Structure, function and engineering of biotechnologically important proteins

Husam Sabah Auhim AL-Maslookhi

A dissertation submitted to Cardiff University in candidature for the degree of Doctor of Philosophy

September 2019

Molecular Biosciences Division

School of Biosciences

Cardiff University

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of **PhD**

Signed _____

Date _____

STATEMENT 2

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 it being submitted concurrently for any other degree or award (outside of any formal collaboration agreement between the University and a partner organisation)

Signed _____

Date _____

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available in the University's Open Access repository (or, where approved, to be available in the University's library and for inter-library loan), and for the title and summary to be made available to outside organisations, subject to the expiry of a University-approved bar on access if applicable.

Signed _____

Date _____

DECLARATION

This thesis is the result of my own independent work, except where otherwise stated, and the views expressed are my own. Other sources are acknowledged by explicit references. The thesis has not been edited by a third party beyond what is permitted by Cardiff University's Use of Third Party Editors by Research Degree Students Procedure.

Signed _____

Date _____

WORD COUNT 63773

(Excluding summary, acknowledgements, declarations, contents pages, appendices, tables, diagrams and figures, references, bibliography, footnotes and endnotes)

Abstract

A knowledge of the molecular structure of proteins is a key part of understanding protein function and rational protein engineering. The insecticidal toxin Vip1Ac1 (vegetative insecticidal protein 1) is a pore-forming toxin, which has potential to act as a biopesticide. First, its molecular structure and function needs to be understood, and how this might contribute to specificity for certain insects. The molecular structure of the monomer Vip1Ac1 (~82 kDa) at 1.47 Å resolution has been solved by X-ray diffraction. The overall fold of Vip1Ac1 is similar to the anthrax protective antigen (PA), which is a receptor binding/translocation component of anthrax toxin. The high similarity amongst the first 3 domains of both Vip1Ac1 and PA in terms of folding, Ca^{+2} binding residues important motifs and functions, indicates that they have the same mode of action, while the differences in domain 4 (receptor-binding domain) explain the specificity of each toxin towards different targets. Activation of Vip1Ac1 by trypsin led to removal of an ~22 kDa N-terminal fragment ($Vip1Ac1_{22}$) and self-association of the ~60 kDa C-terminal fragment ($Vip1Ac1_{60}$) into a ring-shaped oligomer as observed in electron microscopy. To probe complex formation further, a reactive azide group at specific surface residues (Y358 and Y536) was introduced by codon reprogramming and allowed labelling of Vip1Ac1 with fluorescent dye via Click chemistry. This was used to monitor assembly on artificial membranes. The oligomeric state showed binding and the ability to form pores in these membranes.

The restricted functional groups in natural amino acids restricts standard protein engineering. Introduction of new functional groups by incorporation of non-canonical amino acids (ncAAs) expands the protein engineering diversity. Variants of the Venus fluorescent protein with new fluorescent properties (switched off or on), and a heterodimer (sfGFP+Venus) with new spectra properties differ than each monomeric units were generated by incorporation of ncAAs with new functional groups (azide and alkyne) that allow photoactivation and Click chemistry reactions for dimerisation.

Contents

Abbreviations

A standard one letter amino acid code followed by the number of residue has been used to describe amino acids throughout the thesis. Single mutation is denoted as follows: the code of wild type amino acid – number of amino acid – mutant amino acid, in superscript after the mutant protein. For instance, substitution of Histidine (residue number 148) with non-canonical amino acids azF in Venus fluorescent protein has been described as Venus ^{H148azF}.

Abbreviations used throughout the thesis:

λ_{em}	Wavelength of fluorescence emission
λ_{ex}	Wavelength of fluorescence excitation
λ_{max}	Wavelength of maximum absorbance
ε	Molar absorbance
aaRS	Aminoacyl-tRNA-synthetase
azF	*p*-azido-L-phenylalanine
bp	Base pair
°C	Degrees Celsius
CRO	Chromophore
C-terminus	Carboxyl-terminus
CuAAC	Copper-catalysed azide-alkyne cycloaddition
Cyt b_{562}	Cytochrome b_{562}
DBCO-TAMRA	Dibenzocyclooctyne tetramethyl rhodamine fluorescent dye
DMSO	Dimethyl sulfoxide
dNTPs	Mixture of the four deoxyribonuctleotide triphosphates
DTT	Dithiothreitol
Da	Dalton
EDTA	Ethylenediaminetetraacetic acid
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
GST	Glutathione S-transferases
h	Hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Poly Histidine tag
IEX	Ion exchange chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
MR	Molecular replacement
MW	Molecular weight
ncAAs	Non-canonical amino acids

N-terminus	Amine-terminus
OD_{xxx}	Optical density (~absorbance) at xxx nm
PA	Anthrax protective antigen
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein databank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
QY	Quantum yield
RMSD	Root-mean-square deviation
SCO-K	S-cyclooctyne-L-lysine
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SD	Standard deviation
SEC	Size exclusion chromatography
SeMet	Selenomethionine
sfGFP	Superfolder green fluorescent protein
SPAAC	Strain-promoted azide-alkyne cycloaddition
TAE	Tris, acetate, EDTA
TEM	Transmission electron microscopy
TEMED	Tetramethylethylenediamine
TEV	Tobacco etch virus
Tris	Tris-hydroxymethyl-aminomethane
tRNA	Transfer-RNA
v/v	Volume per volume
Vip	Vegetative insecticidal protein
w/v	Weight per volume
(WT) wt	Wild type

Acknowledgment

First of all, I would like to thank my country (Iraq) especially the higher committee for education development in Iraq (HCED) for giving me this great opportunity to study my PhD in Great Britain and for financial support during my study journey. Also, I would like to thank Cardiff University for the provision of research facilities.

I wish to acknowledge my academic supervisors Dr. Dafydd Jones and Prof. Colin Berry, first for accepting me as a student in their research group and for all their help, guidance, advice and support throughout the course of my research. I was lucky enough to be one of your research members, I have learnt a lot from you in terms of the scientific field during the last four years and igniting a passion for protein engineering technologies that will open avenues for my scientific future. I am especially thankful for your time and effort put into proof-reading and improving my thesis. I am much obliged to you, working with both of you was a great experience.

Special thanks for Daniel who helped me a lot during the first year in my lab work. I would also like to thank all our lab members who helped me in big or small ways, thanks to Harley, Ben, Rachel, Ismail and others who work in the DDJ and CB lab for their assistance during my research. A special thanks must, therefore, be made to Jacob who helped me with liposome preparation and cell live imaging and to Dr. Mark Young who helped me with TEM grid preparation and imaging.

I would like to thank Dr. Pierre Rizkallah (School of Medicine, Cardiff University) for helping me with; setting up crystals trials, solving the crystal structures and opening my eyes to understand the principles of protein crystallography. Thanks for your long teaching session and your trip to Diamond for data collection. Also, I would like to thank Dr. Neil Paterson (Diamond Light Source, Harwell Science and Innovation Campus, Didcot, UK) for data collection and analysis of Vip1Ac1 SeMet containing-crystals. I wish to thank Dr. Alan Scott (School of Chemistry, Cardiff University) for his great help with setting up crystals trials for Venus variants using the automated liquid handling systems Oryx.

The work that went into this thesis would not have been possible without the love and support of my wife, a big thanks to the to the love of my life Nisreen who left her family, job and friends, and came with me and stayed alone at home during my long lab work days. Thanks for your sacrifices.

Finally, a big thank to my family that suffered my long absence for at least 5 years. Thanks for your constant support during my PhD journey in the UK, your weekly chats keeping me going on to achieve my aim.

Chapter One

Introduction

1. Introduction

1.1 Protein structure and function relationships

Proteins are considered amongst the most important molecules in biology. They are involved in all biological process ranging from catalysis to energy transduction to communication. Each protein consists of a specific sequence of 20 canonical amino acids; in some rare cases selenocysteine and pyrrolysine, also known as the 21st and 22nd naturally occurring amino acids respectively, can be incorporated (Zhang *et al.* 2005). Protein production occurs via the central dogma in molecular biology. The information determining amino acid sequences is encoded by DNA, which is transcribed into mRNA then translated into a protein. This sequence gives the protein its specific structure, which in most cases defines functions.

Protein folding allows the amino acids from different regions in the primary structure to interact and gives the protein the final functional conformation. For instance, interactions between certain amino acids generate the active site while additional residues interact to form a substrate binding pocket; both combined in an enzyme to catalyse substrate turnover. Changes in the sequence of amino acids involved in the interactions and generation of the active site could lead to loss or change of activity. So, the protein's functions depend directly on the three-dimensional structure, which is determined by its amino acid sequence.

A knowledge of the three-dimensional structure of proteins is a key part of understanding protein function, dynamics, and interactions with other molecules. Still the most common approach to experimentally determine protein structure is using X-ray crystallography techniques. Predicted structures via molecular modelling using a known experimental structure of a homologous protein (the template) is also a popular approach if no structure of the target protein is known. Predicted structures can be very useful tools for probing the protein structure-function relationship but sometimes miss the important details gained from determining the actual structure.

1.2 Protein engineering

The history of protein engineering started from 1978 when Hutchison and his colleagues changed selected bases in a target DNA sequence (Hutchison *et al.* 1978).

Their work introduced the idea of site directed mutagenesis whereby point mutations to change one residue to another. The first protein subjected to protein engineering was β-lactamase in 1982. This is an enzyme produced by bacteria and provides resistance to the β-lactam antibiotics. The active site serine residue was replaced with a cysteine residue to study the enzyme activity (Sigal *et al.* 1982). From this date to the present, protein engineering has played a vital role in advancing molecular biosciences and many proteins have been engineered for different purposes including improving proteins for medical and pharmaceutical applications (Nagaoka *et al.* 2010), food applications (Kapoor *et al*. 2017) and industrial applications (Yasuda *et al.* 2018).

Protein engineering is a powerful tool to study proteins as well as to obtain improved proteins with desirable properties via recombinant DNA technology. Protein engineering takes advantage of the protein sequence-structure-function relationship to improve and/or study their properties. This involves modifying the DNA sequences that encode the protein using targeted mutagenesis methods such as site-directed mutagenesis (SDM) (Smith 1982). For example, amino acids can be substituted, inserted, deleted or even replaced with non-natural amino acids in the context of existing proteins. One powerful technique is to fuse two or more proteins to form hybrid proteins. Such fusions have been used in roles from cell imaging to improving recombinant protein production (Nagaoka *et al.* 2010).

Two main methods have emerged to change proteins to work under different conditions or create proteins with novel functions. The first method is rational design using SDM techniques. This is an effective approach when the mechanism and structure of the protein are well-known. This method allows for deletion /insertion/ replacement of specific amino acids in a target protein (Antikainen and Martin 2005). The second approach, known as directed evolution, is a time-consuming process and uses evolutionary methods that involve random mutagenesis and selection. A central facet is the use of recombinant DNA technology to produce a library of possible variants, then the use of high-throughput screening techniques to search quickly for the variant that offers the best or desirable properties. This approach is an alternative to rational design, especially when the structure and mechanism of the protein are obscure (Nagamune 2017).

Developments in computational techniques and structural genomics, have dramatically increased the information that is accessible, related to sequence, structure, dynamics and evolution of proteins and this has increased our ability to plan targeted protein engineering. This development is expected to increase the applications of protein engineering in the future (Chen and Zeng 2016).

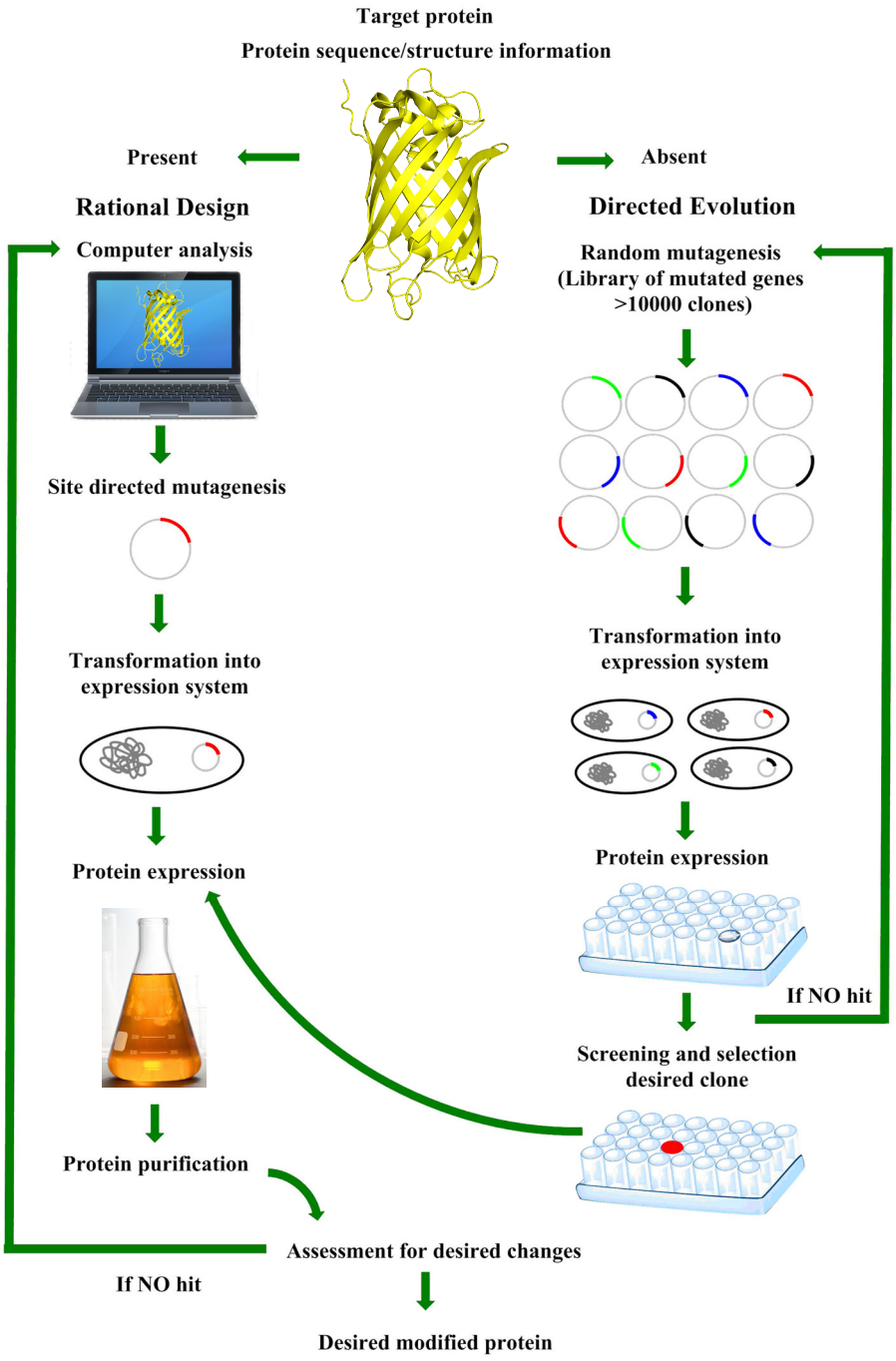

Figure 1.1. Two general strategies for protein engineering.

In this study, rational design has been applied in two separate but linked projects. The first is regarding the structure, function and engineering of the insecticidal toxin, Vip1 (vegetative insecticidal protein 1). The second project relates to the generation of novel auto fluorescent protein variants, including artificial dimers with new useful properties.

1.3 Insecticidal toxins

Insects are the most varied group of multicellular living organisms on the earth, providing multiple ecosystem services such as pollination, decomposition, energy transfer through the food chain and pest control (biocontrol) (Jankielsohn 2018). However, insects are considered one of the greatest agricultural pests around the world and cause severe damage in crops during their life cycle. These pests have a short generation times and a large number of offspring; therefore, they are widely spread in terms of geography, species type and food source. In addition to agricultural pests, some insect species, such as mosquitoes, act as vectors of dangerous human diseases such as lymphatic filariasis, malaria and arboviruses, which include yellow fever, Zika virus, dengue virus and others (Benelli *et al.* 2016). To control the spread of harmful insects and reduce their negative effect, different approaches have been implemented. These methods are based on the use of specific agents (chemical or biological) which cause death or incapacitation of insects. Chemical pesticides have been widely used in the past to control insects, but these insecticides can have low specificity and have a negative impact on the environment and other non-target organisms (Zacharia 2011). Owing to their negative impact, chemicals are increasingly being replaced with biological insecticides (Chattopadhyay *et al.* 2004).

Biological insecticides are environmentally friendly and more specific towards small groups of insects. These insecticides are based on the use of either pathogenic agents or toxins, which cause specific diseases and subsequently death or inactivation of the target insect (Ferré and Escriche 2017). The narrow host ranges of the pathogenic agents (that include viruses, bacteria, fungi and protozoa) make them good candidates as insecticides (Payne 1988). Pesticidal toxins have been a particular focus of attention. Owing to their specificity, each toxin can be used to control a narrow range of insect species or a specific stage of the life cycle. Toxins can also be modified by protein engineering techniques to enhance their toxicity, stability and spectrum of action or to change the specificity from one target to another (Torres-Quintero *et al.* 2018). Finally, the ability of toxins to be

expressed successfully in different economically important plants, allows them to be incorporated into insect-resistant transgenic variants (Arora and Sandhu 2017).

Toxins used in pest control are usually bacterial toxins, with 90% of the commercial toxins derived from *Bacillus thuringiensis* (Bt toxins) (Chattopadhyay *et al.* 2004) and some toxins from other bacteria such as *Xenorhabdus*, *Photorhabdus* (ffrench-Constant *et al.*2007), *Serratia entomophila* and *Lysinibacillus sphaericus* (Pardo-López *et al.* 2013).

