6.2	171 ± 13	406 ± 20	427 ± 20
Relative k_{cat} [1]	**1.00**	**1.37**	**0.18**	**0.31**	**0.74**	**0.78**
K$_M$ (µM)	146 ± 15	183 ± 10	1340 ± 120	106 ± 29	185 ± 26	171 ± 23
Relative affinity [1]	**1.00**	**0.80**	**0.11**	**1.38**	**0.79**	**0.85**
$k$cat/K$_M$ (µM^{-1} s^{-1})	3.77 ± 0.11	4.12 ± 0.061	0.0731 ± 0.11	1.62 ± 0.28	2.19 ± 0.15	2.50 ± 0.14
Relative k_{cat}/ K$_M$ [1]	**1.00**	**1.09**	**0.02**	**0.43**	**0.58**	**0.66**
	Cephalothin					
	Wild type	**TEM[Tyr105AzPhe]**	**TEM[Ala237AzPhe]**	**TEM[Pro174AzPhe]**	**TEM[Trp165AzPhe]**	**TEM[Leu201AzPhe]**
k_{cat} (s^{-1})	78.7 ± 4.9	86.3 ± 4.1	12.1 ± 1.8	38.6 ± 2.3	121 ± 11	58.5 ± 5.3
Relative k_{cat} [1]	**1.00**	**1.10**	**0.15**	**0.49**	**1.54**	**0.74**
K$_M$ (µM)	174 ± 17	212 ± 15	258 ± 54	144 ± 14	221 ± 32	151 ± 23
Relative affinity [1]	**1.00**	**0.82**	**0.67**	**1.21**	**0.79**	**1.15**
$k$cat/K$_M$ (µM^{-1} s^{-1})	0.45 ± 0.11	0.41 ± 0.084	0.047 ± 0.26	0.27 ± 0.11	0.55 ± 0.17	0.39 ± 0.18
Relative k_{cat}/ K$_M$ [1]	**1.00**	**0.91**	**0.11**	**0.60**	**1.22**	**0.87**

[1] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to wild type

Whilst substituting Tyr105 with AzPhe had little impact on the catalytic efficiency of TEM β-lactamase, the same was not true for Ala237. When compared to the wild type enzyme, TEM[Ala237AzPhe] has about 2% of the catalytic efficiency towards ampicillin with both k_{cat} (~5 fold decrease) and K_M (~10 fold increase) disrupted (Table 4-3; Figure 4-6). A similar trend is observed with cephalothin as the substrate. k_{cat} is similarly decreased, however K_M is not so severely disrupted. This results in an enzyme 10-fold less active towards cephalothin than wild type. Incorporation of AzPhe at Ala237 replaces a small amino acid, alanine, with a much larger, aromatic residue. Considering the position of Ala237 in the catalytic pocket (Figure 4-8C), increasing the size of a side chain from a single methyl group to a large phenyl azide is likely to cause large disruptions in the active site. During modeling this was suggested, as mutation to a tyrosine residue at Ala237 appeared to significantly reduce the size of the active site (Figure 3-2C). Additionally, the backbone amine of Ala237 is involved in substrate binding [111], so mutation here (even though the backbone amine is still present) is likely to decrease binding affinity due to movement of the protein backbone to accommodate the larger side chain.

Figure 4-6. **A-D**. Effect of AzPhe-incorporation on TEM β-lactamase. Catalytic efficiency (μM^{-1} s^{-1}) of wild type TEM β-lactamase and five AzPhe-containing variants against (A) ampicillin and (B) cephalothin. Chemical structures of (C) ampicillin and (D) cephalothin. **E**. Michaelis-Menten plot of TEM[Tyr105AzPhe]. **F**. Lineweaver-Burke plot (double reciprocal plot) of TEM[Tyr105AzPhe].

Incorporation of AzPhe at Pro174 also has a negative impact on the catalytic efficiency towards both ampicillin and cephalothin (Table 4-3; Figure 4-6). Against ampicillin, TEM^{Pro174AzPhe} k_{cat} is just over 30% of that of wild type, reducing to 171 s^{-1}, compared to 550 s^{-1} for wild type. This loss of catalytic turnover is somewhat offset by a slight increase in binding affinity which results in an overall catalytic efficiency against ampicillin that is less than half of wild type. When cephalothin is used as the substrate, a similar pattern of results emerges. k_{cat} is halved compared to wild type, while there is a slight decrease in K_M, resulting in an enzyme with 60% of the catalytic efficiency against cephalothin of wild type. Pro174 is located towards the C-terminal end of the Ω-loop (Figure 4-9C), a dynamic region that enables substrate entry into the active site [122]. The flexibility of the Ω-loop is essential for positioning Glu166, a key catalytic residue. Any change in the flexibility of the loop is likely to have a negative effect on the ability of Glu166 to perform its role in both the acylation and deacylation steps of the catalytic mechanism. Proline is a rigid amino acid due to its structure, so it is likely to regulate this flexibility (Pro167 is in another unstructured loop region of the Ω-loop). Mutation of Pro174 to AzPhe is likely to increase the flexibility of the Ω-loop, reducing the efficiency of Glu166 as a proton donor, hence the predominant impact to catalytic turnover.

TEM^{Trp165AzPhe} is one of the more active variants, retaining over half of the activity seen in wild type for ampicillin, and slightly more activity against cephalothin than wild type (Table 4-3; Figure 4-6). Catalytic efficiency is reduced by 144 s^{-1} compared to wild type and K_M is increased by 40 μM, resulting in an enzyme with almost half of the activity of wild type for ampicillin. There is an increase in k_{cat} when cephalothin is used as the substrate, from 79 s^{-1} to 121 s^{-1}, however substrate binding is slightly decreased by over 45 μM compared to wild type. This results in an enzyme with an overall catalytic efficiency against cephalothin that is over 20% higher than wild type. Trp165 is found on the Ω-loop very close to the key residue Glu166 (Figure 4-10A). This puts it in a position where mutation is likely to have a large impact on the ability of the enzyme to breakdown its substrates. Mutation of Trp165 to AzPhe however only has a limited impact on catalytic turnover, reducing hydrolysis of ampicillin and increasing hydrolysis of cephalothin. These changes are only small however, so the prediction of a large impact on catalytic turnover is not realized.

Many mutations in proteins, and specifically enzymes, do not directly change or interact with the active site [108], impacting the activity of enzymes through indirect interactions, or by masking the effects of other deleterious mutations. Leu201 was chosen for mutation as it is not close to the active site (Figure 4-10B), and is not involved in substrate binding, but has been found to mask the effects of other deleterious mutations [107]. TEM[Leu201AzPhe] was one of the more active variants against ampicillin, but not as active as either wild type or TEM[Tyr105AzPhe] (Table 4-3; Figure 4-6). TEM[Leu201AzPhe] retains 78% of the catalytic turnover of ampicillin compared to wild type (427 s^{-1} compared to 550 s^{-1}), and K_M in increased by 25 µM, resulting in an enzyme with two-thirds of the catalytic efficiency of wild type against ampicillin. When cephalothin is used as the substrate only small changes are seen. k_{cat} decreases by 20 s^{-1} only and K_M drops by 20 µM, producing an enzyme with a similar catalytic efficiency against cephalothin to wild type TEM β-lactamase. Leu201 was effectively chosen as a control to demonstrate that the incorporation of AzPhe has little effect on enzyme activity when the residue is far from the active site. As expected, incorporation of AzPhe at Leu201 had little effect on activity as it is far from the active site and doesn't interact with the residues responsible for binding the substrate.

4.2.4.2. Effect of UV irradiation on TEM β-lactamase activity

Exposure of TEM[Tyr105AzPhe] to UV light results in a decrease in activity. With ampicillin as the substrate, both k_{cat} and K_M are reduced by 25% (Table 4-4; Figure 4-7A). This results in an enzyme with a catalytic efficiency 44% lower than the "dark" enzyme. With cephalothin used as the substrate, k_{cat} decreases from 86 s^{-1} to 53 s^{-1} after UV irradiation, which is offset by a slightly increased substrate affinity (K_M decreases by 30 µM). This produces an enzyme that has a catalytic efficiency 71% of that of the dark enzyme (Table 4-4; Figure 4-7B). Upon UV irradiation, the azide group loses molecular nitrogen to form a reactive nitrene species [63], and from there it can go down many reaction pathways, forming many different end products. It is very difficult to predict the outcome of this reaction, especially when considering the variable microenvironments sampled in a protein structure, and the benign (physiological) reaction conditions. Expansion of the phenyl ring to form a 7-membered dihydroazepine ring (Figure 1-3) could decrease the space available for substrates to bind. This change would also impact the SDN loop, a triad of amino

acids involved in maintaining the active site cavity [80]. Asn132 has also been suggested to be involved in substrate binding, and indirectly involved in catalysis [123], so perturbation of this residue would decrease catalytic turnover.

Table 4-4. Effect of UV irradiation on the activity of TEM$^{\text{Tyr105AzPhe}}$ against ampicillin and cephalothin.

	Ampicillin		Cephalothin	
	Dark	UV	Dark	UV
k_{cat} (s^{-1})	753 ± 17	568 ± 25	86.3 ± 4.1	52.6 ± 3.4
Relative k_{cat} [1]	1.00	0.75	1.00	0.61
K_M (µM)	183 ± 10	248 ± 24	212 ± 15	182 ± 18
Relative affinity [1]	1.00	0.74	1.00	1.17
k_{cat}/K_M (µM^{-1} s^{-1})	4.12 ± 0.061	2.29 ± 0.11	0.41 ± 0.084	0.29 ± 0.12
Relative k_{cat}/K_M [1]	1.00	0.56	1.00	0.71

[1] k_{cat}, K_M and k_{cat}/K_M (overall catalytic efficiency) compared directly to the dark form of TEM$^{\text{Tyr105AzPhe}}$
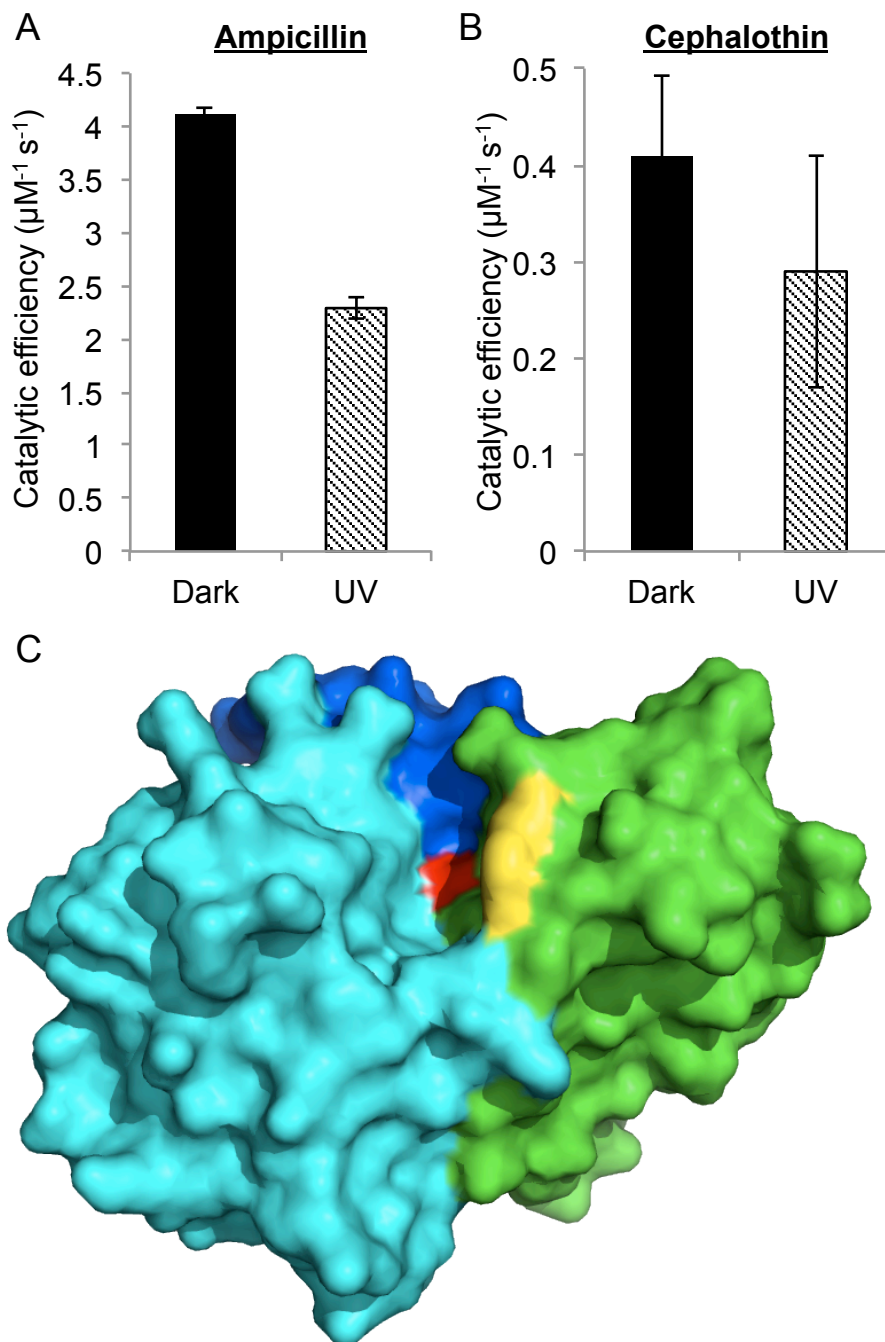
Figure 4-7. Effect of UV irradiation on TEM$^{\text{Tyr105AzPhe}}$. **A** and **B**. Catalytic efficiency (μM^{-1} s^{-1}) of un-irradiated (dark) and irradiated (UV) TEM$^{\text{Tyr105AzPhe}}$ against (A) ampicillin and (B) cephalothin. **C**. Position of Tyr105 (yellow) in TEM β-lactamase (Ser70 is shown in red).

UV irradiation has very little impact on the ability of TEM^{Ala237AzPhe} to break down ampicillin. When this variant is exposed to UV light, k_{cat} is slightly increased, but so is K_M, resulting in an enzyme that has a similar catalytic efficiency to the dark form of the enzyme (Table 4-5; Figure 4-8A). When cephalothin is used as the substrate, the UV irradiated enzyme has a much lower k_{cat}, (5 s^{-1}). However K_M is decreased by over 100 μM, which offsets some of this decrease in catalytic turnover (Table 4-5; Figure 4-8B), resulting in an enzyme that retains 72% of the activity of the dark enzyme against cephalothin. The low activity of this variant results in the errors associated with these data being relatively large, indicating the likelihood that UV irradiation is having little effect on the ability of the enzyme to hydrolyze the two substrates. It is very difficult to predict what happens to the phenyl azide upon irradiation so it is very difficult to present a hypothesis for this effect. It is possible that azide reduction to form an amine at the *para* position of the phenyl ring could contribute to the hydrogen-bonding network that positions cephalothin for catalysis. As there is no increase in ampicillin binding, this proposed interaction may not occur when ampicillin is used as the substrate.

Table 4-5. Effect of UV irradiation on the activity of TEM^{Ala237AzPhe} against ampicillin and cephalothin.

	Ampicillin		Cephalothin	
	Dark	**UV**	**Dark**	**UV**
k_{cat} (s^{-1})	97.8 ± 6.2	116 ± 6.7	12.1 ± 1.8	5.25 ± 0.25
Relative k_{cat} [1]	**1.00**	**1.18**	**1.00**	**0.44**
K_M (μM)	1340 ± 120	1440 ± 110	258 ± 54	157 ± 12
Relative affinity [1]	**1.00**	**0.93**	**1.00**	**1.65**
k_{cat}/K_M (μM^{-1} s^{-1})	0.073 ± 0.11	0.080 ± 0.097	0.047 ± 0.26	0.034 ± 0.090
Relative k_{cat}/K_M [1]	**1.00**	**1.10**	**1.00**	**0.72**

[1] k_{cat}, K_M and k_{cat}/K_M (overall catalytic efficiency) compared directly to the dark form of TEM^{Ala237AzPhe}

Figure 4-8. Effect of UV irradiation on TEM$^{\text{Ala237AzPhe}}$. **A** and **B**. Catalytic efficiency (μM^{-1} s^{-1}) of un-irradiated (dark) and irradiated (UV) TEM$^{\text{Ala237AzPhe}}$ against (A) ampicillin and (B) cephalothin. **C**. Position of Ala237 (yellow) in TEM β-lactamase (Ser70 is shown in red).

When TEM$^{\text{Pro174AzPhe}}$ is irradiated with UV light, there are small, statistically insignificant changes to the activity for both substrates (Table 4-6; Figure 4-9). When ampicillin is used, k_{cat} is reduced by 20% to 136 s^{-1}, but there is no change in K_{M}. When cephalothin was used as the substrate there was a 17% increase in k_{cat} to 45 s^{-1}, but there was also an increase in K_{M}, from 144 µM to 204 µM. These factors combine to produce an enzyme that is 18% less active for ampicillin, and 17% less active for cephalothin than the dark enzyme. During the *in silico* modeling of TEM$^{\text{Pro174AzPhe}}$, several methods of control were suggested. The disruption of an interaction between Arg65 and Asn175 could increase the flexibility of the Ω-loop, disrupting the role of Glu166 in the catalytic process. Alternatively, the incorporation of AzPhe could introduce a new interaction with Gly267 to decrease the flexibility of the Ω-loop. The kinetic data indicate an equally small effect on both substrate hydrolysis and binding, suggesting that the flexibility of the Ω-loop is not greatly altered by UV irradiation of AzPhe.

Table 4-6. Effect of UV irradiation on the activity of TEM$^{\text{Pro174AzPhe}}$ against ampicillin and cephalothin.

	Ampicillin		Cephalothin	
	Dark	**UV**	**Dark**	**UV**
k_{cat} (s^{-1})	171 ± 13	136 ± 10	38.6 ± 2.3	45.3 ± 3.8
Relative k_{cat} [1]	**1.00**	**0.80**	**1.00**	**1.18**
K_M (µM)	106 ± 29	102 ± 27	144 ± 14	204 ± 25
Relative affinity [1]	**1.00**	**1.03**	**1.00**	**0.71**
k_{cat}/K_M (µM^{-1} s^{-1})	1.62 ± 0.28	1.33 ± 0.28	0.27 ± 0.11	0.22 ± 0.15
Relative k_{cat}/K_M [1]	**1.00**	**0.82**	**1.00**	**0.83**

[1] k_{cat}, K_M and k_{cat}/K_M (overall catalytic efficiency) compared directly to the dark form of TEM$^{\text{Pro174AzPhe}}$
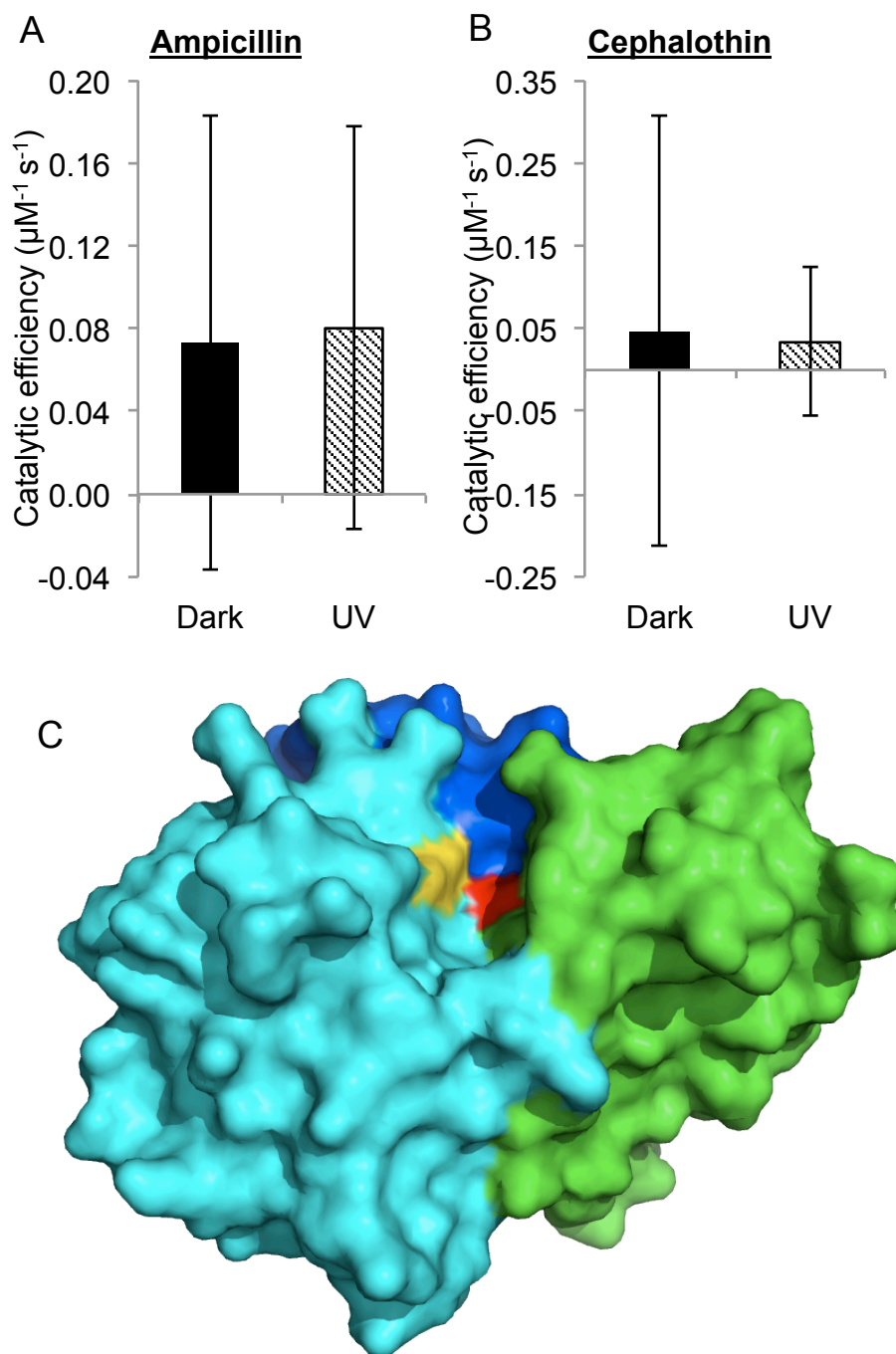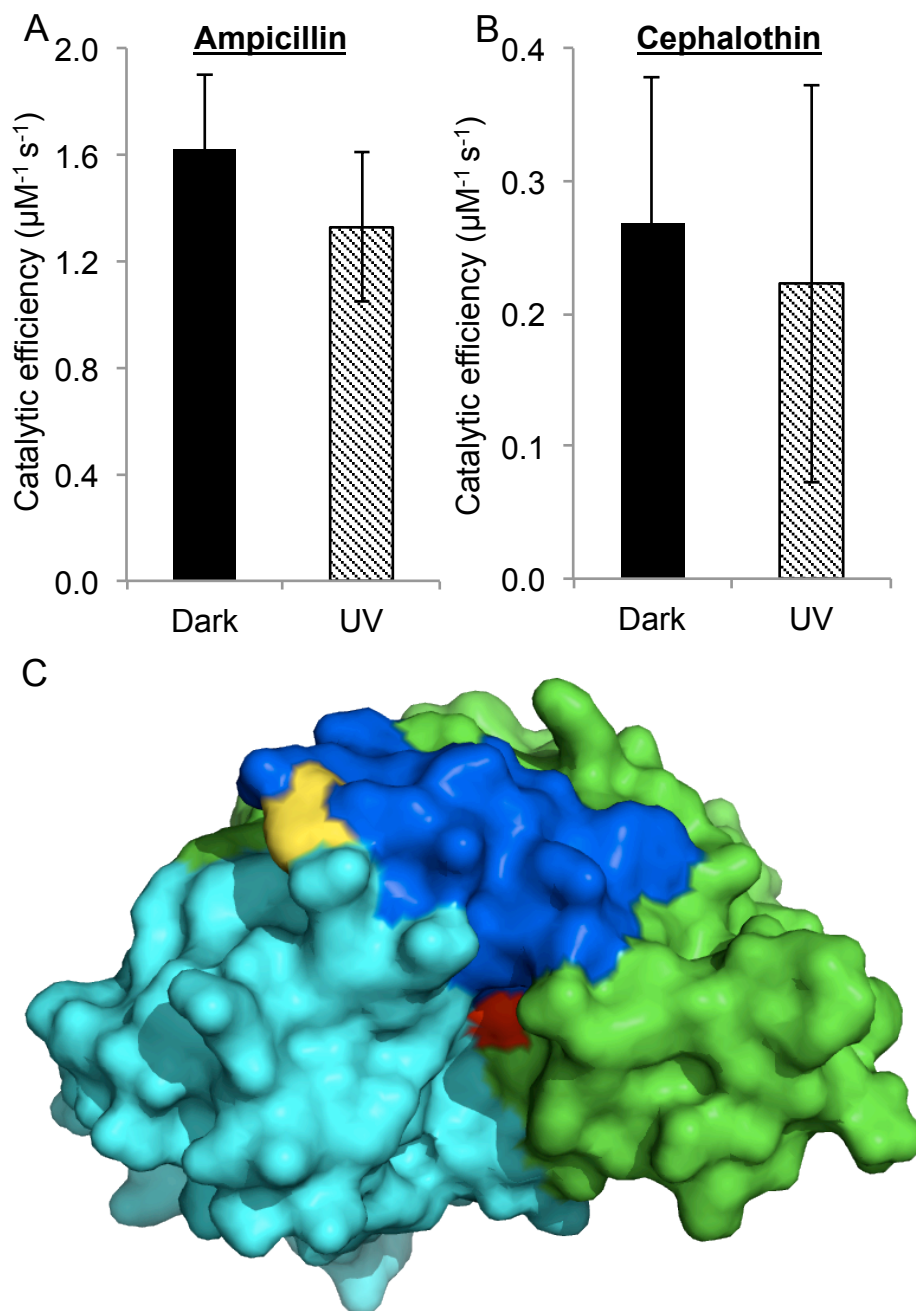
Figure 4-9. Effect of UV irradiation on TEM^{Pro174AzPhe}. **A** and **B**. Catalytic efficiency (μM^{-1} s^{-1}) of un-irradiated (dark) and irradiated (UV) TEM^{Pro174AzPhe} against (A) ampicillin and (B) cephalothin. **C**. Position of Pro174 (yellow) in TEM β-lactamase (Ser70 is shown in red).

UV irradiation has very little effect on the ability of TEM$^{\text{Trp165AzPhe}}$ to hydrolyze ampicillin (Table 4-7). Small decreases in both k_{cat} and K_{M} somewhat offset, causing a slight reduction in the overall catalytic efficiency by 11%. UV irradiation had a larger effect on cephalothin hydrolysis, however all of the decrease seen is due to a 30% decrease in substrate binding. Trp165 does not have a direct role in either substrate hydrolysis or binding, so the limited effect of UV irradiation is not surprising. The mutation of a residue so close to other residues that do have key roles in the catalytic mechanism, particularly Glu166, shows that novel uAAs such as AzPhe can be tolerated in crucial domains of proteins, and not have a significant impact on activity.

Exposure of TEM$^{\text{Leu201AzPhe}}$ to UV light has limited effect on both k_{cat} and K_{M} when ampicillin is the substrate (Table 4-7). Catalytic turnover is reduced by 15% to 364 s^{-1}, and K_{M} is unchanged. When cephalothin is used as the substrate, k_{cat} is increased by almost 30%, whereas binding affinity decreases by over 40% (to 263 µM). These combine to produce an enzyme with 73% of the overall catalytic efficiency of the dark variant. These small changes caused by UV irradiation are unlikely to cause a great change in the active site due to the distance between Leu201 and Ser70.

UV irradiation had no effect on the ability of wild type TEM β-lactamase to hydrolyze both ampicillin and cephalothin. Both k_{cat} and K_{M} were unaffected by UV irradiation, proving UV irradiation did not inherently damage the enzyme's ability to hydrolyze either substrate.

Table 4-7. Effect of UV irradiation on the activity of TEM^{Trp165AzPhe} and TEM^{Leu201AzPhe} against ampicillin and cephalothin.

	TEM^{Trp165AzPhe}			
	Ampicillin		**Cephalothin**	
	Dark	**UV**	**Dark**	**UV**
k_{cat} (s^{-1})	406 ± 20	336 ± 23	121 ± 11	125 ± 13
Relative k_{cat} [1]	**1.00**	**0.83**	**1.00**	**1.03**
K$_M$ (µM)	185 ± 26	172 ± 36	221 ± 32	314 ± 44
Relative affinity [1]	**1.00**	**1.07**	**1.00**	**0.70**
k_{cat}/K$_M$ (µM^{-1} s^{-1})	2.19 ± 0.15	1.95 ± 0.21	0.55 ± 0.17	0.40 ± 0.18
Relative k_{cat}/K$_M$ [1]	**1.00**	**0.89**	**1.00**	**0.72**
	TEM^{Leu201AzPhe}			
	Ampicillin		**Cephalothin**	
	Dark	**UV**	**Dark**	**UV**
k_{cat} (s^{-1})	427 ± 20	364 ± 18	58.5 ± 5.4	74.8 ± 8.4
Relative k_{cat} [2]	**1.00**	**0.85**	**1.00**	**1.28**
K$_M$ (µM)	171 ± 23	171 ± 24	151 ± 23	264 ± 41
Relative affinity [2]	**1.00**	**1.00**	**1.00**	**0.57**
k_{cat}/K$_M$ (µM^{-1} s^{-1})	2.50 ± 0.14	2.12 ± 0.15	0.39 ± 0.18	0.28 ± 0.19
Relative k_{cat}/K$_M$ [2]	**1.00**	**0.85**	**1.00**	**0.73**

[1] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to the dark form of TEM^{Trp165AzPhe}

[2] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to the dark form of TEM^{Leu201AzPhe}
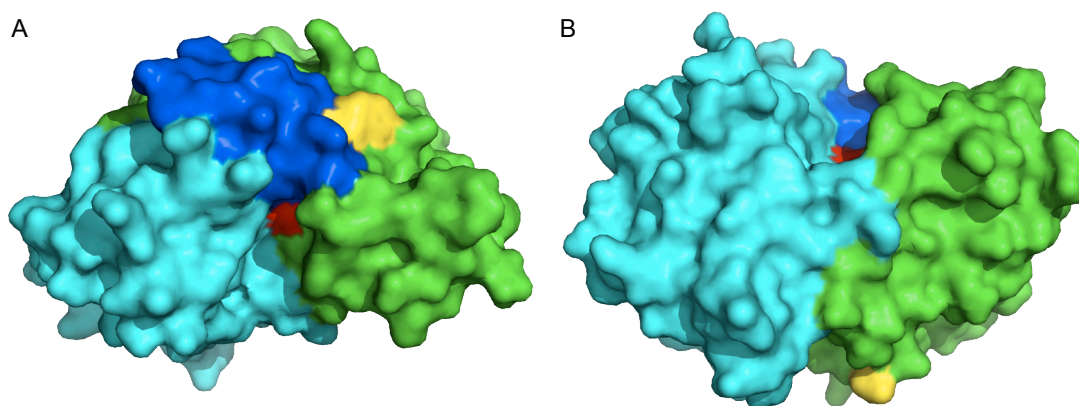
Figure 4-10. **A-B**. Position of (A) Trp165 (yellow) and (B) Leu201 in TEM β-lactamase (Ser70 is shown in red).

4.2.4.3. Effect of SPAAC on TEM β-lactamase activity

The activity of TEM^{Tyr105AzPhe} was largely influenced by Click modification. Modification with **1** resulted in large changes in both k_{cat} and K_M with ampicillin as the substrate (Table 4-8; Figure 4-11A). Catalytic turnover was reduced by 74% and K_M was increased by 38%, indicating that both catalysis and substrate binding had been disrupted. Thus, modification with **1** resulted in a loss of 85% of the catalytic efficiency of the unmodified TEM^{Tyr105AzPhe} variant; a significant difference confirmed by comparison of the 95% confidence intervals for both k_{cat} (719.5-771.9 s^{-1} for unmodified and 172.8-216.0 s^{-1} for modified) and K_M (161.7-203.7 μM for unmodified and 246.2-388.3 μM for modified). Similar changes were seen when **2** was used, with a decrease of 47% in k_{cat}, and a decrease of 45% in substrate affinity, producing an enzyme with a 71% drop in activity compared to the unmodified TEM^{Tyr105AzPhe} variant. With cephalothin as the substrate, there was a similar pattern of results with regards to k_{cat}, but less so with K_M (Table 4-8; Figure 4-11B). When TEM^{Tyr105AzPhe} was modified with **1**, there was a large decrease in turnover, with a 64% decrease in k_{cat}. Catalytic turnover was also decreased upon modification with **2**, but to a lower extent (25%), however there appears to be relatively little effect on cephalothin binding when the enzyme is modified with either **1** or **2**.