1.3.1 Bt toxins

Bacillus thuringiensis (commonly known as Bt) is a typical gram-positive, rod-shaped, spore-forming bacterium. *Bacillus* species can be isolated from a variety of ecosystems (e.g. soils, water, leaves, dead insects, mammals that feed on insects and infected human tissues with severe necrosis (Roh *et al.* 2007 ; Palma *et al.* 2017)). Bt was initially discovered in 1902 from diseased silkworm (*Bombyx mori*) by Shigetane Ishiwatari and in 1915 was characterized from diseased flour moth caterpillars (*Ephestia kuhniella*) by Ernst Berliner (Jisha *et al.* 2013). Bt strains can synthesize various classes of proteins that target invertebrates, mainly insects. Because of the pathogenicity to insects, they have been used for several decades as biological pesticides (Schnepf *et al.* 1998). These insecticidal proteins do not affect humans, vertebrates or plants, but are highly specific to their target insect, and are completely biodegradable (Jisha *et al.* 2013).

1.3.2 Sporulation associated toxins

Bt strains produce parasporal crystals containing crystal (Cry) and cytolytic (Cyt) toxins, (also known as collectively δ-endotoxins), during the stationary growth phase and sporulation, due to the activation of the genes which encode Cry/Cyt proteins. The initiation of transcription of *Cry/Cyt* genes is usually due to sporulation-specific sigma factors that are synthesized specifically during this phase (Tounsi and Jaoua 2002; Sanahuja *et al.* 2011). The Cry proteins comprise the biggest group of insecticidal proteins synthesized by species of *Bacillus*. To date, the Bt Toxin Nomenclature Committee (Crickmore *et al.* 2019) has classified 192 variant families of Cry proteins and 9 variant families of Cyt proteins.

δ-endotoxins have a wide range toxic activity. They are active against some insect species belonging to the orders (Diptera, Lepidoptera, Coleoptera, Hemiptera and Hymenoptera), nematodes and some snails (Palma *et al.* 2014). In general, the mode of action of δ-endotoxins can be summarised in four stages after ingesting the toxins by a target insect: 1) solubilisation of toxin (crystals); 2) proteolytic activation; 3) binding to the specific receptors; 4) cell disruption (Schünemann *et al.* 2014).

1.3.3 Insecticidal proteins produced during vegetative growth

Although Cry toxins have been used successfully as insecticides, some insect species, including pest species belonging to the orders Lepidoptera and Coleoptera, have been found to be highly resistant to Cry toxins. Screening for new toxins to control these insects, led to the discovery of secreted insecticidal proteins produced during the vegetative growth phase, which are known as the vegetative insecticidal proteins (Vip)(Estruch *et al.* 1996) and the secreted insecticidal protein (Sip) (Donovan *et al.* 2006).

1.3.4 Nomenclature of Vip toxins

A system of nomenclature was created by the *Bacillus thuringiensis* Toxin Nomenclature Committee for classification of Bt toxins (Crickmore *et al.* 1998) and is used for all Bt toxins (Cry, Cyt, Vip and Sip proteins) while revisions to the system are underway for the future. According to this system, each toxin is given a four-rank name depending on the percentage amino acid identity with other toxins (Crickmore *et al.* 1998 ; Palma *et al.* 2014), using numbers for the first and fourth ranks, and uppercase and lowercase letters, respectively, for the second and third ranks (Figure 1.2). In this system, Vip (from Vegetative Insecticidal Protein) toxins that have less than 45% amino acid sequence identity are given different numbers in the first rank (e.g. Vip1, Vip2 and Vip3); toxins having less than 78% identity are given different uppercase letters in the second rank (e.g. Vip1A, Vip1B and Vip1C); toxins having less than 95% identity are given different lowercase letters in the tertiary rank (e.g. Vip1Aa, Vip1Ab and Vip1Ac); and, finally, to distinguish between toxins showing more than 95% identity, a quaternary rank denoted by a number is given to each separate submission to the database (e.g. Vip1Aa1, Vip1Aa2 and Vip1Aa3) whether the new entry is different from previous entries or not.

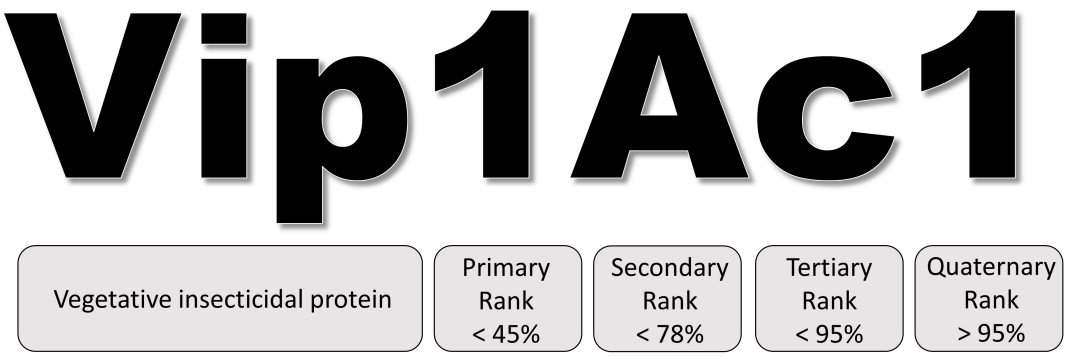

Figure 1.2. Schematic diagram of the nomenclature system for Vip toxins used by Bt Toxin Nomenclature Committee.

1.3.5 Vegetative insecticidal proteins (Vips)

The Vip proteins, despite sharing the name Vip, belong to different structural families. Vip proteins are classified into classes, Vip1, Vip2, Vip3 and the more recent Vip4 on the basis of the degree of amino acid identity (Figure 1.3) (Crickmore *et al.* 2019).

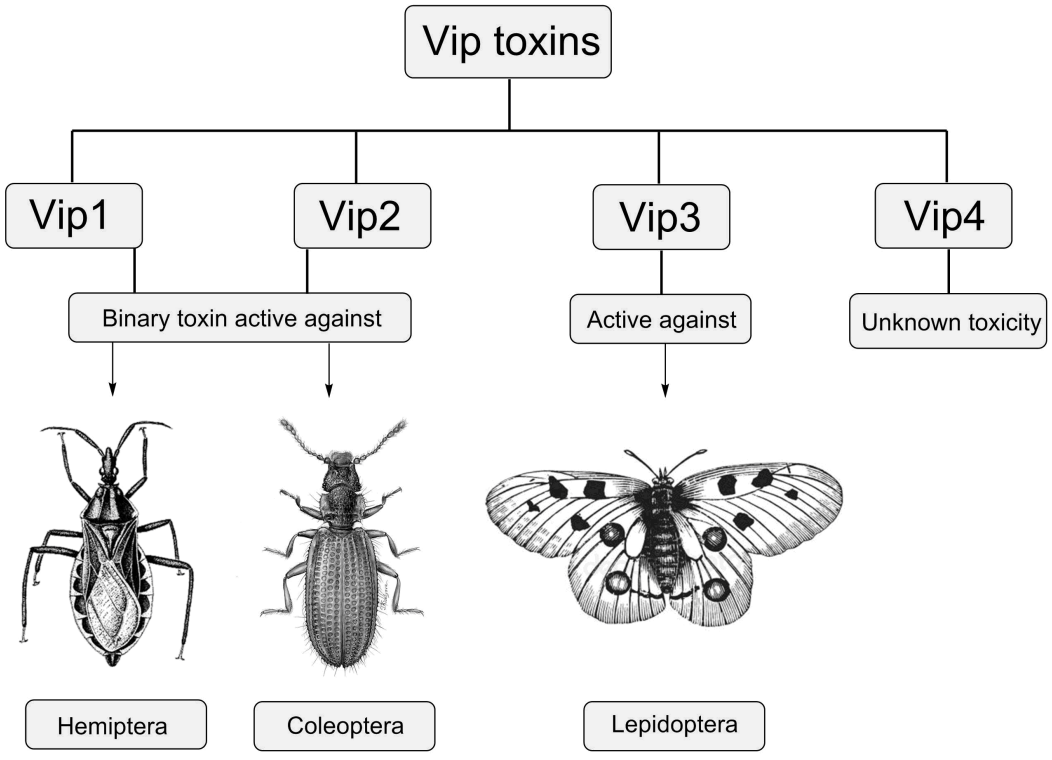

Figure 1.3. Diagram showing the distribution and known host spectrum of different Vip toxins.

1.3.5.1 Vip3

Estruch *et al*. in 1996, described a novel vegetative insecticidal protein, Vip3Aa, from *B. thuringiensis* strain AB88, with a toxic activity against a wide range of lepidopteran larvae including corn ear-worm (*Helicoverpa zea*), fall army-worm (*Spodoptera frugiperda*), black cutworm (*Agrotis ipsilon*), tobacco budworm (*Heliothis virescens*) and beet armyworm (*Spodoptera exigua*). In the same study, they reported another, related, vegetative insecticidal protein Vip3Ab from *B. thuringiensis* strain AB424. Warren (1997) reported another secreted toxin from *B. thuringiensis* strain AB88, which has a toxic effect against *Agrotis ipsilon* (Lepidoptera) and other lepidopteran larvae. This toxin has a molecular weight about 88.5 kDa and was named Vip3 (Warren 1997). These insecticidal proteins are secreted during vegetative growth and have no significant sequence similarity with the Cry proteins that are produced during the sporulation stage.

Vip3 proteins have been used to produce transgenic plants because they have a broad range of toxic activity against economically important lepidopteran pests, eg *Lobesia botrana, Chrysodeixis chalcites, Spodoptera littoralis, Mamestra brassicae* (Palma *et al.*2013). The detailed structure of Vip3 has not yet been reported but a surface topology for the tetrametic form that exists in solution has been produced (Palma *et al.* 2017). The mode of action of Vip3 has been described by Lee *et al.* (2003). Briefly, the Vip3 pro-toxin with a molecular weight of 88 kDa is processed proteolytically into a 62 kDa fragment in the midgut of lepidopteran insects, leading to subsequent lysis of epithelial tissue and the death of the target insect. Interactions with specific receptors present on mid-gut brush border membrane vesicles (BBMV) to form a pore has been demonstrated (Lee *et al.* 2003).

1.3.5.2 Vip1-Vip2 as a binary toxin

In 1997, Warren described two new vegetative toxins from *Bacillus cereus* AB78 that together caused 100% mortality of northern corn rootworm *Diabrotica longicornis barberi* larvae (Coleoptera) and western corn rootworm *Diabrotica virgifera virgifera* (Coleoptera). The two toxins, Vip1 and Vip2, have molecular weights of *circa* 80 and 45 kDa respectively (Warren 1997). It has been found that several isolates of *B. thuringiensis* (Warren 1997), *B. cereus* (Yu *et al.* 2011) and *Brevibacillus laterosporus*

(Schnepf *et al.* 2005; Ruiu 2013) secreted Vip1, Vip2 and Vip3-like vegetative insecticidal proteins during vegetative growth.

1.3.5.3 Mode of action of Vip1-Vip2 binary toxin

Vip1 and Vip2 as individual proteins have shown no toxicity in tests against insects from the orders Coleoptera and Hemiptera. However, Vip1-Vip2 proteins together appear to form a classical binary toxin (AB toxin) and display insecticidal activity against some insects such as northern corn rootworm (*Diabrotica barberi*), western corn rootworm (*Diabrotica virgifera virgifera*) (Warren 1997) and cotton aphid (*Aphis gossypii*) (Shingote *et al.* 2013) but the individual proteins are not toxic to *A. gossypii* (Sattar and Maiti, 2011). The mode of action of Vip1-Vip2 toxins is not fully understood. After ingestion of the toxins by susceptible insects, proteolytic activation of Vip1 occurs in midgut by trypsin-like proteases (Chakroun *et al.* 2016). The activated Vip1 protein has been shown to assemble and create oligomers consisting of seven Vip1 molecules (Leuber *et al.* 2006). In the midgut brush border membrane, the activated toxins are thought to recognize specific receptors and bind either in monomeric or oligomeric form, then allow Vip2 to enter the cytosol via two proposed mechanisms; directly through the channel formed by Vip1 oligomers or by endocytosis of the Vip2-Vip1 oligomer complex as illustrated in Figure 1.4 (Chakroun *et al.* 2016). Warren 1997 reported that Vip1 is important in receptor recognition, finding that replacing Vip1Ab with Vip1Aa led to a loss of toxicity against the Western corn rootworm (*Diabrotica virgifera virgifera*).

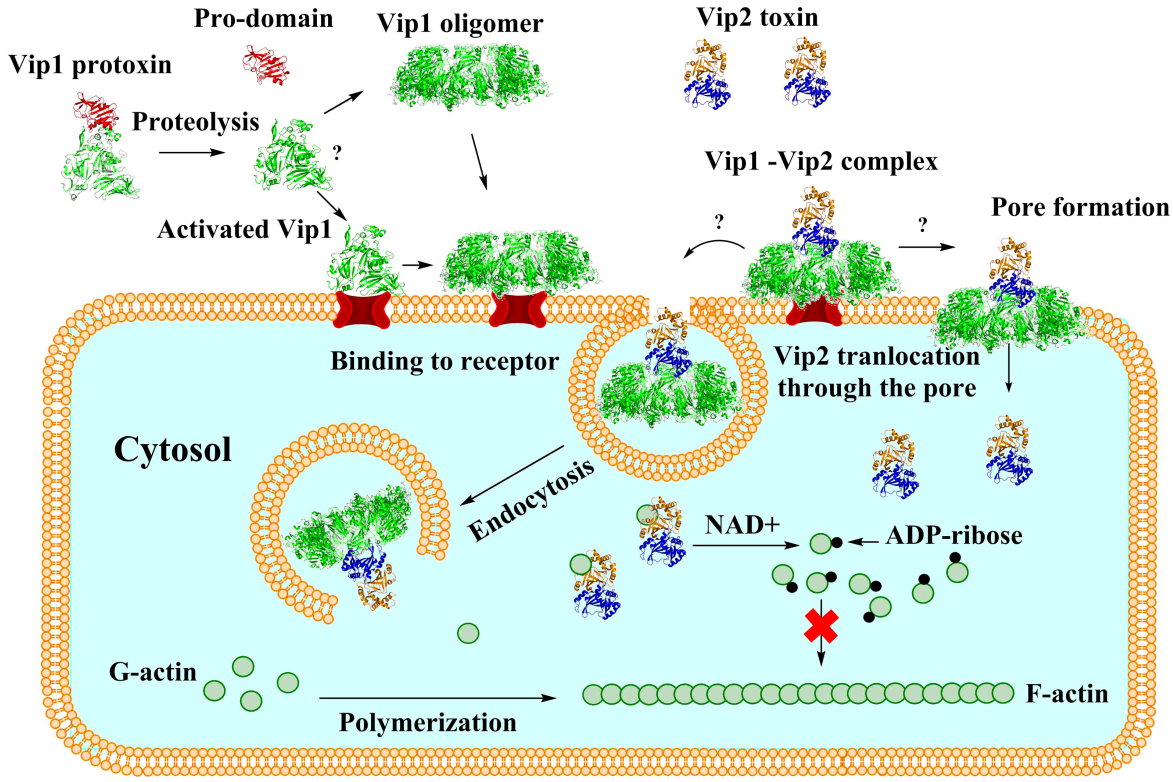

Figure 1.4. Proposed mode of action of Vip1-Vip2 binary toxins.

Several gram-positive, spore-forming bacteria produce synergistic binary toxins, which are called AB toxins. These include *Clostridium spiroforme* toxin CST (CSTa and CSTb; Papatheodorou *et al.* 2012), *Clostridium botulinum* C2 toxin (C2I and C2II; Barth and Aktories 2011), *Clostridium difficile* toxin CDT (CDTa and CDTb; Gerding *et al.* 2013). Component B of these toxins is the receptor binding subunit while component A is the enzymatic subunit (an ADP-ribosyltransferase in the examples above). *Bacillus anthracis* toxins act via protective antigen (PA), which is the B receptor binding subunit and two A subunits; Lethal factor (LF, a Zn metalloproteinase) and Edema factor (EF, an adenylate cyclase). Component B is responsible for toxin specificity and is activated after proteolysis to generate oligomers, these oligomers then form multimeric-receptor complexes and act as docking platforms that subsequently translocate component A into the cytosol (Barth *et al.* 2004). The A components can inhibit cell function in different ways, for instance: mono-ADP-ribosylation of G-actin, which leads to cell death after induction of cytoskeletal disarray (Al-Mohanna *et al.* 1987); inhibiting cell signalling by proteolysis of mitogen-activated protein kinase kinases (MAPKK ; Chopra *et al.* 2003)

as in the case of LF; or forming edema and immunosuppression due to increasing intracellular levels of cyclic AMP (cAMP) as seen for EF (Leppla 1982).

Binary toxins are important for vaccine production for diseases like anthrax because they are virulence factors. They are also useful biological tools for studying a number of cellular functions and delivering heterologous proteins into endosomal, as well as cytosolic, compartments (Barth *et al.* 2004).

1.3.5.4 Vip2

Vip2 is a NAD-dependent ADP-ribosyltransferase with a molecular weight of ~50 kDa. The introduction to many papers on this protein report that it inhibits the functions of the target cells due to modifying the monomeric G-actin at R177 residue and disrupting the integrity of the cytoskeleton by blocking the polymerisation of F-actin but the source references for this assertion (most commonly Warren 1997) makes no reference to this and it is unclear whether there is an evidential basis for this activity or whether it has been proposed based on homology to other ADP-ribosyl transferases. Vip2 has amino acid sequence similarity with the enzymatic components (A) of a number of binary toxins including: *Clostridium spiroforme toxin (CST) CSTa*, *Clostridium botulinum* C2 toxin, *Clostridium difficile* toxin (Han *et al.*1999) and domain 2 of the *Bacillus anthracis* toxin lethal factor (Pannifer *et al.* 2001). The structure of the Vip2 protein (PDB code: 1QS1) and Vip2-NAD complex (PDB code: 1QS2) have been reported by Han *et al.* 1999.

In general, the Vip2 toxin has two nonsymmetrical domains (Figure 1.5A), the N-domain (residues 60-265) and C-domain (266-461). Despite the two domains sharing only 22% sequence identity, they are structurally homologous. The C-domain is the enzymatic domain, which has the cleft for nicotinamide adenine dinucleotide (NAD) binding (Figure 1.5B). This cleft is composed of 6 residues C351, M385, L389, R393, R400 and the catalytic residue E428. Since the C-domain is the enzymatic part of the molecule, the role of the N-domain is not clear, but it might, for instance, play a role in the interactions of this toxin with the receptor binding component B (Vip1) during the translocation of Vip2 into the cytosol of the target cell.

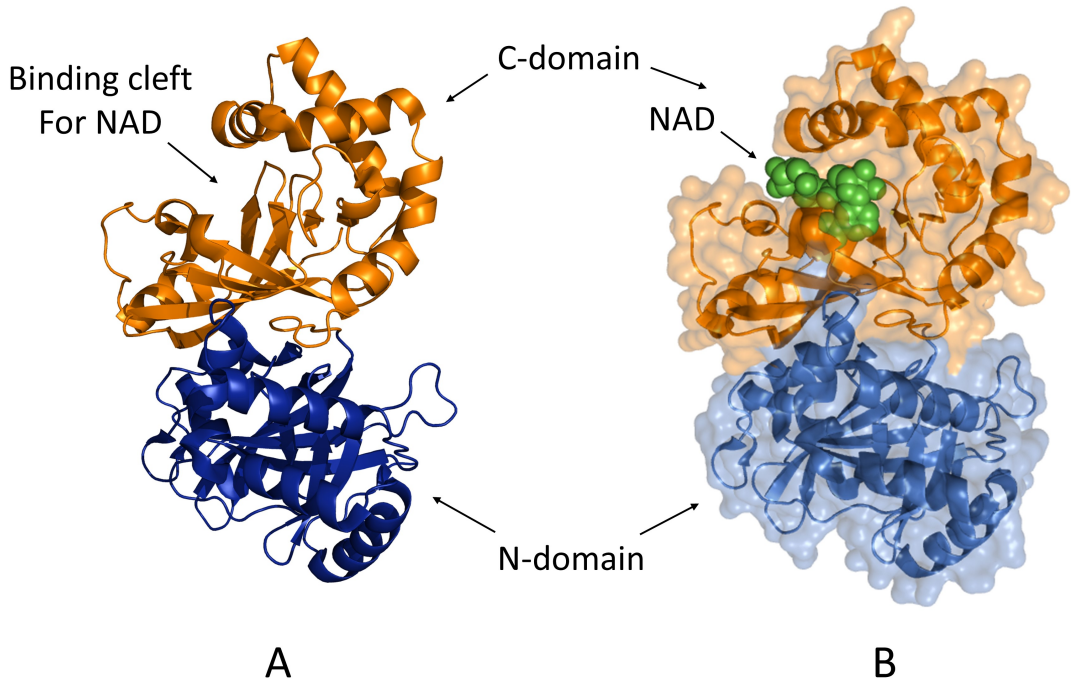

Figure 1.5. Molecular structure of Vip2 toxin.
A. Overall fold, the N-domain coloured blue and the C-domain coloured orange. **B**. Surface structure of Vip2 showing the NAD molecules (green spheres) binding to the cleft in C-domain (PDB code;1QS2)

1.3.5.5 Vip1

Vip1 proteins have a molecular weight of ~95 kDa including a putative N-terminal signal peptide between 31 amino acids (Bi *et al.* 2014) and 49 amino acids (Shi *et al.* 2004) in length. Vip1 proteins have amino acid sequences similar to binding components (B components) of mammalian-active binary toxins: *Bacillus anthracis* toxin protective antigen (PA) (38% identity), *Clostridium botulinum* C2 toxin C2II and *Clostridium spiroforme* toxin CSIb (29% and 33% identity respectively) (Chakroun *et al.*2016). To date, there are 15 individual Vip1 sequences (Vip1Aa1, Vip1Aa2, Vip1Aa3, Vip1Ab1, Vip1Ac1, Vip1Ad1, Vip1Ba1, Vip1Ba2, Vip1Bb1, Vip1Bb2, Vip1Bb3, Vip1Bc1, Vip1Ca1, Vip1Ca2 and Vip1Da1) listed on the Bt Toxin Nomenclature Committee website. The temporary suspension of the nomenclature system for naming of new variants of known toxin classes means that other Vip (and other) toxins are known but not currently listed (eg a Vip1Bb variant isolated by Sellami *et al.* which they named Vip1S (Sellami *et al*. 2016).

The Vip1 protein (Vip1Ac known now as Vip1Ca1) has been expressed as a recombinant protein using the pET28b vector and *E. coli* (BL21(DE3)) as an expression system but most of the protein formed inclusion bodies (Shi *et al.* 2004). The solubility of recombinant Vip1 protein has not been addressed. A model of the structure of Vip1 protein (Vip1S) without any validation parameters was predicted by Sellami *et al.* via the Phyre2 server using a low resolution (2.1 Å) structure of protective antigen (PA) as a template (PDB code:1ACC) (Figure 1.6) with 35% sequence identity (Sellami *et al.* 2016). This predicted structure indicated the potential similarity between Vip1 and PA in terms of folding.

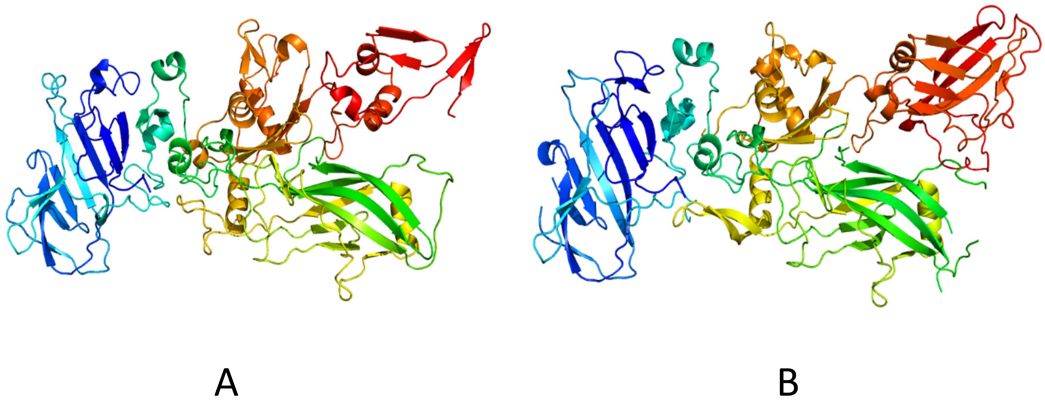

Figure 1.6. Comparison of Vip1S model with PA structure.
A: Phyre2 model of Vip1S as reported by Sellami *et al*. 2016 (adapted from Sellami *et al*. 2016).
B: Structure of PA used as a template to generate the Vip1S model (PDB code:1ACC).

The similarity of Vip1 proteins with other binding components of binary toxins suggests that the mode of action of Vip1 proteins could be similar to those proteins. The putative channel-forming domain in binding components of binary toxins contains two β-strands (of different lengths), connected together via a flexible loop forming β-hairpin of alternating hydrophilic residues facing the lumen of the channel, and hydrophobic residues facing the lipid (Leuber *et al.*2006). In PA, the putative channel-forming region is composed of 24 resides (H328-S354) (Benson *et.al.* 1998); in CSIb, 26 resides (N334-T359); in C2II, 32 resides (N302-T333); and in Vip1Ac (Vip1Ca1) the equivalent region consists of 25 residues (N338-T364) (Leuber *et al.* 2006) as shown in (Figure 1.7).

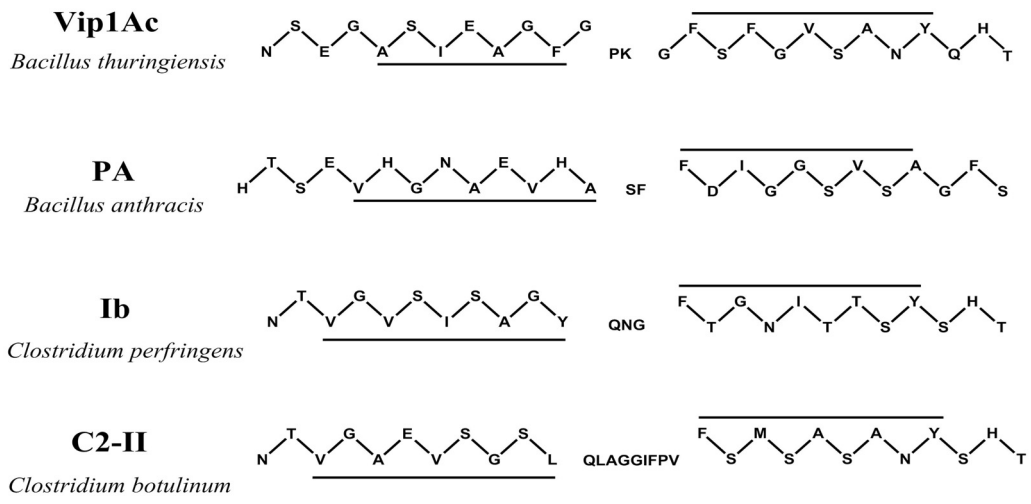

Figure 1.7. Comparison of the two β-strands predicted to be responsible for pore formation in component B of different binary toxins. Residues oriented toward the lipid bilayer are underlined.

1.3.5.6 Vip4

The Vip4 protein (~108 kDa), was reported for the first time in 2010 (NCBI Gen Bank accession number AEB52299) and the Bt toxin Nomenclature Committee (Crickmore *et al.* 2019) lists only one protein in this family, Vip4Aa1. Recently, two proteins were reported by Gomis-Cebolla *et al.*(2018) that belong to the Vip4 family: the first, named Vip4Aa-like1, consists of 873 amino acids (NCBI Gen Bank accession number KY420182); the second, named Vip4Aa-like2, contains 748 amino acids (NCBI Gen Bank accession number KY420193) (Gomis-Cebolla *et al.* 2018) and many variants that appear to be Vip4-like, appear in a range of gram-positive genomes. Vip4 proteins are most closely related to the Vip 1 family of proteins, but the toxicity and host range for most has not been established (Palma *et al.* 2014). In their patent application, Bowen *et al.* (Bowen *et al.*2017) reported a protein which they called TIC5290 (likely to be renamed Vip4Da2) that was active against Coleopteran (*Diabrotica virgifera virgifera*), Lepidopteran (*Spodoptera frugiperda, Helicoverpa zea, Ostrinia nubilalis, Plutella xylostella*) and Hemipteran (*Lygus Hesperus*) pests.

1.4 Auto-fluorescent proteins (AFPs)

1.4.1 General

In nature, some molecules have the ability to emit light, these molecules are known as luminescent molecules (West 1971). The luminescent phenomenon refers to the process of light emission from molecules that are excited by an energy form (Hanuš 2016). Luminescent molecules are classified and named according to the excitation source, with the name starting with a prefix referring to the energy source: for instance, photoluminescence when the excitation source is light energy, as is the case in fluorescent proteins (Fereja *et al.* 2013); and bioluminescence, which is a form of chemiluminescence, that uses a biologically catalysed reaction as an energy source, as happens with luciferases (Marques and Silva 2009).

In general, luminescent molecules are very important tools because the emission of light can be easily detected and measured both *in vivo* and *in vitro*. The real importance of these molecules lies in the rational way researchers have adapted them for use to advance research, especially molecular biosciences and cell biology. Here, I will focus on the fluorescent proteins.

1.4.2 Green Fluorescent Protein GFP

1.4.2.1 General milestones

Green fluorescent protein (GFP) was first discovered in 1962 by Shimomura *et al.* (1962) during their work on the chemiluminescent protein (aequorin) from *Aequorea victoria* jellyfish (Shimomura *et al.* 1962). In 1974, GFP was successfully purified and crystallized by Morise *et al.* (1974) who also measured its absorbance spectrum and fluorescence quantum yield (Morise *et al.*1974). Fourteen years later, the X-ray diffraction patterns for GFP crystals was reported (Perozzol *et al.* 1988) then in 1996 the molecular structure of GFP was solved by two groups independently: protein data bank accession number 1EMA (Ormo *et al.* 1996) and 1GFL(Yang *et al.*1996). Since these initial structures, structures of a range of GFP-mutants have been reported (eg (Brejc *et al.* 1997)).

The important characteristic of GFP is that no additional cofactor is required for fluorescence and the protein has been showed to be functional in a wide variety of

organisms (Chalfie *et al.* 1994). The generation of the GFP chromophore occurs spontaneously without the need for a substrate, co-factor or enzyme activity; it only needs the presence of oxygen. GFP has revolutionized cell biology and molecular biosciences since its gene was cloned by Prasher *et al*. (Prasher *et al.*1992). The ability for GFP to impart useful fluorescence properties has led to its use as a tool to study gene expression, live cell imaging, protein-protein interactions, protein localization (Tsien 1998), fusion proteins (Pédelacq *et al.* 2006) and more recently it has been used in drug delivery (Krishnan *et al.* 2018).

1.5.2.2 Molecular structure of GFP

The primary structure of GFP is composed of 238 amino acids, which give the protein a molecular weight of 27 kDa. The secondary structure comprises 11 β-strands forming a β-barrel with a coaxial helix and long loops. The 11 β-strands forming the β-barrel are connected by a number of loops and short helical segments, which in turn cap both ends of the β-barrel (Figure 1.8). The central helix that is responsible for chromophore (CRO) formation (the functional centre of GFP) (Figure 1.8B) runs through the centre of the β-barrel (Ormo *et al.* 1996, Yang *et al.* 1996). This special arrangement of β-barrel and the capping loops provides protection for CRO from the external environment (Jackson *et al.* 2006).

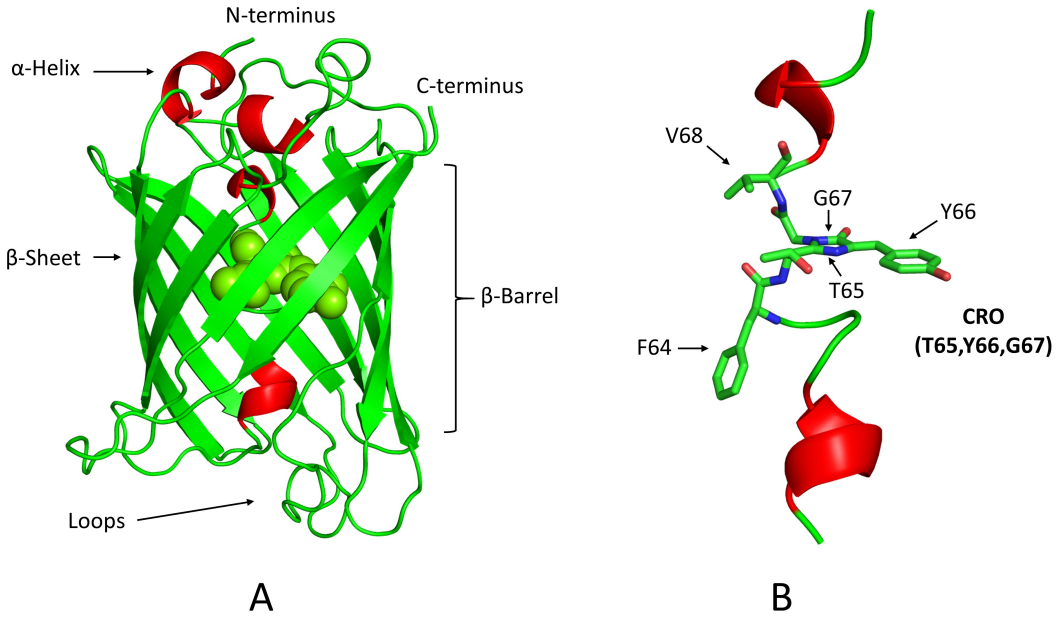

Figure 1.8. Molecular structure of S65T GFP.
A. β-barrel architecture of 11 strand β- strands (dark green) that surrounds a chromophore (light green spheres) (PDB: 1EMA). **B.** Central helix with CRO (β-barrel architecture has been cut away).

A chromophore is a functional group that forms part of a molecule that absorbs UV or visible light to excite electrons to a higher energy state. If the the absorbed light can be emitted, then the chromophore is also fluorescent. The colour emitted is normally at a higher wavelength (lower energy) than the excitation wavelength (Shukla *et al.* 2012). In the original GFP from *A. victoria*, the chromophore consists of the tripeptide (S65, Y66 and G67). While the chromophore is essential for the core fluorescent properties of GFP, the protein as a whole plays an important role. The mechanism of chromophore formation requires the protein to fold to the correct conformation in order for S65 and G67 to come into close proximity with each other. Also, unfolded GFP has little fluorescence due to quenching by the external environment (Jung *et al.* 2005a). The chromophore also makes extensive contacts with other residues in GFP, which ultimately define the precise fluorescence properties and allow protein engineering to alter fluorescence by changing the molecular interactions. Among the additional residues that play a vital role in GFP fluorescence are V61, T62, Q69, Q94, R96, N146, H148, R168, Q183, T203, S205 and E222 (Tsien 1998; Figure 1.9).