When modified with **2**, the effects on the enzyme kinetics for ampicillin are very similar with the effect of UV irradiation, suggesting the amine group of **2** could be performing a similar role to the potential amine formed through UV irradiation of the azide group. However, the positions of the proposed amine groups would be quite different in the 3D structure of the enzyme, and so would likely make different interactions. The modification of TEM^{Tyr105AzPhe} with **1** had a similar pattern of results to modification with **2**, however with larger decreases, particularly to k_{cat}. TEM^{Tyr105AzPhe} modified with **1** had half the catalytic turnover of TEM^{Tyr105AzPhe} modified with **2** for both ampicillin and cephalothin. These decreases, in proportion with the size of the modification, suggest the size of the modification is the determining factor in k_{cat} at this position. Modification with both **1** and **2** had a very similar impact on ampicillin binding, however neither had an impact on cephalothin binding. This result suggests that while the two modifications have an effect on substrate catalysis, irrelevant of the specific substrate used, it only has an effect on ampicillin binding.

Table 4-8. Effect of Click modification on the activity of TEM^{Tyr105AzPhe} against ampicillin and cephalothin.

	Ampicillin			Cephalothin		
	Unmodified	1	2	Unmodified	1	2
k_{cat} (s^{-1})	753 ± 17	194 ± 11	399 ± 31	86.3 ± 4.1	31.1 ± 2.2	64.9 ± 3.1
Relative k_{cat} [1]	1.00	0.26	0.53	1.00	0.36	0.75
K$_M$ (µM)	183 ± 10	317 ± 35	332 ± 51	212 ± 15	198 ± 21	184 ± 13
Relative affinity [1]	1.00	0.58	0.55	1.00	1.07	1.16
k_{cat}/K$_M$ (µM^{-1} s^{-1})	4.12 ± 0.061	0.61 ± 0.12	1.20 ± 0.17	0.41 ± 0.084	0.16 ± 0.13	0.35 ± 0.085
Relative k_{cat}/K$_M$ [1]	1.00	0.15	0.29	1.00	0.39	0.87

[1] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to the unmodified form of TEM^{Tyr105AzPhe}

Figure 4-11. Effect of Click modification on TEM^{Tyr105AzPhe}. Catalytic efficiency (µM^{-1} s^{-1}) of unmodified TEMT^{yr105AzPhe}, and TEM^{Tyr105AzPhe} modified with **1** and **2** against (A) ampicillin and (B) cephalothin.

When TEM$^{\text{Ala237AzPhe}}$ was modified using Click chemistry, k_{cat} was decreased for both substrates, however the effect on K_{M} was different for the two substrates. With ampicillin as substrate, k_{cat} was decreased while substrate affinity was increased, when either **1** or **2** was used to modify the enzyme (Table 4-9; Figure 4-12A). With cephalothin as the substrate, the impact on k_{cat} is not so severe and there was no increase in substrate binding, with slight decreases seen with both **1** and **2** (Table 4-9; Figure 4-12B). These changes are very small and the errors associated with these data suggest there is no significant impact of Click chemistry on the activity of TEM$^{\text{Ala237AzPhe}}$.

The fall in k_{cat} correlates well with the size of the modification for both the hydrolysis of ampicillin and cephalothin as the biggest decrease was observed after modification with **1**. This suggests the size of the modification is playing the key role in disrupting the catalytic mechanism. The proximity of Ala237 to the active site (Figure 4-8C) means that any modification is likely to sit in the cleft formed by the two domains, disrupting the interaction between the substrates and the residues involved in catalysis; the larger the modification, the larger this disruption. The increase in the binding affinity of ampicillin – but not cephalothin – when this variant is modified with **1** and **2** suggests a specific interaction between the modifications and ampicillin that doesn't form when cephalothin is used as the substrate. The increase is present when the enzyme is modified with both the modifications, suggesting it may be the DBCO group itself, or the carbonyl group interacting with ampicillin as these groups are shared by the two modifications.

Table 4-9. Effect of Click modification on the activity of TEM^{Ala237AzPhe} against ampicillin and cephalothin.

	Ampicillin			Cephalothin		
	Unmodified	1	2	Unmodified	1	2
k_{cat} (s^{-1})	97.8 ± 6.2	15.0 ± 2.1	21.3 ± 1.4	12.1 ± 1.8	6.70 ± 0.84	10.6 ± 1.4
Relative k_{cat} [1]	1.00	0.15	0.22	1.00	0.56	0.88
K_M (µM)	1340 ± 117	736 ± 180	418 ± 56	258 ± 54	329 ± 55	294 ± 52
Relative affinity [1]	1.00	1.82	3.20	1.00	0.78	0.88
k_{cat}/K_M (µM^{-1} s^{-1})	0.073 ± 0.11	0.020 ± 0.28	0.051 ± 0.15	0.047 ± 0.26	0.020 ± 0.21	0.036 ± 0.22
Relative k_{cat}/K_M [1]	1.00	0.28	0.70	1.00	0.44	0.77

[1] k_{cat}, K_M and k_{cat}/K_M (overall catalytic efficiency) compared directly to the unmodified form of TEM^{Ala237AzPhe}
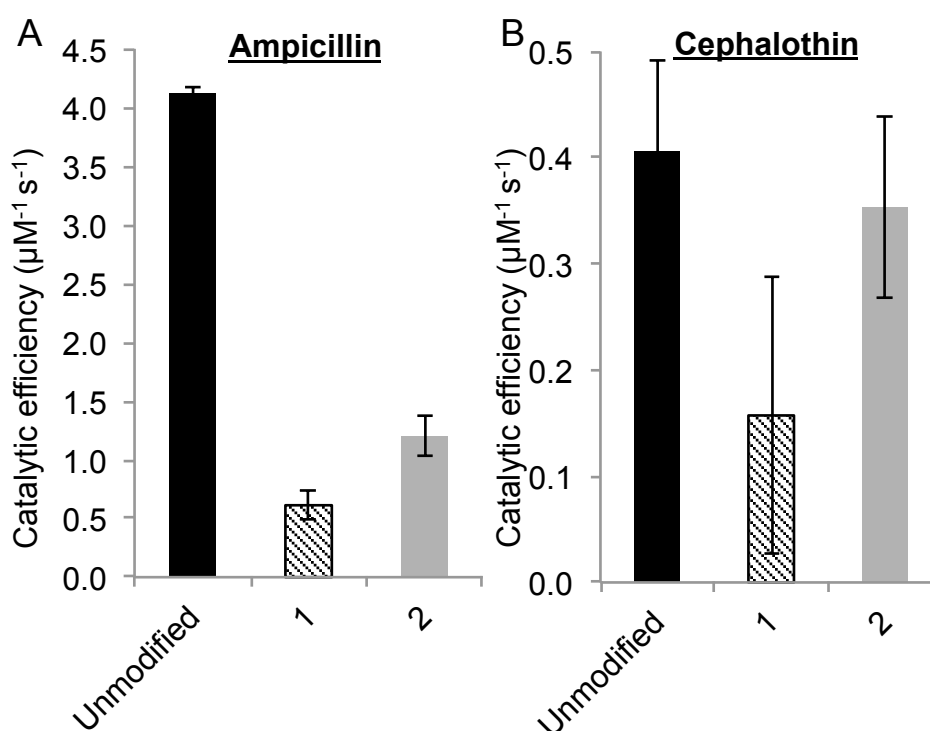

Figure 4-12. Effect of Click modification on TEM^{Ala237AzPhe}. Catalytic efficiency (µM^{-1} s^{-1}) of unmodfied TEM^{Ala237AzPhe}, and TEM^{Ala237AzPhe} modified with **1** and **2** against (A) ampicillin and (B) cephalothin.

When TEM^{Pro174AzPhe} was modified, both positive and negative modulation of enzyme activity was observed depending on which modification was used. When **1** was used, the catalytic turnover of ampicillin was decreased by 56% but binding affinity was increased by 52% (Table 4-10; Figure 4-13A). These produce an enzyme with 66% of the overall catalytic efficiency for ampicillin as the unmodified enzyme. When cephalothin was used as the substrate, there was no effect on k_{cat}, but K_M was increased by 20 μM (Table 4-10; Figure 4-13B). When **2** was used to modify the enzyme, catalytic turnover was more than doubled to 360 s^{-1} when ampicillin was used as the substrate, and K_M was decreased to 83 μM, increasing the binding affinity of ampicillin. These two factors combine to produce an enzyme that when modified with **2** has nearly three times the activity of unmodified TEM^{Pro174AzPhe}. Comparison of the 95% confidence intervals shows only the difference in k_{cat} (143.9-197.7 s^{-1} for unmodified and 334.9-385.7 s^{-1} for modified) is significant; the change in K_M is not. With cephalothin as the substrate, k_{cat} was similarly increased, however there was a 90 μM increase in K_M. This produced an enzyme with an overall catalytic efficiency 59% higher than the unmodified enzyme.

The effect of SPAAC on ampicillin turnover is particularly interesting as the use of the larger modification results in a decrease in k_{cat}, whereas the smaller modification results in an increase. For TEM^{Tyr105AzPhe} and TEM^{Ala237AzPhe}, the decrease in catalytic efficiency was proportional to the size of modification used, however for TEM^{Pro174AzPhe}, the two modifications cause opposite effects, which suggests it is the chemical differences between the two modifications that is causing the difference in activity. **1** has a large, hydrophobic dye structure whereas **2** only has a small amine group on the end of a short linker. The original mutation to TEM^{Pro174AzPhe} caused a decrease in activity, possibly caused by an increase in Ω-loop flexibility caused by the removal of a proline from the loop. This increased flexibility would have a negative impact on the positioning of Glu166, a crucial residue in the catalytic mechanism. The increase in activity back to wild type levels caused by modification with **2** suggests that the Ω-loop now has decreased flexibility, resulting in better positioning of Glu166 and an increase in catalytic turnover. The formation of a hydrogen bond between the amine present in **2** and the side chain of Gln269 could act as an anchor, tethering the Ω-loop to another region of the protein, increasing the stability of the Ω-loop (Figure 4-13C-D).

Table 4-10. Effect of Click modification on the activity of TEM^{Pro174AzPhe} against ampicillin and cephalothin.

	Ampicillin			Cephalothin		
	Unmodified	1	2	Unmodified	1	2
k_{cat} (s^{-1})	171 ± 13	74.3 ± 4.2	360 ± 12	38.6 ± 2.3	39.4 ± 4.1	99.4 ± 7.0
Relative k_{cat} [1]	1.00	0.43	2.11	1.00	1.02	2.58
K$_M$ (µM)	106 ± 29	69.3 ± 17	82.7 ± 11	144 ± 14	165 ± 27	233 ± 24
Relative affinity [1]	1.00	1.52	1.28	1.00	0.87	0.62
k_{cat}/K$_M$ (µM^{-1} s^{-1})	1.62 ± 0.28	1.07 ± 0.25	4.36 ± 0.14	0.27 ± 0.11	0.24 ± 0.19	0.43 ± 0.12
Relative k_{cat}/K$_M$ [1]	1.00	0.66	2.69	1.00	0.89	1.59

[1] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to the unmodified form of TEM^{Pro174AzPhe}

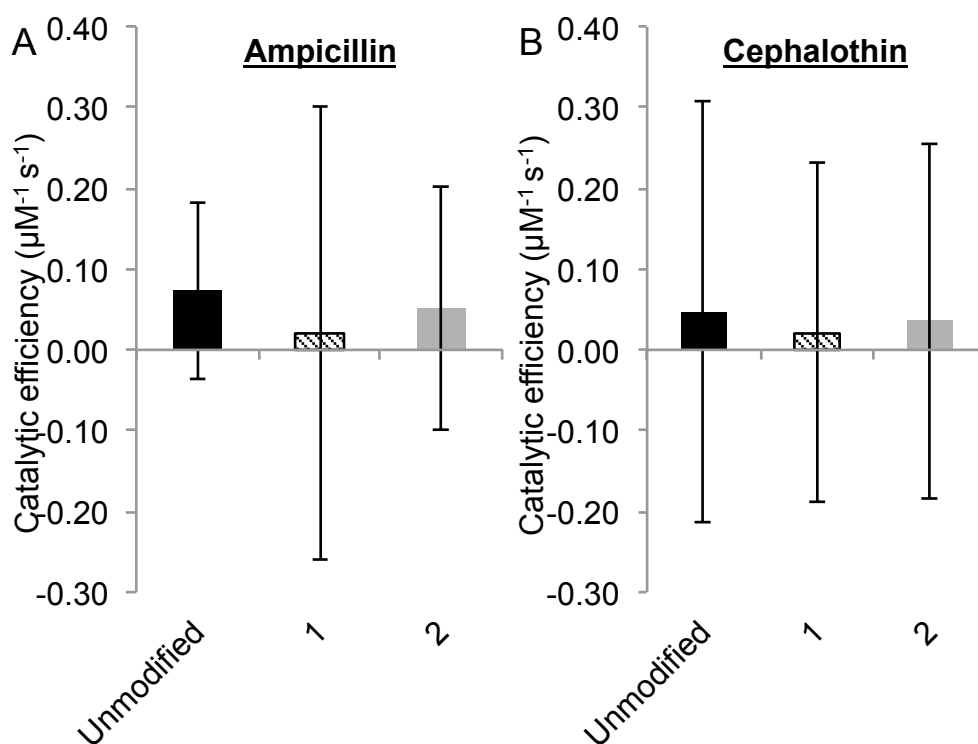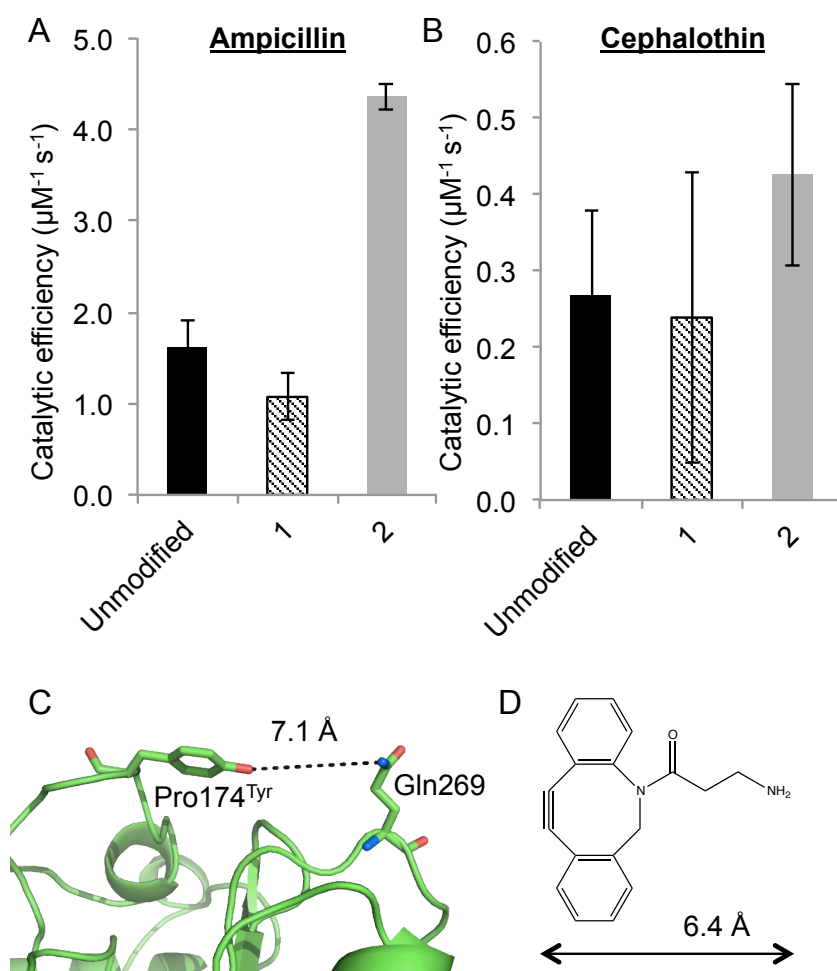
Figure 4-13. Effect of Click modification on TEM^{Pro174AzPhe}. **A-B**. Catalytic efficiency (µM^{-1} s^{-1}) of unmodfied TEM^{Pro174AzPhe}, and TEM^{Pro174AzPhe} modified with **1** and **2** against (A) ampicillin and (B) cephalothin. **C**. Distance between the side chains of Gln269 and a tyrosine modeled at position 174. **D**. The size of DBCO-amine.

Post-translational modification of TEM^{Trp165AzPhe} via SPAAC caused little change in the breakdown of ampicillin (Table 4-11). Modifying the protein with **1** caused an increase in k_{cat}, and a decrease in substrate binding, resulting in an enzyme with 87% of the catalytic efficiency of the unmodified variant. Modification with **2** had a much smaller effect on both k_{cat} and K_M (although a similar pattern was observed). This produced an enzyme with the same overall catalytic efficiency as the unmodified variant. Click chemistry had a larger effect on cephalothin hydrolysis (Table 4-11). Modification with **1** results in a 29% decrease in k_{cat} to 87 s^{-1} and a 24% decrease in substrate binding, resulting in a near two-fold decrease in overall catalytic efficiency. Modification with **2** decreased k_{cat} by 14% to 105 s^{-1}, but K_M was virtually unaffected. After the initial decrease in activity caused by the incorporation of AzPhe, post-translational modification of AzPhe through Click chemistry (and by UV irradiation) does not result in large changes in overall catalytic efficiency against ampicillin. Trp165 is not directly involved in the catalysis of ampicillin, however it is very close to residues that are (e.g. Glu166). These two characteristics of TEM^{Trp165AzPhe} make it an ideal candidate for associating the enzyme to carbon allotropes such as graphene (see Chapter 5). These materials are sensitive to electrostatic changes caused by the function of a variety of proteins [124]. The defined and oriented coupling of TEM β-lactamase to graphene via SPAAC at a residue so close to the active site, but still remaining active, would present an ideal opportunity to monitor the effect of different substrates and inhibitors on the active site.

Post-translational modification of TEM^{Leu201AzPhe} with **1** has little effect on enzyme activity for both ampicillin and cephalothin (Table 4-11). When ampicillin is used as the substrate, both catalytic turnover and substrate binding are reduced by little over 10%, k_{cat} falling to 373 s^{-1}, and K_M increasing to 193 μM, however when cephalothin is used, neither the k_{cat} nor the K_M is significantly changed. Modification with **2** affects the hydrolysis of both ampicillin and cephalothin in a similar manner. Catalytic turnover is decreased for both substrates, down to 308 s^{-1} for ampicillin and 52 s^{-1} for cephalothin. Substrate binding is slightly increased however, up 18% to 145 μM for ampicillin and 31% to 115 μM for cephalothin. These results show that modification of the protein does not result in a uniform change in activity, regardless of where the protein is being modified.

Table 4-11. Effect of Click modification on the activity of TEM^{Trp165AzPhe} and TEM^{Leu201AzPhe} against ampicillin and cephalothin.

	TEM^{Trp165AzPhe}					
	Ampicillin			Cephalothin		
	Dark	1	2	Dark	1	2
k_{cat} (s^{-1})	406 ± 20	572 ± 19	497 ± 38	121 ± 11	86.6 ± 6.0	105 ± 9.0
Relative k_{cat} [1]	1.00	1.41	1.22	1.00	0.71	0.86
K$_M$ (µM)	185 ± 26	302 ± 20	223 ± 34	221 ± 32	289 ± 29	225 ± 28
Relative affinity [1]	1.00	0.61	0.83	1.00	0.76	0.98
k_{cat}/K$_M$ (µM^{-1} s^{-1})	2.19 ± 0.15	1.90 ± 0.075	2.23 ± 0.17	0.55 ± 0.17	0.30 ± 0.12	0.46 ± 0.21
Relative k_{cat}/K$_M$ [1]	1.00	0.87	1.02	1.00	0.54	0.84
	TEM^{Leu201AzPhe}					
	Ampicillin			Cephalothin		
	Dark	1	2	Dark	1	2
k_{cat} (s^{-1})	427 ± 20	373 ± 26	308 ± 20	58.5 ± 5.4	62.6 ± 4.2	52.2 ± 6.2
Relative k_{cat} [2]	1.00	0.87	0.72	1.00	1.07	0.89
K$_M$ (µM)	171 ± 23	193 ± 22	145 ± 28	151 ± 23	153 ± 17	115 ± 24
Relative affinity [2]	1.00	0.88	1.18	1.00	0.99	1.31
k_{cat}/K$_M$ (µM^{-1} s^{-1})	2.50 ± 0.14	1.93 ± 0.13	2.13 ± 0.21	0.39 ± 0.18	0.41 ± 0.13	0.45 ± 0.24
Relative k_{cat}/K$_M$ [2]	1.00	0.77	0.85	1.00	1.06	1.17

[1] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to unmodified TEM^{Trp165AzPhe}

[2] k_{cat}, K$_M$ and k_{cat}/K$_M$ (overall catalytic efficiency) compared directly to unmodified TEM^{Leu201AzPhe}

SPAAC had very little effect on the ability of wild type TEM β-lactamase to hydrolyze both ampicillin and cephalothin. Both k_{cat} and K_M were unaffected by modification with both **1** and **2**, showing SPAAC does not have an effect on enzyme catalysis when there is no handle within the protein for modification.

4.3. Conclusions

4.3.1. Kinetic analysis of TEM β-lactamase

Full kinetic analysis of AzPhe-containing TEM β-lactamase variants revealed the protein activity can be regulated using chemical modifications through defined post-translational modifications facilitated by uAA incorporation within a protein. While most of the post-translational modifications reduced enzyme activity (such as a drop in k_{cat}, or an increase in K_M), there were instances where activity was enhanced through improved turnover and/or substrate affinity. This proves that the designed incorporation of a uAA, combined with site-specific modification is a potentially powerful and versatile approach to protein modification.

4.3.1.1. TEM^{Tyr105AzPhe}: enzyme inhibition through post-translation modification

The initial incorporation of AzPhe at position 105 had no negative effect on the catalytic activity of the enzyme. UV irradiation had a small impact on the overall catalytic efficiency, reducing the k_{cat} against both substrates. However, it was post-translational modification by SPAAC that had the largest impact on enzyme activity. Modification of TEM^{Tyr105AzPhe} with **1** resulted in an 85% drop in activity against ampicillin, and a 60% drop against cephalothin. Modification with **2** resulted in smaller decreases, suggesting that the effect of Click chemistry is dependent on the size of the chemical adduct added to the protein. This effect of near total inhibition provides an excellent model for the use of novel post-translation modifications in the modulation of protein activity. By essentially switching off the enzyme after modification, a novel switch has been incorporated into this variant of TEM β-lactamase.

4.3.1.2. TEM^{Ala237AzPhe}: reduction of activity through AzPhe-incorporation

TEM^{Ala237AzPhe} was only 2% as active as wild type TEM β-lactamase. This massive reduction in activity was the result of changing a small side chain, such as that found in alanine, for a large, aromatic side chain so close to the active site. UV irradiation of TEM^{Ala237AzPhe} had very little effect on the ability of the enzyme to hydrolyze ampicillin and had only a small impact on the catalytic turnover of cephalothin. When TEM^{Ala237AzPhe} was modified via SPAAC, activity against both substrates was reduced further. This variant highlights the need for good mutant design and strategic placement of a novel uAA. By introducing a large side chain so near to the active site,

activity was essentially destroyed, rendering any novel post-translational modifications moot. Computer-aided modeling of potential mutations and even the modeling of the novel post-translational modifications can aid this design process, and increase the likelihood of achieving novel protein control.

4.3.1.3. TEM^{Pro174AzPhe}: positive modulation of activity through Click chemistry

The effects of SPAAC on TEM^{Pro174AzPhe} were variable, depending on which adduct was used to modify the protein. Modification with **2** resulted in a large increase in activity against both substrates, particularly ampicillin. The modification of TEM^{Tyr105AzPhe} resulted in a large decrease in activity when modified with the same chemical adduct. By causing both an increase and a decrease in activity in two different variants of TEM β-lactamase using the same chemical adduct (DBCO-amine − **2**) displays the versatility of these novel post-translational modifications. Achieving both positive and negative control over enzyme activity using a small chemical modification highlights the impact that combining uAAs and novel post-translational modifications can have on protein function and control.

4.3.1.4. TEM^{Trp165AzPhe}: defined and oriented association with carbon allotropes

The limited effect of AzPhe incorporation, and subsequent post-translational modification via both UV irradiation and Click chemistry, at a position so close to the active site offers another facet to the use of uAAs and novel post-translational modifications, namely the opportunity to tether TEM β-lactamase to surfaces. The proximity of Trp165 to the active site means that the catalytic residues and the surface will be in close proximity, opening up the potential for measuring catalytic processes by using the electronic properties of carbon sp^2-bonded materials, such as graphene and carbon nanotubes (see Section 5.2.3.). Novel post-translational modifications could then be used to enable analysis of the catalytic mechanisms of enzymes, with single molecule detail.

4.3.1.5. TEM^{Leu201AzPhe}: no negative impact caused by modification far from active site

Both UV irradiation and SPAAC had very little effect on the activity of TEM^{Leu201AzPhe}, providing a control to show that modification does not result in universal decreases in activity. This shows that mutations to cause a desired effect,

whether that be an increase or decrease in activity, can be designed, and are not just random events caused by the modification of the protein.

4.3.2. Utility of in silico modeling of AzPhe-containing variants

In Chapter 3 the use of *in silico* modeling to hypothesize the potential effects of mutating specific residues to tyrosine (used as a structural homologue of AzPhe) was described. This process mutated a specific residue to tyrosine and using energy minimization processes determined the likely orientation of the new side chain and any impact on the orientation on other side chains. These *in silico* models were then used to hypothesize effects on the activity of the enzyme caused by the incorporation of the uAA and the subsequent modification of the enzyme by both UV irradiation and Click chemistry. Through this process ten positions were selected for mutation based on the likelihood of them impacting the activity of the enzyme both positively and negatively.

One of these ten positions selected, Tyr105, was not mutated *in silico* as the residue was already a tyrosine. The position of this residue in the protein alone was enough reason to select this position. Present in the domain interface that forms the active site, mutation of a tyrosine to AzPhe was suggested to not have a large impact on activity due to the structural similarities between the two amino acids, however any modifications to AzPhe at this positions would likely have an effect on activity due to its proximity to the active site. These two suggestions were both proved to be correct, as mutation to AzPhe resulted in no decrease in activity. However, upon post-translational modification, especially via Click chemistry, large decreases in activity were seen.

Conversely, the modeling of tyrosine at Ala237 suggested a large impact on activity. Ala237 samples a similar position to Tyr105 in that it forms part of the domain interface. The crucial difference between these two residues however is the size of the native amino acid, tyrosine being structurally similar to AzPhe, whereas alanine is not. The introduction of tyrosine at this position had a large impact on the size of the cleft available for substrate entry into the active site, predicting a large impact on activity. This prediction was again correct, as TEM^{Ala237AzPhe} retained only 2% of the

activity of wild type against ampicillin. The activity was so low, that post-translational modification only had a negligible impact on activity.

Although these two variants represent times the *in silico* modeling proved to accurately predict the effect of mutation, there were times where the suggestions inferred from the models were incorrect. Modeling of a mutation at Trp165 suggested a potential negative impact on activity due to the proximity of this residue to Glu166, one of the key residues in the catalytic mechanism. Although the exchange of tryptophan for tyrosine/AzPhe represents a decrease in side chain size, any modification through Click chemistry was thought to likely have a large impact on the precise positioning of Glu166. This prediction proved to be false, as modification of TEM$^{\text{Trp165AzPhe}}$ by both UV irradiation and Click chemistry resulted in very little impact on enzyme activity.

Two residues on the Ω-loop were also selected for mutation based on the results of the *in silico* modeling. Four residues on the Ω-loop, Arg164, Asp176, Arg178 and Asp179, all interact to help stabilize the structure and help position Glu166 for catalysis. All four of these positions were mutated to tyrosine *in silico* to determine which, if any, may result in useful mutations. This process suggested that mutating Arg164 or Asp179 might result in a variant that destroy the interactions initially, but then be able to reform these interactions via UV-induced photo-crosslinking of the phenyl azide side chain. The first part of this hypothesis proved to be correct, as both variants were completely inactive, suggesting the destruction of these interactions on the Ω-loop results in an inactive enzyme. However the second part proved to be incorrect, as no activity could be restored upon UV irradiation of either variant.

Overall, the *in silico* modeling approach used in this study proved to be successful as many of the variants chosen for further study did result in the effects on enzyme activity initially suggested by the models. There were some occasions where effects suggested by the models were not observed experimentally, however this is be to expected in any *in silico* modeling approach. Of the ten mutations originally chosen from the modeling, all were studied using nitrocefin activity analyses, and four of them (Leu201 was chosen later) were studied in more detail using full kinetic analysis against two more clinically relevant substrates. Half the variants initially chosen

yielded interesting results, indicating the success of the *in silico* modeling approach in identifying potential mutations.

4.3.3. uAA incorporation

One of the major points when considering uAA incorporation into proteins is the efficiency at which they are incorporated. Much work has been done over a number of years to overcome the problem of poor uAA incorporation, and continually increase the efficiency of this technique. Original attempts to incorporate uAAs into proteins were not very successful, and even when proteins were produced containing simple uAAs, the yields of these proteins were not very high [125]. The optimization of the process of uAA incorporation has gone a long way to making this technique a viable method. However there remains some debate over the orthogonality of the engineered amino-acyl tRNA-synthetases (aaRSs) used to incorporate uAAs during cellular translation [41, 42]. This debate has led to the introduction of a number of technologies that aid the orthogonality of this system without the further engineering of the translational machinery, such as whole genome engineering, release factor removal and ribosome engineering.

One of the first problems with the current system is the use of the amber stop codon, TAG, as the reprogrammed codon. It is the least used stop codon in the *E. coli* genome, so it is the obvious choice for reprogramming, however it is still the natural stop codon for ~7% of genes in *E. coli*. This means that the uAA will be incorporated at the natural C-terminus of these proteins instead of translation termination. This also results in read through of the stop codon, and can result in elongation products, which may have downstream effects on cellular function. In recent years, whole genome engineering has been used to produce strains of *E. coli* that have had every amber stop codon replaced in their genome [33, 34]. This removes the natural function of TAG as a stop codon, so that TAG becomes another codon that can be assigned an amino acid. By introducing the orthogonal tRNA/aaRS pair into the cell, the TAG can be assigned to a uAA.

Removal of the natural function of the TAG codon as a stop codon also allows release factor 1 (RF1) to be deleted from the genome [126]. *E. coli* have two release factors whose role during translation is to decode the three stop codons. RF1 recognizes UAA

and UAG and RF2 recognizes UAA and UGA. By removing the TAG/UAG codon from the genome, RF1 can be removed as the remaining two stop codons are served by RF2. This now allows a reprogrammed TAG codon system that increases uAA efficiency through reduced premature translation termination, and allows multiple uAAs to be incorporated into the same target [127]. This latter aspect of uAA incorporation was always a problem using traditional methods as the introduction of two or more uAAs into a protein dramatically reduced target protein production due to premature termination [128].

Another avenue that has been exploited is ribosomal engineering. The lack of trinucleotide codons available for the incorporation of uAAs means quadruplet codons are being used more and more, however these are not decoded efficiently by ribosomes. Engineering the ribosome by mutating two residues to form an extended pocket where a four-base tRNA anticodon may bind, increases the decoding efficiency and allows the incorporation of uAAs in response to several quadruplet codons [53]. This technology not only increases the efficiency at which uAAs can be incorporated into proteins, but also allows two or more different uAAs into the same protein through the use of multiple blank codons. The availability of *E. coli* strains that have had the amber stop codon removed, or have an engineered ribosome that can decode multiple quadruplet codons increases the efficiency of uAA incorporation, and thus furthers the application of this technology.

4.4. Summary

This chapter introduced the technique of incorporating uAAs into proteins and the effects on enzyme activity that can be caused by the expansion of the genetic code. The effect on activity was variable depending on the position of AzPhe in the protein. Some positions were highly tolerated, with no impact on activity, whereas other positions had much reduced levels of activity, or even no activity. Even though uAA incorporation is a relatively new technique, a lot of work has been completed to increase the efficiency and reliability of this process, resulting in a technique that can be used to introduce novel functionality and reactivity into proteins reliably and repeatedly. This chapter briefly discusses the limitations of this technique, and how technological advances such as whole genome codon-replacement and ribosome engineering are improving this technique by removing the natural function of the

amber stop codon, or increasing the number of codons available for uAA incorporation.