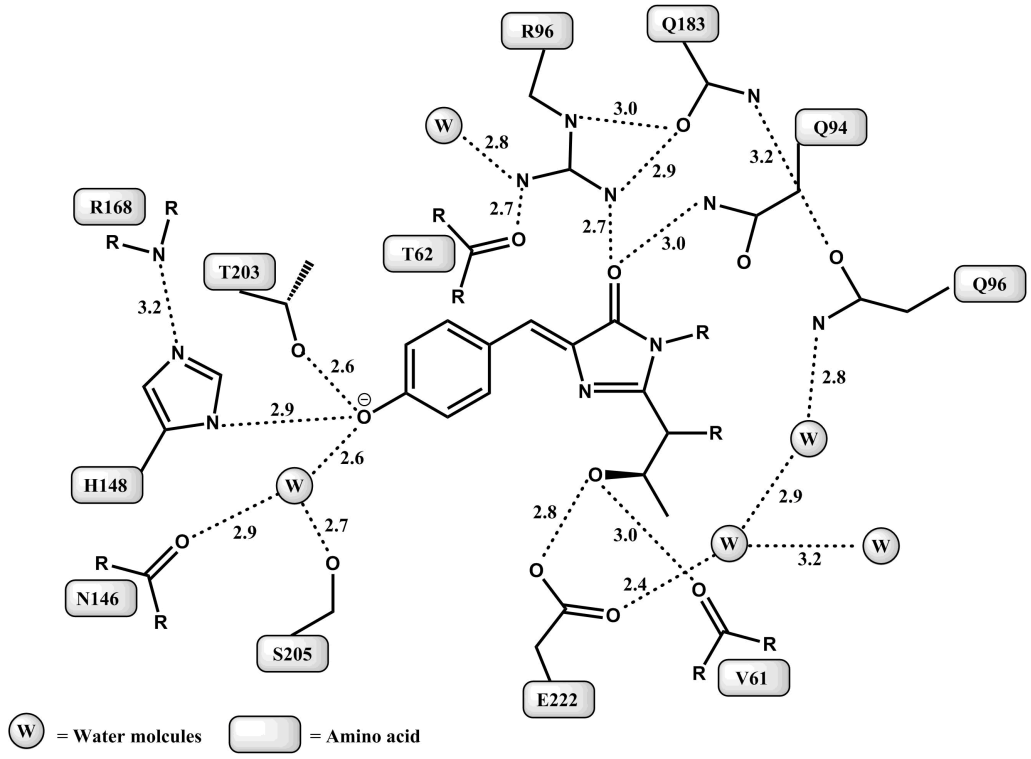

Figure 1.9. Chromophore environment residues of S65T GFP.
Mature chromophore with important adjacent residues (labelled as one letter code with residue number) and water molecules. Hydrogen bonds are denoted by dotted lines and the distance shown in angstroms. Adapted and redrawn from (Tsien 1998).

1.5.2.3 Chromophore biosynthesis in GFP

The generation of the GFP chromophore has been studied in some detail, which had has led several groups to propose a pathway to its synthesis (Heim *et al.*1994; Tsien, 1998; Ma *et al.* 2010; Grigorenko *et al.* 2017). The pathway comprises three distinct chemical processes (or stages): cyclisation, oxidation and dehydration that lead to covalent rearrangement of chromophore's residues to produce the mature chromophore. These reactions are preceded by protein folding, which is required to bring the 3 contigous residues (chromophore's residues; S65, Y66 and G67) into the right configuration for the intial cyclisation reaction. There are two proposed pathways for chromophore maturation.

The first pathway called "cyclisation, dehydration and oxidation" was proposed by Getzoff's group (Barondeau *et al.* 2003; Barondeau *et al.* 2005; Wood *et al.* 2005). In this pathway, after main chain cyclisation between G67 and S/T65, the carbonyl carbon of S65 undergoes dehydration followed by oxidation of the Cα-Cβ bond in T66 (Figure 1.10A). The second pathway called "cyclisation, oxidation and dehydration" was proposed by Heim *et al.* (Heim *et al.*1994) and Wachter's group (Rosenow *et al.* 2004; Sniegowski *et al.* 2005). In this pathway, the oxidation occurs after cyclisation followed by dehydration, and the mature chromophore may be present in two charge forms, which absorb at two different wavelengths (400 nm and 480 nm) (Figure 1.10B). The second pathway is believed to be more likely than the first because the first was based on studies carried out under anaerobic conditions while the actual chromophore maturation occurs under aerobic conditions *in vivo*. Additionally, based on the protonation state of the OH group in the Y66 phenol ring, the CRO in GFP can exist in two states: neutral state (CRO-A), which absorbs at approximately 395 nm and anionic state (CRO-B) absorbs at approximately 485 nm (Tsien 1998).

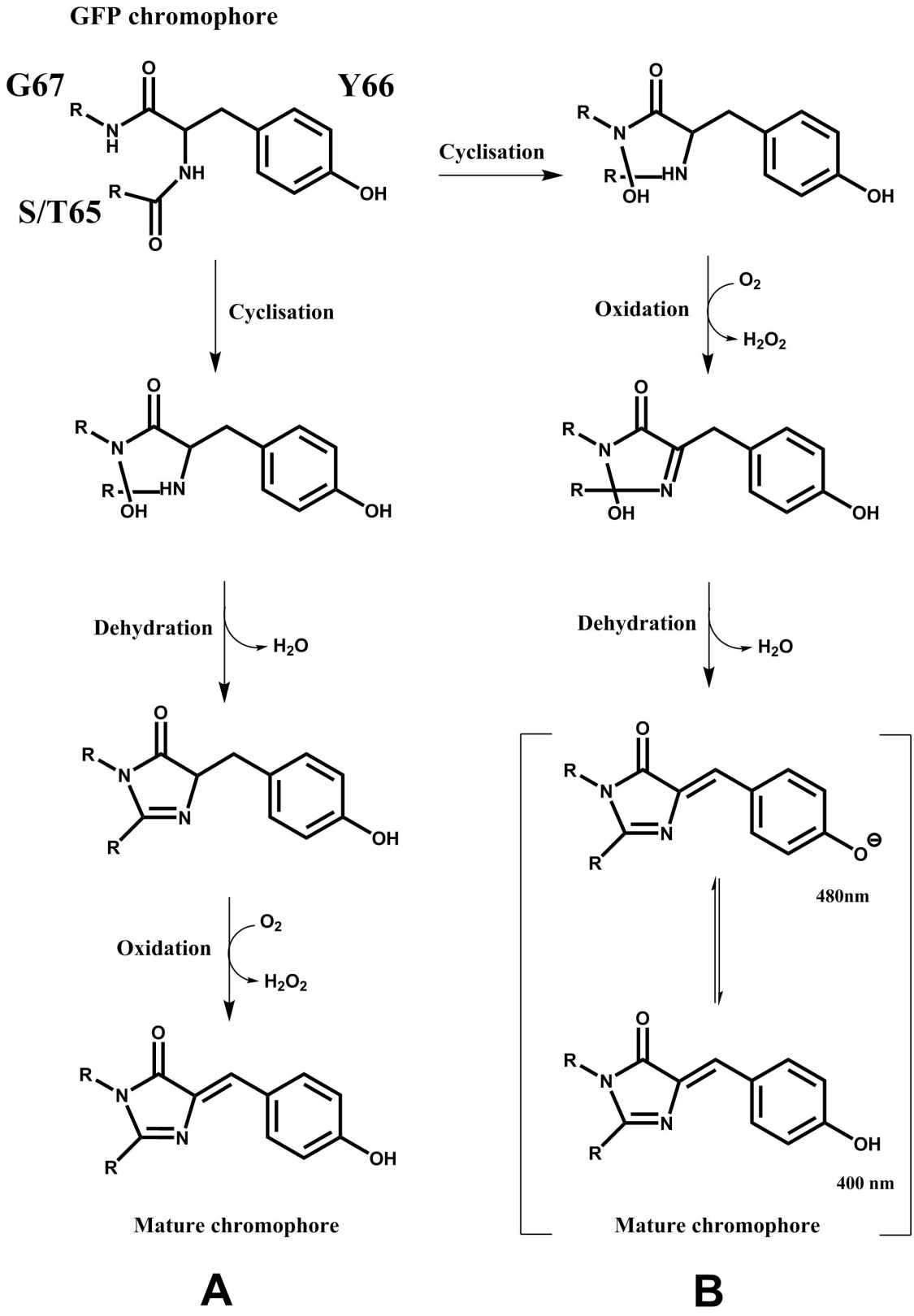

Figure 1.10. Proposed mechanisms for chromophore maturation.
A. Cyclisation, dehydration and oxidation pathway for chromophore maturation proposed by Getzoff's group. **B.** cyclisation, oxidation and dehydration pathway proposed by Wachter's group.

1.5.2.4 GFP engineering

GFP is an important tool in molecular biosciences. Since its discovery, considerable attention has been paid to engineering this protein to improve its use for one of its main applications, as a genetically encoded protein tag, through the generation of new colour variants or to enhance its properties in general (e.g. stability, maturation kinetics). The first point mutation S65T was made by Heim *et al.* in 1995, increasing the fluorescence intensity and photostability of GFP (Heim *et al.* 1995), followed by another point mutation F64L to generate enhanced GFP (EGFP) (Cormack *et al.* 1996) that has both mutations (S65T and F64L) and has improved folding efficiency and 100-fold increased fluorescence intensity at 37°C. In addition, excitation occurred at a single and preferred wavelength of *circa* 490 nm, which is of lower energy and less damaging to living cells. These properties make the EGFP a workhorse of cell biological applications.

Misfolding of the existing variants of GFP when used as reporters of gene expression (fusion proteins), encouraged protein engineers to improve GFP and generate new variants. Eventually a new variant superfolder GFP (sfGFP) was generated using directed evolution by Pédelacq *et al.* (Pédelacq *et al.* 2006). They started by combining mutations present in three existing useful GFP variants (EGFP, folding reporter GFP (Waldo *et al.* 1999) and cycle 3 variant GFP (Crameri *et al.*1996)). Using this initial protein as a starting point sfGFP was subjected to multiple rounds of directed evolution with six new additional mutations observed in the final version (S30R, Y39N, N105T, Y145F, I171V and A206V). This results in a protein with high resistance to chemical denaturation, with rapid and robust folding even when fused to poorly folded proteins. In the studies described in this thesis, sfGFP was used to incorporate non-canonical amino acids (ncAAs) at surface residues and used to generate a heterodimer with another fluorescent protein (Venus).

To engineer GFP to generate new colour variants that will expand the applications available to fluorescent proteins (e.g. two protein labelling and imaging, protein-protein interactions), research has targeted chromophore residues and chromophore environment residues. By changing the chemical nature of residues in or surrounding the chromophore, the inherent fluorescent properties can be altered in terms of the maximum excitation and emission wavelengths. By taking this approach, four different coloured versions were generated: blue fluorescent protein (BFP) with the substitution Y66H (in the

chromophore)/Y145F; cyan fluorescent protein (CFP) with the chromophore substitution Y66W (Heim *et al.*1994; Heim and Tsien, 1996); yellow fluorescent protein (YFP) with the substitutions T203Y/S65G (Ormo *et al.* 1996) and red fluorescent protein R10-3 with the substitutions F46L/F64L/V68N/V163A/K162Q/I167V/I171L (Mishin *et al.* 2008). This work is focused on yellow fluorescent protein so I will focus on variants related to YFP.

1.5.3 Yellow fluorescent protein (Venus)

Yellow fluorescent protein (YFP) is one of the GFP-derived fluorescent proteins generated by Ormo and co-workers in 1996. They found that the T203Y mutant was responsible for red-shifted excitation compared to wtGFP (Ormo *et al.* 1996): the newly introduced Y203 stacked on top of the chromophore phenol ring generating a π-π stacking interaction (also known as pi stacking), which is a chemical term referring to non-covalent interactions between aromatic rings, which lead to a reduction in the energy required to excite the chromophore, resulting in the red shift. Starting with this basic YFP, 2 further yellow variants were generated by the same study: variant 1, with mutations T203Y/F64L/S65G/S72A (excites at 513 nm, emit at 525 nm); and variant 2 with mutations T203Y/S65G/V68L/S72A (excites at 513 nm, emit at 527 nm). The excitation and emission wavelengths of YFP are significantly different than GFP and are easily distinguishable using appropriate filters in fluorescence microscopy. However, these early versions of yellow fluorescent protein were highly sensitive to acidic environments and chloride ions, so an improvement of these versions to reduce environmental sensitivity was desired (Griesbeck *et al.* 2001).

In 2002, Nagai and his co-workers studied the effect on some of the mutations known to improve GFP might affect YFP. They started with a widely used version of YFP, termed EYFP (T203Y/ S65G/V68L/S72A) (Miyawaki *et al.*1999). The introduction of 4 common folding mutations (Siemering *et al.*1996) into EYFP resulted in the generation of a well-folded version called super-EYFP (EYFP-F64L /M153T /V163A /S175G). They then introduced the improved chromophore maturation mutation F46L (Cormack *et al.* 1996) into super-EYFP to generate the SEYFP-F46L variant, which was more tolerant of exposure to chloride ions and acidic environments (by preventing ion access to the binding site). It also folds and forms the chromophore more efficiently at 37°C than YFP and EYFP. Because of the bright yellowish light of SEYFP-F46L variant

similar to the planet Venus (the brightest object in the night-time sky except for the moon) they called SEYFP-F46L variant "Venus" (Nagai *et al.* 2002).

The molecular structure of Venus was reported by Rekas *et al*. in 2002. In general, it has the same overall β-barrel structure as wtGFP and other GFP-derived fluorescent proteins (Figure 1.11A). The chromophore and the surrounding interaction residues are very similar to the original GFP and sfGFP (V61, T62, Q69, Q94, R96, N146, H148, R168, Q183, S205 and E222) (Figure 1.12) with two exceptions: the critical green to yellow mutation, T203Y that generates the π-π stacking interactions with chromophore phenol ring; and a T65G mutation (Figure 1.11B). In the studies presented in this thesis, Venus has been engineered with ncAAs to generate new versions with useful properties for use in cell biology and to generate a heterodimer with sfGFP.

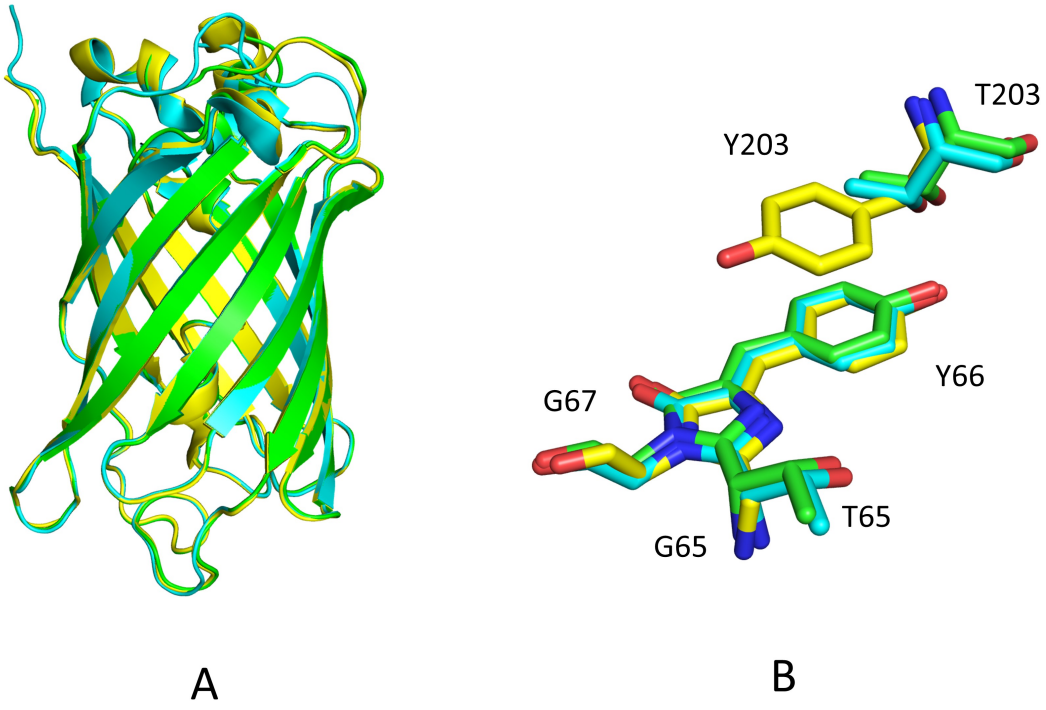

Figure 1.11. Structure alignment between wtGFP, sfGFP and Venus.
A. General structure alignment between wt GFP (green), sfGFP (cyan) and Venus (yellow). **B.** Chromophore and 203 residue alignment showing the difference in residue 65 and residue 203 that generate π-π stacking interactions with the chromophore phenol ring.

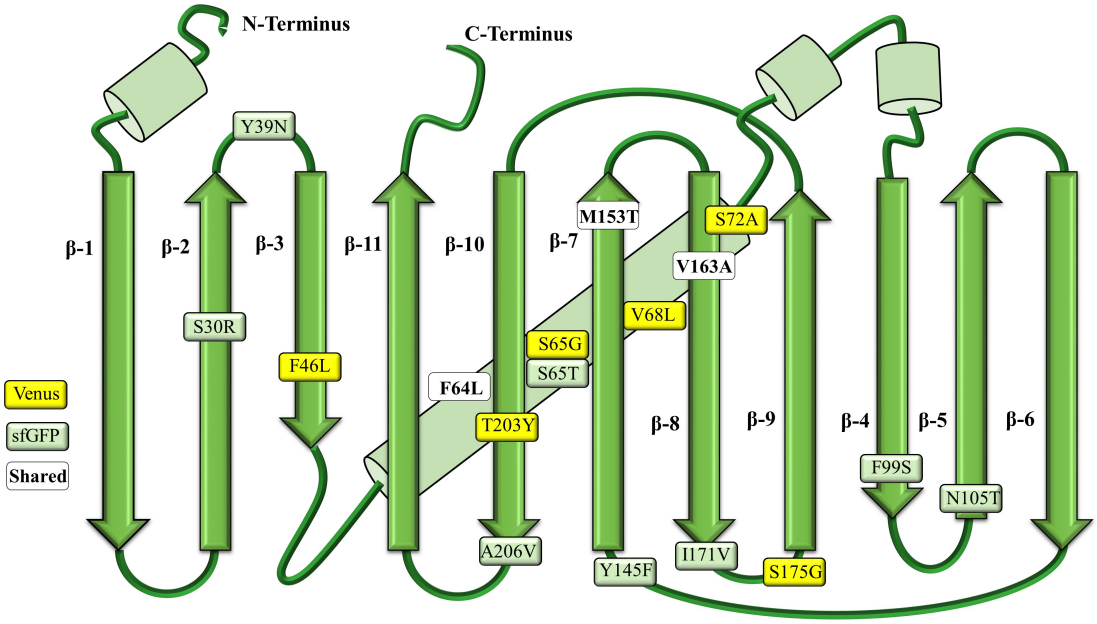

Figure 1.12. Topological map of GFP showing the mutations of sfGFP and Venus.
The green arrows correspond to the 11 β-strands, which are numbered from the N- to C-terminus. The light green cylinders correspond to the α-helices. The yellow rectangles represent mutations in Venus, while the light green rectangles represent mutations of sfGFP. The white rectangles represent shared mutations of both versions.

1.6 Protein oligomerisation

In some proteins, the final functional conformation requires assembly of more than one protein subunit (polypeptide chains) in a process known as oligomerisation to form the quaternary structure (such as haemoglobin, which consists of two α and two β polypeptide chains (Foot *et al.* 2009); complex structures such as collagen; and dimers such as transcription factors and cell receptors (Ahnert *et al.* 2015)). In fact, most of the proteins in biological systems work as part of a complex with other proteins or molecules and rarely work in isolation (Marianayagam *et al.* 2004). Formation and stabilisation of protein oligomers occurs spontaneously via different mechanisms, which leads to change in protein conformation and generates crucial functions. The most common mechanism is the formation of compatible non-covalent interactions, which include hydrogen bonds, hydrophobic interactions and electrostatic attraction (Larsen *et al.* 1998). Other mechanisms include binding two protein subunits via shared metal cofactors (e.g. insulin;

Adams *et al.* 1969), domain swapping (such as swapped C-terminal helix in cytochrome c (Hirota *et al.* 2010)) and formation of covalent interactions such as disulphide bridges between protein monomers (Nilsson *et al.* 2004).

Oligomerisation or dimerisation of proteins are key parts of the regulation of many biological processes, including enzyme activity, ion channels, transcription factors and receptors (Marianayagam *et al.* 2004), and are also involved in toxin activity and pathogenicity, especially in pore-forming toxins that require assembly of multiple units of the monomers to form a β-barrel pore channel (Los *et al.* 2013). Larger proteins have benefits compared to smaller ones in terms of stability, resistance to denaturation and degradation, complexity and cooperative functionality (eg multiple active sites) (Monod *et al.* 1964; Larsen *et al.* 1998). Oligomerization of multiple protein subunits helps with the control of translation errors. The small protein subunit with an error can more easily be discarded before assembling than a big, complex protein (Goodsell *et al.*2000). This also helps to minimise the gene size: a protein composed of 1000 amino acids requires a long gene (about 3 kb long) or could made up of 2 subunits (500 amino acids with gene sizes about 1.5 kb) or 4 subunits (250 amino acids with gene size about 750 bp) (Mei *et al.* 2005).

1.6.1 Generation of artificial protein oligomers

Generation of novel proteins with new properties, which do not exist in nature, is a major area of interest within the field of protein engineering. Creating artificial oligomers could successfully enhance or produce a new protein structure and function that can be used to solve medical, scientific and technological problems (Ljubetič *et al.*2017). A variety of techniques have been implemented to create protein oligomers. These include; formation disulphide bridges (Leibly *et al.* 2015), linking via a metal ion (Song and Tezcan, 2014), helix-helix interactions (Huang *et al.* 2014), β-sheet-β-sheet interactions (Khakshoor *et al.* 2010), and domain-swapped oligomers (Kobayashi and Ryoichi, 2017).

Among the techniques mentioned above, formation disulphide bridges is a common technique for artificial dimerisation because it requires only a single mutation in the target proteins by introducing a cysteine residue and can be used with a wide range of proteins. In 2017, Thamri *et al.* described the formation of an artificial dimer via disulphide bridging using a small synthetic cationic antimicrobial peptide (pep-1037, KRFRIRVRV)

after introducing a cysteine at the N-terminus (cys-pep1037) (Thamri *et al.* 2017). The newly formed dimer (HOOC-VRVRIRFRKcys-cys-KRFRIRVRV-COOH) showed high activity against the microorganisms tested, up to 60 times higher than the monomer peptide (pep -1037). This technique is a powerful method with small peptides or proteins that lack cysteine in their natural sequence.

However, the main drawback of this technique is misfolding of some proteins by forming disulphide bridges between introduced mutant cysteine and existing cysteine residues. Another issue with this technique, it is that it may be suitable for work *in vitro* only, due to the reducing environment of the cytosol of living cells (Van Amsterdam *et al.* 2002). Furthermore, it does not allow for easy construction of defined heterodimers due to the same functional group linking the two proteins together (A + B = AB, AA, BB mixture). Using ncAAs to introduce novel reactivity into a specific part of a protein overcomes these issues for artificial dimerisation (Hendrickson *et al.* 2004).

1.7 Incorporating ncAAs into protein

Natural amino acids have restricted reactive chemistry including; carboxyl, sulphydryl, hydroxyl, amide groups and amines. This makes conventional protein engineering limited to the use of these chemical groups, which are common amongst the proteome and within individual proteins so reducing specificity to terms of target and residue. Incorporating ncAA (amino acid analogues with a functional group) into a protein is a powerful technique to introduce new and useful chemistry that does not exist in nature. This, in turn, would allow the engineering of proteins for different purposes including but not limited to studying protein structure (using ncAA containing heavy atom to improve phase resolution) (Xie *et al.* 2004 ; Kobayashi *et al.* 2005), studying protein-protein interaction via photocrosslinking (ncAA containing photoreactive groups), visualization, localization and intermolecular interactions of protein via biophysical probes (ncAA containing fluorescent groups) (Xie and Schultz 2006) and introducing a unique tag for bioorthogonal reaction such as azides, alkynes and tetrazines (Lang and Chin 2014a).

To incorporate ncAAs into protein, expansion of the genetic code is required (Wang *et al.* 2001). By far the most common approach is the reprogramming of stop codons and their recruitment to encode ncAA incorporation. Among the three stop codons in DNA (amber, ochre and opal), the amber stop codon (TAG) has proved the most successful as a reprogrammed codon for ncAA incorporation because it is least prevalent compared to other stop codons in both prokaryotes and eukaryotes. In normal protein synthesis, there are different tRNAs and aminoacyl tRNA synthetases (aaRS), which recognise, carry and incorporate an amino acid in response to the genetic code in mRNA. Transforming a stop codon to a genetic code for ncAA required an engineered tRNA-aaRS pair. This engineered pair should be orthogonal, with no interaction with any of the endogenous tRNA/aaRS. Providing the cell with this engineered tRNA-aaRS pair and ncAA of interest allows ncAA incorporation. The engineered aaRS recognises, carries and attaches the ncAA onto the cognate engineered tRNA, which will then be utilised by the ribosome an incorporated into the growing peptide chain on recognition of the stop codon (rather than termination of protein synthesis) (Figure 1.13).

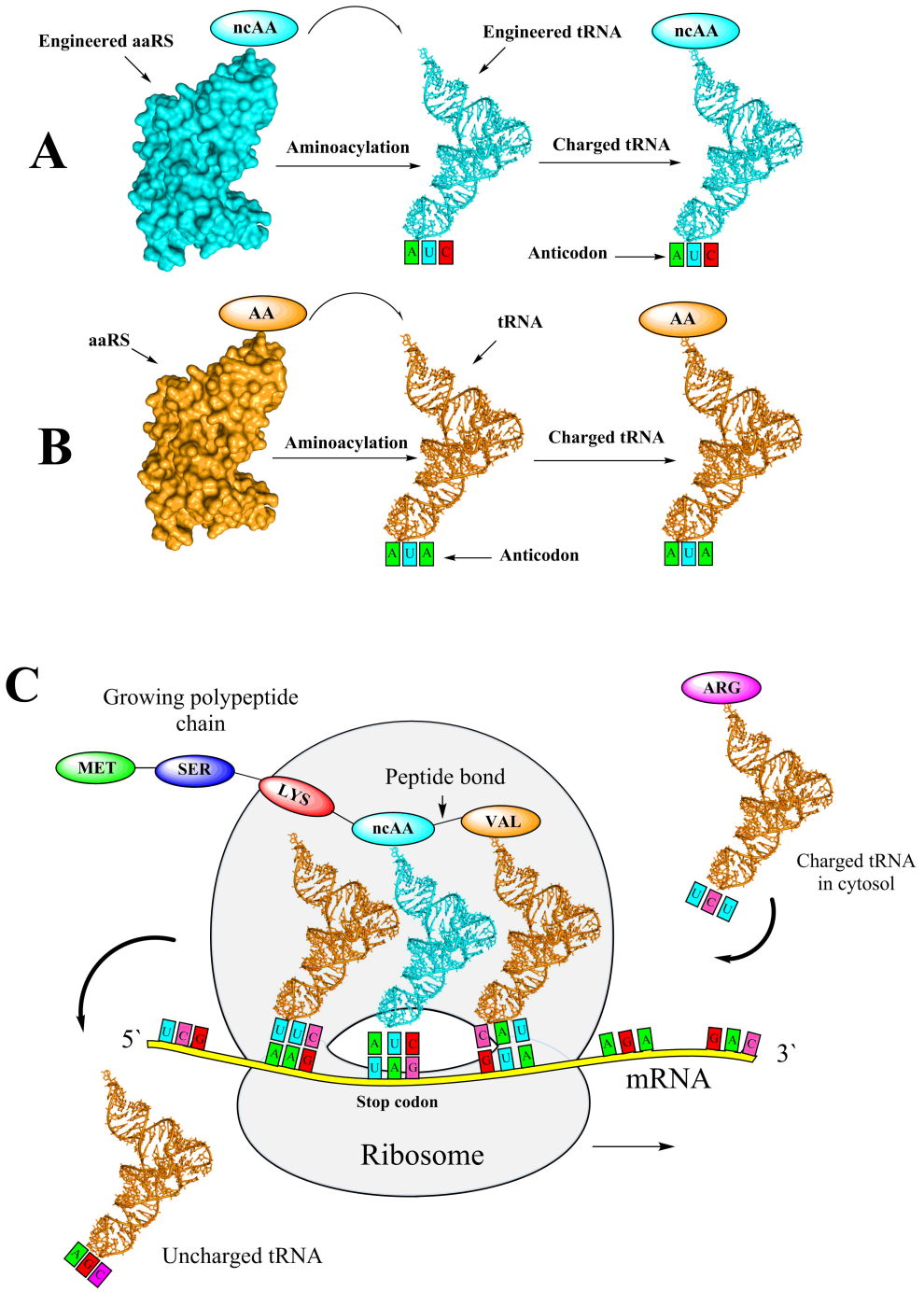

Figure 1.13. Incorporation of ncAAs.
A. The aminoacylation process of engineered tRNA-aaRS pair (cyan) leading to charging of the engineered tRNA with ncAA. **B.** The aminoacylation process of wild-type tRNA-aaRS pair (orange) leading to charging of the tRNA with amino acid. **C.** Incorporation of ncAA into growing polypeptide chain during protein synthesis in the ribosome. The ribosome moves from 5` to 3` along the mRNA. Charged tRNAs with canonical amino acid (three letter code) enter and recognise their cognate codon on mRNA in the ribosome, providing the canonical amino acid for the growing polypeptide chain, then exist the ribosome. The engineered tRNA charged with ncAA enters the ribosome and recognises the amber stop codon and presents the ncAA for the growing polypeptide chain. Protein synthesis continues as normal and termination occurs at a stop codon other than an amber codon.

This technique (amber codon suppression technology) has been successfully used to incorporate a wide range of ncAAs in both prokaryotic and eukaryotic cells (Wang *et al.* 2012; Lang *et al.* 2014b). In this project, three different ncAAs, as shown in Figure 1.14, have been used: two tyrosine derivatives; 4-Iodo-L-phenylalanine, which is a ncAA containing a heavy atom for phasing improvement and *p*-azido-L-phenylalanine (azF) (a ncAA containing an azide group); and a lysine derivative, s-cyclooctyne-L-lysine (SCO-K) for bioorthogonal modification via Click chemistry. This ncAA contains a strained alkyne group that reacts specifically with azide groups to generate a stable triazole linkage (Figure1-15).

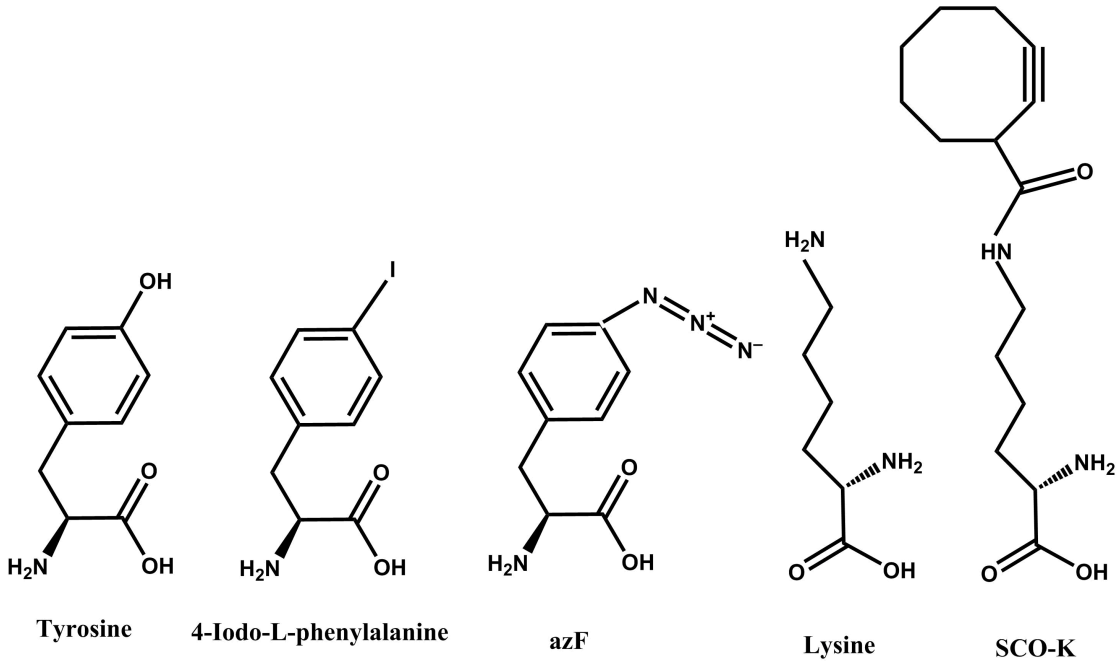

Figure 1.14. Structure of canonical amino acid and its ncAAs derivatives.
Line drawing of canonical amino acid Tyrosine and its ncAA derivatives; 4-Iodo-L-phenylalanine and *p*-azido-L-phenylalanine (azF). Canonical amino acid Lysine and its ncAA derivative, s-cyclooctyne-L-lysine (SCO-K).

1.7.1 *p*-azido-L-phenylalanine (azF)

p-azido-L-phenylalanine (azF) is one of the most commonly used ncAA in protein engineering owing to its characteristics (Reddington *et al.* 2013): (1) bioorthogonal chemistry via Click chemistry; (2) photochemical reactivity; (3) IR signature. The new properties do not impose a large change in bulk as the hydroxyl group of tyrosine (-OH)

is replaced by only 3 nitrogens (-N_3). As a result, it is also well tolerated in proteins, including within the packed interior (Figure 1.14).

1.7.1.1 Selective reactivity of (azF) via Click chemistry

Introducing an azide group in a specific part of a protein allows for bioorthogonal Click chemistry (bioorthogonal reaction refers to any chemical reaction that can occur inside of living systems without interfering with native biochemical processes) with target molecules *in vivo* and *in vitro* by stepwise chemical reactions called Click chemistry, in which the azide group reacts with carbon-carbon triple bond-containing (alkyne) molecules via a cycloaddition reaction to produce a stable triazole linkage between the conjugated molecules (Figure 1.15). Click chemistry is a term first described by the Sharpless group in 2001 (Kolb *et al.* 2001). It refers to a wide range of chemical reactions that are characterized by excellent selectivity, producing high yields of desired products with almost no by-products, quickly and robustly, and compatible with physiological conditions so that they can be performed *in vivo* and *in vitro* (Kolb *et al.* 2001; Singh *et al.* 2016).

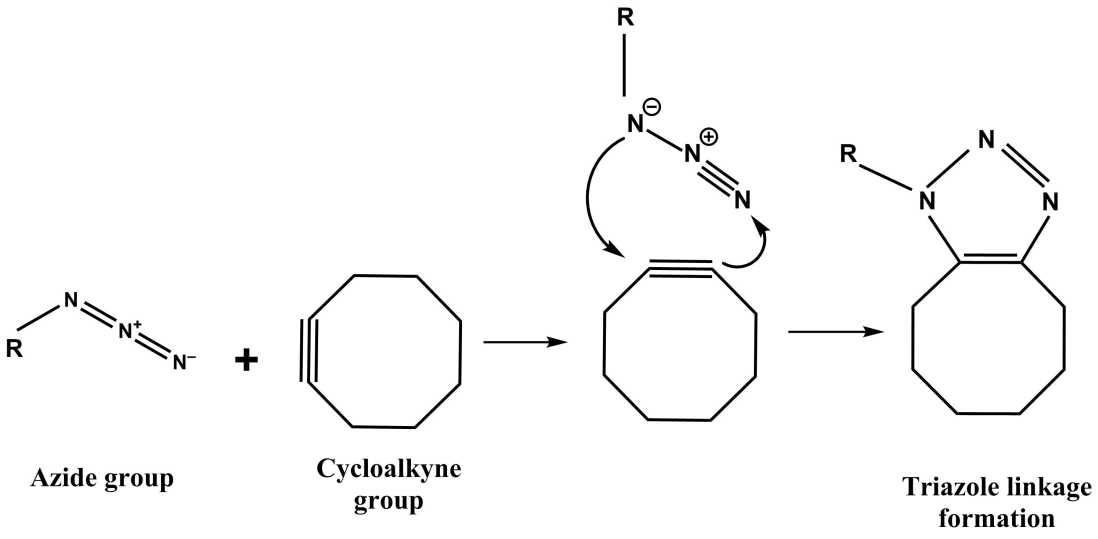

Figure 1.15. Strain-promoted azide-alkyne cycloaddition (SPAAC).
Azide group reacts with strained cyclic alkynes and forms a stable triazole linkage.