The incorporation of uAAs into TEM β-lactamase and their subsequent purification enabled in depth kinetic analyses to be performed on AzPhe-containing enzyme variants. Using this method, the impact of AzPhe incorporation could be determined to a detailed level. The incorporation of AzPhe produced a variety of effects on enzyme activity, some residues retaining a high level of activity when mutated to AzPhe and some becoming much less active. Incorporation of AzPhe at TEM^{Ala237AzPhe} resulted in a change in substrate preference, this variant having a lower K_M when cephalothin was used as the substrate compared to ampicillin, the more natural substrate for this enzyme. The subsequent modification of the azide group in these enzyme variants, especially using Click chemistry, brought about much larger changes in activity. The post-translational modification of TEM^{Tyr105AzPhe} resulted in large decreases in activity, likely due to the proximity of this residue to the active site. Other residues positioned further from the catalytic residues had a much smaller effect on activity. The modification of TEM^{Pro174AzPhe} using Click chemistry had interesting effects on ampicillin hydrolysis. This variant lost some activity upon incorporation of AzPhe however after Click chemistry modified this residue to include an amine group, activity was restored to a level that exceeded wild type.

The variety of the effects caused by AzPhe incorporation and modification highlight the ability of genetic code expansion and uAAs to incorporate novel functions and novel control into proteins. The variety of modifications available to biochemists for use in Click chemistry and SPAAC means there is an almost limitless supply of opportunity to incorporate desired changes. This work used a rational design approach as a way to introduce mutations that would result in changes in activity. The *in silico* approach used helped the identification of interesting variants that might lead to novel changes in activity upon both incorporation of the uAA and subsequent modification. However, many of the predicted effects proved to be smaller in practice. This highlights the complex interplay between residues in proteins, particularly in the active site of enzymes, and highlights the difficult nature of designing novel post-translational effects into proteins using uAAs.

5. Using AzPhe-mediated non-natural post-translational modification to introduce new functionality into proteins

5.1. Introduction

The incorporation of *p*-azido-L-phenylalanine (AzPhe) into defined positions in a target protein enables the use of novel post-translational modifications that can modulate the function and control of a protein (see Chapters 4 and 6). However, these novel post-translational modifications can also be used to introduce additional functionality. The immobilization of proteins onto surfaces is a technique that has revolutionized the study of protein-protein interactions through the construction of high-throughput nanoarrays that allow the simultaneous analysis of large numbers of proteins [129], and the study of protein-substrate interactions, such as small molecule screening in drug discovery [130, 131]. As well as increasing the numbers of proteins that can be studied quickly, protein immobilization can also enable single-molecule studies [132, 133]. Original methods of attaching proteins to surfaces involved physical adsorption, whereby a protein would attach to a surface via unspecific interactions such as hydrophobic or electrostatic interactions [134]. This technique was replaced by covalent attachment via the functional side chains of particular residues such as lysines and cysteines, which improved the strength of the interaction but did not address specificity, as proteins (and the entire proteome) generally have multiple surface exposed side chains of this type. The use of AzPhe and Click chemistry can help overcome this lack of specificity. The incorporation of a single azide group at a defined position in the target protein via the use of AzPhe, and subsequent site-specific post-translational modification, enables attachment to a surface with increased specificity with regards to the orientation of the protein on the surface. This defined orientation is crucial in the surface-based study of enzymes, as access to active site and other allosteric binding sites must be consistent between different enzymes and enzyme variants.

This chapter introduces two methods used to attach TEM β-lactamase to a surface. The first involves the post-translational modification of AzPhe with a functionalised pyrene group. This pyrene group is able to interact strongly with sp^2-bonded materials such as graphene via a π-π stacking interaction [135]. Graphene is a highly useful

material as not only does it have a high strength-to-weight ratio, but it also has conductive properties that can aid the analysis of proteins [136]. The second method is through the use of DNA origami. DNA origami [137] is a technique whereby single stranded DNA is designed to self hybridize into complex 2- or 3-dimensional shapes, creating a scaffold that can be used to immobilize proteins. DNA is an ideal material to attach proteins to surfaces as the nature of DNA and its complementarity means proteins can be organized upon a surface. The protein organization benefits offered by DNA, and the protein orientation benefits offered by AzPhe and Click chemistry, can combine to form a highly useful method for surface-based protein analyses.

When Click chemistry is being used to modulate the function of a protein, or introduce new functionality, the efficiency at which these chemical modifications occur (and what determines this efficiency in terms of the protein microenvironment) is of the utmost importance. If only a small percentage of the protein population is modified, the utility of the modification will be compromised. In this chapter the final modification efficiency of each AzPhe-containing TEM β-lactamase variant, and the rate at which each variant is modified is investigated, and using these results, the effects of incorporation site on the modification efficiency is discussed.

5.2. Results
5.2.1. Effect of AzPhe position on Click modification efficiency

Click chemistry (and in particular strain-promoted azide-alkyne cyclo-addition – SPAAC) is able to modulate protein function, and even introduce new functionality into a variety of proteins via the defined, covalent triazole linkage between the azide group of AzPhe and an alkyne group. This cyclo-addition reaction, between the only azide group in the protein and the ring strained cyclo-octyne offers more specificity and selectivity over other methods of post-translational modification such as lysine or cysteine modification. However, the efficiency of this novel modification must reach a high level if the technique is to prove useful. If the majority of protein remains unmodified, then any effect that may be caused by the modification will effectively be lower and suffer from background effects caused by unmodified protein. To determine the effect of different positions on the efficiency of the SPAAC reaction, a coloured dye (**1** – Figure 4-4) was used to modify TEM β-lactamase. A comparison of

the absorbance by the protein and by the dye can then be used to calculate the amount of protein that has been modified with the dye.

5.2.1.1. Final modification efficiency

It has previously been observed for sfGFP that residue position can determine the final extent to which a protein population is modified [110]. To see if this effect is repeated in TEM β-lactamase, the final extent of modification and efficiency was determined (Table 5-1). The most efficiently modified variant was TEM[Tyr105AzPhe], which reached a final efficiency of 83%. Three variants were modified to a similar extent to each other; TEM[Ala237AzPhe], TEM[Trp165AzPhe] and TEM[Leu201AzPhe]. These three variants were approximately 50% modified, the highest modification efficiency being 60% for TEM[Trp165AzPhe]. The second variant that modified to ~50% was TEM[Ala237AzPhe], which reached a final modification of 55%. TEM[Leu201AzPhe] was modified to a similar extent again, reaching a final efficiency of 47%. TEM[Pro174AzPhe] was modified to the lowest extent of all the variants tested, with a final modification efficiency of 33%.

Table 5-1. Modification efficiency, surface exposure and flexibility of the five AzPhe-containing TEM β-lactamase variants calculated using the absorbance of a DBCO-modified dye

	Efficiency of SPAAC modification (%)	Surface exposure (Å^2) [1]	B-factor (structure) [2]
TEM[Tyr105AzPhe]	83	113	22.8 (loop)
TEM[Trp165AzPhe]	60	74	8.5 (loop)
TEM[Ala237AzPhe]	55	26	5.5 (β-sheet)
TEM[Leu201AzPhe]	47	107	15.9 (α-helix)
TEM[Pro174AzPhe]	33	54	12.2 (loop)

[1] Surface exposure was calculated using NetSurfP [109] to indicate the area of a particular residue that is accessible to solvent

[2] B-factor used to indicate the flexibility of a particular residue, or the likelihood of a residue sampling a single, fixed orientation

5.2.1.2. Modification rate

Determining the extent to which each AzPhe-containing TEM β-lactamase variant was modified showed that all the variants were modified differently, some reached a high level of modification whereas others did not. Another aspect to SPAAC

efficiency is modification rate. Post-translational mechanisms of protein control rely on a rapid event so to control protein activity quickly. If the reaction rate between the protein-borne azide and the cyclo-octyne is slow, the utility of the approach could be diminished, as protein activity requires rapid modulation. To determine the extent of modification over time and to determine at which point the reaction reaches completion, two methods were used. For those variants susceptible to changes in activity, enzyme activity was measured at several time points up to 24 hours after the addition of the DBCO adduct to determine modification rate. However, as not all the TEM β-lactamase variants are affected equally, an additional approach was used to directly measure modification. The absorbance of the conjugated dye was measured via fluorescence imaging of SDS-PAGE and subsequent analysis using image software. The combination of these two methods gives a good representation of the rate at which each variant is modified, and reaches its final point.

As with the final extent of modification, the rate of modification was different for each variant. In general, a higher rate of modification corresponded with a higher final modification efficiency and *vice versa*. TEM^{Tyr105AzPhe} was modified to the highest efficiency of all the variants (Table 5-1), and it did so at the fastest rate (Figure 5-1A). After 10 minutes, there was a noticeable decrease in activity (from 360 μM min^{-1} with no modification to 186 μM min^{-1}), and after 4 hours (116 μM min^{-1}), there was no further decrease in activity. Analysis of the SDS-PAGE gel showed that the fluorescence of the modified protein increased to almost completion after just 10 minutes of incubation with the dye (Figure 5-1D). Fluorescence increased slightly between 10 and 60 minutes, but after this time no increases in fluorescence are detected, suggesting modification had reached its final point after only 60 minutes.

TEM^{Ala237AzPhe} reached a final modification efficiency of over 50%, however the rate of this modification was slower than TEM^{Tyr105AzPhe}. Activity dropped after every time point sampled, including the final time point used at 24 hours (Figure 5-1B). Analysis of SDS-PAGE showed increases in fluorescence up to 2 hours and then no further increases after this time (Figure 5-1E). Although there are no changes in fluorescence after 2 hours, the decrease in activity suggests modification is either not complete at this point or that secondary conformational changes occur after modification that result in changes in activity. TEM^{Trp165AzPhe} is largely unaffected by modification, and

this is seen here, with little change in activity even after 24 hours of incubation with the chemical adjunct (Figure 5-1C). The initial rate of the variant before post-translational modification is 490 μM min^{-1}, and after 24 hours of incubation with the modification this only decreases to 433 μM min^{-1}. Analysis of the SDS-PAGE gel showed increases in fluorescence all the way up to 6 hours, with a slight increase up to 24 hours (Figure 5-1F). After two hours incubation with the dye there is very little modification occurring. It is only after 4 hours that there are large increases in fluorescence.

Figure 5-1. Modification over time of TEM β-lactamase with a DBCO-modified dye. **A**, **B** and **C**. Nitrocefin activity analyses of (A) TEM^{Tyr105AzPhe}, (B) TEM^{Ala237AzPhe} and (C) TEM^{Trp165AzPhe}. **D**, **E** and **F**. Fluorescence analysis of an SDS-PAGE gel with modified (D) TEM^{Tyr105AzPhe}, (E) TEM^{Ala237AzPhe} and (F) TEM^{Trp165AzPhe}. Coomassie stained gels show equal loading in each lane.

TEM^{Pro174AzPhe} was modified with the lowest efficiency of all the TEM β-lactamase variants, reaching a final modification efficiency of 33% (Table 5-1). The rate of modification for this variant was also the slowest. The activity of the protein barely decreased at all after an hour of incubation with the dye (from 125 µM min^{-1} to 119 µM min^{-1}), with inhibition clearly observed after several hours of incubation (Figure 5-2A). Analysis of gel fluorescence shows that there was no clear signal until after 4 hours of incubation with the dye (Figure 5-2C). Fluorescence continues to increase after 6 hours, suggesting that the reaction may not have reached its completion after 24 hours. The activity of TEM^{Leu201AzPhe} was not affected by modification by the dye, as no decrease in activity was seen until after 24 hours of incubation, from 450 µM min^{-1} to 382 µM min^{-1} (Figure 5-2B). Analysis of gel fluorescence showed steady but small increases in fluorescence up to 6 hours, with very little increase after that (Figure 5-2D). This suggests that the modification takes between 6 and 24 hours to reach a final modification efficiency of 47%.
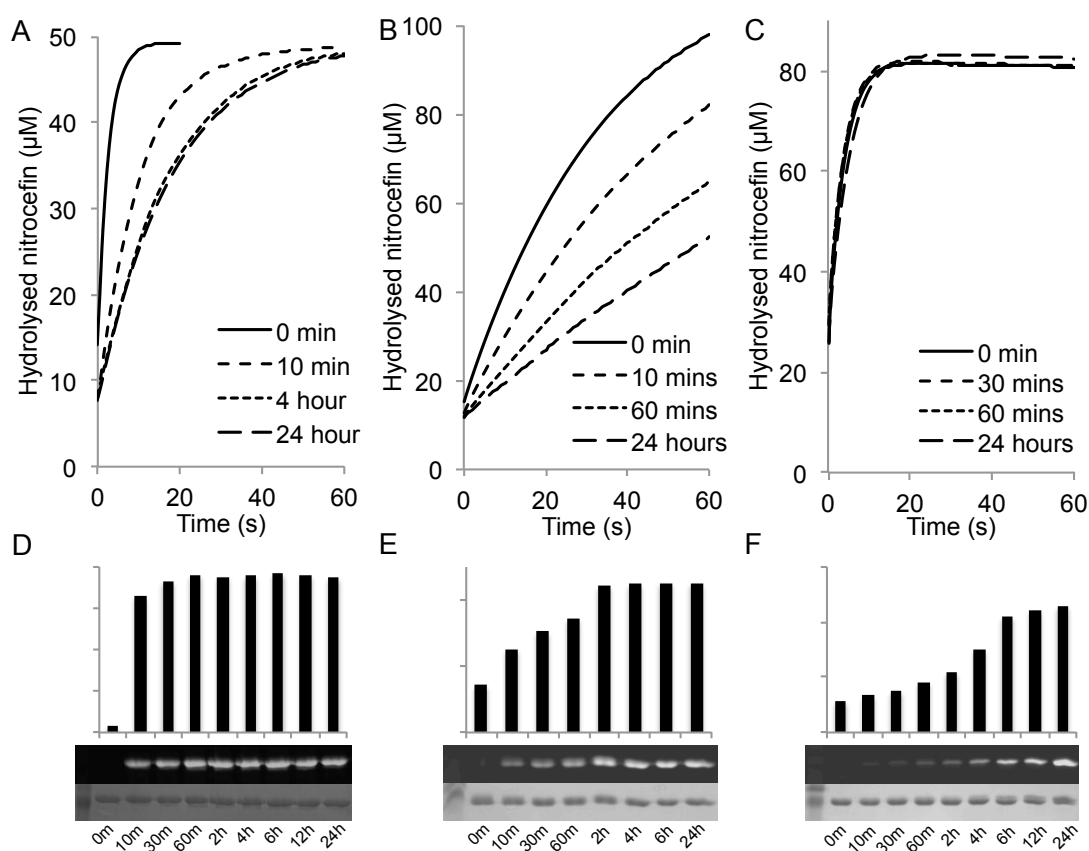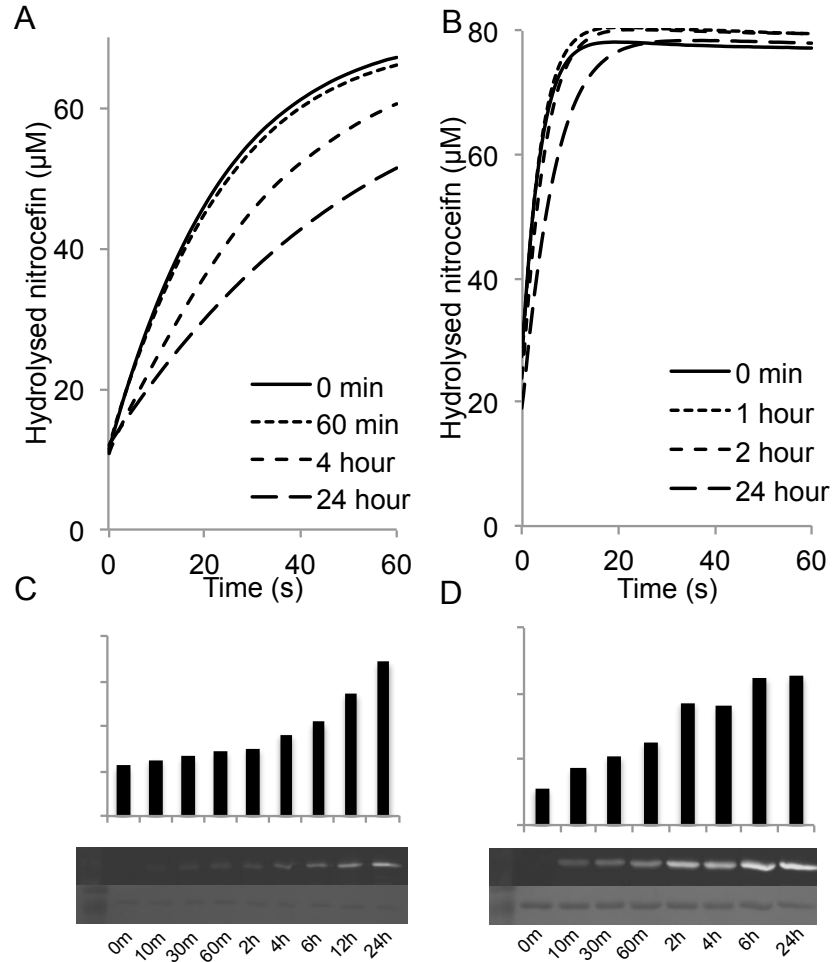
Figure 5-2. Modification over time of TEM β-lactamase with a DBCO-modified dye.
A and **B**. Nitrocefin activity analyses of (A) TEM^{Pro174AzPhe} and (B) TEM^{Leu201AzPhe}. **C** and **D**. Fluorescence analysis of an SDS-PAGE gel with modified (C) TEM^{Pro174AzPhe} and (D) TEM^{Leu201AzPhe}. Coomassie stained gels show equal loading in each lane.

5.2.2. Attaching TEM β-lactamase to DNA origami

The use of DNA as a tool to organize proteins on a surface was first suggested in 1982 [138], and then developed into the method known as DNA origami in 2006 [137]. The predictable nature of base pairing means DNA can be designed to form corners and junctions rather than double-stranded duplexes. This enables the DNA, with the aid of "staple" strands that solidify the structure, to form larger, more complex shapes (Figure 5-3). This characteristic of DNA means it can be used as a scaffold for the study of various biomolecules, through the use of complimentary DNA oligonucleotides. These DNA oligonucleotides can be designed to hybridize to the single stranded DNA origami "tile" at a single position, thereby specifically defining the position of a single biomolecule. The nature of DNA complementarity means several DNA strands can be designed to bind to the tile at different places, enabling the construction of protein networks such as signaling cascades or enzyme pathways.

ssDNA   +   88 small ssDNA fragments ("staples")   =
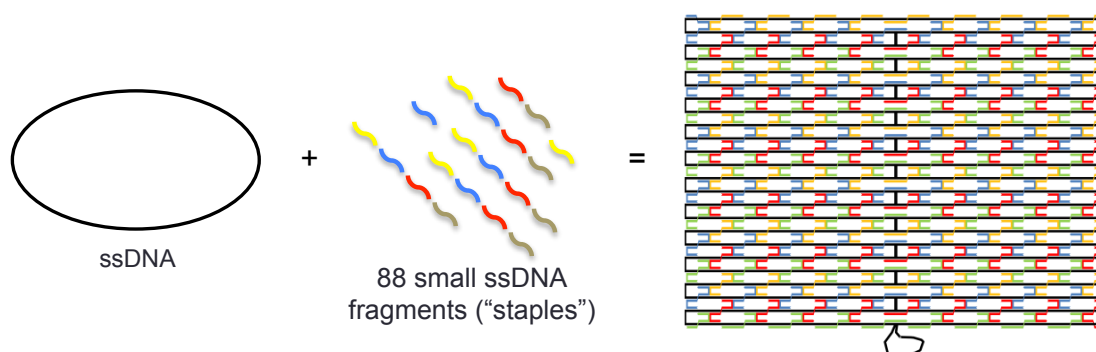
Figure 5-3. Schematic of DNA origami. Single stranded DNA folds, with the aid of small "staple" strands into a complex 2- or 3-dimensional structure.
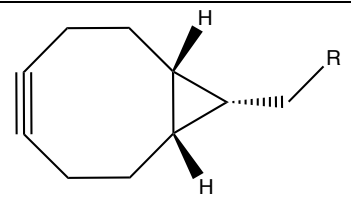
The conjugation of biomolecules to the DNA oligonucleotide is an important step in the success of DNA origami. This conjugation has been previously achieved through biotin-streptavidin interactions [139], histidine-NTA interactions [140, 141] and direct protein-DNA (aptamer) interactions [142]. The use of SPAAC to attach the protein to the oligonucleotide strands offers several benefits over the methods previously used. The use of AzPhe to conjugate the protein to a DNA oligonucleotide via SPAAC removes the need for large molecules such as biotin or a His-tag to be introduced to the protein of interest. Additionally, the site-specific nature of AzPhe-incorporation allows the orientation of a protein on the DNA origami tile to be precisely defined, as

well as the location of the protein on the surface. These benefits mean AzPhe-DNA interactions are an ideal method to isolate a protein on a DNA tile.

5.2.2.1. Modifying TEM β-lactamase with DNA oligonucleotides

Two DNA oligonucleotides were used, both modified at the 5' end with a cyclo-octyne for reaction with the azide group of AzPhe. These oligonucleotides are designed to bind to the DNA origami tile in different ways, and as such differ in the length of the nucleic acid chain, and therefore their molecular weight (Table 5-2). One of the strands (B5900) is designed to hybridize to one of the staple strands that are used to construct the tile, and the second DNA strand (B5963) is an extended staple, used to help the folding and formation of the DNA tile. Five TEM β-lactamase variants were modified with both oligonucleotides (Figure 5-4). Gel mobility shift assays were used to assess the protein-oligonucleotide conjugates. The migration of the protein-DNA conjugate through the SDS-PAGE gel varied depending on both the DNA oligonucleotide used, and the residue modified. Modification of TEM β-lactamase at TEM^{Pro174AzPhe} and TEM^{Trp165AzPhe} had the largest effect on the migration having a larger apparent molecular weight than other TEM β-lactamase variants modified at other positions (Figure 5-4). The pattern of migration is consistent for both DNA oligonucleotides used. Wild type was also incubated with both DNA oligonucleotides but no band shift was observed confirming that modification with the oligonucleotides had not occurred.

Table 5-2. The structure of the cyclo-octyne conjugated to the 5' end of the DNA staple oligonucleotides used and the sequences and molecular weights of those DNA strands.

	B5900. GGTGGAGTGATGGATGGGAG
	Molecular weight – 6.7 kDa
	B5963. TTTTTTACTCACATTAATTGCGACCTGTCGTGCCAGCT
	Molecular weight – 11.9 kDa
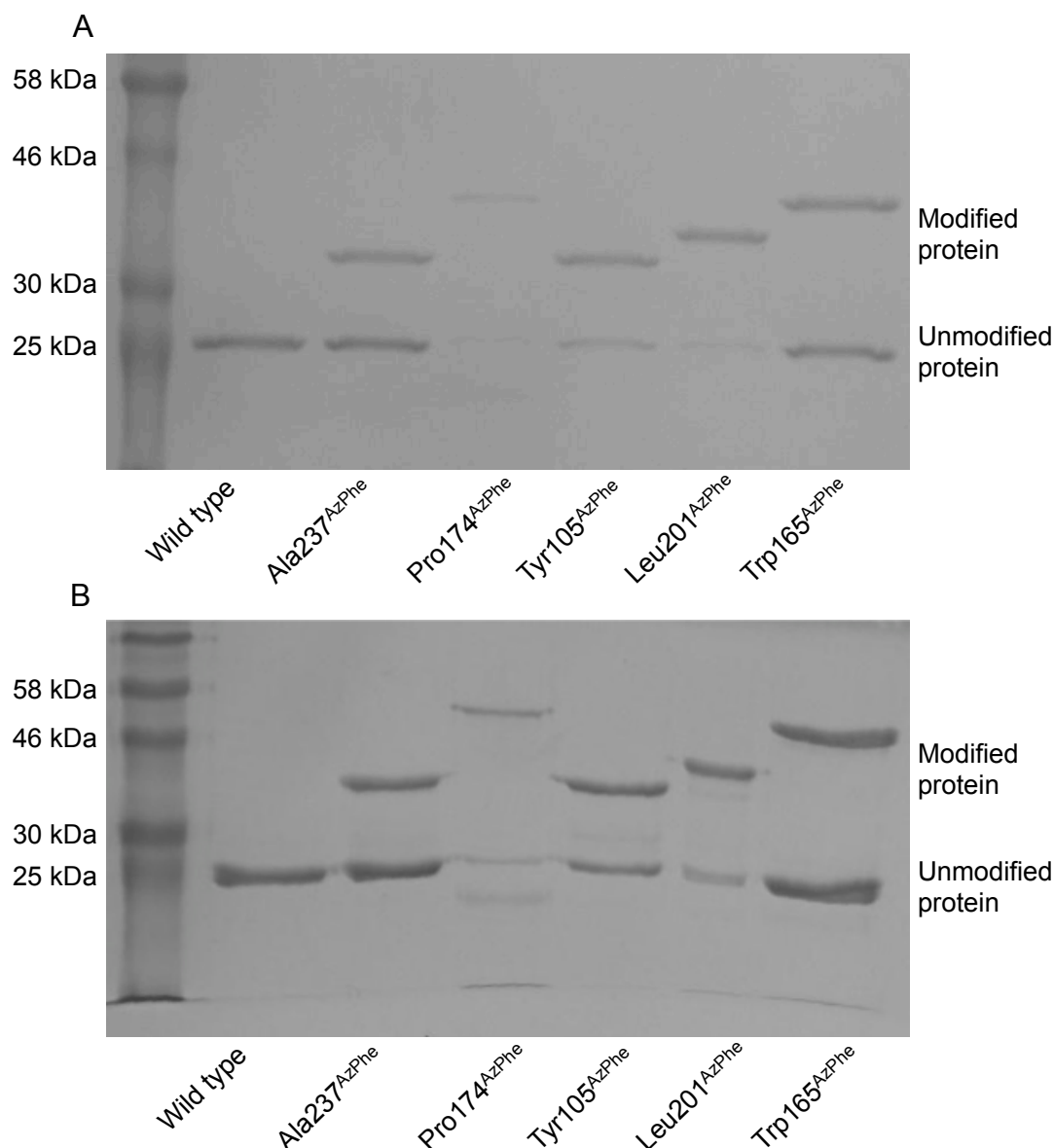
Figure 5-4. SDS-PAGE analysis of five AzPhe-containing TEM β-lactamase variants modified with (A) B5900 and (B) B5963.

As observed in Chapter 4, modification of TEM β-lactamase can influence activity, so the activity of each variant modified with the two DNA oligonucleotides was also determined via nitrocefin activity assays (Figure 5-5). The activity of TEM$^{\text{Tyr105AzPhe}}$ was most affected by modification with the two DNA oligonucleotides, reducing the initial rate from 180 μM min^{-1} to 30 μM min^{-1} for B5900 and 20 μM min^{-1} for B5963 (Figure 5-5A). This shows that the size of the DNA oligonucleotide conjugated to the protein does not have a great effect on the activity of the modified enzyme, as modification with both oligonucleotides resulted in similar decreases. The activity of

TEM^{Pro174AzPhe} was not as effected by the modifications, modification with B5900 reduced the initial rate from 100 µM min^{-1} to 80 µM min^{-1} and modification with B5963 reduced the initial rate to 75 µM min^{-1} (Figure 5-5C). Although the activity of TEM^{Ala237AzPhe} was decreased when modified with both of the DNA oligonucleotides (Figure 5-5E), the initial rate was lower when TEM^{Ala237AzPhe} was modified with the shorter staple strand (35 µM min^{-1}) rather than the extended staple strand (55 µM min^{-1}). Modification at Trp165 (Figure 5-5B) and Leu201 (Figure 5-5D) with either DNA oligonucleotide had very little impact on the activity of the enzyme.
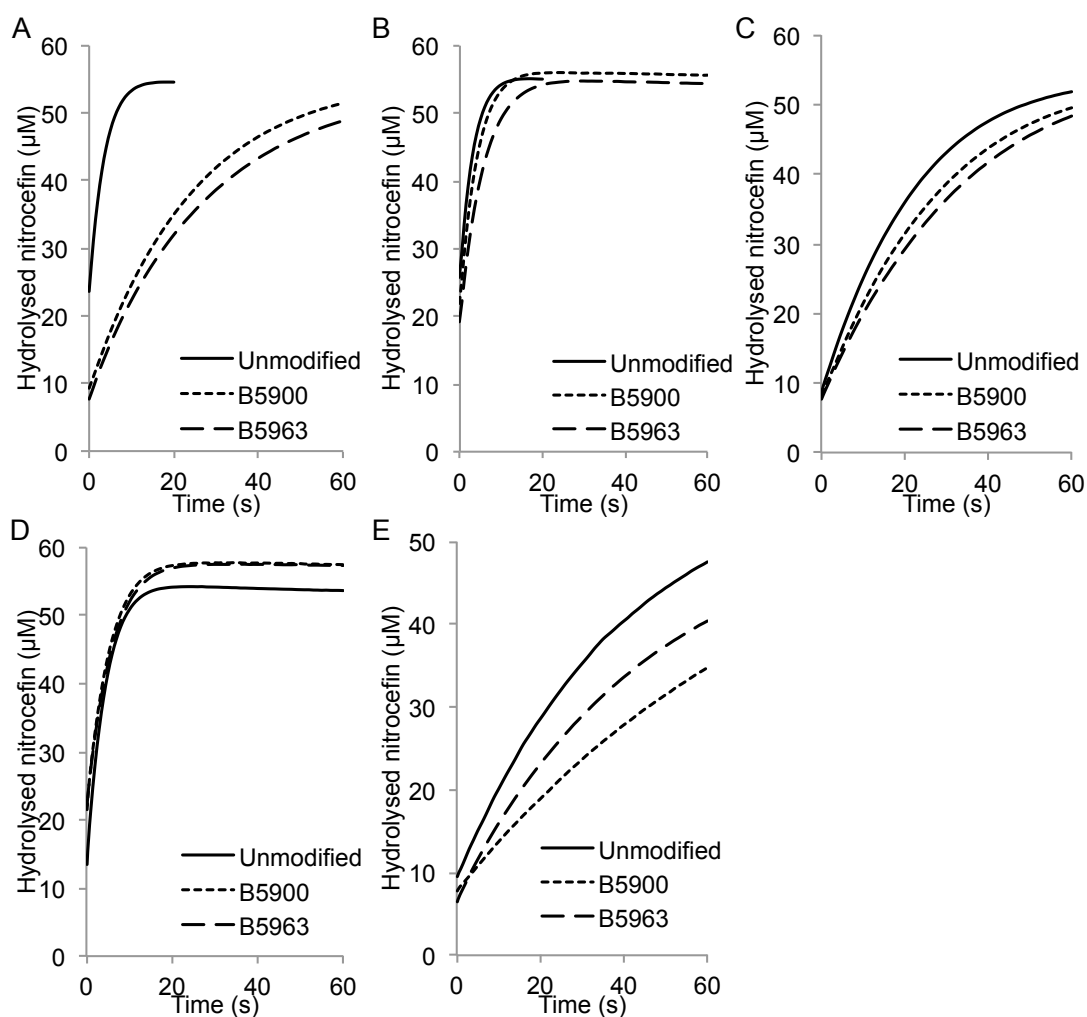
Figure 5-5. Nitrocefin activity analysis of five TEM β-lactamase variants modified with the two DNA staple strands. The activity of (A) TEM^{Tyr105AzPhe}, (B) TEM^{Trp165AzPhe}, (C) TEM^{Pro174AzPhe}, (D) TEM^{Leu201AzPhe} and (E) TEM^{Ala237AzPhe} was determined before modification, and after 24 hours incubation with the two DNA strands.