The azide-alkyne cycloaddition reaction can be achieved by two methods. Copper-catalysed azide-alkyne cycloaddition (CuAAC) was described by Rostovtsev and co-workers. In this method, the azide group reacts with terminal alkynes in the presence of

copper to generate a triazole linkage (Figure1.16) (Rostovtsev *et al.* 2002). Despite the successful use of CuAAC for attaching fluorescent probes, purification tags and long-chain polyethylene glycol (PEG) into target molecules (Link and Tirrell, 2003; Dirks *et al.* 2005), it is unworkable in living cells due to the requirement for Cu^{+}, which is highly toxic to living cells and can damage proteins (Agard *et al*. 2004).

The second method for the azide-alkyne cycloaddition reaction, called strain-promoted azide-alkyne cycloaddition (SPAAC), was first reported by the Bertozzi group. The azide group in this method reacts quickly and selectively with strained cyclic alkynes instead of terminal alkynes, without the need for copper, to form a stable triazole linkage as shown in Figure 1.16 (Baskin *et al.* 2007; Ornelas *et al.* 2010). This method is a metal-free Click chemistry and has been widely used for bioorthogonal modification *in vitro* such as cell surface labelling of glycoproteins (Baskin *et al.* 2007) and *in vivo*, such as imaging in Zebrafish (Laughlin *et al.* 2008). In the studies described in this thesis, SPAAC has been used for bioorthogonal modification.

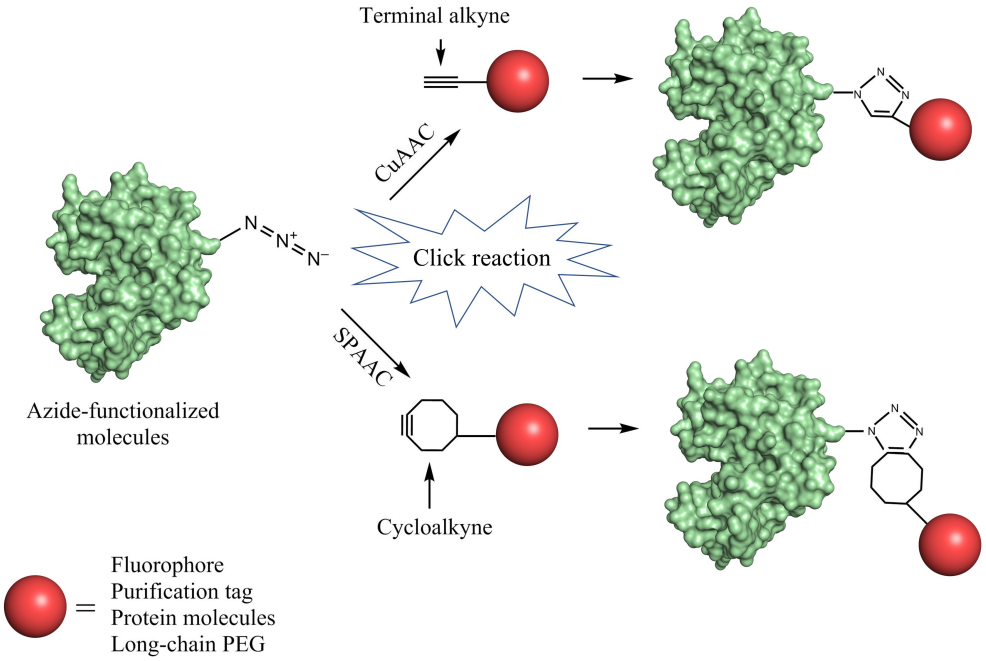

Figure 1.16. Azide-alkyne cycloaddition reaction via click chemistry.
Top: The azide group in azide-functionalized molecule reacts with a terminal alkyne in the other molecule in the presence of Cu by copper-catalysed azide-alkyne cycloaddition (CuAAC) reaction to form a stable triazole linkage between two molecules**. Bottom:** The azide group can react with strained cyclic alkynes by strain-promoted azide-alkyne cycloaddition (SPAAC) to link two molecules via a stable triazole linkage.

1.7.1.2 Photochemistry of phenyl azides

Phenyl azides (sometimes called aryl azides) as photoactivatable groups have been widely used as cross-linkers to probe protein-phospholipid interactions (Gubbens *et al.* 2009) and protein-protein interactions (Chen *et al.*1994), photoaffinity labels (Cai *et al.* 1993), and light-induced activation for protein ligand immobilization (Nahar *et al.* 2001). It would be useful to introduce this photoactivatable chemistry at a specific position in proteins for different uses. The photochemistry of phenyl azides lies in the ability to generate a singlet reactive nitrene radical after activation by UV-visible light and loss of N_2 molecular nitrogen.

The nitrene radical is a key part of the phenyl azides' chemistry, which can undertake different reaction pathways depending on the surrounding environments (Bräse *et al.* 2005; Morris *et al.* 2013) (Figure 1.17). It can undergo a reduction reaction to generate a phenylamine, which can be utilised as a hydrogen bond acceptor/donor group (Figure 1.17 pathway 1). The nitrene radical can react with carbon double bonds via addition reactions (Schrock and Schuster 1984)(Figure 1.17 pathway 2), which can be used to link with carbon nanotubes (Zaki *et al.* 2018). It can also form an amide crosslink by insertion in a hydrocarbon sidechain (Figure 1.17 pathway 3). Additionally, it can undertake reversible ring internalisation and expansion, forming an unsaturated heterocycle moiety (azepine) (Hartley *et al.* 2016), which is an important functional group in chemistry that can be involved in different reactions such as cycloaddition and metal carbonyl complexation (Paquette *et al.* 1969; Farah *et al.* 2009)(Figure 1.17 pathway 4).

Insertion of genetically encoded, photoactivable chemistry such as a phenyl azide at a specific position in a protein is a powerful tool for optogenetics. Optogenetics is a growing biological technique in which biomolecules have been genetically modified to be controlled by light (Kushibiki *et al.* 2014). The photoreaction of incorporated azF could alter the local environment around the insertion position and the overall protein properties (Reddington *et al.* 2015).

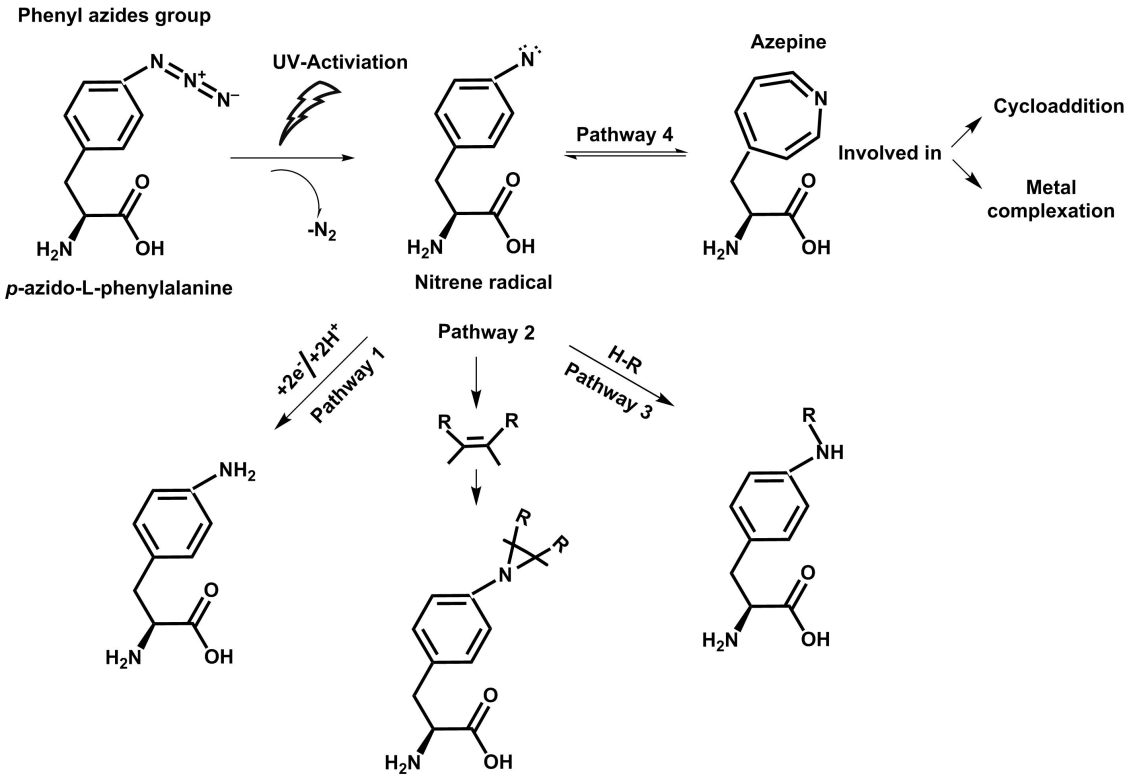

Figure 1.17. Photoreactivity of phenyl azides.

1.8 General aims of the thesis

1.8.1 General aims of the vegetative insecticidal protein project

The vegetative insecticidal protein 1 and 2 (Vip1Ac1; pore-forming toxin and Vip2Ae3; ADP-ribosyl transferase) belong to the AB binary toxin family, which includes anthrax toxin and iota like binary toxins (mammalian toxins) and has potential to act as a biopesticide. First, its molecular structure and function needs to be understood, and how this might contribute to specificity for certain insects. The specificity of the toxin is determined by toxin structure and receptors on the target cells. Therefore, studying the molecular structure of toxins and understanding their mode of action was the central aim of this project, which could open avenues to engineer toxins to produce novel variants with desirable properties and could help to understanding other AB binary toxins. To achieve this, a variety of protein engineering and molecular technologies were applied. For structural protein studies and other analyses, high amounts of soluble protein that can be purified under native conditions are required. This can be achieved using DNA

recombinant technology, and this was the first aim in this project (chapter 3). To determine the molecular structure of a protein using X-ray crystallography, high quality protein crystals need to be prepared to allow data collection, which can be used to solve the molecular structure, and this was the second aim (chapter 4). After successful recombinant protein production and structural determination, studies of the properties of Vip1Ac1 and Vip2Ae3, such as the activation process, oligomerisation, protein-protein interaction, pore formation in artificial membranes, interactions with lipid and electron microscopy studies was the third aim (chapter 5).

1.8.2 General aims of the fluorescent proteins engineering project with ncAAs

Traditional protein engineering is limited by the restricted functional groups in canonical amino acids. The addition of new functional groups at desirable positions in the protein by genetic incorporation of non-canonical amino acids (ncAAs) via reprogramming of stop codons increases the diversity of protein engineering. The main aim of this project was to use the ncAAs to engineer and modify the protein genetically and post-translationally to generate new useful variants. To achieve these goals, the auto fluorescent Venus protein was chosen as a target to generate variants with new fluorescent properties (which are important in imaging techniques) by modifying the protein genetically with ncAAs containing a photoactivable group (azide group) around the functional centre of the protein (chapter 6). For posttranslational protein modification, sfGFP and Venus were chosen as a monomer target to generate and analyse a functionally linked heterodimer between sfGFP and Venus via Click chemistry using modified protein with ncAAs containing compatible Click chemistry group (azide group in Venus and strained alkyne group in sfGFP: chapter 7).

Chapter Two

Material and Methods

2 Materials and methods

2.1 Materials

2.1.1 Chemicals

Antibiotics were supplied by Melford and made as 1000x concentrated stock solutions using MilliQ water except Chloramphenicol and Tetracycline (Fluka Biochemika, Sigma-Aldrich) using ethanol. Tetracycline was prepared as 5 mg/ml stock, sterilized by filtration through 22 µm syringe filter units (Fisher Scientific, Loughborough, Leicestershire, UK) and kept at -20°C. The final working concentrations of each antibiotic were: Ampicillin 100 µg/ml, Kanamycin 30 µg/ml, Chloramphenicol 30 µg/ml and Tetracycline 25 µg/ml. Routine laboratory chemicals were purchased from Fisher Scientific UK Ltd. (Loughborough, UK) and Sigma Chemicals Ltd. (Poole, Doreset, UK). Suppliers of specialised chemicals are mentioned in the appropriate Sections of the text. All solutions and buffers were prepared using MilliQ water and sterilized by filtration using 22 µm filter units.

Non-canonical amino acids (ncAAs) 4-iodo-L-phenylalanine and *p*-azido-L-phenylalanine (Bachem, Weil am Rhein, Germany) were prepared by dissolving 0.145 g and 0.1 g respectively of ncAAs in 0.5 ml of 1 M NaOH. These were then added to 1 l of the broth media at a final concentration of 0.5 mM. S- cyclooctyne-L-lysine (Sichem, Bremen, Germany) was prepared as a 100 mM stock solution in 0.2 M NaOH / 15% DMSO and kept at -80°C as described in the manufacturer's instructions. The stock was diluted 1:4 with 1 M HEPES buffer prior to addition to 1 l of the broth media at a final concentration of 0.5 mM. Selenomethionine (ACROS ORGANICS, USA) was added as a powder (0.196 g) to1 l of the broth media to give a final concentration of 1 mM.

2.1.2 Media

LB Agar (Melford) and 2xYT (Sigma) were prepared according to the manufacturer's instructions using ultra-pure water. M9 salt medium was composed of (88 mM Na_2HPO_4, 22 mM KH_2PO_4, 3.4 mM NaCl, 18 mM NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$). All media were sterilized by autoclaving at 121ºC, cooled to 40ºC then supplemented with appropriate antibiotic. Defined medium for methionine biosynthesis inhibition and incorporation of selenomethionine was composed of M9 salt medium

supplemented with 0.5% glucose (sterilized by filtration), 100 mg/l each of (phenylalanine, lysine and threonine), 50 mg/l each of (valine, leucine and isoleucine), 2mg/l each of Thiamine (Sigma) and Biotin (Duchefa Biochemie) and 1 mM of selenomethionine.

2.1.3 Bacterial strains

Three different *Escherichia coli* strains were used. *E.coli* DH5α electrocompetent cells (New England Biolabs (NEB)) for cloning and plasmid amplification. *E.coli* BL21(DE3) chemically competent cells (NEB) and Top 10^{TM} electrocompetent cells (Invitrogen, Paisley, UK) for protein expression. Genotypes of competent cells shown in Table 2.1

Table 2.1 Genotypes of *E. coli* competent cells.

***E. coli* competent strain**	**Genotype**
DH5α	F$^-$ *endA1 glnV44 thi1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*Δ M15 Δ(*lacZYA-argF*) U169, hsdR17($r_K^- m_K^+$), λ$^-$
BL21(DE3)	F$^-$ *ompT gal dcm lon hsdS$_B$* ($r_B^- m_B^-$) λ (DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$] $_{K\text{-}12}$(λ^S)
Top 10^{TM}	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$*

2.1.4 Vectors for protein expression

Different vectors were used for recombinant protein expression in this study as shown in Table 2.2

Table 2.2 Vectors used throughout this study.

Vectors	**Encoding gene and purpose**	**Source**
pET-28b (Kanamycin ^{R})	Encoding Vip1Ac1 protein containing N-terminal His-tag and signal peptide (original clone)	Donated from Prof Primitivo Caballero, Public University of Navarra, Pamplona, Spain
pGEX-4T-2 (Ampicillin ^{R})	Encoding glutathione S-transferase (GST) as soluble fusion partner	Donated from Dr Colin Berry, Cardiff School of Biosciences, Cardiff University, Cardiff, UK

pBAD (Ampicillin ^{R})	Encoding wild-type yellow fluorescent protein (Venus)	Purchased from Addgene
pRK793 (Ampicillin ^{R}) and (Chloramphenicol ^{R})	Used for His_6-tobacco etch virus (S219V) (TEV) protease expression.	Donated from DDJ group, Cardiff School of Biosciences, Cardiff University, Cardiff, UK
pBAD (Ampicillin ^{R})	Encoding single TAG mutation in super-folder green fluorescent protein (sfGFP)	
pET22b (Ampicillin ^{R})	Encoding Cytochrome b_{562} (Cyt b_{562}) as soluble fusion partner	
pDULE (Tetracycline ^{R})	Encoding an engineered orthogonal tRNA and tRNA synthatase, used for facilitated incorporation of tyrosine ncAA derivatives in bacterial cells	
pEVOL (Chloramphenicol ^{R})	Encoding an engineered orthogonal tRNA and tRNA synthatase, used for facilitated incorporation of lysine ncAA derivatives in bacterial cells	

2.2 Molecular Biology and recombinant DNA methods

2.2.1 Oligonucleotide Primers

Serial Cloner 2-6-1 and Amplifx software were used for *in silico* design of oligonucleotide primers, and for general sequence manipulation, including virtual restriction digests, ligations and sequence alignments. Primers were synthesized by Eurofins MWG Biotech and Integrated DNA Technologies; a full list of primers is presented in Table 2.3. Annealing temperatures for primers were determined by Tm Calculator available on the Thermo Fisher Scientific website (https://www.Thermofisher.com/us/en/home/brands/thermo-scientific/molecularbiology/molecular-biology learning-center/molecular-biologyresource-library/thermo-scientific-web-tools/tmcalculator).

Table 2.3 Sequences of primers used in this study.

Restriction sites are underlined, sequences encoding the TEV cleavage site are in italics, TAG mutation in bold and His-tag in italics and bold.