5.2.2.2. Approaches to separate modified TEM β-lactamase from unmodified

For the attachment of the modified enzymes to the DNA origami tile, it is important to separate the modified proteins from the unmodified. As there was a large degree of separation between modified and unmodified enzymes on an SDS-PAGE gel, size exclusion chromatography was initially used to try and separate modified from unmodified protein (Figure 5-6A). A small peak eluted from the column before the main peak, suggesting that the modified protein was purified from the unmodified, however SDS-PAGE analysis of the two peaks revealed there was no separation, both the modified and unmodified proteins eluting in the main peak. Ion exchange chromatography was then used to separate the proteins using the negative charge of the DNA strand to bind to an anion exchange column (Figure 5-6B). Successful separation of modified and unmodified protein was confirmed using SDS-PAGE (Figure 5-7).
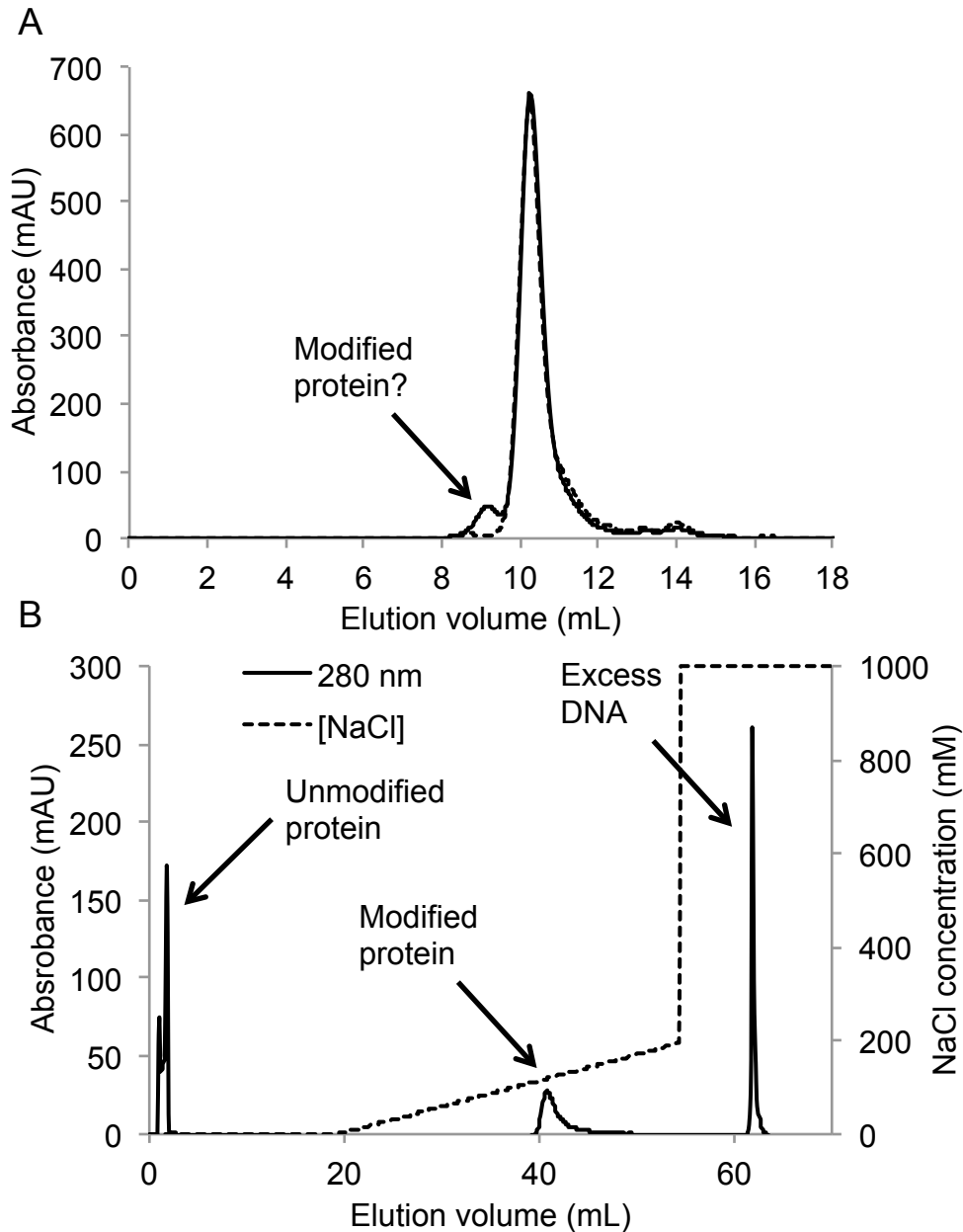
Figure 5-6. Separation of unmodified TEM β-lactamase from protein modified with the DNA staple strand. **A**. Size exclusion chromatography of TEM$^{\text{Tyr105AzPhe}}$ modified with B5900. The small peak prior to the main peak did not contain modified protein after SDS-PAGE analysis (dashed line is unmodified TEM$^{\text{Tyr105AzPhe}}$). **B**. Ion exchange chromatography of TEM$^{\text{Tyr105AzPhe}}$ modified with B5900. The negative charge of the DNA strand bound strongly to an anion exchange column and was successfully separated from the unmodified protein.
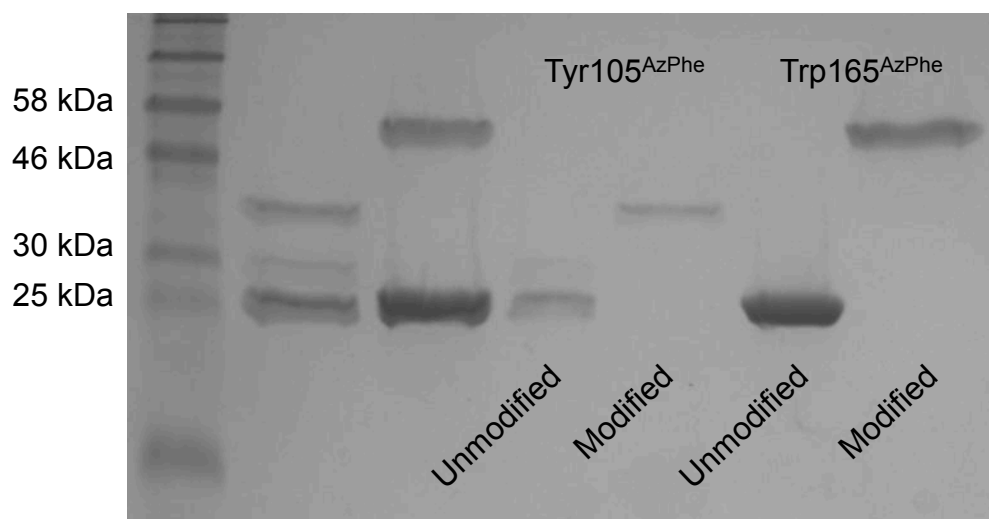
Figure 5-7. SDS-PAGE analysis of the ion exchange chromatography of modified TEM^{Tyr105AzPhe} and TEM^{Trp165AzPhe}. The first two lanes are a mixture of modified and unmodified TEM^{Tyr105AzPhe} and TEM^{Trp165AzPhe} respectively. These were then separated into different fractions using ion exchange chromatography.

5.2.3. Attaching TEM β-lactamase to carbon allotropes

Since the isolation and description of graphene in 2004 [143], numerous applications for the material have been suggested, from its use in touch screens and solar cells [144, 145], to a use in gas detection [146]. One potential use for graphene in a biological context is the immobilization of biomolecules onto a surface and their subsequent analysis [147, 148]. The conductive properties of graphene and other carbon allotropes means they are very sensitive to electronic changes in the proteins, and can communicate these changes to electrodes very quickly and efficiently [124]. This means that tiny changes in biomolecules such as charge and electrostatic changes that occur in protein-protein interactions, or protein-substrate interactions can be detected. Therefore, graphene is seen as an ideal material to which proteins can be tethered in a specific and defined manner, enabling the analysis of single molecules, and thus the measurement of individual biological events. One way of adhering proteins to a surface like graphene is through the use of pyrene, a planar, polycyclic hydrocarbon that can interact with the surface of graphene via a very strong π-π stacking interaction. The attachment of proteins to carbon allotropes has been accomplished using pyrene by coating a carbon nanotube with pyrene and then subsequently attaching a protein via cysteine chemistry [149]. Although this method allowed defined placement of the protein, it required extensive molecular biology to

115

remove all natural cysteines from the protein. The use of AzPhe removes this need, as only one will be present in the target protein, whilst maintaining the ability to specifically design the orientation of a protein on the surface. By functionalizing the pyrene group with a dibenzyl cyclo-octyne (Figure 5-8), it becomes SPAAC reactive and can be used to modify the unique azide group in the protein via AzPhe.

5.2.3.1. Modifying TEM β-lactamase with pyrene

To attach TEM β-lactamase to the surface of a carbon allotrope it first must be modified with pyrene. However pyrene itself cannot react with the azide group so it must be functionalized with a DBCO (dibenzyl cyclo-octyne) group to enable SPAAC between the azide group of AzPhe and the alkyne of the DBCO. Pyrene was synthesized using DBCO-amine (as used in Chapters 4 and 6) and pyrene butanoic acid, succinmidyl ester (Sam Reddington, personal communication – Figure 5-8). This yielded an azide-reactive pyrene-based chemical modification that can be used to reactive specifically with AzPhe.
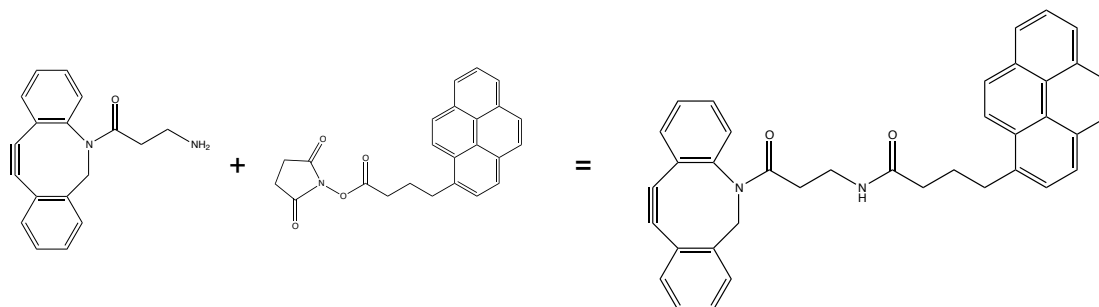
Figure 5-8. The synthesis of DBCO-pyrene. DBCO-amine and pyrene butanoic acid, succinmidyl ester were used to synthesize DBCO-pyrene that is able to react with the azide group of AzPhe.

Two TEM β-lactamase variants that contained AzPhe near the active site were then modified with the DBCO-pyrene, TEM^{Tyr105AzPhe} and TEM^{Trp165AzPhe}. Tyr105 forms one side of the active site whereas Trp165 is present on the Ω-loop, a crucial structure that sits over the active site (Figure 5-9A). Modification with pyrene resulted in two opposite effects, depending on where the DBCO-pyrene was conjugated to the protein. Modification of TEM β-lactamase at TEM^{Tyr105AzPhe} with DBCO-pyrene resulted in a decrease in activity, whereas modification at TEM^{Trp165AzPhe} had no effect on the activity of the enzyme (Figure 5-9B-C). The modification of the variants was

confirmed using the absorption of the pyrene group, which has a distinctive λ_{max} at 346 nm (Figure 5-9D). These two positions exhibit differential effects of SPAAC and are excellent candidates for immobilization on graphene, as they are both close to the active site and in a position to link the charge and electrostatic changes that occur in the active site during catalysis and/or inhibitor binding to the conductive properties of graphene.
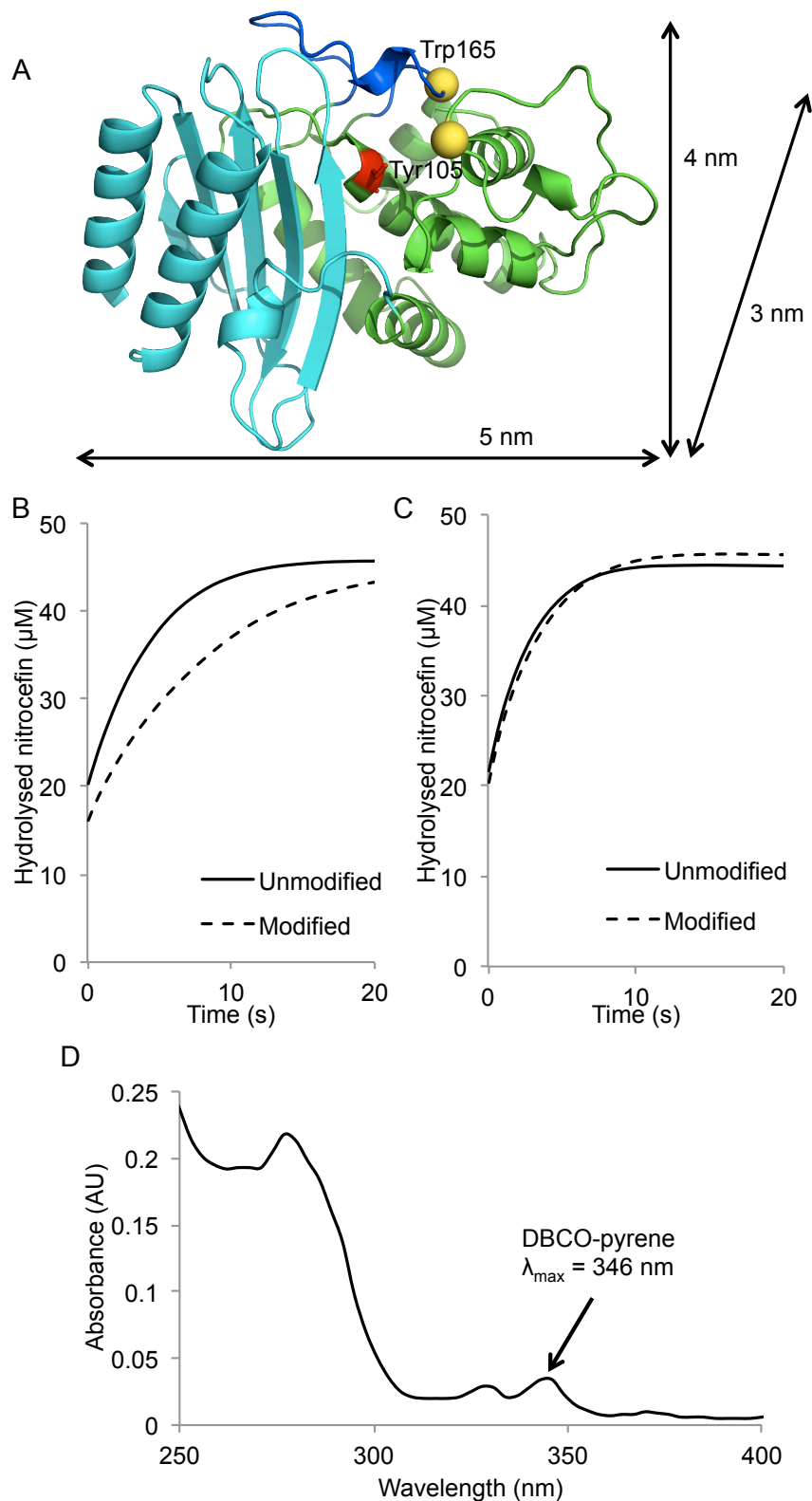
Figure 5-9. Modification of TEM β-lactamase with DBCO-pyrene. **A**. Positions of Tyr105 and Trp165 in TEM β-lactamase and the dimensions of the protein. **B** and **C**. Nitrocefin hydrolysis activity analysis of (B) TEM$^{\text{Tyr105AzPhe}}$ and (C) TEM$^{\text{Trp165AzPhe}}$ modified with DBCO-pyrene (Figure 5-8). **D**. Absorption spectrum of TEM$^{\text{Trp165AzPhe}}$ modified with DBCO-pyrene (λ_{max} = 346 nm).

5.2.3.2. Attaching TEM β-lactamase to graphite

To mimic the graphene surface, highly ordered pyrolytic graphite (HOPG) was used in initial experiments with TEM^{Tyr105AzPhe}. Two main methods were used to form a stable interaction between the pyrene and the HOPG, binding DBCO-pyrene to the surface with no protein, and binding protein already modified with the DBCO-pyrene. Two further methods were then used to incubate the protein with HOPG, drop-casting whereby 20 µL of protein is dropped onto the HOPG surface and immersion where the HOPG is submerged in 100 µL of protein. Protein concentration was varied from 0.01 µM to 5 µM and incubation time was varied from 5 – 30 minutes. Wash conditions were also varied. The surfaces were then imaged using atomic force microscopy (AFM) courtesy of A. Zaki, Martin Elliot and Emyr MacDonald, Cardiff University School of Physics.

Of all the combinations used, the best results were seen when pyrene was laid down on the surface prior to modification with the protein. 0.2 mM pyrene was incubated with the HOPG for 15 minutes and then the surface was washed with 100% DMSO, then rinsed with water and dried with N_2 gas. The HOPG with pyrene was then immersed in 100 µL TEM^{Tyr105AzPhe} and wild type TEM β-lactamase (both 1 nM) for 10 minutes and then rinsed with water and dried with N_2 gas as before. AFM can measure height profiles much more accurately than it can measure lateral dimensions, so in the analysis of the surfaces, the heights of objects found on the surface were primarily analyzed, rather than the 2-dimensional size of the object. Initially the surface labeled with only the pyrene was imaged (Figure 5-10A). The height profile of this image shows the expected height and pattern of pyrene laid down on the HOPG surface. When proteins were incubated with the surface, the height profiles of suspected proteins matched the expected values of between 3 and 4 nm (Figure 5-10B-C). When wild type TEM β-lactamase was incubated with the surface, no objects were seen on the surface (Figure 5-11).
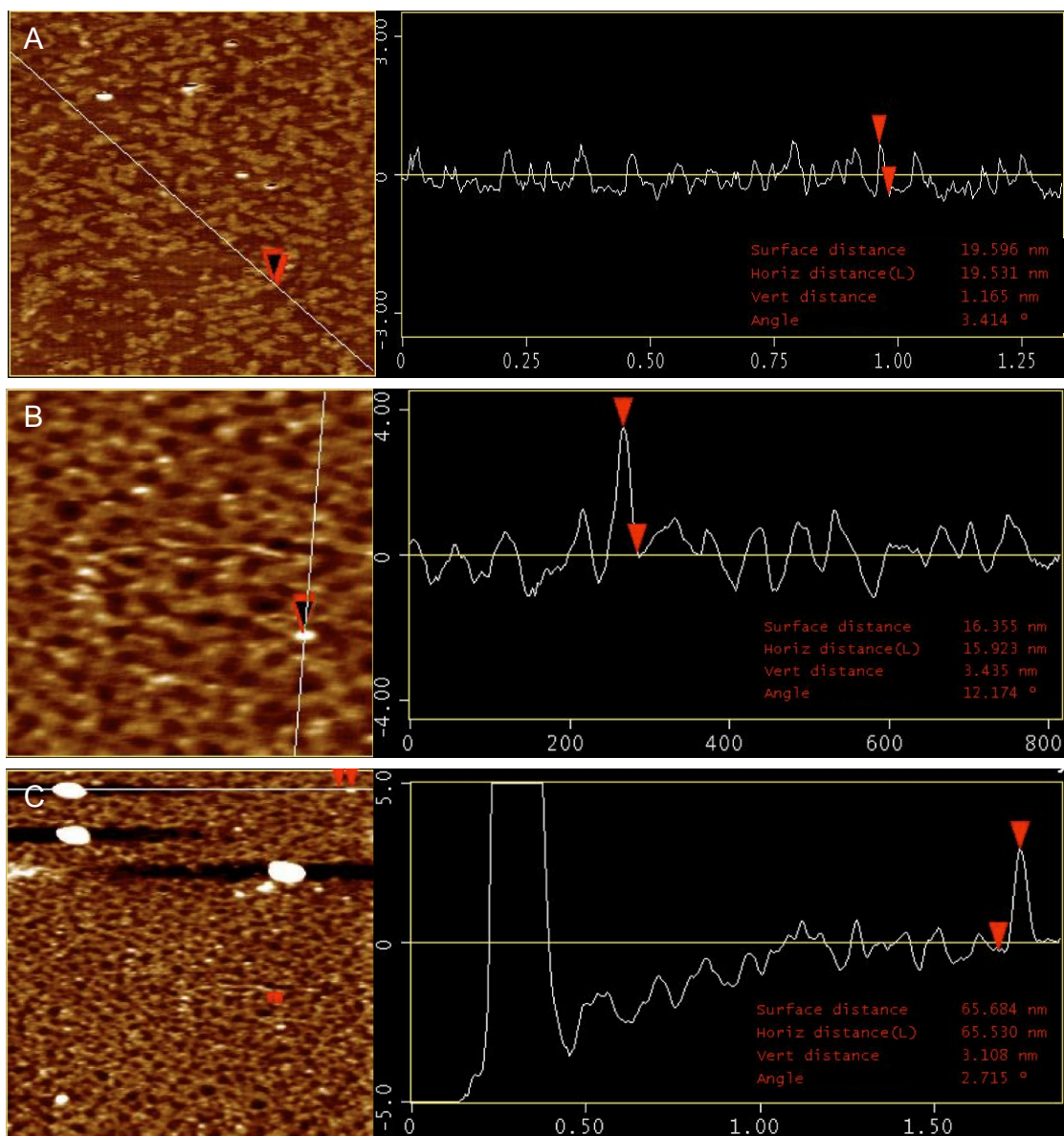
Figure 5-10. AFM images and height profiles of pyrene and TEM^{Tyr105AzPhe} on HOPG. **A**. Image and height profile of DBCO-pyrene laid down on HOPG prior to protein binding. **B** and **C**. Images and height profiles of TEM^{Tyr105AzPhe} bound to the HOPG pre-labelled with DBCO pyrene.
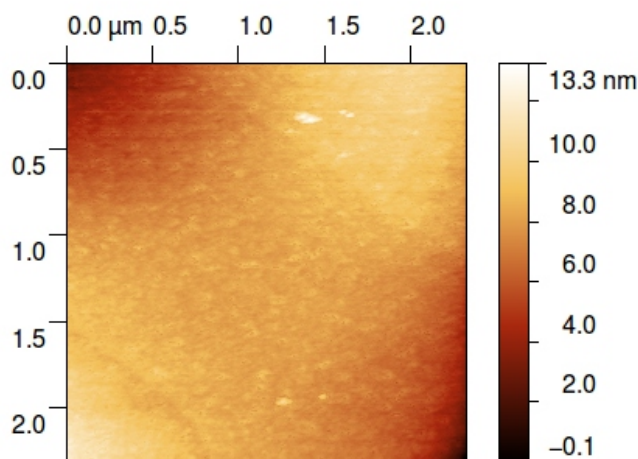
Figure 5-11. AFM image of wild type TEM β-lactamase incubated with HOPG.

5.2.3.3. *Attaching TEM β-lactamase to graphene*

Once TEM β-lactamase was successfully attached to HOPG, two variants, TEM^{Tyr105AzPhe} and TEM^{Trp165AzPhe}, were then used to bind to graphene. For binding to graphene, TEM β-lactamase already modified with the DBCO-pyrene was used to attach to the surface, so as not to cover the graphene surface with pyrene, desensitizing the surface. Drop-casting with 5 nM TEM^{Trp165AzPhe} pre-modified with DBCO-pyrene resulted in single proteins being imaged on the surface, with the expected height of 3-4 nm (Figure 5-12A-B). TEM^{Tyr105AzPhe} was also attached to the surface using the same method and the same protein concentration (Figure 5-12C). Modified protein shows excellent persistence on the surface, remaining attached to the graphene after multiple scans using AFM tapping mode. The bond between the pyrene and graphene is therefore very strong, and is not easily disrupted via mechanical forces. When wild type in incubated with graphene, proteins are knocked off the surface after a few scans (Figure 5-12D), showing the interaction between the pyrene and the graphene is stronger than other, non-specific interactions between the protein and the surface.
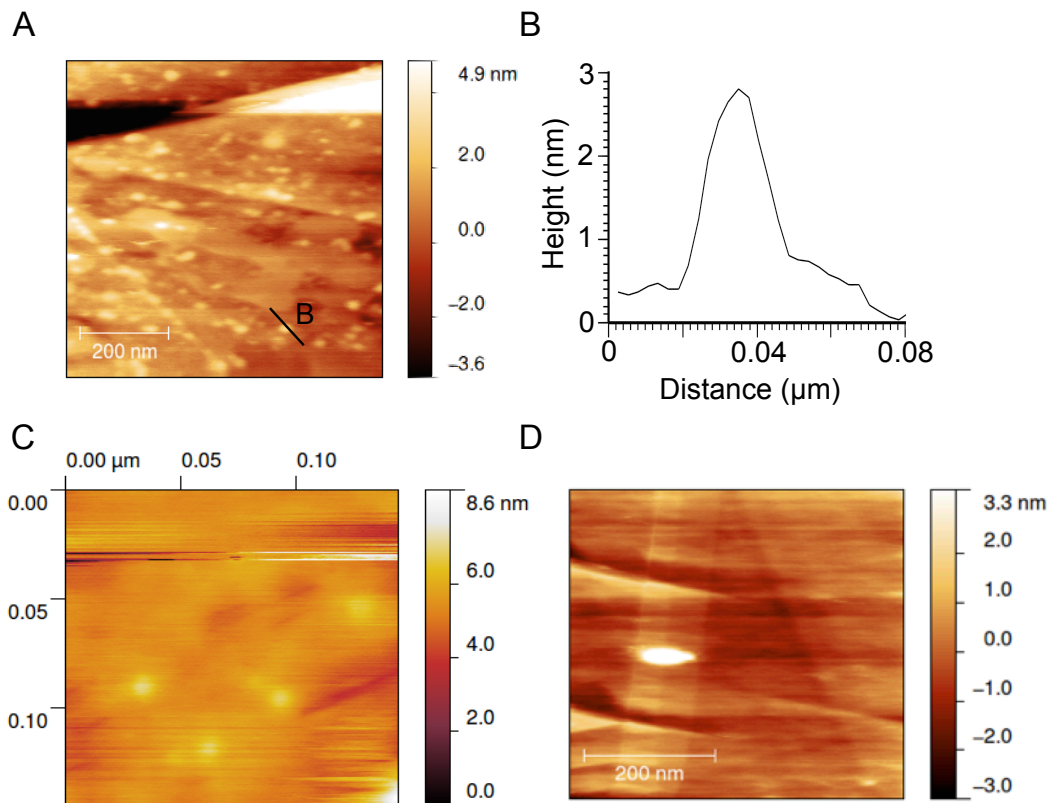
Figure 5-12. AFM images of TEM β-lactamase on graphene. **A-B**. AFM image (A) and height profile (B) of TEM$^{\text{Trp165AzPhe}}$ pre-modified with DBCO-pyrene on graphene. **C**. AFM image of TEM$^{\text{Tyr105AzPhe}}$ pre-modified with DBCO-pyrene on the surface of graphene. **D**. AFM image of wild type TEM β-lactamase incubated with graphene.

5.3. Discussion

5.3.1. Effect of residue position on SPAAC efficiency

Post-translational modifications using AzPhe via SPAAC have been shown to be able to modulate the function of enzymes, and also introduce new functionality to proteins [150, 151]. However, the efficiency of these reactions has been shown to vary substantially between different proteins, and even different variants of the same protein [110]. Proteins are highly complex and dynamic biomolecules that sample a wide range of microenvironments across their structure due to the different residues (and chemistry) that make up a protein at each point of their structure. It is very difficult to predict the influence of these unique microenvironments on the SPAAC reaction, mainly because very little work has been done in this area. By modifying several AzPhe-containing variants of a protein and analyzing the rate, and final modification efficiency, the effect of these microenvironments can be evaluated.

TEM$^{\text{Tyr105AzPhe}}$ was modified at the quickest rate and to the highest extent of all the TEM β-lactamase variants sampled. Tyr105 is a surface exposed residue (113 Å^2) that is located in the cleft formed between the two domains of TEM β-lactamase and forms a key part of the substrate binding pocket. The two closest residues to the side chain of Tyr105 are Met129 and Ser130, one hydrophobic, and one hydrophilic residue (Figure 5-13A). Other residues within 4 Å of Tyr105 include Glu104, Asp131 and Asn132. Residues Ser130, Asp131 and Asn132 are more commonly known as the "SDN" loop and are known to have a function in interacting with substrates [80]. The position of Tyr105, so close to a feature of a protein that is designed to accept substrates, is likely to be more accommodating to chemical adjuncts that are trying to react with the azide group of AzPhe, in a manner that is not entirely dissimilar to substrates trying to gain access to the active site. This could contribute to the high final modification efficiency seen, and the high rate of reaction.

Ala237 is another residue that is found in the active site of TEM β-lactamase. The final modification efficiency when AzPhe was incorporated at Ala237 was just over 50%, and the rate of reaction was slower than that of TEM$^{\text{Tyr105AzPhe}}$. Ala237 is relatively buried (26 Å^2), however the size difference between alanine and AzPhe means that on mutation the side chains are likely to sample different microenvironments, so the azide group of AzPhe is likely to be in the cleft between

the two domains, increasing the effective surface exposure of the residue. Ala237 is also flanked by two small glycine residues (Figure 5-13B) that could provide extra space for the DBCO-adduct to approach the side chain of AzPhe, potentially increasing the modification efficiency, despite the limited surface exposure.

TEM$^{\text{Trp165AzPhe}}$ is also modified to a good level, reaching a maximum efficiency of 60%. Trp165 is a large amino acid that is surface exposed (74 Å^2) with one face of the indole side chain facing the solvent and one facing the protein. It is present on the Ω-loop, a loop structure that sits over the active site, and is adjacent to a catalytic residue (Glu166). Despite the proximity of Trp165 to the active site, particularly Glu166, modification at this residue did not greatly affect the activity of the enzyme. The closest residues to the side chain of Trp165, Asn136 and Thr140, both have polar groups on their side chains, which could disrupt the interaction between AzPhe and the hydrophobic DBCO group (Figure 5-13C), however the surface exposed nature of the residue is likely to promote accessibility aiding modification.
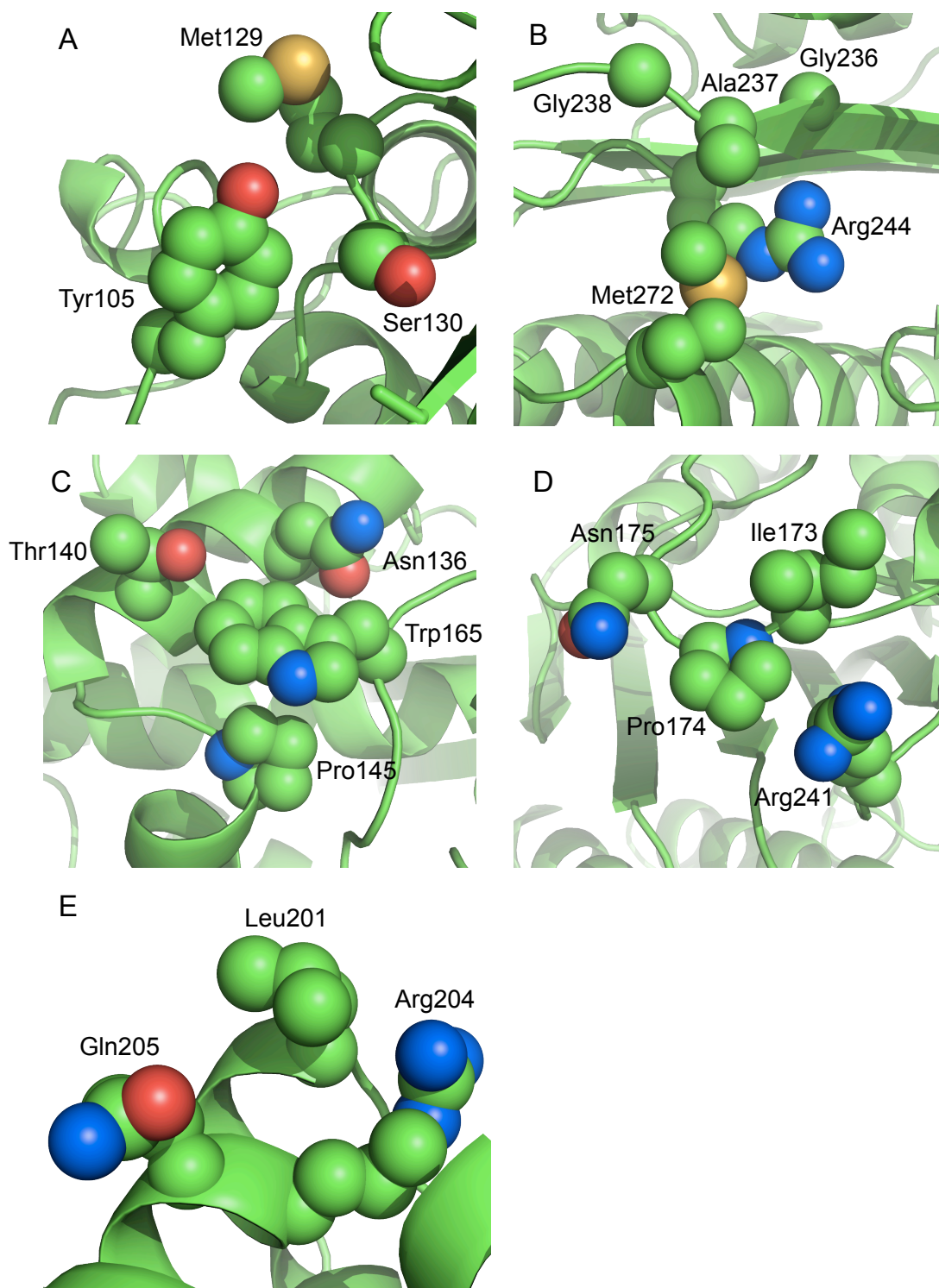
Figure 5-13. The surrounding residues and microenvironments of (A) Tyr105, (B) Ala237, (C) Trp165, (D) Pro174 and (E) Leu201.