Name	5- Sequence -3	Tm (˚C)
Remove His-tag and signal sequence from Vip1Ac1		
Vip1-RHS-F	GACAACAAAATCAATCAGGATTCT	57
Vip1-RHS-R	CATGGTATATCTCCTTCTTAAAGTTAA	57
Internal sequences primers for Vip1Ac1		
Vip1AcF3	GAATTTTTAGCGAAAGCATCG	57
Vip1AcR1	TTGGTTTGTTGTTATAGTACAATAATCC	57
Insert *Vip1Ac1* into *GST* Gene Fusion System		
Vip1GstF	GAATTCTCGCAGGATCCGACAACAAAATCAATCAGGATTCT	60
Vip1GstR	GGAGAATTCTTATTTTTTTACTGCTGACACTTCTGTAA	60
Adding TEV site and restriction sites into cytochrome b_{562}		
Cyt-b562 F	GCCATGGGATCCTCCAGAATTCTCGAGCACCACCACC	60
Cyt-b562 R	*GCCTTGAAAATAGAGGTTTTC*ACGATACTTCTGGTGATAGGC	60
Fusion Vip1Ac1 with cytochrome b_{562}		
Vip1pET-22bF	GAATTCGCTCCAGGATCCAGACAACAAAATCAATCAGGATTC	60
Vip1GstR	GGAGAATTCTTATTTTTTTACTGCTGACACTTCTGTAA	60
Remove cytochrome b_{562} and link its signal sequence with Vip1Ac1		
Vip1-RHS-F	GACAACAAAATCAATCAGGATTCT	57
CytoR	TGCGGCAAACGACG	57
Adding N-terminal His-tag and TEV site to Vip1Ac1		
Vip-TEV-F	*GAAAACCTCTATTTTCAAGGC*GACAACAAAATCAATCAGGATT	60
Vip-His-tag-R	***GTGGTGGTGGTGGTGGTG***TGCGGCAAACGACG	60
Vip1Ac1 amber stop codon mutation at codon 358		
F358 mutant	GCAGGA**TAG**TTAAATGCAAATGTT	57
R358 mutant	TGACGCAGTATTAAGTTGCGTA	57
Vip1Ac1 amber stop codon mutation at codon 536		
F536 mutant	TTATTG**TAG**TATAACAACAAACCAATTTA	57
R536 mutant	TCCATCTGTTTCTTTTATCTCTTC	57
Isolate *Vip2Ae3* gene from original strain H29.3		
F-VIP2-cyto	TCGCTCCAGGATCCTGAACATTTAAATCTAAATT	60
R-VIP2-cyto	GCTCGAATTCTTAATCTGTTAATAATGTTGCATC	60

Continuous

Name	5- Sequence -3	Tm (˚C)
Adding N-terminal His-tag, remove Cytochrome b_{562} and link its signal sequence with Vip2Ae3		
F-vip2-his	***CATCATCATCAT***GAAAACCTCTATTTTCAAGGC	60
R-vip2-his	***ATGATGATG***TGCGGCAAACGACG	60
Venus amber stop codon mutation at codon 148		
Venus148 F	AAC AGC **TAG** AACGTCTATATCACC	57
Venus148 R	GTAGTTGTACTCCAGCTTGTGC	57
Venus amber stop codon mutation at codon 204		
Venus 204 F	AGC TAC **TAG** TCC GCCCTGAGCA	57
Venus 204 R	CAGGTAGTGGTTGTCGGGCA	57
Venus amber stop codon mutation at codon 203		
Venus 203 F	AGC**TAG**CAGTCCGCCCTGAG	60
Venus 204 R	CAGGTAGTGGTTGTCGGGCA	60
Venus amber stop codon mutation at codon 145		
Venus 145 F	TACAAC**TTC**AACAGCTAGAACGTCTATATCA	60
Venus 145 R	CTCCAGCTTGTGCCCCAG	60
Venus amber stop codon mutation at codon 66		
Venus 66 F	GGC**TAG**GGCCTGCAGTGCT	57
Venus 66 R	CAGGGTGGTCACGAGGGTG	57
Primers used for pBAD vector sequences		
pBAD-F	ATGCCATAGCATTTTTATCC	57
pBAD-R	GATTTAATCTGTATCAGG	57

2.2.2 Amplification of DNA by polymerase chain reaction (PCR).

Q5 DNA polymerase (NEB) was used to amplify genes and for whole vector site-directed mutagenesis (SDM). General reaction mixtures and PCR conditions for Q5 DNA polymerase are shown in Table 2.4. Go Taq G2 Green Master Mix (Promega, USA) was used for colony screening using PCR to identify clones with the desired insert. Colonies were cultured in 5 ml 2xYT broth for 4-6 hours, then 50 µl of each colony were boiled for 5 min and 2 µl was transferred into the PCR reaction mixture. General reaction mixture and PCR conditions for Go Taq amplifications are shown in Table 2.5. Sterile PCR tubes were used for all PCR, which was carried out in a Flexigene thermocycler (Techne).

Table 2.4 Composition of PCR Mixture and PCR conditions

Materials			**Volume in µl**		
Q5 Reaction buffer (5x)			10		
dNTPs (10 mM)			1		
Forward Primer (10 µM/ µl)			1		
Reverse Primer (10 µM/ µl)			1		
Template DNA (2 ng/µl)			3		
Q5 DNA polymerase			0.5		
DMSO			0.5		
Ultra-pure water			33		
Total reaction mixture			50		
PCR conditions					
Initial denaturation	30 reaction cycles			Final Extension	Hold
	Denaturation	Annealing	Extension		
30 s (98˚C)	10 s (98˚C)	30 s (X˚C) *	3 min (72˚C)	5 min (72˚C)	5 min (4˚C)
* X= Tm used was the lowest Tm of the primer pairs as listed in Table 2.2					

Table 2.5 Composition of PCR Mixture and PCR conditions

Materials			Volume in µl		
Go Taq G2 Green Master Mix			50		
Forward Primer (10 µM/ µl)			1		
Reverse Primer (10 µM/ µl)			1		
Ultra-pure water			48		
Total reaction mixture			100		
Aliquot 9.5 µl into 10 PCR tubes and add 2 µl of boiled bacterial culture					
PCR conditions					
Initial denaturation	27 reaction cycles			Final Extension	Hold
	Denaturation	Annealing	Extension		
30 s (98˚C)	10 s (98˚C)	30 s (X˚C) *	2 min (72˚C)	5 min (72˚C)	5 min (4˚C)
* X= Tm used according to the primer pairs listed in Table 2.2					

2.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was used to analyse and separate DNA samples, mostly using a 1% (w/v) agarose gel. Agarose gels were prepared by dissolving 1 g of agarose powder (Bioline, London) in 100 ml of 1x TAE (40 mM Tris-acetate, 1 mM EDTA, pH 8.8) (Invitrogen, Paisley, UK), then supplemented with 4 µl ethidium bromide (0.5 µg/ml). Loading dye (2 µl of 6x, NEB) was used to mix with 10 µl of samples and run against 3 µl of marker ladder (1 kb DNA ladder, NEB). Electrophoresis was performed at 130 V for 30 min. DNA bands were visualized with UV light in a GelDoc-It UV-Transilluminator (Ultra-Violet Products Ltd, Cambridge, UK).

2.2.4 DNA purification and isolation

Principle

Under high salt concentrations, DNA is absorbed onto a silica membrane; DNA can then be eluted (after contamination has been washed away) with a low salt buffer as described in the manufacturer's protocol.

2.2.4.1 Extraction and purification of plasmid DNA from bacterial cells

Plasmid DNA was purified and isolated from the bacterial cells using QIAquick Miniprep Kit (Qiagen, West Sussex, UK). Cells were harvested from 10 ml of overnight bacterial culture (incubated at 37°C, 200 rpm) by centrifugation at 5000 x*g* for 20 min, then the pellet was resuspended and lysed using an alkaline lysis solution supplied with the kit. Plasmid DNA was extracted using the manufacturer's procedures.

2.2.4.2 Purification of PCR product

After PCR, a QIAquick PCR purification kit (Qiagen) was used to purify the PCR product according the manufacturer's procedures.

2.2.4.3 Extraction and purification of DNA from agarose gel

To isolate a specific DNA band from an agarose gel after digestion with restriction enzymes, a QIAquick gel extraction kit (Qiagen) was used. A sterile scalpel was used to cut out the band of interest after visualization with a UV-Transilluminator. The gel

fragment was then solubilized in a buffer supplied with the Kit at 50ºC to release the DNA for purification according the manufacturer's procedures.

2.2.4.4 DNA purification from enzymatic reactions

The QIAquick MinElute Kit (Qiagen) was used to purify and isolate DNA from enzymatic reactions such as DNA phosphorylation and ligation reactions.

2.2.5 DNA quantification

DNA concentration was determined using a NanoDrop ND 1000 Spectrophotometer (Thermo Fisher Scientific). The appropriate buffer was used as a blank, then the DNA concentration was calculated from measured absorption at 260 nm according to the Beer-Lambert law. In addition to the Spectrophotometer, agarose gel electrophoresis (Section 2.2.3) was used to quantify DNA samples by comparison of the intensity of DNA bands with the known concentration of DNA marker (NEB).

2.2.6 Restriction endonuclease digestion

The production of isolated DNA fragments after PCR will have blunt ends and require digestion with restriction enzymes to produce cohesive ends. The vector also requires digestion with the same restriction enzymes to make linear DNA with complementary sticky ends in preparation for ligation. *Bam*HI (NEB) and *Eco*RI (NEB) were used to digest the PCR products and vector. Reaction mixtures and conditions are shown in Table 2.6. After the second restriction digest, the reaction was stopped by running the DNA on an agarose gel and extracting and purifying from the gel as described in Section 2.2.4.3.

Table 2.6 Reaction mixture and conditions for restriction endonuclease digestion

Purified PCR product digestion		**Vector digestion**	
Materials	Volume µl	Materials	Volume µl
PCR product	16	Vector (150-50 µg/µl)	16
NEB buffer 2.1	2	NEB buffer 2.1	2
BSA	1	BSA	1
*Bam*HI enzyme	1	*Bam*HI enzyme	1
Incubate for 60 min at 37°C		Incubate for 60 min at 37°C	
Then add 1 µl *Eco*RI and incubate for 60 min at 37°C		Then add 1 µl *Eco*RI and incubate for 60 min at 37°C	

2.2.7 Phosphorylation of DNA

The 5' end of DNA fragments needs to be phosphorylated before ligation and recircularization of vector during SDM. T4 polynucleotide kinase (NEB) was used for phosphorylation of DNA in quick ligase buffer (66 mM Tris-HCl, 1 mM dithiothreitol, 7.5% polyethylene glycol (PEG 6000), 10 mM $MgCl_2$, 1 mM ATP, pH 7.6) (NEB). The presence of ATP allows the ligation step to performed without changing the buffer. The total volume of reaction mixture, 21 µl (10 µl of 2 X quick ligase buffer, 10 µl of purified DNA, 1 µl of T4 polynucleotide kinase (10 units/µl)) was incubated for 30 min at 37°C.

2.2.8 Ligation

The linear vector generated by SDM needs to be recircularised and Quick Ligase (NEB) was used. After phosphorylation was completed (Section 2.2.7), 1 µl of DNA quick ligase (NEB) was added directly to the reaction mixture and incubated for 8 min before the reaction was stopped by DNA purification as described in Section 2.2.4.4. For the insertion of a gene of interest into a vector, an insert: vector molar ratio of 3:1 was used, using the following formula to convert of molar ratios to mass ratios:

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

A control reaction was carried out as following:

Materials	Ligation	Control 1	Control 2	Control 3
2 X Quick ligase buffer	√	√	√	√
Vector	√	----	√	√
Insert	√	√	----	√
Quick ligase	√	√	√	----

2.2.9 Construction of expression vectors

2.2.9.1 Construction of Vip1Ac1 expression vectors

Different constructs were generated in order to optimise production of Vip1Ac1 protein in soluble form using different primers as shown in Table 2.2. Restriction endonuclease digestion was performed as described in Section 2.2.6. The primer pair, Vip1AcF1 and Vip1AcR1 was used to delete the N-terminal His-tag and signal sequence from the *Vip1Ac1* gene in pET-28b (the original *Vip1Ac1* gene construct used in this study contained N-terminal His-tag and signal sequence). The SignalP 4.1 Server (http: //www. cbs . dtu .dk/services /SignalP/) was used to predict the signal peptide. The primer pair, Vip1GstF with a *Bam*HI restriction site and Vip1GstR with an *Eco*RI restriction site were used to amplify the *Vip1Ac1* gene without signal sequence, and the product was then inserted into the pGEX-4T-2 vector between the *Bam*HI and *Eco*RI sites to form a fusion product with the *glutathione S-transferase* gene (*GST-Vip1Ac1*). The primer pair, Cyt-b562 F with *Bam*HI and *Eco*RI restriction sites and Cyt-b562 R encoding a tobacco etch virus (TEV) cleavage site were used to add a TEV cleavage site and *Bam*HI and *Eco*RI restriction sites at the C-terminus of the *Cyt* b_{562} gene. The primer pair, Vip1pET-22bF with a *Bam*HI restriction site and Vip1GstR with an *Eco*RI restriction site were used to amplify the *Vip1Ac1* gene and fuse it with *Cyt* b_{562} gene to form a fusion gene (*Cyt* b_{562}-*Vip1Ac1*) in the pET22b vector. The primer pair, Vip1AcF2 and CytoR were used to remove Cyt b_{562} and link the signal sequence of Cyt b_{562} with Vip1Ac1. To facilitate the purification, N-terminal His-tag and TEV cleavage site were added to the latest construct using primer pair Vip-TEV-F and Vip-His-tag-R.

2.2.9.2 Construction of Vip2Ae3 expression vector

To clone the *Vip2* gene without its signal sequence, the SignalP 4.1 Server was used to predict the signal peptide. The gene was amplified using primer pair F-VIP2-cyto and R-VIP2-cyto (Table.2.2) from the original *B. thuringiensis* strain H29.3 (donated by Prof Primitivo Caballero, Public University of Navarra, Pamplona, Spain). This gene was fused with *Cyt* b_{562} gene (containing TEV cleavage site, *Bam*HI and *Eco*RI restriction sites at the C-terminus, previously constructed as described in Section 2.2.9.1) to form a fusion gene (*Cyt* b_{562}-*Vip2Ac1*) in the pET22b vector. The primer pair F-vip2-his and R-vip2-his (Table.2.2), were used to add a His-tag at the N-terminus, remove Cyt b_{562} and link the signal sequence of Cyt b_{562} with Vip2Ac1.

2.2.9.3 Generation of amber codon (TAG) mutants in recombinant proteins

To incorporate ncAAs into proteins, an amber stop codon (TAG) needs to be generated at the position of interest. Whole vector SDM PCR (Section 2.2.2) was used to introduce the TAG codon using appropriate primer pairs as listed in Table 2.2. Two TAG mutants were generated in the *Vip1Ac1* gene in pET22b: first at amino acid 358 to produce the clone pET22b-Vip1^{358TAG}, which was then used as template to introduce the second TAG at amino acid 536.

The pBAD vector carrying the wild-type *Venus* gene was used as a template to introduce TAG in Venus protein at different positions (66, 145, 148, 203 and 204). After PCR was completed (Section 2.2.2), the PCR products were analysed and purified as described in Sections 2.2.3 and 2.2.4.2 respectively. The purified PCR products were subject to phosphorylation and ligation steps to recircularize the vector as described in Sections 2.2.7 and 2.2.8 respectively. The DNA was cleaned up as described in Section 2.2.4.4 and transformed into electrocompetent cells (Section 2.2.10.2). Colonies were picked up and cultured in 10 ml broth media and vector DNA extracted as described in Section 2.2.4.1 and submitted for sequence conformation (Section 2.2.11). The pBAD vector carrying either *sfGFP*^{148TAG} or *sfGFP*^{204TAG} gene were donated from the DDJ lab.

2.2.10 Bacterial transformation

2.2.10.1 Preparation of competent cells

Competent cells were prepared from the original competent cell strains (Table 2.1). Cells were grown on agar plates without antibiotic overnight at 37˚C, then single colony was transferred into 1l of 2xYT medium and incubated at 37˚C until the OD reached between 0.4-0.6. Cells were harvested by centrifugation at 5000 x*g* (Beckman Coulter-rotor JLA-16.25) for 20 min at 4˚C. All the materials and solutions used were autoclaved and pre-chilled at 4˚C.

For chemically competent cell preparation, cells were washed once with sterilized water, harvested and resuspended in 100 ml of $CaCl_2$ (100 mM) and kept in ice for 30 min. Cells were then harvested by centrifugation and resuspended in 50 ml of 85 mM $CaCl_2$, 15% glycerol. Cells were harvested and resuspended again in 1-2 ml of 85 mM $CaCl_2$, 15% glycerol and stored as 50 µl volumes in microfuge tubes.

For electrocompetent cell preparation, cells were washed twice with sterilized water and once with 10% glycerol by centrifugation at 5000 x*g* (Beckman) for 20 min at 4℃. Cells were finally resuspended in 1-2 ml of 10% glycerol and divided into 50 µl volumes in microfuge tubes.

The 50 µl samples of cells for both chemically and electrocompetent cells were snap frozen in liquid nitrogen and kept at -80℃.

2.2.10.2 Electroporation

Vectors were transformed into DH5α electrocompetent cells for vector production for further use or transformed into Top10^{TM} electrocompetent cells for recombinant protein production (L-arabinose operon expression system). Electroporation was performed using a Flowgen Cellject electroporator. Electrocompetent cells (50 µl) were thawed on ice and 3 µl of vector (from ligation reactions) was added to the cells and mixed gently before transfer to an electroporation cuvette (VWR; 2mm cap). Cells were pulsed with 2500 V then 400 µl of 2xYT medium was added and incubated for 1 h at 37˚C to allow the cells to recover. The cells were then plated on LB agar plates supplemented with the appropriate antibiotic, and the plates were incubated overnight at 37˚C.

2.2.10.3 Heat shock

For recombinant protein production, the required vectors were transformed into chemically competent *E. coli* BL21(DE3). Aliquots of cells (50 µl) were thawed on ice, before 3 µl of vector (20-50 ng) was added to the cells and mixed gently, then incubated on ice for 30 min. Cells were incubated for 30 s in a water bath at 42˚C to heat shock and were immediately transferred to ice for 4 min. Fresh medium (400 µl 2xYT medium) was added and incubated for 1 h at 37˚C for recovery. The transformed cells were plated on LB agar plates supplemented with the appropriate antibiotic, and the plates were incubated overnight at 37˚C.