TEM^Leu201AzPhe is found far from the active site, and as expected, modification here has been shown to have very little effect on activity. This variant was modified to a similar extent as TEM^Trp165AzPhe and TEM^Ala237AzPhe, reaching a highest final modification efficiency of 47% and a lowest of 35%. Leu201 is surface exposed (107 Å^2) and found at the end of a α-helix so the DBCO would have no problem gaining access to the side chain, however it is surrounded by charged residues (Figure 5-13E), which may reduce the ability of the hydrophobic DBCO group to approach the azide.

The position that was modified the least of all the variants tested was TEM^Pro174AzPhe, another variant that is present on the Ω-loop. Modification of this variant only reached a final efficiency of 33%. This residue is located on the opposite side of the Ω-loop to Trp165 and the active site. It is partially surface exposed (54 Å^2) with one face of the cyclic ring facing the solvent, and surrounded by a mixture of hydrophobic and hydrophilic residues (Figure 5-13D). On the surface, it samples a similar microenvironment to both Tyr105 and Trp165, however it is modified to a much lesser extent. In Chapter 4 I hypothesized that the mutation of Pro174 to AzPhe removed a regulator of the flexibility of the Ω-loop. By increasing the flexibility, the activity of the protein was negatively affected. Here, by increasing the flexibility of the loop, the modification efficiency has been negatively affected. The modification of Trp165 reaches a higher final efficiency, as there is still a proline at position 174 (and one at position 167), modulating the flexibility of the loop as a whole, holding Trp165 in place so that it can be modified. The increase in flexibility caused by removal of the proline at position 174 could result in the lower modification efficiency.

By ranking the variants in the order of most modified to least modified (TEM^Tyr105AzPhe > TEM^Trp165AzPhe > TEM^Ala237AzPhe > TEM^Leu201AzPhe > TEM^Pro174AzPhe) we can try and determine some of the factors that may influence modification by looking at the characteristics that these variants have in common. The flexibility sampled by TEM^Pro174AzPhe could explain why this variant is modified so little, however Tyr105 is also located in the middle of a large loop region of the protein with no defined secondary structure. However this position is likely to be held in place much more effectively due to its proximity to the active site, whereas Pro174 is situated in an inherently flexible domain of the protein. Two of the residues modified

with the highest final efficiency, Tyr105 and Trp165, are surface exposed, whereas Ala237 is similarly modified but less surface exposed (26 Å^2 compared to 113 Å^2 and 74 Å^2 for Tyr105 and Trp165 respectively). The change in size between AzPhe and alanine would most likely increase the surface exposure from 26 Å^2, but would likely not be comparable to Tyr105. Leu201 is the second most exposed residue sampled, but this residue is modified to a lesser extent. There appears to be no pattern in the surrounding residues of each variant; most variants have a variety of residues around them with a combination of polar and non-polar groups. These findings underline the difficulty in predicting which positions may be modified to a great extent and which would be modified poorly as there is no single factor that correlates with the findings.

5.3.2. Using SPAAC to attach proteins to surfaces

The immobilization of proteins on surfaces is an ideal method to study proteins as it unlocks the potential of measuring single molecules. By studying many millions of proteins in solution, individual effects on substrate binding and catalysis can be averaged out and lost. By focusing on the smallest possible number of individual proteins, a more detailed analysis of temporal processes can be investigated. Protein immobilization has been previously achieved via a variety of methods, many of which rely on the intrinsic properties of a protein, such as hydrophobic areas on its surface to adsorb to a surface, or a lysine residue on its surface. Whilst these methods can attach a protein to a surface, it is also wholly unspecific, as proteins often have numerous hydrophobic regions or inherent functional groups on their surface. In previous cases, all other surface exposed cysteine residues (a relatively uncommon amino acid) have been removed from the target protein [149]. However this is an extensive undertaking, and can introduce significant changes to the protein's structure and function as cysteine residues often play important structural and functional roles.

By using specific unnatural amino acids with unique reactive groups, such as AzPhe, the need for such changes to the target protein is removed. Using this method, only one mutation was introduced into TEM β-lactamase, which can be designed to have little or no effect on activity, as is the case for TEM$^{\text{Trp165AzPhe}}$. The use of novel post-translational modifications has been previously used to tether proteins [65] and cells [152] to a surface, however these methods failed to benefit from the novel reactivity of AzPhe and the functional significance of graphene. The combination of AzPhe and

SPAAC offers many benefits over existing methods of protein immobilization. By using different adducts that react with AzPhe in the same manner, and by using different surfaces with different properties, different strategies for immobilizing proteins on a surface can be exploited using a single mutation.

The modification of AzPhe at two defined positions of TEM β-lactamase with a functionalized pyrene group (which allows proteins to interact with carbon allotropes such as graphene and carbon nanotubes (CNTs)) resulted in different effects on the activity of the enzyme. Modification at TEM$^{\text{Tyr105AzPhe}}$ with DBCO-pyrene resulted in a slight decrease in activity whereas modification at TEM$^{\text{Trp165AzPhe}}$ had no effect. These results are the same as those seen with other DBCO-based adducts (Section 4.2.4.3.). TEM$^{\text{Trp165AzPhe}}$ in particular represents an ideal opportunity to interface TEM β-lactamase with graphene or CNTs, making use of the conductive properties of these materials. The limited effect on activity means the catalytic process is not interfered with, and the proximity of the mutation to the active site means any charge changes during catalysis can alter the electronic signature of graphene or CNTs. AFM imaging of tightly bound proteins on the surface showed single molecules of expected dimensions. The successful interfacing of TEM β-lactamase with both graphite and graphene in a defined manner using pyrene represents the first step towards this technology, allowing precise and repeatable measurements of atomic level details in biological processes such as enzyme-substrate binding and protein-protein interactions. The ability to analyze single molecules in such a manner would increase our understanding of such events greatly.

The modification of AzPhe-containing variants of TEM β-lactamase with a DNA oligonucleotide containing a cyclo-octyne opens up the possibility of attaching an enzyme to a DNA-based surface with high levels of specificity and organization. DNA origami uses the complementarity of DNA bases to organize proteins on a surface. By attaching a DNA oligonucleotide to a protein in a defined position, the protein can be immobilized on the surface in a highly specific and defined manner, and the position of the tethered protein can be controlled with a high degree of precision. As a DNA origami "tile" is a large piece of single stranded DNA (~7 kb), DNA strands can be designed that anneal to different parts of the surface, allowing multiple proteins to be tethered to the surface at the same time in a designed and

organized manner [153]. The modification of these AzPhe-containing variants with two different DNA oligonucleotides with limited effects on activity in some cases, and the successful separation of modified protein from unmodified, opens the door for the assembly of enzyme pathways or signaling cascades on a DNA origami tile. If two different enzymes in a specific reaction pathway, or two different kinases in a signaling cascade are modified with different DNA oligonucleotides, they can be designed to reside in series on a surface, thus recreating the pathway or cascade *in vitro*, so that the product of one reaction is fed directly into the active site of the next protein in the enzyme pathway, or a chemical signal is propagated along a series of receptor proteins.

5.4. Summary

This chapter introduced the addition of novel functionality to proteins via the use of unnatural amino acids and Click chemistry. By incorporating AzPhe into TEM β-lactamase and modifying the azide with two different chemical adjuncts, two different methods of attaching a protein to a surface were explored. The modification of TEM β-lactamase with pyrene enabled the protein to be deposited in a highly defined and specific manner onto a surface that can be used to detect the changes in a protein caused by substrate binding or protein-inhibitor interactions. By exchanging the pyrene group in the chemical modification with a DNA strand, the AzPhe-containing protein can then be attached to a DNA origami "tile" that allows precise organization of multiple proteins on a single surface. The effect of modification of TEM β-lactamase with two different DNA strands was determined, and then the modified protein was purified from the unmodified protein.

Post-translational modifications are an important facet to protein function, and they rely on an ability to affect protein activity or function immediately. Although SPAAC is experimentally compatible with biological systems, if the reactions do not occur rapidly enough, their biological relevance will be negligible. In some cases, this was clearly evident and outlines why it is important to design potential variants. The effect of the site of AzPhe incorporation on SPAAC efficiency was explored, and the factors that may determine the success of a reaction were discussed. This chapter highlights that SPAAC is an adaptable technique that can be used to not only affect the function of a protein, but can introduce novel methods of protein analysis.

6. The impact of AzPhe incorporation and modification on sfGFP structure and function

6.1. Introduction

The incorporation of unnatural amino acids (uAAs), especially *p*-azido-L-phenylalanine (AzPhe), has been shown to be a useful tool for introducing new functional properties into proteins [154]. In Chapter 4, the use of AzPhe as a tool for modulating enzyme activity using novel post-translational control was demonstrated. In this chapter, the generality of the approach is explored by transferring the idea to a second protein, superfolder GFP (sfGFP). sfGFP is an intrinsically fluorescent protein (see Section 1.7) that has been shown to tolerate the incorporation of uAAs [87]. AzPhe in particular has proved to be particularly useful, introducing photo-activation, photo-deactivation and fluorescence shifts through covalent modification of AzPhe [23]. The structure and fluorescence properties of sfGFP have been characterized with the roles of many residues, both within the chromophore itself and those residues surrounding it contributing towards fluorescence, investigated. This chapter focuses on the incorporation of AzPhe into sfGFP at His148 (Figure 6-1), a position that is involved in the complex hydrogen-bonding network that interacts with the chromophore and defines the fluorescence properties of the protein [88]. It details the effect of AzPhe incorporation, and the effect of subsequent UV irradiation on the fluorescence characteristics of sfGFP. It goes on to assess the effect of post-translational modifications via Click chemistry on sfGFP fluorescence. To provide a detailed molecular insight into how AzPhe incorporation at His148 affects fluorescence and how the subsequent photo-control is exerted, the protein structure in the dark and irradiated states was solved by X-ray crystallography. The original sfGFP structure [92], together with recent uAA-containing variants [23], have provided an excellent insight into the effects of UV irradiation on AzPhe in sfGFP, and the precise mechanisms by which novel photo-control can influence the behaviour of sfGFP fluorescence.
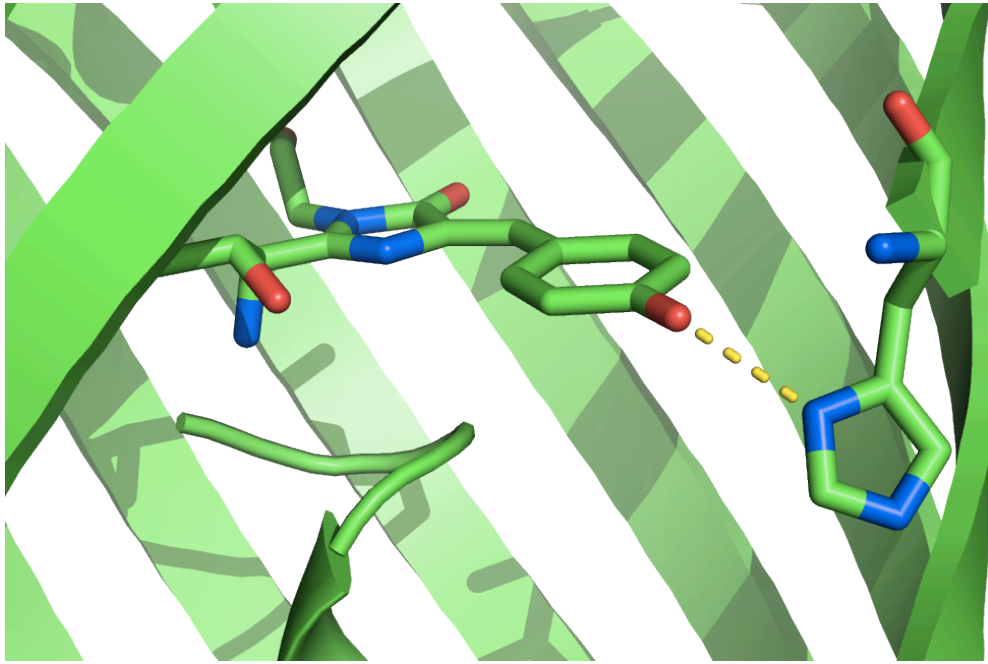
Figure 6-1. Interaction between Tyr66 of the chromophore and His148

6.2. Results

6.2.1. Production and purification of AzPhe-containing sfGFP

A detailed description of the approach used to produce proteins containing uAAs can be found in Sections 1.5 and 4.2.1., and the methods are outlined in Section 2.4.1.. Briefly, *E. coli* Top10 cells were transformed with two plasmids, pBAD containing the sfGFP gene with a TAG codon inserted at His148, and pDULE containing the translational machinery required for uAA incorporation. sfGFP[His148AzPhe] was then purified in two steps. The first step was Nickel-affinity chromatography using a hexa-histidine tag (His-tag) at the C-terminus of the protein. During purification, protein elution was monitored at 280 nm, to monitor all proteins in the cell lysate, and at 485 nm, for the specific monitoring of sfGFP elution. Upon application of the imidazole gradient, monitoring at 280 nm was stopped to avoid any potential photolysis of the phenyl azide group in AzPhe (Figure 6-2A). Secondly, the protein was purified using anion exchange chromatography, with elution only monitored at 485 nm (Figure 6-2B). After purification, SDS-PAGE was used to determine the purity of sfGFP[His148AzPhe]. The protein was then concentrated into PBS and buffer exchanged using a 3000 MWCO centrifugal concentrator.
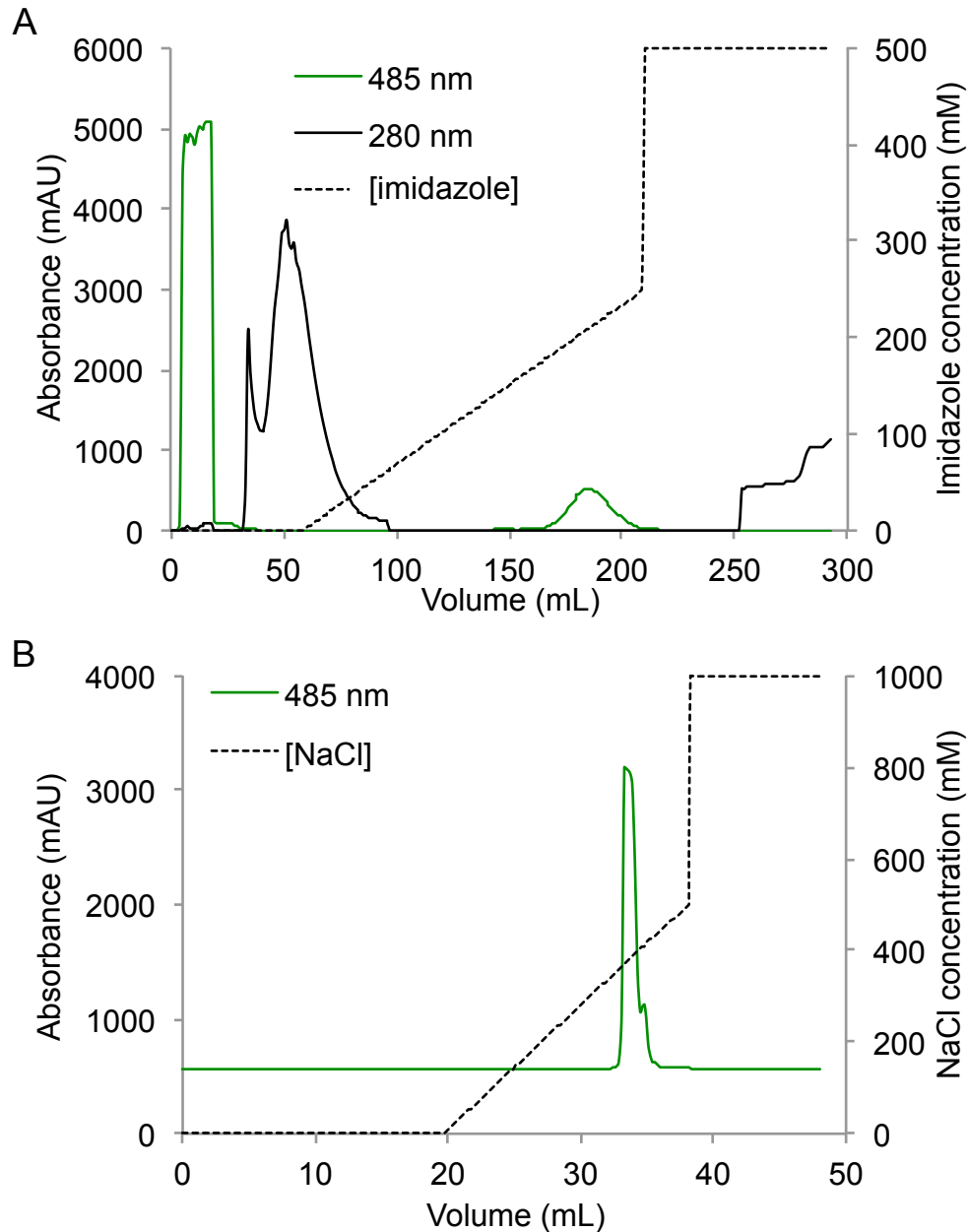
Figure 6-2. Purification of sfGFP^{His148AzPhe}. **A**. Nickel-affinity chromatography purification elution profile. Absorbance at 280 nm monitoring protein elution (black line) and absorbance at 485 nm monitoring sfGFP elution (green line) on application of an imidazole gradient (dashed line). **B**. Anion exchange chromatography elution profile. Absorbance at 485 nm monitoring sfGFP elution (green line) on application of a NaCl gradient (dashed line).

6.2.2. Spectral analysis of sfGFP^{His148AzPhe}

The chromophore of GFP can exist in two states that normally correlate with two different excitation peaks, 395 or 475 nm, corresponding to the neutral phenol or the deprotonated phenolate form of the chromophore respectively [155]. Only one emission peak exists at 511 nm, corresponding to the phenolate form. The generation of sfGFP through directed evolution results in the peak at 485 nm dominating (ε = 49000 M^{-1} cm^{-1} [23]), and the peak at 395 nm only representing a small shoulder in the excitation spectrum (Figure 6-3A). When AzPhe is incorporated at His148, a position that hydrogen bonds to Tyr66 in the wild type structure, the two excitation peaks switch intensities, with the phenol peak at 400 nm (red shifting 5 nm from 395 nm) becoming the predominant peak (ε = 35600 M^{-1} cm^{-1}), and the 485 nm peak virtually disappearing (Table 6-1; Figure 6-3A). There is a small peak at 500 nm but is much less intense than either the 400 nm peak in sfGFP^{His148AzPhe}, or the 485 nm peak in wild type sfGFP. There is no change in the emission spectra of sfGFP^{His148AzPhe}.

Table 6-1. Effect of UV irradiation on the spectral properties of sfGFP^{His148AzPhe}

Variant	State	λ_{max} (nm)	λ_{em} (nm)	ε (M^{-1} cm^{-1})	Quantum yield	Brightness	
						(M^{-1} cm^{-1})	% sfGFP
His148^{AzPhe}	Dark	400	511	35600	0.64	22800	62
	Dark	500	511	20300	0.46	9400	25
	UV	400	511	12500	0.37	4600	13
	UV	490	511	40400	0.48	19500	53
Wild type [1]	-		485	511	49000	0.75	36800	100

[1] Values taken from [23]
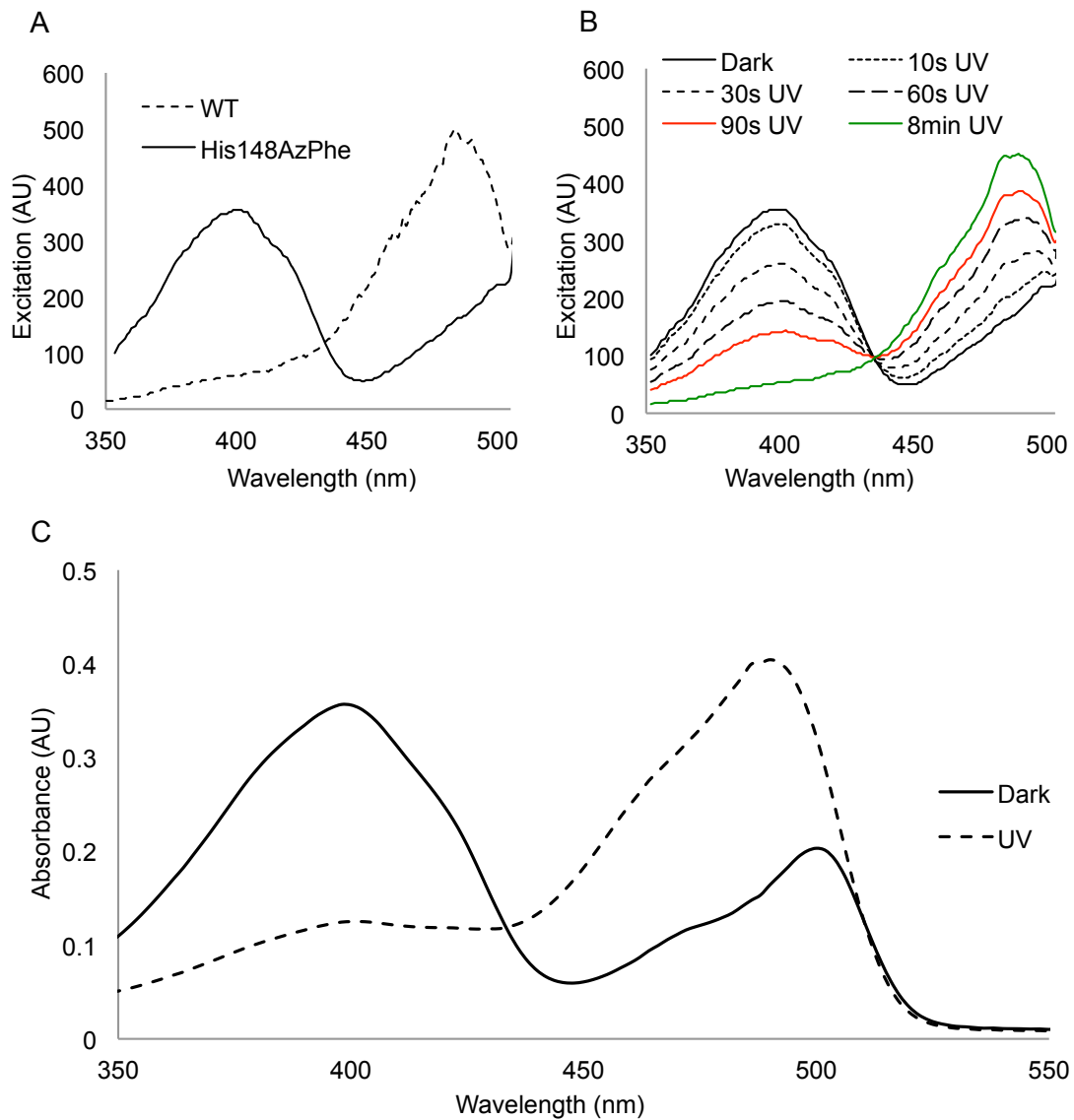
Figure 6-3. Effect of UV irradiation of the fluorescence spectra of sfGFP^{His148AzPhe}. **A**. Comparison of the excitation spectra of wild type sfGFP and sfGFP^{His148AzPhe} prior to irradiation (emission was monitored at 511 nm). **B**. Effect of UV irradiation on the excitation spectrum of sfGFP^{His148AzPhe}. The protein was irradiated for 8 minutes, and fluorescence was measured at multiple time points (emission was monitored at 511 nm). **C**. Absorption spectra of sfGFP^{His148AzPhe} before (dark) and after (UV) UV irradiation.

Upon UV irradiation of sfGFP^{His148AzPhe}, the intensity of the two peaks switch, resembling the fluorescence of wild type sfGFP. The peak at 400 nm becomes less intense and a peak appears at 485 nm that grows more intense after more UV exposure (Table 6-1; Figure 6-3B). After 10 seconds of UV irradiation, the peak at 400 nm decreases in intensity, and fluorescence at ~490 nm increases, but no clear peak emerges. As exposure time increases, a clear peak around 485 nm starts to emerge. After 8 minutes of UV irradiation, the peak at 400 nm is barely evident, whereas the peak at 485 nm has reached its highest intensity ($\varepsilon = 40400$ M^{-1} cm^{-1}). Beyond 8 minutes of irradiation, no further increase in excitation at 485 nm or decrease at 400 nm is observed. UV irradiation also causes no change in the emission wavelength, which remains at 511 nm, but does have an effect on the intensity. When wild type sfGFP is irradiated with UV light, there is no change in the fluorescence spectrum after 8 minutes of exposure.

The azide group (N$_3$) of AzPhe can also be used to attach novel chemical entities to the protein via strain-promoted azide-alkyne cyclo-addition (SPAAC). The azide group can undergo highly specific and selective cyclo-addition reactions with alkyne groups, resulting in a stable triazole link [156]. Two different modifications; **1**, a large, fluorescent rhodamine dye based on the commonly used probe Texas Red (λ_{em} = 600 nm, $\varepsilon = 97000$ M^{-1} cm^{-1}), and **2**, an amine attached to a short linker, were used to determine the effect of novel post-translational modifications on the fluorescence of sfGFP (Table 6-2; Figure 6-4). When sfGFP^{His148AzPhe} is modified with either of the two different adducts, two different effects occur depending on which modification is used. Modification with **1** results in the main excitation peak remaining at 400 nm, indicating the modification is having little effect on the chromophore environment as the phenol form of the chromophore predominates (Figure 6-4A). Click modification with **1** does provide sfGFP with a compatible FRET (Forster resonance energy transfer) partner. His148 is located near the chromophore of sfGFP so modification at this site with a fluorescent dye whose spectral characteristics overlap with that of sfGFP (**1** has an excitation maxima at 584 nm) could result in FRET. Upon excitation of sfGFP^{His148AzPhe} at both 400 nm and 485 nm, there is detectable emission at 610 nm, which corresponds with the emission characteristics of the dye used (Table 6-2; Figure 6-4B). When sfGFP^{His148AzPhe} is excited at 400 nm, the emission at 610 nm is much less intense than the emission when the protein is excited at 485 nm.

Modification with **2** has a more direct and significant effect on sfGFP function (Table 6-2; Figure 6-4A). Modification of sfGFP[His148AzPhe] with **2** results in a switch in the main excitation from 400 nm to 485 nm. The excitation spectrum resembles that of the irradiated sfGFP[His148AzPhe], and of wild type sfGFP.

Table 6-2. Effect of SPAAC on the spectral properties of sfGFP[His148AzPhe]

Variant	λ_{max} (nm)	λ_{em} (nm)	ε (M^{-1} cm^{-1})	Quantum yield	Brightness	
					(M^{-1} cm^{-1})	% sfGFP
His148[AzPhe]	400	511	35600	0.64	22800	62
	500	511	20300	0.46	9400	25
His148[AzPhe][1 a]	400	511	19400	0.50	9600	26
	490	511	29300	0.26	7700	21
His148[AzPhe][2 b]	400	511	24600	0.23	5600	15
	490	511	37400	0.49	18300	50
Wild type [1]	485	511	49000	0.75	36800	100

[1] Values taken from [23]

[a] sfGFP[His148AzPhe] modified with **1** (Figure 6-2)

[b] sfGFP[His148AzPhe] modified with **2** (Figure 6-2)
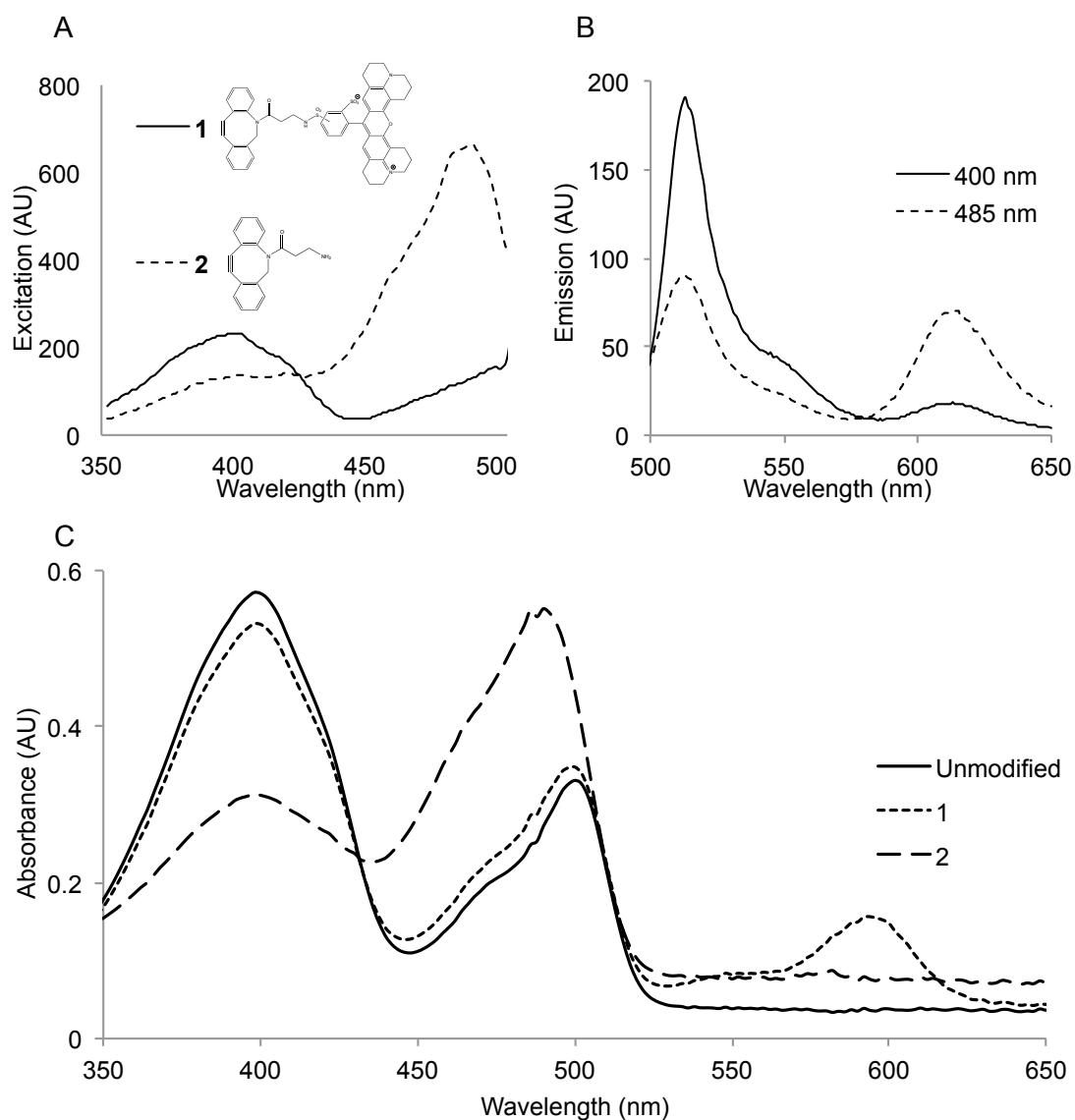
Figure 6-4. Effect of SPAAC on the fluorescence spectra of sfGFP$^{\text{His148AzPhe}}$. **A**. Comparison of the excitation spectra of unmodified sfGFP$^{\text{His148AzPhe}}$ and sfGFP$^{\text{His148AzPhe}}$ modified with **1** and **2**. **B**. Emission spectra of sfGFP$^{\text{His148AzPhe}}$ modified with **1**, excited at both 400 nm and 485 nm. **C**. Absorption spectra of unmodified sfGFP$^{\text{His148AzPhe}}$ and sfGFP$^{\text{His148AzPhe}}$ modified with **1** and **2**.

6.2.3. X-ray crystallography of sfGFP^{His148AzPhe}

6.2.3.1. Protein crystallization

Key to understanding how photolysis of AzPhe in sfGFP exerts its effects on fluorescence, it is important to define the molecular events that occur upon conversion from the dark to the irradiated state. Furthermore, understanding these changes will allow us to use the photochemistry of AzPhe to control the activity of proteins more generally. Therefore, the crystal structures of the dark and irradiated states of sfGFP^{His148AzPhe} were determined.

Protein was produced and purified as previously stated and concentrated to 10 mg/mL. Other AzPhe-containing variants of sfGFP had been previously crystallized [23], so the successful screen conditions for those studies were used as a starting point. Initial screens varied pH (7-9) and ammonium sulphate concentration (50-90% saturation – 2-3.7 M) in 50 mM Tris, but yielded no crystals. The screen was then extended to cover a wider range of pH (4-9), with MMT (Malic acid, MES, Tris) buffer replacing Tris. This screen yielded two crystals at a single condition; pH 8.5, 2.52 M $(NH_4)_2SO_4$ (Figure 6-4A). 1 mM ethylene glycol was then added as a cryo-protectant prior to crystal harvesting. One of the two crystals was harvested in the dark so that the azide group remained intact, while the second crystal was irradiated for 1 hour *in crystallo*.
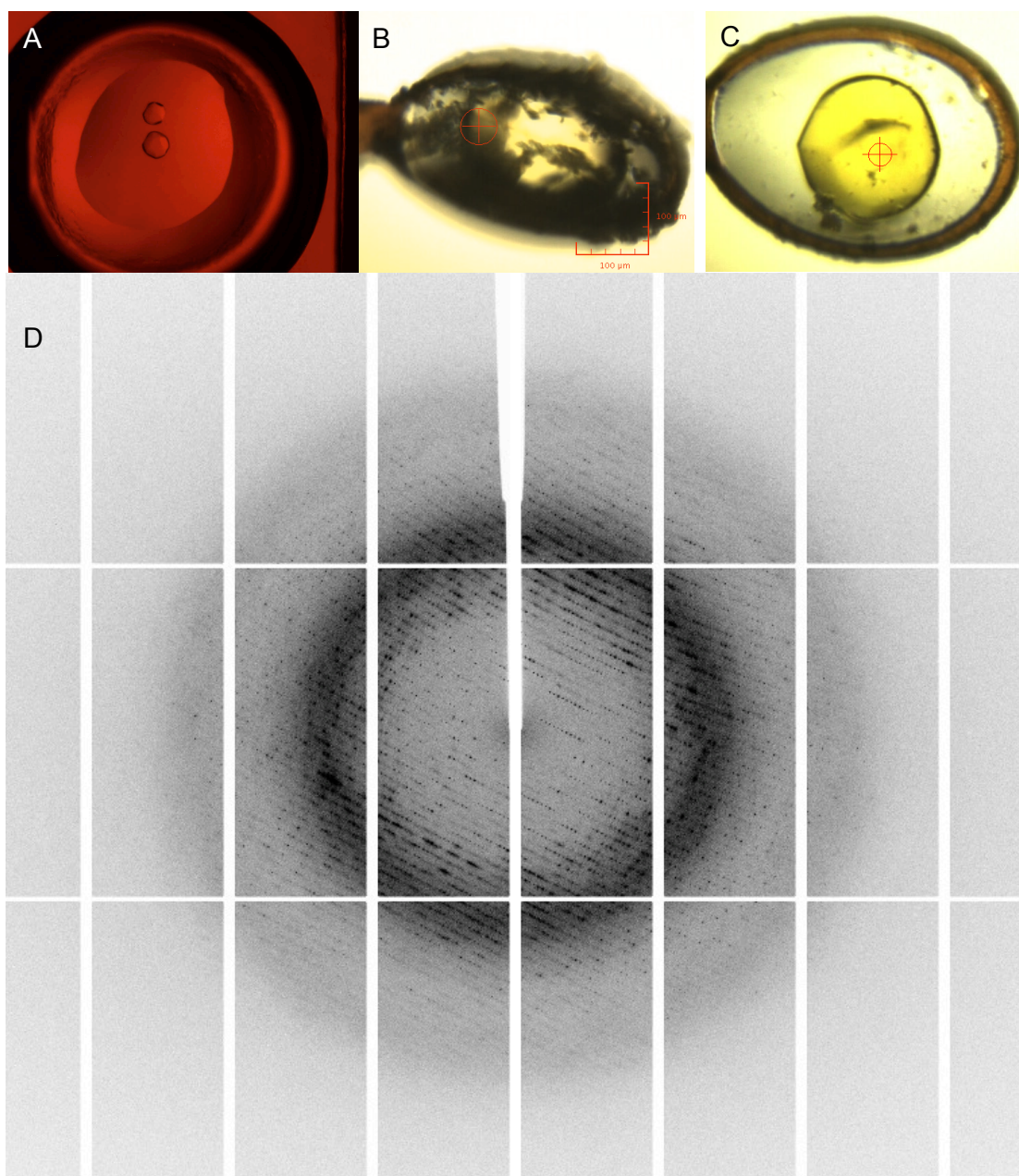
Figure 6-5. Protein crystals and X-ray diffraction pattern. **A**. The two crystals in the tray, visualised under red light. Images of (B) dark sfGFP^{His148AzPhe} and (C) UV irradiated sfGFP^{His148AzPhe} on the beamline (red crosshair). **D**. Diffraction pattern generated from X-ray irradiation of dark sfGFP^{His148AzPhe} (2.0 Å).

6.2.3.2. Structure determination and refinement

Diffraction data (Figure 6-5D; Table 6-3) were collected at the Diamond Light Source (Harwell, UK) by Dr. Pierre Rizkallah (Cardiff Medical School, Cardiff University). Data were reduced using the XIA2 package [102], assigned a space group using POINTLESS [103], scaled using SCALA [103] and merged using TRUNCATE [104]. Structures were solved by molecular replacement with PHASER, using a previously determined sfGFP structure (PDB code 2B3P – [92]). Structures were then adjusted manually using COOT and refined by TLS restrained refinement using RefMac. Initially, the structures were assigned the tetragonal space group P 4_1 2_1 2, however this was reassigned to P4$_3$ 2_1 2 during refinement for both crystals. Both the dark and the irradiated crystals contained two molecules in the asymmetric unit, and were determined to a resolution of 2.0 Å (dark state) and 2.1 Å (irradiated state).

Table 6-3. Crystal structure refinement statistics of dark state and irradiated state sfGFP^{His148AzPhe}

	DARK	LIGHT
Data collection		
Wavelength	0.92	0.92
Beamline	Diamond I04-1	Diamond I04-1
Space group	P4(3)2(1)2	P4(3)2(1)2
a (Å)	135.6	135.14
b (Å)	135.6	135.14
c (Å)	69.23	69.56
Resolution range	42.88 - 2.03	56.24 - 2.14
Total reflections measured (outer shell)	600948 (42383)	525718 (37806)
Unique reflections (outer shell)	42237 (3082)	36124 (2621)
Completeness (%) (outer shell)	100 (100)	100 (100)
I/σ (outer shell)	16.7 (3.9)	22.9 (4.0)
R (merge)[1] (%) (outer shell)	10.3 (72.6)	7.9 (68.8)
B(iso) from Wilson (Å^2)	39.4	42.2
Refinement statistics		
Protein atoms excluding H	4079	4002
Solvent molecules	350	274
R-factor[2] (%)	17.2	17.1
R-free[3] (%)	20.4	20.7
Rmsd bond lengths (Å)	0.018	0.018
Rmsd angles (°)	2.029	2.059
Ramachandran plot statistics		
Core region (%)	97.3	97.03
Allowed region (%)	2.47	2.97
Additionally allowed region (%)	0	0
Disallowed region (%)	0.22	0

[1] $R_{merge} = \Sigma_h \Sigma_j (I_{hj} - <I_j>) / \Sigma_h I_{hj}$

[2] $R_{factor} = (\Sigma_h |F_{h,obs} - Fh,calc|) / (\Sigma_h Fh,obs)$

[3] R_{free} is calculated from a set of 5% randomly selected reflections that were not included in the refinement

6.2.3.3. Comparison of sfGFP^{His148AzPhe} dark state and wild type

Mutating His148 to AzPhe does not affect the ability of the protein to fluoresce, so no major changes at the whole protein level are expected. This was confirmed when comparing the general structure to that of sfGFP (Figure 6-6A). When the solved structure for the dark state of sfGFP^{His148AzPhe} is compared with wild type sfGFP (PDB code 2B3P), the two structures are very similar, with RMSD values of 0.432 Å, 0.418 Å and 0.432 Å for Cα, backbone atoms and all atoms respectively for one of the molecules in the asymmetric unit. The other molecule in the asymmetric unit gave essentially the same values (0.474 Å, 0.461 Å and 0.474 Å). These values represent small deviations away from the structure of the wild type protein, but no gross changes.

Figure 6-6. Structure of the dark form of sfGFP^{His148AzPhe} **A**. Backbone alignment of dark state sfGFP^{His148AzPhe} (cyan) with wild type sfGFP (PDB code 2B3P – green). **B-C**. B factor representation of (B) wild type sfGFP (PDB code 2B3P) and (C) the dark state of sfGFP^{His148AzPhe}. B-factors are shown in a relative scale based on thickness (thin = low, thick = high) and colour (blue = low, red = high). The chromophore and His148/His148^{AzPhe} is shown as sticks on both structures.

B-factors (otherwise known as the Debye-Waller factors or the temperature factors) are a good indicator of flexibility in a particular region of a protein or peptide. Comparison of per-residue B-factors in the dark state compared to that of the same residues in the wild type showed small but significant differences (Figure 6-6B-C). Five regions of the wild type sfGFP structure have high B-factors. The C-terminus has a high degree of flexibility, a common trait shared by many proteins. Four other regions also exhibit a high degree of flexibility, Thr49 to Lys52, Asp155-Asn159, Glu172-Gly174 and Lys209-Arg215. These regions of sfGFP comprise loop regions on the outside of the protein, primarily at the top or bottom of the β-barrel. When these regions are directly compared to the dark state of sfGFP[His148AzPhe], the four internal regions and the C-terminus have lower B-factors in the dark state of sfGFP[His148AzPhe] than the wild type, particularly Glu172-Gly174 and Lys209-Arg215. One region of sfGFP[His148AzPhe], Ile188-Val193, has increased B-factors compared to wild type sfGFP (Figure 6-6C).

Although RMSD values and B-factors give a good indication of the gross structural changes to the whole protein caused by the incorporation of AzPhe, it is subtle conformational changes caused by the incorporation of AzPhe that likely gives rise to the functional effects. His148 in sfGFP interacts with the hydroxyl group of Tyr66 in the chromophore via the Nδ of the imidazole ring [92]. Mutation to AzPhe removes this interaction, as there is no electron-accepting group at an equivalent position (Figure 6-7A). Another residue that interacts directly with Tyr66 is the hydroxyl group in the side chain of Thr203. The side chain of Thr203 in sfGFP[His148AzPhe] is rotated and shifted so that the hydroxyl group in the side chain is no longer in a position where it can interact with the hydroxyl group of Tyr66 (Figure 6-7B). In wild type sfGFP, the distance between the two hydroxyl groups of Tyr66 and Thr203 is 2.7 Å, easily within the range of a hydrogen bond (2.3 – 3.2 Å). After AzPhe is incorporated into the protein, the distance between these two chemical groups increases to 4.8 Å, 1.6 Å over the upper limit for a hydrogen bond.
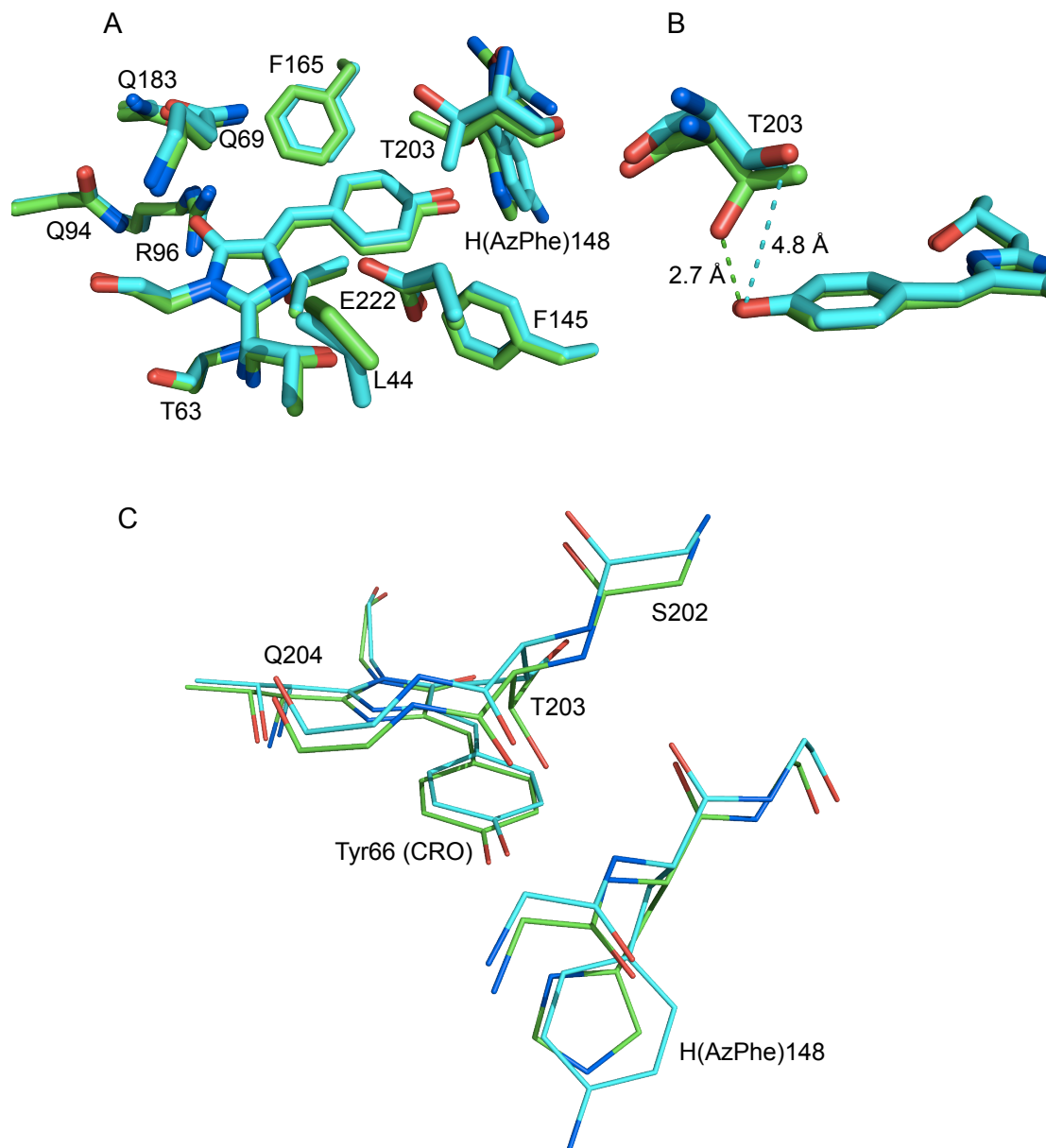
Figure 6-7. Effect of AzPhe incorporation on sfGFP structure. **A**. An alignment of the residues surrounding the chromophore in wild type sfGFP (green) and dark sfGFP$^{\text{His148AzPhe}}$ (cyan). **B**. The change in distance between the hydroxyl groups of Tyr66 and Thr203. **C**. The shift in the backbone position of β10 caused by AzPhe incorporation.

Thr203 resides in strand β10 and His148 is in β7, two β-sheets that lie close to each other in the folded protein. In sfGFP, Thr203 and His148 interact with each other via two backbone hydrogen bonds (the carbonyl of Thr203 and the amide of His148, and the carbonyl of His148 and the amide of Thr203). These interactions between the two β-sheets help stabilize the structure of the protein and maintain the precise positioning

of the side chain of Thr203. The incorporation of AzPhe at position 148 and the resulting increase in the size of the side chain causes a shift in the positioning of these two β-sheets, disrupting the precise orientation of Thr203 (Figure 6-7C). This slight shift upwards results in the side chain of Thr203 rotating, taking the hydroxyl out of reach of Tyr66 and breaking the hydrogen bond network. The removal of the hydrogen bonds between Tyr66 and His148, and between Tyr66 and Thr203 disrupts the hydrogen bond network that promotes the phenolate state (exciting at 485 nm). Thus, this provides a rationale for the predominant peak in the spectrum to 400 nm, which corresponds to the neutral phenol form of Tyr66.

6.2.3.4. Comparison of sfGFP^{His148AzPhe} dark state and the irradiated state

An alignment of the dark and irradiated states of sfGFP^{His148AzPhe} shows that there are no large changes to the structure of the protein after an hour of *in crystallo* UV irradiation. Direct alignment of the two molecules in the asymmetric unit of the dark and irradiated states gave RMSD values of 0.330 Å for all atoms, 0.326 Å for backbone atoms and 0.325 Å and Cα atoms only (Figure 6-8A). These values represent very little change between the two structures. When the irradiated state of sfGFP^{His148AzPhe} is compared with wild type, similarly small changes are seen overall. One molecule of the asymmetric unit gives RMSD values of 0.384 Å, 0.373 Å and 0.384 Å for Cα atoms, backbone atoms and all atoms respectively. The second molecule gives slightly higher values; 0.416 Å, 0.420 Å and 0.416 Å. When B-factors are compared, the areas of high flexibility in the irradiated state correlate with the dark state (Figure 6-8B-C). The same unstructured loop regions at the top and bottom of the β-barrel show some flexibility and the majority of the β-barrel is highly ordered. The N- and C-termini of both structures are similarly flexible, however two fewer residues were observed at the C-terminus in the irradiated state than in the dark state due to decreased electron density in this region. Of the four regions mentioned earlier, Asp155-Asn159, Glu172-Gly174, Ile188-Val193 and Lys209-Arg215, the irradiated state is only slightly more flexible. The Asn155-Asn159 loop occupies a slightly different orientation in the two crystal structures but the similarly high B-factors indicate this region is highly flexible and any single conformation in a crystal does not represent a final orientation.
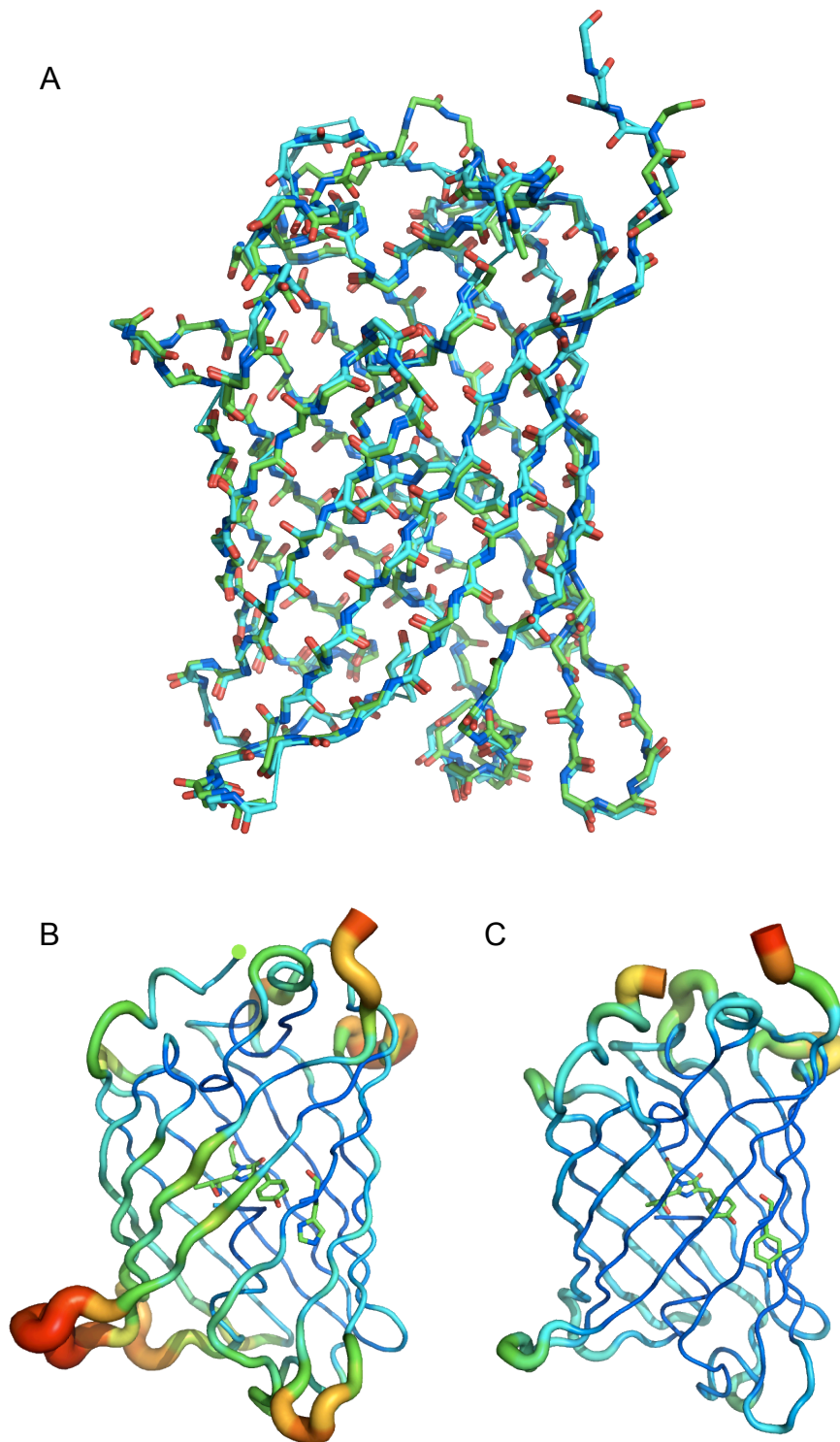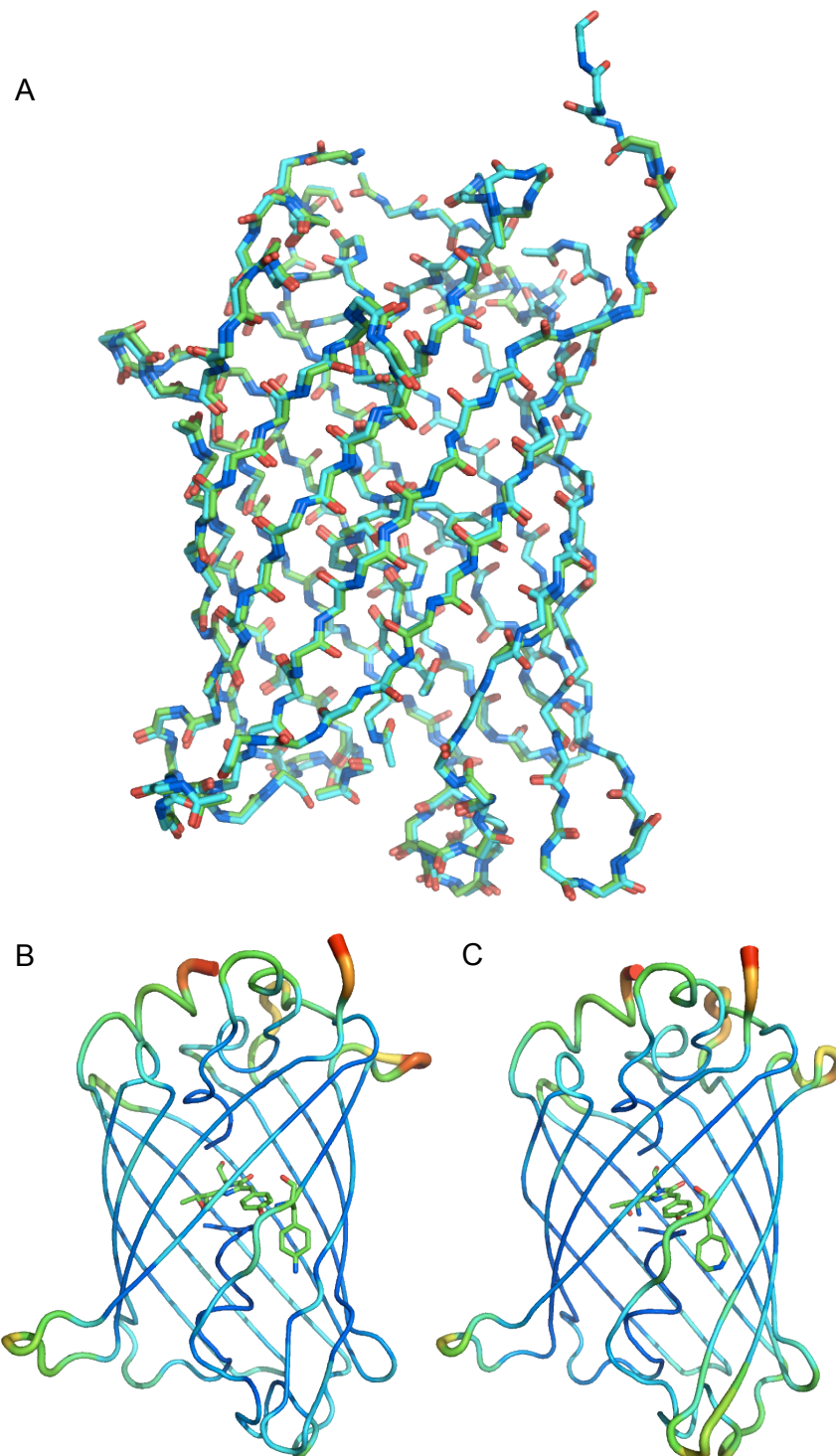
Figure 6-8. Structure of the irradiated form of sfGFP^{His148AzPhe} **A**. Backbone alignment of dark state sfGFP^{His148AzPhe} (cyan) with irradiated sfGFP^{His148AzPhe}. **B-C**. B factor representation of (B) the dark state of sfGFP^{His148AzPhe} and (C) the irradiated state of sfGFP^{His148AzPhe}. B-factors are shown in a relative scale based on thickness (thin = low, thick = high) and colour (blue = low, red = high). The chromophore and sfGFP^{His148AzPhe} is shown as sticks on both structures.

Irradiation of the azide group with UV light results in the loss of molecular nitrogen and the formation of a reactive nitrene group. This radical can then proceed down several reaction pathways and result in several different structures (see Section 1.5). When looking at the electron density around the side chain of AzPhe at position 148 in both the molecules in the asymmetric unit, there is a clear difference between the two (Figure 6-9). In one of the molecules (Mol-1) there is clear density around the *para* position of the phenyl ring, which is absent in Mol-2 (Figure 6-9). One potential end point of the reactions involving the nitrene radical is the formation of an amine group at the *para* position so this structure was modeled into Mol-1. However, it should be noted that even in the dark state, no electron density is seen beyond the Nα of the azide group, which may indicate the azide group has no fixed orientation. In Mol-2, there is no clear density at the *para* position, however the electron density surrounding the phenyl ring itself is larger (Figure 6-9B). A second reaction pathway is the formation of a 7-membered ring (a dihydroazepine ring) so this was modeled into Mol-2. The difference in the density that surrounds the side chain at position 148, and the subsequent difference in the modeling had a slight effect on the surrounding residues as well. The B-factors of Asn146 and Ser147 are slightly higher when the 7-membered ring is modeled in Mol-2, compared to an NH_2 group at the *para* position in Mol-1.
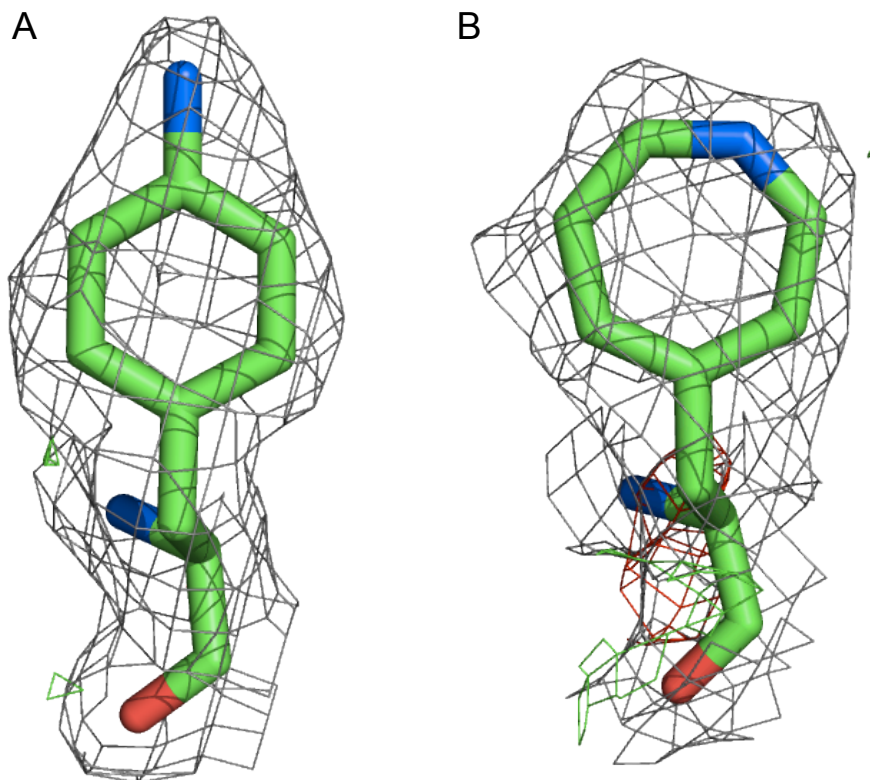
Figure 6-9. Difference in the electron density surrounding the side chain of AzPhe in the two molecules of the UV irradiated sfGFP$^{\text{His148AzPhe}}$ structure. **A**. Amine group at the *para* position modeled into Mol-1 of irradiated sfGFP$^{\text{His148AzPhe}}$. **B**. 7-membered ring modeled into Mol-2 of irradiated sfGFP$^{\text{His148AzPhe}}$.

The main difference between the sfGFP structure (PDB code 2B3P) and the dark state of sfGFP$^{\text{His148AzPhe}}$ was the orientation of the side chain of Thr203 and this is thought to contribute to the change in the fluorescence spectrum of the AzPhe variant. In the irradiated state of sfGFP$^{\text{His148AzPhe}}$, the side chain of the threonine is still rotated compared to sfGFP so the hydroxyl group cannot form a direct interaction with Tyr66. However, there is a water molecule in between the two residues that connects the hydroxyl of Tyr66 and the backbone carbonyl of Thr203 (Figure 6-10A) to reform the hydrogen-bonding network that potentially promotes the phenolate form of Tyr66 This water molecule is present in both Mol-1 and Mol-2 in the asymmetric unit, and the side chain of Thr203 occupies the same conformer in both.

Figure 6-10. Effect of UV irradiation on Thr203 **A**. Electron density of water molecule bridging Tyr66 and the backbone carbonyl of Thr203. **B-C**. Comparison of the positions of β10 in dark sfGFP^His148AzPhe (green) and UV irradiated sfGFP^His148AzPhe when modeled with (B) an amine group and (C) a 7-membered ring at position 148.
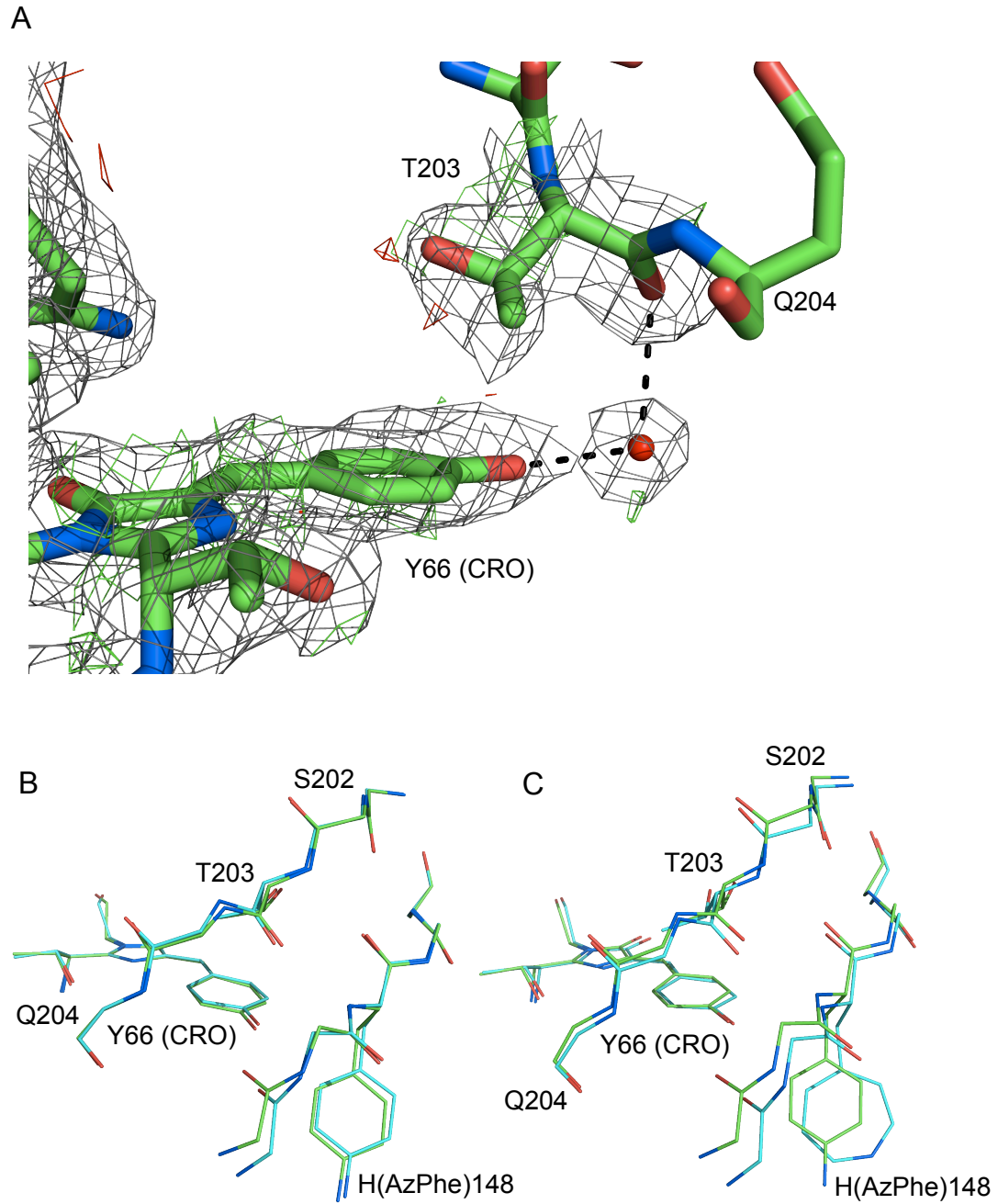
The irradiation of AzPhe results in a shift in the position of the backbones of β7 (containing position 148) and β10, which allows the interaction between Tyr66 and Thr203 to be reformed (Figure 6-10B-C). The original effect of the incorporation of AzPhe resulted in a large shift upwards, breaking the hydrogen bond between the Tyr66 and Thr203. The second shift caused by irradiation of the side chain results in a small shift in the opposite direction, bringing Thr203 closer to Tyr66, enabling the formation of a hydrogen bond between the two residues, via a water molecule. There is a difference in the position of the β-strands depending on what is modeled into the electron density surrounding position 148. When an NH_2 group is modeled into the *para* position of the phenyl ring (Mol-1), there is only a very small shift in towards the chromophore (0.3 A). However, when the 7-membered ring is modeled in (Mol-2), the shift is doubled (0.6 A).

6.3. Discussion

6.3.1. Molecular basis of the spectral shift of sfGFP^{His148AzPhe}

While replacing His148 does not destroy fluorescence, it changes the excitation profile significantly from that of sfGFP (Table 6-1; Figure 6-3A). The change in major excitation peak from 485 nm in sfGFP to 400 nm in sfGFP^{His148AzPhe} represents a shift back to that originally seen in the first GFP protein discovered in *A. victoria* [86]. This protein had a complex excitation spectrum with the major peak at 395 nm and a minor peak at 475 nm, indicating a mix of the phenol and phenolate forms of the chromophore, with the former the predominant. The mutations introduced into sfGFP resulted in a simplification of the fluorescence spectrum, notably a switch in peak intensities, with the phenolate form now being the dominant form of the chromophore, with the proton shuttled through a local charge transfer network. [92]. In wild type sfGFP, Tyr66 is kept in the phenolate form by a hydrogen-bonding network, with His148 a major contributor [88]. The mutation of histidine to AzPhe removes this interaction with Tyr66, thus disrupting the proton relay system around the chromophore preventing deprotonation in the ground state (Figure 6-11).
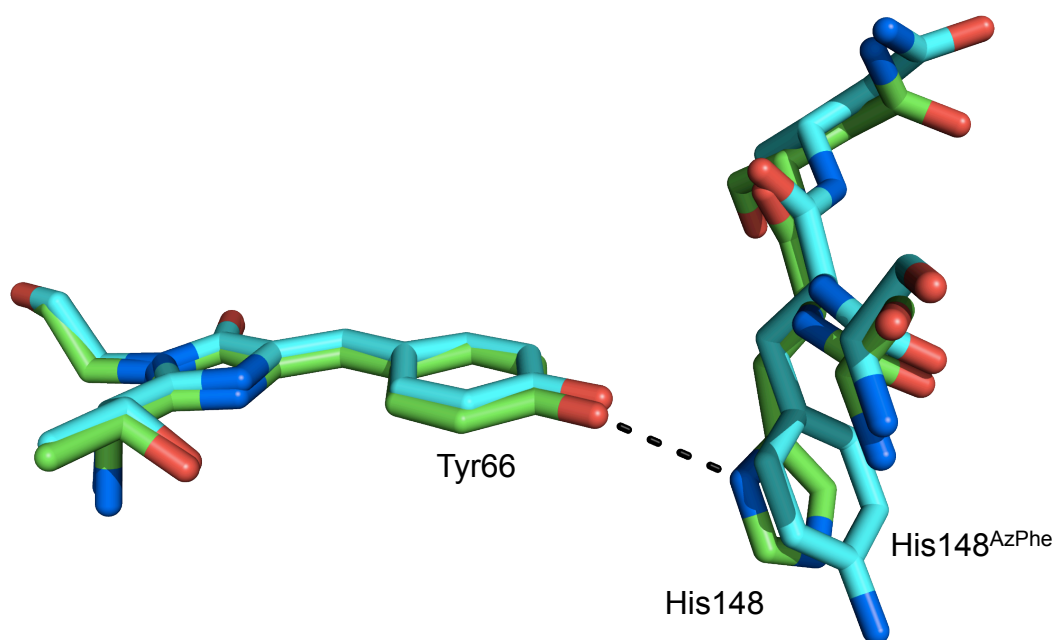
Figure 6-11. Removal of a hydrogen bond between Tyr66 and His148 upon incorporation of AzPhe.

In addition to the interaction between Tyr66 and His148, Tyr66 also forms a direct interaction with the hydroxyl side chain of Thr203 in sfGFP. In the dark form of sfGFP^{His148AzPhe}, the side chain of Thr203 has been rotated so that the distance between the two hydroxyl groups is increased by two angstroms (Figure 6-7B). The incorporation of AzPhe and the resultant increase in side chain size and bulk causes a shift in the positioning of the two β-sheets that contain His(AzPhe)148 and Thr203. This shift in backbone position, and rotation of the side chain increasing the distance between the two hydroxyl groups removes the interaction between Thr203 and the chromophore, potentially pushing Tyr66 into its phenol form. This contributes to the changed excitation spectrum of sfGFP^{His148AzPhe}, as two of the interactions involving Tyr66 that maintain the chromophore predominantly in its phenolate form have been removed. Tyr66 has a third interaction in sfGFP with a water molecule. This interaction is also present in the dark form of sfGFP^{His148AzPhe}.

On irradiation of sfGFP^{His148AzPhe}, the two peaks representing the phenol and the phenolate forms of Tyr66 switch intensity so that the predominant peak is now at 485 nm. This is similar to the excitation and emission profile of wild type sfGFP. This suggests that the hydrogen bond and charge transfer network between Tyr66 and His148, and between Tyr66 and Thr203 are re-established. It is the Nδ of His148 that interacts with Tyr66 in the chromophore and no matter which reaction pathway the azide group undergoes upon UV irradiation, there is no possibility of forming an equivalent interaction with AzPhe. Therefore the irradiation of AzPhe is not going to produce a direct interaction with the chromophore. The electron density that surrounds the side chain of AzPhe suggests that there are no interactions with the chromophore via water molecules either. This suggests that on irradiation of the phenyl azide group, the final photochemical product is not directly altering the interaction with Tyr66.

The other disrupted interaction in the dark state is between the chromophore and Thr203. In the irradiated state of sfGFP^{His148AzPhe}, the side chain of Thr203 does not change orientation compared to the dark state, so there are no interactions between the side chain of Thr203 and the chromophore. However there is a slight shift in the position of β10, the β-sheet where Thr203 resides. This small shift downwards reduces the effect of the original shift caused by the incorporation of AzPhe and

allows the backbone carbonyl group of Thr203 to interact with the hydroxyl group of Tyr66 via a water molecule (Figure 6-10A). This backbone shift allows the hydroxyl group of Tyr66 to form one of the two interactions it makes with other residues in the wild type structure. The reformation of this interaction is likely to push the chromophore towards its phenolate state, which excites at 485 nm, and from its phenol form, reducing the excitation at 400 nm.

6.3.2. Effect of SPAAC on sfGFP^{His148AzPhe} fluorescence

When Click chemistry is used to modulate fluorescence, there are two different outcomes depending on the molecule used to modify the protein (Table 6-2; Figure 6-4A). When **1** is used to modify the protein, the predominant peak remains at 400 nm whereas when the protein is modified with **2**, the wavelength of the predominant peak switches to 485 nm in a result similar to the effect of UV irradiation. Through the structural studies we know that the incorporation of AzPhe disrupts the hydrogen-bonding network around Tyr66, pushing the chromophore into its phenol state, and thus exciting predominantly at 400 nm, and we know UV irradiation partially reforms this hydrogen-bonding network, which results in the shift in excitation wavelength back to 485 nm. These details allow us to hypothesize that modification with **2** is having a similar impact on the hydrogen-bonding network, as the phenotypic result of modification with **2** is very similar to modification with UV light. The structure of **2** contains an amine group at the end of a short linker that connects it to the dibenzyl cyclo-octyne (DBCO) group that is used to react with the azide group during SPAAC. This amine group would be able to participate in a hydrogen-bonding network provided it was close enough to another side chain. The side chain of AzPhe at position 148 is near the surface of the protein with the azide group pointing away from the core of the protein. The short linker, part of the structure of **2**, is likely to be flexible so it is possible that the amine group could bend back towards the protein surface and contribute, directly or indirectly, to the hydrogen-bonding network around Tyr66. Without any crystal structure of the modified protein, or of any computer modeling or molecular dynamics simulations of the modification, it is very difficult to know precisely how the chemical modification is interacting with the chromophore and how it is effecting the hydrogen-bonding network.

Modification with **1** does not result in the same shift in fluorescence, the predominant peak remains at 400 nm. The sheer size of the modification suggests it would remain outside the β-barrel that protects the chromophore from the surrounding environment. From here it would be unable to impact the hydrogen-bonding network that occurs inside the β-barrel. This would leave the chromophore in its phenol state, exciting at 400 nm. However, the position of AzPhe incorporation means the fluorophore, which is part of the structure of **1**, is close enough to the chromophore of the protein for energy transfer. The quantum yield of sfGFP^{His148AzPhe} modified with **1** is lower than the dark state of the same variant when excited at both 400 nm and 485 nm (Table 6-1), suggesting that some of the energy has been transferred to the fluorophore in **1**, which emits at 610 nm. The excitation of sfGFP^{His148AzPhe} modified with **1** at 400 nm or 485 nm results in the emission of light energy at 511 nm, which is then used as the excitation energy for the fluorophore of **1** (excitation maxima 584 nm), resulting in emission at 610 nm. When the protein is excited at 400 nm, there is only a small emission at 610 nm. There is an increase in emission at 610 nm when the modified protein is excited at 485 nm. As the wavelength of excitation is so close to the emission of sfGFP (511 nm), it is likely some of the emission at 610 nm is caused by direct excitation of the fluorophore by the light source, not by energy transfer.

An approximate relative FRET efficiency, calculated experimentally using the fluorescence spectra was calculated as 9% using Equation 2 (Section 2.8.3.). This used the emission profile of the sfGFP chromophore (donor) and the fluorophore of **1** (acceptor) when excited at 400 nm, as excitation at 485 nm is likely to result in some direct excitation of the acceptor fluorophore. Theoretical calculations were impossible due to the unknown distance between the two fluorophores. Therefore, only a small amount of energy transfer is seen between the chromophore and the fluorophore. The difference between the emission maxima of sfGFP, 511 nm, and the excitation maxima of the fluorophore, 584 nm is quite large, and as a result the efficiency of energy transfer between the two is probably quite low. Similarly, the linker between the DBCO and the fluorophore is quite flexible, which would limit the efficiency at which energy transfer occurs due to the lack of a consistent orientation between the two chromophores [157]. Without any modeling of the modification it is very difficult to estimate the exact distance between the two chromophores, however it is likely to be too large to result in high levels of energy transfer.

6.3.3. Effects of UV irradiation on AzPhe

The sensitivity of azide groups to UV light has been widely studied over many years, however much of this understanding relies on the use of extreme temperatures and reaction conditions to achieve specific products, such as using methanol to push the reaction to form a nitrenium ion [63]. The incorporation of the azide group into proteins via an amino acid, and the study of these proteins removes such extreme conditions as a possibility, and therefore removes much of the control when managing the reactivity of azide groups. This means that the result of irradiating an azide group in a protein setting is difficult to predict, with multiple factors affecting the outcome, such as buffer conditions and the residue microenvironment. Recently, sfGFP containing the azide group has been crystallized and this has aided the prediction of potential reaction pathways and shown the variety of outcomes that are possible [23].

The change in excitation spectra in the dark form of sfGFP^{His148AzPhe} due to the incorporation of AzPhe, and the further change in excitation due to irradiation of the azide group have been explained in molecular terms thanks to crystallization of the protein, however the exact effect on AzPhe remains unclear. The difference in the electron density that surrounds the phenyl ring of AzPhe, particularly the *para* position, in the two molecules of irradiated sfGFP^{His148AzPhe} raise questions over the end result of UV irradiation on the azide group. In one of the molecules, clear density can be seen around the *para* position, whereas the second molecule has no such density. In this molecule the density that corresponds to the phenyl ring is slightly larger, which would accommodate a 7-membered ring. These two molecules were modeled differently during refinement and no problems were encountered during the process, indicating both structures were tolerated. It is possible that there are multiple outcomes during the irradiation process, however as both molecules are sampling the same environment this is unlikely. It is also possible, due to the *in crystallo* irradiation conditions, that the two different structures represent two different points of the reaction. The first step in the reaction after UV irradiation is the loss of molecular nitrogen to form a nitrene structure that closely resembles the electron density seen in the first molecule of the asymmetric unit. It is possible that upon freezing of the crystal for data collection, the reaction was trapped at this stage. The second molecule then samples a further stage of the reaction pathway, at which point the nitrene radical has joined the phenyl ring to form a 7-membered ring (the dihydroazepine ring). The

positioning of β10 in the two models also suggests this may be the case. When the density around the *para* position is filled with an amine group (the first step), there is only a small shift in the backbone position of Thr203. When the 7-membered ring is modeled in (the second step), there is a larger shift in the backbone position, bringing Thr203 into closer proximity to the chromophore. These structures could represent two stages of the reduction of the azide group, bringing Thr203 closer to Tyr66.

6.4. Summary

In this chapter I have introduced the control of protein fluorescence using two different types of novel post-translational modification. Superfolder GFP (sfGFP) has a well-known excitation and emission spectra and has been shown previously to be sensitive to single amino acid changes [158]. The incorporation of AzPhe into sfGFP resulted in a shift in excitation wavelength from 485 nm in sfGFP to 400 nm in sfGFP^{His148AzPhe}. This corresponds in a change in the chromophore, from the phenolate form of Tyr66 to its neutral phenol form. This change was reversed upon UV irradiation of the azide group. Crystallography was used to try and determine the precise molecular effects of UV irradiation on AzPhe itself and on the sfGFP chromophore. In the dark form of sfGFP^{His148AzPhe}, two hydrogen bonds that form between Tyr66 of the chromophore and His148 and Thr203 were removed, causing the chromophore to exist predominantly in its phenol state. In the irradiated state, the hydrogen bond between Tyr66 and Thr203 was reformed via a water molecule, shifting the chromophore back to its phenolate form, moving the main excitation peak to 485 nm. Of the two molecules in the irradiated crystal structure, the electron density around AzPhe differed. One showed density around the *para* position and one accommodated a larger, 7-membered ring. These two structures were modeled separately, and could highlight two steps of the photolysis reaction. Click chemistry had a variable effect on excitation depending on the modification used. When a modification containing an amine group was used, the excitation was the same as that of the irradiated form, exciting primarily at 485 nm. When another modification was used that had no amine groups, but replaced it with a large fluorophore, the main excitation peak remained at 400 nm. For all of the post-translational modifications used, the emission spectra of the protein were unchanged, remaining at 511 nm.

7. Discussion

Unnatural amino acids (uAAs) are an ideal method to introduce new functionality into proteins as they can be designed and tailored to suit very specific needs. They can be used to modulate the function of a protein, or even access novel functionality not normally present in the protein. In this thesis, the uAA *p*-azido-L-phenylalanine (AzPhe) was used to modulate the activity of TEM β-lactamase (Chapter 4), a clinically relevant enzyme that is the most widespread resistance marker [159]. Beyond incorporation of the uAA, two novel methods of protein control, UV irradiation and Click chemistry, were used to post-translationally modify the activity of the enzyme, with both positive and negative effects. As well as affecting activity, the incorporation of AzPhe and novel post-translational modification via Click chemistry was used to enable the defined and oriented immobilization of the protein onto two different, biologically relevant surfaces (Chapter 5). The interaction between pyrene and graphene was used to interface TEM β-lactamase with an electronically sensitive surface that could be used to analyze tiny changes in the enzyme active site [136]. Strain-promoted azide-alkyne cyclo-addition (SPAAC), a form of Click chemistry, was also used to attach DNA oligonucleotides to the protein to enable attachment of the protein to a DNA origami "tile". This technology allows multiple proteins to be immobilized in defined positions on a single surface [153]. In addition to the control of TEM β-lactamase, AzPhe was also incorporated into sfGFP and modified to modulate the protein's intrinsic fluorescence (Chapter 6). X-ray crystallography was used to determine the precise molecular impact of AzPhe incorporation and the effect of UV irradiation on the novel side chain.

7.1. The efficiency of uAA incorporation

The use of uAAs present many benefits over other synthetic biology approaches, namely its versatility. Hundreds of uAAs have been designed and tailored to suit individual purposes, such as mimicking phosphorylation [20, 21] or enabling photo-control [160], resulting in a vast library of novel amino acids that can be used to introduce novel functionality and control into proteins. There are many ways to incorporate uAAs into proteins, from chemical synthesis and cell-free synthesis through to chemical modification of existing amino acids, however the best method is

co-translational incorporation via an expanded genetic code. Producing proteins *in vivo* is cheaper, and results in far greater yields compared to chemical synthesis. The key question that remains however is the efficiency of co-translational uAA incorporation. Translation has evolved into a highly reliable and accurate cellular process and the adaptation of this system is a difficult task. Many different cellular machines interact to convert DNA to protein, and the precise interplay between these molecules makes "hijacking" this process a difficult task to accomplish without negatively affecting global protein translation.

Much work has been done to overcome the issues associated with co-translational incorporation of uAAs and ensure that they are incorporated with the utmost reliability and accuracy. The choice of which codon to reassign to the uAA presents the first problem. Despite TAG/UAG being the least common codon in the E. coli genome, it is still the stop codon for 7% of the genes, so reassigning this to a sense codon for uAAs causes extended translation in these proteins. Also, by using a trinucleotide codon, the number of codons that can be reassigned is limited to one, as all of the other 63 sense codons play a key role during translation. Several methods have been used to eliminate each of these problems. Whole genome engineering approaches have been used to remove all natural TAG codons from the *E. coli* genome [33, 34], and release factor-1 has been removed to recode TAG to a sense codon [126, 127]. Both of these approaches have increased the efficiency of uAA incorporation in response to TAG, but do not address the issue of limited codon choice. To address this issue, four-base codons have been used along with an engineered ribosome that keep TAG as a stop codon [53]. The change to using four-base codons is a good way of not affecting any of the trinucleotide codons, and increasing the number of uAA codons that can be used simultaneously.

A second issue is the fidelity of the evolved amino-acyl tRNA synthetases (aaRSs) that are used to incorporate the uAAs co-translationally. A positive and negative selection strategy has been used to generate a library of evolved aaRSs that are capable of incorporating uAAs in response to the amber stop codon, however detailed studies have thrown questions over their efficiency and fidelity [41, 42]. The structural differences between large uAAs and the natural amino acids that are normally incorporated by these tRNA/aaRS pairs often results in issues with

recognition and binding, reducing the efficiency of co-translational uAA incorporation [161]. This problem is being eased by the expansion of the template aaRSs used to incorporate uAAs. Previously, only tyrosyl aaRSs were evolved to incorporate uAAs, however more and more aaRSs are being generated using other aaRSs as the building blocks. However, the efficiency of these evolved aaRSs at incorporating uAAs remains some way off what is seen with endogenous aaRSs. Computer-based modeling approaches may aid the directed evolution of these aaRSs by identifying sites within the protein which may increase the binding efficiency of uAAs whilst decreasing the binding efficiency of the natural amino acid that partners the particular aaRS [162]. Work has been done on other factors that influence the charging of tRNAs, such as elongation factors (EF-Tu) and their interactions with both the tRNA [163] and the uAA [164]. It has also been suggested that a designer EF-Tu, alongside the designer tRNA and aaRS could increase the efficiency of tRNA charging and uAA incorporation [51].

Although the efficiency of uAA incorporation is not comparable to endogenous amino acid incorporation, the benefits of using uAAs far outweigh the negatives. The ability to control protein function through novel post-translational modifications is a powerful technique that can be used in many proteins and in many organisms. It offers huge variety in terms of both the uAA that can be used, and the method of control that can be accessed, such as light and chemical modification.

7.2. Comparison of UV irradiation and Click chemistry as a means to introduce protein modulation via uAAs

The use of uAAs opens up novel methods of post-translational protein control, and the use of *p*-azido-L-phenylalanine (AzPhe) in particular opens up two such methods. The phenyl azide group present in AzPhe is photo-reactive, and can undergo cyclo-addition reactions with alkynes [68]. Light is a suitable method to post-translationally modify proteins, as it is non-invasive, can be controlled spatially and temporally with high precision and is cheap. When AzPhe is exposed to UV light, molecular nitrogen is lost from the azide group resulting in a reactive nitrene radical. This can then undergo numerous reactions and form several different chemical products, the result of which can be a very difficult task to predict in a protein environment. The cyclo-addition of azides and alkynes (SPAAC) can be used to modify AzPhe in a highly

defined and specific manner, and can be used to mimic the process of natural post-translational modifications such as phosphorylation whereby a specific residue is modified chemically to either increase or decrease protein activity. The versatility of this technique is virtually unmatched as an alkyne group can be incorporated into almost any chemical modification, enabling SPAAC.

In this thesis both UV light and SPAAC were used to modify proteins, and both resulted in changes to protein function and control. UV light had a very limited effect on the activity of TEM β-lactamase; of the ten AzPhe-containing variants of TEM β-lactamase studied, UV light caused no large changes in activity. UV irradiation had a more significant impact on the fluorescence of an AzPhe-containing variant of sfGFP however. UV irradiation caused a change in the excitation spectra of sfGFP, shifting the λ_{max}. X-ray crystallography was used to determine the precise effects of UV irradiation and discovered a subtle change in protein structure caused by both the initial incorporation of AzPhe and the subsequent irradiation of the side chain. Modification via SPAAC on the other hand had a much larger impact on the activity of TEM β-lactamase. Modification of AzPhe via SPAAC resulted in both increases and decreases in enzyme activity, and was then used to attach the protein onto a surface in a defined and controlled manner.

Although it would appear from the results in this thesis that Click chemistry is by far a more powerful approach to modify proteins post-translationally via uAAs, light can offer many useful approaches. Many other uAAs not used in this work have been designed that utilize light to change their structure, from photo-caged tyrosines [20] and serines [21] to photo-isomerizable residues [17] that can alternate between *cis* and *trans* bond formations. The use of light to control proteins is a growing field, and light sensitive uAAs have a role to play in this growth. The uAA *p*-benzoyl-L-phenylalanine in particular has been extensively used in studying protein-protein interactions [24, 165] and protein-DNA interactions [25] due to its photo-reactive side chain and predictable photochemistry. The variety of photo-sensitive uAAs available means the use of light to impart novel post-translational control over protein behaviour and activity is a very useful technique.

7.3. Use of Click chemistry to modulate protein function

The versatility and variability presented by Click chemistry make it an ideal method to induce novel post-translational control over proteins. The Huisgen dipolar cyclo-addition reaction [166] between azides and alkynes is one of the most popular variations of Click chemistry due to the stability of the azide group [66]. Original forms of azide-alkyne cyclo-addition required a copper catalyst, and this method was widely used to modify proteins *in vitro*. The use of this reaction is prevalent in both drug discovery and medicinal chemistry [167, 168], such as the use of Click chemistry to generate enzyme inhibitors using *in situ* proteins [169, 170]. Copper-catalyzed azide-alkyne cyclo-addition has also been used to label cell surfaces via the specific interaction between chemicals containing an alkyne group, and a membrane protein harbouring a genetically encoded azide. This technique has been used to label bacterial cell surfaces [171] and enable to the site-specific PEGylation of proteins on the cell surface [172]. The ability to genetically encode the azide means azide-alkyne cyclo-addition reactions can be performed *in vivo*, if the need for the cytotoxic copper catalyst could be alleviated. This was achieved with the introduction strain-promoted azide-alkyne cyclo-addition (SPAAC) [173]. SPAAC has been used to successfully attach proteins to surfaces [65] and enable the *in vivo* imaging of biomolecules [174, 175]. The combination of uAAs and SPAAC *in vivo* represents a great step forward in the progress of this technology, and the versatility of this novel chemical genetic approach means many more applications are likely to be found in the near future.

7.4. The use of SPAAC to attach proteins to surfaces in a defined orientation

The desire for the simultaneous analysis of many biomolecules, or the enhanced detail gained from analyzing individual biomolecules, has led to the development of surface-based protein analysis. The ability to immobilize many thousands of biomolecules on a surface has greatly increased the speed at which many protein-based studies, such as small molecule screening, can be completed. Conversely, the ability to immobilize a single protein onto a surface and to measure a single biomolecule increases the insight we can gain into the atomic events that occur within proteins. Protein immobilization has been historically achieved using protein adsorption or lysine/cysteine modification, both of which result in heterogeneous orientation and positioning of the biomolecule on the surface. The use of SPAAC, and

the benefits of specificity and selectively that it offers, can result in the homogeneous immobilization of proteins onto a surface with a single, defined orientation.

In this thesis, AzPhe and SPAAC were used to attach TEM β-lactamase onto graphene, a carbon-based sp^2-bonded material [143], and two DNA oligonucleotides, that can be used to attach the protein onto DNA origami, a 2D structure formed using single-stranded DNA [137]. To achieve this, only one mutation was required, utilizing the versatility of SPAAC to enable immobilization on both surfaces. The modification of AzPhe at two defined positions of TEM β-lactamase with a functionalized pyrene group resulted in the immobilization of the protein on graphene. The conductive properties of graphene and carbon nanotubes (CNTs) make them hugely valuable tools in protein analysis, as it can be used to detect atomic changes caused by protein-protein interactions [176] or enzyme catalysis [177, 178]. AFM (atomic force microscopy) imaging of proteins on the surface showed single molecules of expected dimensions, indicating the successful interfacing of TEM β-lactamase with graphene in a defined manner. The pyrene-graphene interaction is an ideal method to attach proteins to graphene, as it is a very strong interaction that does not disrupt the electronic properties of graphene. Direct insertion of the protein onto the graphene surface via the photo-reactivity of AzPhe is also a possible avenue of attachment, as azides have previously been used to functionalize graphene [179]. This method would provide a more intimate link between the protein and the graphene, and change the conductive properties of the graphene surface [180].

The modification of TEM β-lactamase with two DNA oligonucleotides is the first step in attaching a protein to DNA origami. The nature of DNA complementarity means proteins can be attached to the DNA origami "tile" in a defined position. By modifying an active enzyme with two different oligonucleotides, the first step towards creating enzyme pathways *in vitro* via SPAAC has been successfully completed. This approach (without the use of SPAAC) has been used to site-specifically immobilize two different enzymes onto a surface with a controlled distance between them [153]. The use of DNA as a surface to construct protein circuits is a growing field with exciting potential [181].

163

7.5. Future work

Unnatural amino acid incorporation can be applied to many proteins, so much of the work done in this thesis can be repeated in other proteins. The introduction of uAAs into other enzymes or other proteins such as membrane proteins and their subsequent control using light and Click chemistry would be an interesting addition to the work presented here. To build upon the work in this thesis, protein crystallography could be used to determine the exact effects of uAA incorporation and novel post-translational modification on TEM β-lactamase. X-ray crystallography was successfully used to define the precise molecular events that led to the spectral change seen in sfGFP^{His148AzPhe} caused by the UV irradiation of AzPhe (Chapter 6), and this technique could be duplicated to determine the effects of Click chemistry. SPAAC modification caused novel changes in two AzPhe-containing TEM β-lactamase variants, namely the large decrease in activity seen when TEM^{Tyr105AzPhe} was modified, and the increase in activity seen when TEM^{Pro174AzPhe} was modified with a DBCO-amine. SPAAC also resulted in an interesting effect on the fluorescence spectrum of sfGFP^{His148AzPhe} when modified with the same DBCO-amine. Crystallization and determination of the structures of these three modified proteins could help explain the precise effects on TEM β-lactamase and sfGFP caused by modification via SPAAC.

The immobilization of TEM β-lactamase onto graphene lays the foundation for further work to build on these advances (Chapter 5). The immobilization of proteins onto graphene enables the measurement of single molecules using the conductive properties of the surface. The interactions between TEM β-lactamase and different substrates/inhibitors could be defined using this technology and tiny atomic differences between their binding could be detected. This could provide highly sensitive measurement of protein-protein and protein-ligand interactions. Additionally, the immobilization of an enzyme onto DNA origami opens up the prospect on constructing enzyme pathways *in vitro* on a 2D or 3D surface. The defined placement of multiple enzymes on a single DNA origami "tile" would require the modification of proteins with different DNA oligonucleotides, a technique displayed in this thesis (Chapter 5). The isolation of these modified proteins onto the surface would then enable the construction of these protein networks.

Chapter 3 introduced a method whereby potential uAA incorporation sites were modeled *in silico* to help determine possible effects. Although the method used in this thesis was relatively basic, relying on the similar structure of tyrosine to imply the potential effects of UV irradiation or Click chemistry, the advances made in mutation modeling and specifically uAA mutation modeling allows increased complexity to be introduced *in silico* [182]. The power of this method is therefore greatly enhanced and can be used to help create novel protein variants that can have designed effects upon novel post-translational incorporation and modification.

7.6. Summary

Chapter 3 presented the use of *in silico* modeling as a means to determine which sites for uAA incorporation were most likely to produce novel effects caused by the post-translational modification of the azide side chain in AzPhe via UV irradiation and/or Click chemistry. It detailed the steps taken to replace the natural amino acid with tyrosine, the closest available structural analogue of AzPhe, and the interpretation of any changes to the precise orientation of residue side chains around the mutation site. Using this method, ten positions were selected for the incorporation of AzPhe in TEM β-lactamase. Chapter 3 also detailed the molecular biology required to adapt an expression vector for the over-expression of TEM β-lactamase containing uAAs.

Chapter 4 introduced the production of TEM β-lactamase containing uAAs and the coarse control over protein function gained by using uAAs; when the uAA was not provided in the growth medium, no protein activity was observed. The chapter went on to describe the steps taken to purify the proteins so that detailed kinetic analyses could be performed on the enzymes. The measurement of k_{cat}, K_M and overall catalytic efficiency of five AzPhe-containing variants was used to determine the precise effects of uAA incorporation on enzyme catalysis, and the subsequent effects of UV irradiation and Click chemistry. This revealed a variant (TEM^{Tyr105AzPhe}) that had no change in activity upon uAA incorporation, but large changes in activity after modification, a variant whose activity was greatly affected by uAA incorporation (TEM^{Ala237AzPhe}), a variant whose ability to hydrolyze its substrates increased upon the specific modification of the uAA (TEM^{Pro174AzPhe}) and two variants (TEM^{Trp165AzPhe} and TEM^{Leu201AzPhe}) that had limited effects on enzyme activity before and after modification.

Chapter 5 continued the use of Click chemistry as a means to modify proteins post-translationally, but transferred the technique into another area. SPAAC was used to modify TEM β-lactamase with two different modifications that can enable immobilization on two different surfaces; modification with pyrene allowed association with graphene, and modification with a DNA oligonucleotide represents the first step toward immobilization on a DNA origami "tile". TEM β-lactamase associated with both graphite and graphene via pyrene modification was successfully imaged using AFM.

Chapter 6 introduced the effects of uAA incorporation and novel post-translational modifications on a second protein, sfGFP. AzPhe was incorporated into sfGFP at a single position (His148) and the effects of its incorporation and subsequent modification using both UV irradiation and Click chemistry were determined via fluorescence spectroscopy. This found that incorporation of AzPhe at His148 resulted in a change in the excitation spectrum of the protein, changing the λ_{max} from 485 nm to 400 nm. Modification of AzPhe using UV irradiation then resulted in a shift in λ_{max} back to 485 nm. Modification via Click chemistry had different results depending on which chemical modification was used. Modification with a large, fluorescent dye had no effect on the fluorescence of the protein but did result in FRET. Modification with DBCO-amine resulted in the same effect as UV irradiation, changing the λ_{max} back to 485 nm. The dark and irradiated states of sfGFP^{His148AzPhe} were then crystallized so that the precise molecular events that led to the change in excitation spectrum could be determined. This found a disruption in the hydrogen-bonding network that involves the chromophore, His148 and Thr203 caused by incorporation of AzPhe. This hydrogen-bonding network was then partially reformed upon photolysis of the azide group, explaining the changes in fluorescence.

8. References

1. Crick, F., *Central Dogma of Molecular Biology.* Nature, 1970. **227**(5258): p. 561-&.

2. Sutcliffe, J.G., *Nucleotide-Sequence of Ampicillin Resistance Gene of Escherichia-Coli Plasmid Pbr322.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(8): p. 3737-3741.

3. Hunter, T., *Why nature chose phosphate to modify proteins.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. **367**(1602): p. 2513-2516.

4. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein.* Biochemistry, 1997. **36**(22): p. 6786-6791.

5. Srinivasan, G., C.M. James, and J.A. Krzycki, *Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA.* Science, 2002. **296**(5572): p. 1459-1462.

6. Bock, A., et al., *Selenocysteine - the 21st Amino-Acid.* Molecular Microbiology, 1991. **5**(3): p. 515-520.

7. Lagerkvist, U., *2 out of 3 - Alternative Method for Codon Reading.* Proceedings of the National Academy of Sciences of the United States of America, 1978. **75**(4): p. 1759-1762.

8. Walsh, C., *Posttranslational modification of proteins : expanding nature's inventory*2006, Englewood, Colo.: Roberts and Co. Publishers. xxi, 490 p.

9. Bartova, E., et al., *Histone modifications and nuclear architecture: A review.* Journal of Histochemistry & Cytochemistry, 2008. **56**(8): p. 711-721.

10. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.

11. Rudd, P.M. and R.A. Dwek, *Glycosylation: Heterogeneity and the 3D structure of proteins.* Critical Reviews in Biochemistry and Molecular Biology, 1997. **32**(1): p. 1-100.

12. Wang, L. and P.G. Schultz, *A general approach for the generation of orthogonal tRNAs.* Chemistry & Biology, 2001. **8**(9): p. 883-890.

13. Sakamoto, K., et al., *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells.* Nucleic Acids Research, 2002. **30**(21): p. 4692-4699.

14. Schultz, K.C., et al., *A genetically encoded infrared probe.* Journal of the American Chemical Society, 2006. **128**(43): p. 13984-13985.

15. Chin, J.W., et al., *Addition of p-azido-L-phenylaianine to the genetic code of Escherichia coli.* Journal of the American Chemical Society, 2002. **124**(31): p. 9026-9027.

16. Chin, J.W., et al., *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, 2002. **99**(17): p. 11020-11024.

17. Bose, M., et al., *The incorporation of a photoisomerizable amino acid into proteins in E-coli.* Journal of the American Chemical Society, 2006. **128**(2): p. 388-389.

18. Deiters, A., et al., *A genetically encoded photocaged tyrosine.* Angewandte Chemie-International Edition, 2006. **45**(17): p. 2728-2731.

19. Wang, L., et al., *Addition of the keto functional group to the genetic code of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(1): p. 56-61.

20. Xie, J.M., L. Supekova, and P.G. Schultz, *A genetically encoded metabolically stable analogue of phosphotyrosine in Escherichia coli.* Acs Chemical Biology, 2007. **2**(7): p. 474-478.

21. Park, H.S., et al., *Expanding the Genetic Code of Escherichia coli with Phosphoserine.* Science, 2011. **333**(6046): p. 1151-1154.

22. Riggsbee, C.W. and A. Deiters, *Recent advances in the photochemical control of protein function.* Trends in Biotechnology, 2010. **28**(9): p. 468-475.

23. Reddington, S.C., et al., *Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence.* Angewandte Chemie-International Edition, 2013. **52**(23): p. 5974-5977.

24. Chen, H.T., L. Warfield, and S. Hahn, *The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex.* Nature Structural & Molecular Biology, 2007. **14**(8): p. 696-703.

25.   Lee, H.S., R.D. Dimla, and P.G. Schultz, *Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid.* Bioorganic & Medicinal Chemistry Letters, 2009. **19**(17): p. 5222-5224.

26.   Banghart, M., et al., *Light-activated ion channels for remote control of neuronal firing.* Nature Neuroscience, 2004. **7**(12): p. 1381-1386.

27.   Mourot, A., et al., *Rapid optical control of nociception with an ion-channel photoswitch.* Nature Methods, 2012. **9**(4): p. 396-U114.

28.   Kent, S.B.H., *Chemical Synthesis of Peptides and Proteins.* Annual Review of Biochemistry, 1988. **57**: p. 957-989.

29.   Wang, Q., A.R. Parrish, and L. Wang, *Expanding the Genetic Code for Biological Studies.* Chemistry & Biology, 2009. **16**(3): p. 323-336.

30.   Watson, J.D., *Molecular biology of the gene.* 6th ed2008, San Francisco, CA: Pearson Education/Benjamin Cummings. xxxii, 841 p.

31.   Brennan, T. and M. Sundaralingam, *Structure of Transfer-Rna Molecules Containing Long Variable Loop.* Nucleic Acids Research, 1976. **3**(11): p. 3235-3251.

32.   Young, T.S. and P.G. Schultz, *Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon.* Journal of Biological Chemistry, 2010. **285**(15): p. 11039-11044.

33.   Isaacs, F.J., et al., *Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement.* Science, 2011. **333**(6040): p. 348-353.

34.   Mukai, T., et al., *Codon reassignment in the Escherichia coli genetic code.* Nucleic Acids Research, 2010. **38**(22): p. 8188-8195.

35.   Liu, D.R., T.J. Magliery, and P.G. Schultz, *Characterization of an 'orthogonal' suppressor tRNA derived from E-coli tRNA(2)(Gln).* Chemistry & Biology, 1997. **4**(9): p. 685-691.

36.   Liu, D.R. and P.G. Schultz, *Progress toward the evolution of an organism with an expanded genetic code.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(9): p. 4780-4785.

37.   Wang, L., et al., *A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

38.   Kobayashi, T., et al., *Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion.* Nature Structural Biology, 2003. **10**(6): p. 425-432.

39.   Brick, P., T.N. Bhat, and D.M. Blow, *Structure of Tyrosyl Transfer-Rna Synthetase Refined at 2.3-a Resolution - Interaction of the Enzyme with the Tyrosyl Adenylate Intermediate.* Journal of Molecular Biology, 1989. **208**(1): p. 83-98.

40.   Santoro, S.W., et al., *An efficient system for the evolution of aminoacyl-tRNA synthetase specificity.* Nature Biotechnology, 2002. **20**(10): p. 1044-1048.

41.   Antonczak, A.K., et al., *Importance of single molecular determinants in the fidelity of expanded genetic codes.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(4): p. 1320-1325.

42.   Young, D.D., et al., *Synthetase polyspecificity as a tool to modulate protein function.* Bioorganic & Medicinal Chemistry Letters, 2011. **21**(24): p. 7502-7504.

43.   Santoro, S.W., et al., *An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli.* Nucleic Acids Research, 2003. **31**(23): p. 6700-6709.

44.   Anderson, J.C. and P.G. Schultz, *Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression.* Biochemistry, 2003. **42**(32): p. 9598-9608.

45.   Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(20): p. 7566-7571.

46.   James, C.M., et al., *The amber codon in the gene encoding the monomethylamine methyltransferase isolated from Methanosarcina barkeri is translated as a sense codon.* Journal of Biological Chemistry, 2001. **276**(36): p. 34252-34258.

47.   Fekner, T. and M.K. Chan, *The pyrrolysine translational machinery as a genetic-code expansion tool.* Current Opinion in Chemical Biology, 2011. **15**(3): p. 387-391.

48.   Chatterjee, A., H. Xiao, and P.G. Schultz, *Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid*

mutagenesis in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(37): p. 14841-14846.

49. Chatterjee, A., et al., *A Tryptophanyl-tRNA Synthetase/tRNA Pair for Unnatural Amino Acid Mutagenesis in E.coli.* Angewandte Chemie-International Edition, 2013. **52**(19): p. 5106-5109.

50. Dale, T., L.E. Sanderson, and O.C. Uhlenbeck, *The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid.* Biochemistry, 2004. **43**(20): p. 6159-6166.

51. Doi, Y., et al., *Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system.* Journal of the American Chemical Society, 2007. **129**(46): p. 14458-14462.

52. Chapman, S.J., J.M. Schrader, and O.C. Uhlenbeck, *Histidine 66 in Escherichia coli Elongation Factor Tu Selectively Stabilizes Aminoacyl-tRNAs.* Journal of Biological Chemistry, 2012. **287**(2): p. 1229-1234.

53. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome.* Nature, 2010. **464**(7287): p. 441-444.

54. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-967.

55. Bianco, A., et al., *Expanding the genetic code of Drosophila melanogaster.* Nature Chemical Biology, 2012. **8**(9): p. 748-750.

56. Greiss, S. and J.W. Chin, *Expanding the Genetic Code of an Animal.* Journal of the American Chemical Society, 2011. **133**(36): p. 14196-14199.

57. Albayrak, C. and J.R. Swartz, *Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation.* Nucleic Acids Research, 2013. **41**(11): p. 5949-5963.

58. Goerke, A.R. and J.R. Swartz, *High-Level Cell-Free Synthesis Yields of Proteins Containing Site-Specific Non-Natural Amino Acids.* Biotechnology and Bioengineering, 2009. **102**(2): p. 400-416.

59. Xie, J.M. and P.G. Schultz, *Adding amino acids to the genetic repertoire.* Current Opinion in Chemical Biology, 2005. **9**(6): p. 548-554.

60. Albayrak, C. and J.R. Swartz, *Direct Polymerization of Proteins.* Acs Synthetic Biology, 2014. **3**(6): p. 353-362.

61. Shrestha, P., M.T. Smith, and B.C. Bundy, *Cell-free unnatural amino acid incorporation with alternative energy systems and linear expression templates.* New Biotechnology, 2014. **31**(1): p. 28-34.

62. Ye, S.X., et al., *Tracking G-protein-coupled receptor activation using genetically encoded infrared probes.* Nature, 2010. **464**(7293): p. 1386-U14.

63. Voskresenska, V., et al., *Photoaffinity Labeling via Nitrenium Ion Chemistry: Protonation of the Nitrene Derived from 4-Amino-3-nitrophenyl Azide to Afford Reactive Nitrenium Ion Pairs.* Journal of the American Chemical Society, 2009. **131**(32): p. 11535-11547.

64. Grunbeck, A., et al., *Mapping the Ligand-Binding Site on a G Protein-Coupled Receptor (GPCR) Using Genetically Encoded Photocrosslinkers.* Biochemistry, 2011. **50**(17): p. 3411-3413.

65. Seo, M.H., et al., *Controlled and Oriented Immobilization of Protein by Site-Specific Incorporation of Unnatural Amino Acid.* Analytical Chemistry, 2011. **83**(8): p. 2841-2845.

66. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click chemistry: Diverse chemical function from a few good reactions.* Angewandte Chemie-International Edition, 2001. **40**(11): p. 2004-+.

67. Jang, S., et al., *Development of a Simple Method for Protein Conjugation by Copper-Free Click Reaction and Its Application to Antibody-Free Western Blot Analysis.* Bioconjugate Chemistry, 2012. **23**(11): p. 2256-2261.

68. Best, M.D., *Click Chemistry and Bioorthogonal Reactions: Unprecedented Selectivity in the Labeling of Biological Molecules.* Biochemistry, 2009. **48**(28): p. 6571-6584.

69. Milles, S., et al., *Click Strategies for Single-Molecule Protein Fluorescence.* Journal of the American Chemical Society, 2012. **134**(11): p. 5187-5195.

70. Matagne, A., J. Lamotte-Brasseur, and J.M. Frere, *Catalytic properties of class A beta-lactamases: efficiency and diversity.* Biochemical Journal, 1998. **330**: p. 581-598.

71. Cricco, J.A. and A.J. Vila, *Class B beta-lactamases: The importance of being metallic.* Current Pharmaceutical Design, 1999. **5**(11): p. 915-927.

72. Medeiros, A.A., *Beta-Lactamases.* British Medical Bulletin, 1984. **40**(1): p. 18-27.

73.   Palzkill, T. and D. Botstein, *Identification of Amino-Acid Substitutions That Alter the Substrate-Specificity of Tem-1 Beta-Lactamase.* Journal of Bacteriology, 1992. **174**(16): p. 5237-5243.

74.   Jelsch, C., et al., *Crystal-Structure of Escherichia-Coli Tem1 Beta-Lactamase at 1.8-Angstrom Resolution.* Proteins-Structure Function and Genetics, 1993. **16**(4): p. 364-383.

75.   Minasov, G., X.J. Wang, and B.K. Shoichet, *An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation.* Journal of the American Chemical Society, 2002. **124**(19): p. 5333-5340.

76.   Meroueh, S.O., et al., *Ab initio QM/MM study of class A beta-lactamase acylation: Dual participation of Glu166 and Lys73 in a concerted base promotion of Ser70.* Journal of the American Chemical Society, 2005. **127**(44): p. 15397-15407.

77.   Leung, Y.C., et al., *Site-Directed Mutagenesis of Beta-Lactamase-I - Role of Glu-166.* Biochemical Journal, 1994. **299**: p. 671-678.

78.   Strynadka, N.C.J., et al., *Molecular-Structure of the Acyl-Enzyme Intermediate in Beta-Lactam Hydrolysis at 1.7 Angstrom Resolution.* Nature, 1992. **359**(6397): p. 700-705.

79.   Mustafi, D., A. Sosa-Peinado, and M.W. Makinen, *ENDOR structural characterization of a catalytically competent acylenzyme reaction intermediate of wild-type TEM-1 beta-lactamase confirms glutamate-166 as the base catalyst.* Biochemistry, 2001. **40**(8): p. 2397-2409.

80.   Jacob, F., et al., *Role of the Conserved Amino-Acids of the Sdn Loop (Ser130, Asp131 and Asn132) in a Class-a Beta-Lactamase Studied by Site-Directed Mutagenesis.* Biochemical Journal, 1990. **271**(2): p. 399-406.

81.   Lenfant, F., R. Labia, and J.M. Masson, *Replacement of Lysine-234 Affects Transition-State Stabilization in the Active-Site of Beta-Lactamase Tem1.* Journal of Biological Chemistry, 1991. **266**(26): p. 17187-17194.

82.   Therrien, C., et al., *Roles of amino acids 161 to 179 in the PSE-4 Omega loop in substrate specificity and in resistance to ceftazidime.* Antimicrobial Agents and Chemotherapy, 1998. **42**(10): p. 2576-2583.

83. Cantu, C., W.Z. Huang, and T. Palzkill, *Cephalosporin substrate specificity determinants of TEM-1 beta-lactamase.* Journal of Biological Chemistry, 1997. **272**(46): p. 29144-29150.

84. Maveyraud, L., et al., *Structural basis of extended spectrum TEM beta-lactamases - Crystallographic, kinetic, and mass spectrometric investigations of enzyme mutants.* Journal of Biological Chemistry, 1996. **271**(18): p. 10482-10489.

85. Petrosino, J.F. and T. Palzkill, *Systematic mutagenesis of the active site omega loop of TEM-1 beta-lactamase.* Journal of Bacteriology, 1996. **178**(7): p. 1821-1828.

86. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from Luminous Hydromedusan, Aequorea.* Journal of Cellular and Comparative Physiology, 1962. **59**(3): p. 223-&.

87. Baldwin, A.J., et al., *Expanded chemical diversity sampling through whole protein evolution.* Molecular Biosystems, 2009. **5**(7): p. 764-766.

88. Tsien, R.Y., *The green fluorescent protein.* Annual Review of Biochemistry, 1998. **67**: p. 509-544.

89. Cubitt, A.B., et al., *Understanding, Improving and Using Green Fluorescent Proteins.* Trends in Biochemical Sciences, 1995. **20**(11): p. 448-455.

90. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved Green Fluorescence.* Nature, 1995. **373**(6516): p. 663-664.

91. Cormack, B.P., R.H. Valdivia, and S. Falkow, *FACS-optimized mutants of the green fluorescent protein (GFP).* Gene, 1996. **173**(1): p. 33-38.

92. Pedelacq, J.D., et al., *Engineering and characterization of a superfolder green fluorescent protein (vol 24, pg 79, 2005).* Nature Biotechnology, 2006. **24**(9): p. 1170-1170.

93. Wachter, R.M., et al., *Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein.* Structure with Folding & Design, 1998. **6**(10): p. 1267-1277.

94. Ormo, M., et al., *Crystal structure of the Aequorea victoria green fluorescent protein.* Science, 1996. **273**(5280): p. 1392-1395.

95. Arpino, J.A.J., P.J. Rizkallah, and D.D. Jones, *Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 angstrom Resolution Reveals Alternative Conformations for Glu222.* Plos One, 2012. **7**(10).

96. Studier, F.W., *Protein production by auto-induction in high-density shaking cultures.* Protein Expression and Purification, 2005. **41**(1): p. 207-234.

97. Guex, N. and M.C. Peitsch, *SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling.* Electrophoresis, 1997. **18**(15): p. 2714-2723.

98. Schwede, T., et al., *SWISS-MODEL: an automated protein homology-modeling server.* Nucleic Acids Research, 2003. **31**(13): p. 3381-3385.

99. Scott, W.R.P., et al., *The GROMOS biomolecular simulation program package.* Journal of Physical Chemistry A, 1999. **103**(19): p. 3596-3607.

100. Laemmli, U.K., *Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4.* Nature, 1970. **227**(5259): p. 680-&.

101. Edwards, W.R., et al., *Regulation of beta-Lactamase Activity by Remote Binding of Heme: Functional Coupling of Unrelated Proteins through Domain Insertion.* Biochemistry, 2010. **49**(31): p. 6541-6549.

102. Winter, G., *xia2: an expert system for macromolecular crystallography data reduction.* Journal of Applied Crystallography, 2010. **43**: p. 186-190.

103. Evans, P., *Scaling and assessment of data quality.* Acta Crystallographica Section D-Biological Crystallography, 2006. **62**: p. 72-82.

104. Bailey, S., *The Ccp4 Suite - Programs for Protein Crystallography.* Acta Crystallographica Section D-Biological Crystallography, 1994. **50**: p. 760-763.

105. Neylon, C., *Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution.* Nucleic Acids Research, 2004. **32**(4): p. 1448-1459.

106. Banerjee, S., et al., *Role of the Omega-loop in the activity, substrate specificity, and structure of class A beta-lactamase.* Biochemistry, 1998. **37**(10): p. 3286-3296.

107. Marciano, D.C., et al., *Genetic and Structural Characterization of an L201 P Global Suppressor Substitution in TEM-1 beta-Lactamase.* Journal of Molecular Biology, 2008. **384**(1): p. 151-164.

108. Sideraki, V., et al., *A secondary drug resistance mutation of TEM-1 beta-lactamase that suppresses misfolding and aggregation.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(1): p. 283-288.

109. Petersen, B., et al., *A generic method for assignment of reliability scores applied to solvent accessibility predictions.* Bmc Structural Biology, 2009. **9**.

110. Reddington, S.C., E.M. Tippmann, and D.D. Jones, *Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry.* Chemical Communications, 2012. **48**(67): p. 8419-8421.

111. Knox, J.R., *Extended-Spectrum and Inhibitor-Resistant Tem-Type Beta-Lactamases - Mutations, Specificity, and 3-Dimensional Structure.* Antimicrobial Agents and Chemotherapy, 1995. **39**(12): p. 2593-2601.

112. Dubus, A., et al., *Catalytic Mechanism of Active-Site Serine Beta-Lactamases - Role of the Conserved Hydroxy Group of the Lys-Thr(Ser)-Gly Triad.* Biochemical Journal, 1994. **301**: p. 485-494.

113. Bos, F. and J. Pleiss, *Multiple Molecular Dynamics Simulations of TEM beta-Lactamase: Dynamics and Water Binding of the Omega-Loop.* Biophysical Journal, 2009. **97**(9): p. 2550-2558.

114. Vakulenko, S.B., et al., *Effects on substrate profile by mutational substitutions at positions 164 and 179 of the class A TEMpUC19 beta-lactamase from Escherichia coli.* Journal of Biological Chemistry, 1999. **274**(33): p. 23052-23060.

115. Schleif, R., *Regulation of the L-arabinose operon of Escherichia coli.* Trends in Genetics, 2000. **16**(12): p. 559-565.

116. Jones, D.D., *Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 beta-lactamase to an amino acid deletion.* Nucleic Acids Research, 2005. **33**(9).

117. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.

118. Miyada, C.G., L. Stoltzfus, and G. Wilcox, *Regulation of the Arac Gene of Escherichia-Coli - Catabolite Repression, Auto-Regulation, and Effect on*

Arabad Expression. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1984. **81**(13): p. 4120-4124.

119. Fisette, O., S. Gagne, and P. Lague, *Molecular Dynamics of Class A beta-lactamases-Effects of Substrate Binding.* Biophysical Journal, 2012. **103**(8): p. 1790-1801.

120. Atanasov, B.P., D. Mustafi, and M.W. Makinen, *Protonation of the beta-lactam nitrogen is the trigger event in the catalytic action of class A beta-lactamases.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(7): p. 3160-3165.

121. Diaz, N., et al., *Molecular dynamics simulations of the TEM-1,beta-lactamase complexed with cephalothin.* Journal of Medicinal Chemistry, 2005. **48**(3): p. 780-791.

122. Fisette, O., et al., *TEM-1 Backbone Dynamics-Insights from Combined Molecular Dynamics and Nuclear Magnetic Resonance.* Biophysical Journal, 2010. **98**(4): p. 637-645.

123. Osuna, J., et al., *Substitution of Asp for Asn at Position-132 in the Active-Site of Tem Beta-Lactamase - Activity toward Different Substrates and Effects of Neighboring Residues.* Journal of Biological Chemistry, 1995. **270**(2): p. 775-780.

124. Shao, Y.Y., et al., *Graphene Based Electrochemical Sensors and Biosensors: A Review.* Electroanalysis, 2010. **22**(10): p. 1027-1036.

125. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

126. Johnson, D.B.F., et al., *Release Factor One Is Nonessential in Escherichia coli.* Acs Chemical Biology, 2012. **7**(8): p. 1337-1344.

127. Johnson, D.B.F., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites.* Nature Chemical Biology, 2011. **7**(11): p. 779-786.

128. Ayyadurai, N., et al., *A facile and efficient method for the incorporation of multiple unnatural amino acids into a single protein.* Chemical Communications, 2011. **47**(12): p. 3430-3432.

129. MacBeath, G. and S.L. Schreiber, *Printing proteins as microarrays for high-throughput function determination.* Science, 2000. **289**(5485): p. 1760-1763.

130. Cooper, M.A., *Optical biosensors in drug discovery.* Nature Reviews Drug Discovery, 2002. **1**(7): p. 515-528.

131. Sevecka, M. and G. MacBeath, *State-based discovery: a multidimensional screen for small-molecule modulators of EGF signaling.* Nature Methods, 2006. **3**(10): p. 825-831.

132. Muller, D.J., et al., *Observing structure, function and assembly of single proteins by AFM.* Progress in Biophysics & Molecular Biology, 2002. **79**(1-3): p. 1-43.

133. Saleh, O.A., et al., *Single-molecule manipulation measurements of DNA transport proteins.* Chemphyschem, 2005. **6**(5): p. 813-818.

134. Rabe, M., D. Verdes, and S. Seeger, *Understanding protein adsorption phenomena at solid surfaces.* Advances in Colloid and Interface Science, 2011. **162**(1-2): p. 87-106.

135. Bjork, J., et al., *Adsorption of Aromatic and Anti-Aromatic Systems on Graphene through pi-pi Stacking.* Journal of Physical Chemistry Letters, 2010. **1**(23): p. 3407-3412.

136. Salavagione, H.J., *Promising alternative routes for graphene production and functionalization.* Journal of Materials Chemistry A, 2014. **2**(20): p. 7138-7146.

137. Rothemund, P.W.K., *Scaffolded DNA origami for nanoscale shapes and patterns.* Abstracts of Papers of the American Chemical Society, 2006. **231**.

138. Seeman, N.C., *Nucleic-Acid Junctions and Lattices.* Journal of Theoretical Biology, 1982. **99**(2): p. 237-247.

139. Numajiri, K., et al., *Stepwise and reversible nanopatterning of proteins on a DNA origami scaffold.* Chemical Communications, 2010. **46**(28): p. 5127-5129.

140. Goodman, R.P., et al., *A Facile Method for Reversibly Linking a Recombinant Protein to DNA.* Chembiochem, 2009. **10**(9): p. 1551-1557.

141. Shen, W.Q., et al., *NTA Directed Protein Nanopatterning on DNA Origami Nanoconstructs.* Journal of the American Chemical Society, 2009. **131**(19): p. 6660-+.

142. Rinker, S., et al., *Self-assembled DNA nanostructures for distance-dependent multivalent ligand-protein binding.* Nature Nanotechnology, 2008. **3**(7): p. 418-422.

143. Novoselov, K.S., et al., *Electric field effect in atomically thin carbon films.* Science, 2004. **306**(5696): p. 666-669.

144. Blake, P., et al., *Graphene-based liquid crystal device.* Nano Letters, 2008. **8**(6): p. 1704-1708.

145. Wang, X., L.J. Zhi, and K. Mullen, *Transparent, conductive graphene electrodes for dye-sensitized solar cells.* Nano Letters, 2008. **8**(1): p. 323-327.

146. Schedin, F., et al., *Detection of individual gas molecules adsorbed on graphene.* Nature Materials, 2007. **6**(9): p. 652-655.

147. Dong, X., et al., *A graphene nanoribbon network and its biosensing application.* Nanoscale, 2011. **3**(12): p. 5156-5160.

148. Song, B., et al., *Graphene on Au(111): A Highly Conductive Material with Excellent Adsorption Properties for High-Resolution Bio/Nanodetection and Identification.* Chemphyschem, 2010. **11**(3): p. 585-589.

149. Choi, Y.K., et al., *Single-Molecule Lysozyme Dynamics Monitored by an Electronic Circuit.* Science, 2012. **335**(6066): p. 319-324.

150. Tian, H., et al., *Bioorthogonal Fluorescent Labeling of Functional G-Protein-Coupled Receptors.* Chembiochem, 2014. **15**(12): p. 1820-1829.

151. Yang, H., et al., *A General Method for Artificial Metalloenzyme Formation through Strain-Promoted Azide-Alkyne Cycloaddition.* Chembiochem, 2014. **15**(2): p. 223-227.

152. Amir, L., et al., *Surface Display of a Redox Enzyme and its Site-Specific Wiring to Gold Electrodes.* Journal of the American Chemical Society, 2013. **135**(1): p. 70-73.

153. Fu, J.L., et al., *Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures.* Journal of the American Chemical Society, 2012. **134**(12): p. 5516-5519.

154. Zhang, W.H., G. Otting, and C.J. Jackson, *Protein engineering with unnatural amino acids.* Current Opinion in Structural Biology, 2013. **23**(4): p. 581-587.

155. Heim, R., D.C. Prasher, and R.Y. Tsien, *Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(26): p. 12501-12504.

156. van Berkel, S.S., et al., *Metal-free triazole formation as a tool for bioconjugation.* Chembiochem, 2007. **8**(13): p. 1504-1508.

157. Dosremedios, C.G. and P.D.J. Moens, *Fluorescence Resonance Energy-Transfer Spectroscopy Is a Reliable Ruler for Measuring Structural-Changes in Proteins - Dispelling the Problem of the Unknown Orientation Factor.* Journal of Structural Biology, 1995. **115**(2): p. 175-185.

158. Zacharias, D.A. and R.Y. Tsien, *Molecular biology and mutation of green fluorescent protein.* Methods Biochem Anal, 2006. **47**: p. 83-120.

159. Petrosino, J., C. Cantu, and T. Palzkill, *beta-Lactamases: Protein evolution in real time.* Trends in Microbiology, 1998. **6**(8): p. 323-327.

160. Farrelli, I.S., et al., *Photo-cross-linking interacting proteins with a genetically encoded benzophenone.* Nature Methods, 2005. **2**(5): p. 377-384.

161. Hohsaka, T., et al., *Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system.* Biochemistry, 2001. **40**(37): p. 11060-11064.

162. Nadarajan, S.P., et al., *An in silico approach to evaluate the polyspecificity of methionyl-tRNA synthetases.* Journal of Molecular Graphics & Modelling, 2013. **39**: p. 79-86.

163. Schrader, J.M., S.J. Chapman, and O.C. Uhlenbeck, *Understanding the Sequence Specificity of tRNA Binding to Elongation Factor Tu using tRNA Mutagenesis.* Journal of Molecular Biology, 2009. **386**(5): p. 1255-1264.

164. Mittelstaet, J., A.L. Konevega, and M.V. Rodnina, *A Kinetic Safety Gate Controlling the Delivery of Unnatural Amino Acids to the Ribosome.* Journal of the American Chemical Society, 2013. **135**(45): p. 17031-17038.

165. Krishnamurthy, M., et al., *Caught in the Act: Covalent Cross-Linking Captures Activator-Coactivator Interactions in Vivo.* Acs Chemical Biology, 2011. **6**(12): p. 1321-1326.

166. Huisgen, R., *Kinetics and Reaction-Mechanisms - Selected Examples from the Experience of 40 Years.* Pure and Applied Chemistry, 1989. **61**(4): p. 613-628.

167. Kolb, H.C. and K.B. Sharpless, *The growing impact of click chemistry on drug discovery.* Drug Discovery Today, 2003. **8**(24): p. 1128-1137.

168. Tron, G.C., et al., *Click chemistry reactions in medicinal chemistry: Applications of the 1,3-dipolar cycloaddition between azides and alkynes.* Medicinal Research Reviews, 2008. **28**(2): p. 278-308.

169. Krasinski, A., et al., *In situ selection of lead compounds by click chemistry: Target-guided optimization of acetylcholinesterase inhibitors.* Journal of the American Chemical Society, 2005. **127**(18): p. 6686-6692.

170. Manetsch, R., et al., *In situ click chemistry: Enzyme inhibitors made to their own specifications.* Journal of the American Chemical Society, 2004. **126**(40): p. 12809-12818.

171. Link, A.J. and D.A. Tirrell, *Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition.* Journal of the American Chemical Society, 2003. **125**(37): p. 11164-11165.

172. Deiters, A., et al., *Site-specific PEGylation of proteins containing unnatural amino acids.* Bioorganic & Medicinal Chemistry Letters, 2004. **14**(23): p. 5743-5745.

173. Agard, N.J., J.A. Prescher, and C.R. Bertozzi, *A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems.* Journal of the American Chemical Society, 2004. **126**(46): p. 15046-15047.

174. Baskin, J.M., et al., *Copper-free click chemistry for dynamic in vivo imaging.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(43): p. 16793-16797.

175. Ning, X.H., et al., *Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions.* Angewandte Chemie-International Edition, 2008. **47**(12): p. 2253-2255.

176. Star, A., et al., *Electronic detection of specific protein binding using nanotube FET devices.* Nano Letters, 2003. **3**(4): p. 459-463.

177. Besteman, K., et al., *Enzyme-coated carbon nanotubes as single-molecule biosensors.* Nano Letters, 2003. **3**(6): p. 727-730.

178. Patolsky, F., Y. Weizmann, and I. Willner, *Long-range electrical contacting of redox enzymes by SWCNT connectors.* Angewandte Chemie-International Edition, 2004. **43**(16): p. 2113-2117.

179. Liu, L.H., M.M. Lerner, and M.D. Yan, *Derivitization of Pristine Graphene with Well-Defined Chemical Functionalities.* Nano Letters, 2010. **10**(9): p. 3754-3756.

180. Park, J. and M.D. Yan, *Covalent Functionalization of Graphene with Reactive Intermediates.* Accounts of Chemical Research, 2013. **46**(1): p. 181-189.

181. Seeman, N.C., *Nanomaterials Based on DNA*. Annual Review of Biochemistry, Vol 79, 2010. **79**: p. 65-87.

182. Mills, J.H., et al., *Computational Design of an Unnatural Amino Acid Dependent Metalloprotein with Atomic Level Accuracy.* Journal of the American Chemical Society, 2013. **135**(36): p. 13393-13399.