2.2.11 Sequencing

All the constructed vectors were sequenced to verify the mutations and gene inserts using the Eurofins Genomics DNA sequencing service.

2.2.12 Preparation of *E. coli* dimethyl sulphoxide (DMSO) stocks

E. coli cells with desired expression vectors were stored at -80˚C for long term storage in 8 % DMSO. From 5 ml of overnight bacterial culture grown in 2xYT medium with the appropriate antibiotic, 920 µl was transferred into sterile Eppendorf tube, then 80 µl of DMSO was added and mixed thoroughly using a vortex and kept at -80˚C. DMSO stocks were used to prepare starter cultures, by transfer of a scraping from the part of frozen cell-DMSO mixture into 10 ml 2xYT broth containing the appropriate antibiotic, using a sterile loop.

2.3 Recombinant protein expression and purification methods

2.3.1 Optimizing of Vip1Ac1 protein expression in *E. coli*

Starter cultures of 5 ml 2xYT broth (supplemented with the appropriate antibiotic) were inoculated with a single bacterial colony and incubated with shaking (200 rpm) at 37℃ overnight. Next day, 100 µl from the starter culture was transferred into 10 ml 2xYT medium and grown in preparation for induction. A variety of IPTG concentrations, cell culture attenuance (D_{600}) levels, incubation temperatures and induction times were used to optimize conditions for maximum protein expression from each construct. Recombinant protein expression was detected by SDS-PAGE analysis (Section 2.4.1).

2.3.2 Expression of wild type Vip1Ac1, Vip2Ae3 and Venus proteins in *E. coli*

A starter culture of 10 ml 2xYT broth supplemented with ampicillin was inoculated with a single bacterial colony carrying the gene of interest and incubated with shaking (200 rpm) at 37˚C overnight. Next day, 1 l of 2xYT broth supplemented with ampicillin was inoculated with a 1/100 dilution from the starter culture and incubated at 37˚C until a D_{600} of 0.7 (± 0.1) was reached. Then the culture was induced for protein expression by adding either (i) IPTG to a final concentration of 0.1 mM and incubating at 20˚C for 16 h for Vip1Ac1 and Vip2Ae3, or (ii) adding 0.1% arabinose and incubating at 25˚C for 16 h for Venus expression. Induced cells producing either Vip1Ac1 or Vip2Ae3 were lysed as described in Section 2.3.6.2. Induced cells containing Venus were lysed as described in Section 2.3.6.3 and recombinant protein expression was detected by SDS-PAGE analysis (Section 2.4.1).

2.3.3 Expression of C-terminal His-tag tobacco etch virus (TEV) protease in *E. coli*

A starter culture of 10 ml 2xYT broth supplemented with ampicillin and chloramphenicol was inoculated with a single bacterial colony containing the pRK793 vector carrying the gene encoding TEV and incubated with shaking (200 rpm) at 37˚C overnight. The 10 ml starter culture was used to inoculate 1 l culture, which was incubated at 37˚C until an OD_{600} of 0.7 (± 0.1) was reached, when the culture was induced by the addition of IPTG to a final concentration of 1 mM. The culture was then incubated at 30˚C for 16 h (Tropea *et al.* 2009). Induced cells were lysed as described in Section 2.3.6.3 and recombinant protein expression was detected by SDS-PAGE analysis (Section 2.4.1).

2.3.4 Expression of recombinant protein containing ncAAs in *E. coli*

Incorporation of ncAAs required two vectors in an expression system, one vector carrying the gene of interest with desired TAG mutation (Section 2.2.9.3) and another vector carrying the genes of an engineered tRNA/aaRS pair for the specific ncAA (Table 2.2; Figure 2.1). Both vectors with two different antibiotic selectivities were mixed in equal concentrations and transformed into appropriate expression competent cells as described in Section 2.2.10. Cells were then cultured on LB agar plates containing both

antibiotics for selection of successfully co-transformed colonies. Incorporation of ncAA was assessed depending on the role of amber stop codon as signal to stop protein translation process. Three expression cultures (10 ml each one) containing appropriate antibiotics were inoculated with 0.1 ml of expression *E. coli* (containing two plasmids for ncAA incorporation). Culture number one was used as a negative control for expression and contained ncAA but no IPTG, so that no recombinant protein should be produced. Culture number two was used as a positive control for the presence of the amber stop codon (TAG) in the gene and contained IPTG but no ncAA so that no full-length protein was produced as the amber stop codon works as a stop signal for protein translation and only truncated protein is produced. The last culture was used as the positive control for incorporating ncAA and contained ncAA and IPTG to produce full-length protein, indicating that the amber stop codon is functioning as signal for ncAA incorporation. All cultures were incubated at the appropriate expression temperature and time for the protein of interest, followed by SDS-PAGE analysis (Section 2.4.1).

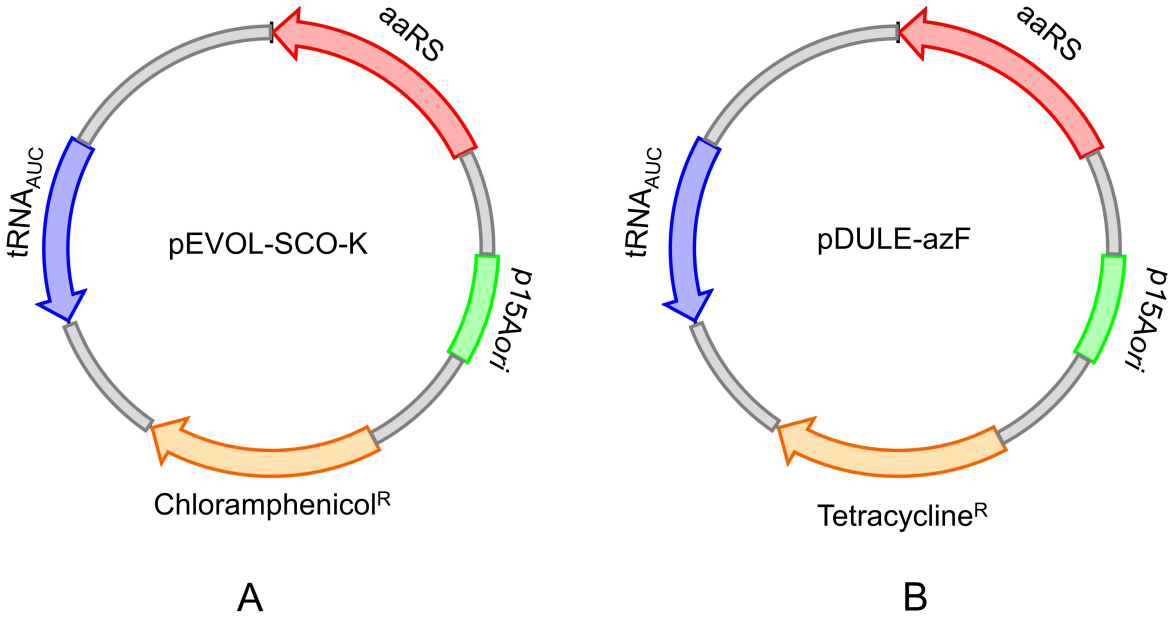

Figure 2.1. Vectors map used for ncAA incorporation.

Vector map showing the key vector elements of **A.** pEVOL vector for SCO-K incorporation and **B.** pDULE vector for azF incorporation.

2.3.4.1 Expression of Vip1Ac1 containing ncAAs

A starter culture of 10 ml 2xYT supplemented with ampicillin and tetracycline was inoculated with a single bacterial colony containing two vectors (pET22b-Vip1$^{358\text{-}536TAG}$ and pDULE (Hammill *et al*. 2007)) and incubated with shaking (200 rpm) at 37˚C overnight. 1 l of 2xYT containing 0.1% arabinose (for pDULE vector induction), 0.5 mM of desired ncAA and supplemented with ampicillin and tetracycline was inoculated with a 1/100 dilution from the starter culture and incubated at 37˚C until a D_{600} of 0.7 (± 0.1) was reached. The culture was then induced for Vip1Ac1 protein expression by adding IPTG to 0.1 mM and incubation at 20˚C for 20 h. Induced cells were harvested and lysed as described in Section 2.3.6.2. Incorporation of ncAA was detected by SDS-PAGE analysis (Section 2.4.1). When azF was used, all stages were performed in the dark to prevent photolysis of azF by ambient light.

2.3.4.2 Expression of Venus containing ncAA(azF)

A starter culture of 10 ml 2xYT broth supplemented with ampicillin and chloramphenicol was inoculated with a single bacterial colony containing two vectors (pBAD-Venus^{TAG} and pDULE (Hammill *et al*. 2007)) and incubated with shaking (200 rpm) at 37˚C overnight. 2xYT medium (10 ml,100 ml and 500 ml) containing 0.5 mM of azF and supplemented with ampicillin and tetracycline was inoculated with a 1/100 dilution from the starter culture and incubated at 37˚C until a D_{600} of 0.7 (± 0.1) was reached, when the culture was induced by the addition of 0.2% arabinose and incubated at 25˚C for 20 h. Induced cells were harvested, and lysed as described in Section 2.3.6.3. Incorporation of azF was detected by SDS-PAGE analysis (Section 2.4.1). All the stages were carried out in the dark to prevent photolysis of azF by ambient light.

2.3.4.3 Expression of sfGFP containing ncAA (SCO-K)

A starter culture of 10 ml 2xYT supplemented with ampicillin and chloramphenicol was inoculated with a single bacterial colony containing two vectors (pBAD-sfGFP$^{148 \text{ or } 204TAG}$ and pEVOL (Plass *et al*.2011)) and incubated with shaking (200 rpm) at 37˚C overnight. The 2xYT medium (10 ml,100 ml and 500 ml) containing 0.5 mM of SCO-K and supplemented with ampicillin and chloramphenicol was inoculated with a 1/100 dilution from the starter culture and incubated at 37˚C until a D_{600}

of 0.7 (± 0.1) was reached, when the culture was induced by the addition of 0.2% arabinose and incubated at 37˚C for 20 h. Induced cells were harvested, and lysed as described in Section 2.3.6.3. Incorporation of SCO-K was detected by SDS-PAGE analysis (Section 2.4.1)

2.3.5 Expression of Vip1Ac1 containing selenomethionine (SeMet)

A starter culture of 5 ml 2xYT broth supplemented with ampicillin was inoculated with a single bacterial colony containing the pET22b vector carrying the gene encoding Vip1Ac1 and incubated with shaking (200 rpm) at 37˚C overnight. Next day, 2 ml of the starter culture were spun down at 5000 x*g* for 5 min at 4˚C. Cells were washed twice with M9 salt medium and used to inoculate 1 l of M9 salt medium containing (glucose, thiamine and biotin) as detailed in Section 2.1.2 and incubated at 37˚C until a D_{600} of 0.7 (± 0.1) was reached, when methionine biosynthesis was inhibited by adding 6 different amino acids (phenylalanine, lysine, threonine, valine, leucine and isoleucine) and selenomethionine, as described in Section 2.1.2. After 15 min selenomethionine incorporation was induced by adding IPTG to a final 0.1 mM and incubation at 20˚C for 20 h. Induced cells were lysed as described in Section 2.3.6.2 and protein expression was detected by SDS-PAGE analysis (Section 2.4.1) and by mass spectrometry. Selenomethionine incorporation efficiency was calculated as following:

$$\text{Incorporation efficiency } \% = \frac{(MW^{\text{SeMet-experimental}}) - (MW^{\text{Wt-experimental}})}{(MW^{\text{SeMet-theoretical}}) - (MW^{\text{Wt-theoretical}})} \times 100$$

Here, $MW^{\text{SeMet-experimental}}$ and $MW^{\text{Wt-experimental}}$ are the experimental molecular weights of protein containing selenomethionine and wild-type protein respectively, as determined by mass spectrometry. $MW^{\text{SeMet-theoretical}}$ and $MW^{\text{Wt- theoretical}}$ are the theoretical molecular weights of protein containing selenomethionine and wild-type protein respectively, calculated by protein calculator (www.scripps.edu/~cdputnam/protcalc.html).

2.3.6 Cell lysis and determination of recombinant protein solubility

2.3.6.1 Cell lysis by sonication

Cells containing induced recombinant protein were harvested by centrifugation at 5000 x*g* for 20 min at 4˚C. Cell pellets from 10 ml culture were suspended in 1 ml of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) for Ni-affinity purification or PBS lysis buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.3) for GST-affinity purification. Then the cell mixture was lysed using 0.25 mg/ml lysozyme, PMSF was added to 1 mM and benzonase nuclease (Sigma) was added to 20 µg/ml. Samples were incubated on ice for 30 min, and then sonicated for 2 min at 200–300 W with 1 min pause for two cycle of sonication. The lysate was centrifuged at 10,000 x*g* at 4˚C for 30 min. The supernatant (containing the soluble protein) was collected and the pellet (insoluble protein) was resuspended in 1 ml lysis buffer, then 10 µl of each sample were analysed via SDS-PAGE (Section 2.4.1).

2.3.6.2 Periplasmic extraction

Extraction of periplasmic proteins was performed by the periplasmic shock method as follows. The cells were harvested by centrifugation at 5000 x*g* for 20 min (Beckman). Harvested cells were suspended in TSB buffer (Tris-HCl 30 mM, EDTA 1 mM, sucrose 20%, pH 8.0; $1/15^{th}$ culture volume) then incubated on ice for 15 min with gentle agitation. The cells were harvested by centrifugation at 5000 x*g* for 20 min. The supernatant was removed, and the pellet was re-suspended in$1/15^{th}$ culture volume of ice-cold 5 mM $MgSO_4$ containing 1 mM PMSF and 1 mM EDTA. After shaking for 15 min in an ice bath, the mixture was centrifuged at 5000 x*g* for 20 min and the supernatant containing periplasmic proteins was collected, then analysed by SDS-PAGE (Sections 2.4.1).

2.3.6.3 Cell lysis by French Press

Cells harvested by centrifugation at 5000 x*g* for 20 min (Beckman) were resuspended in 1/25 volume original culture volume of 50 mM Tris-HCl buffer pH 8.0 containing protease inhibitors (complete, Roche, 1 tablet in 50 ml). The cell suspension (maximum 20 ml) was lysed under pressure (1250 psi) using a French pressure cell press. The lysate

was collected and centrifuged at 20,000 rpm (Beckman Coulter- rotor JA-25.50) for 40 min to separate a clear lysate from cell debris.

2.3.8 Protein purification

An ÄKTA purifier FPLC was used for protein purification in conjunction with appropriate protein purification columns (GE Healthcare). Proteins were monitored by absorption at 280 nm and analysed by SDS-PAGE. Protein samples were concentrated to 2 ml using centrifugal concentrator units (MWCO: 10 kDa and 3 kDa, Merck Millipore) by centrifugation at 5000 x*g* before loading onto the columns.

2.3.8.1 Nickel affinity chromatography (gravity column)

Protino^{R} Ni-TED 2000 Packed Columns, 1 ml bed volume with 5 mg protein binding capacity (Machery-Nagel, Germany) and 1 ml of HisPur^{TM} Cobalt Superflow agarose (20 mg/ml binding capacity, Thermo scientific) poured in liquid chromatography column (0.1 cm x 10 cm , Sigma) were used to purify polyhistidine-tagged proteins according the manufacturer's procedures apart from using Tris buffer instead of using the buffer provided. The column was equilibrated by adding 4 bed volumes of equilibration-wash buffer (50 mM Tris-HCl, pH 8) and was then allowed to drain by gravity. Periplasmic extraction or cleared cell lysate was added to the pre-equilibrated column and allowed to drain by gravity. After that, the column was washed twice with 4 bed volumes of equilibration-wash buffer before polyhistidine tagged protein was eluted with 4 bed volumes of elution buffer (50 mM Tris- HCl, 250 mM imidazole, pH 8). Three elution fractions were collected and analysed by SDS-PAGE.

2.3.8.2 Ion Exchange Chromatography (IEX)

Before using IEX, the net charge of Vip1Ac1 protein was determined using the Bachem website (http://www.bachem.com/service-support/peptide-calculator) in order to choose the right column (cation or anion exchange). A HiTrap Q HP 5 ml column connected to the ÄKTA purifier was used for IEX, using the following buffers: 50 mM Tris-HCl pH 8 as starting buffer (buffer A) and 50 mM Tris-HCl, 500 mM NaCl pH 8 (buffer B) to generate the salt gradient over 0-500 mM NaCl. The column was equilibrated with buffer A. A 5 ml volume of protein sample was then loaded onto the column using a 5 ml loop. The column was washed with 4 column volumes of buffer A

before target protein was eluted with a linear gradient of 0-500 mM NaCl applied over 6 column volumes. The elution of protein was monitored by absorption at 280 nm and elution fractions were collected and analysed by SDS-PAGE.

2.3.8.3 Size Exclusion Chromatography (SEC)

A Hiload™ 16/60 Superdex™ S200 pg column (preparative grade, 120 ml bed volume, separation range 10 kDa to 600 kDa) was used for purification of Vip1Ac1, Vip2Ae3 and TEV enzyme as well as for separation the activated oligomer form from the monomer form of Vip1. A Hiload™ 16/600 Superdex™ S75 pg (preparative grade, 120 ml bed volume, separation range 3 kDa to 75 kDa) was used for Venus variants, sfGFP variants and hetero-dimer (Venus-sfGFP) purification. Prior to the application of protein, the column was equilibrated with Tris-HCl buffer (50 mM pH 8). A 2 ml protein sample was loaded onto the column using a 2 ml loop. Samples were run through the column at 0.5 ml per min, the elution of protein was monitored by absorption at 280 nm and 1-0.5 ml fractions were collected. Collected fractions were analysed by SDS-PAGE to assess the purity of the target protein and pure fractions were pooled for further studies.

2.4 Protein analysis methods

2.4.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS–PAGE was performed in a similar manner to the classical approach described by Laemmli (Laemmli 1970). A gel was prepared and run using the mini-PROTEIN electrophoresis system (Bio-Rad). Normally, 12.5% polyacrylamide gels were used; the composition of separating and stacking gels are shown in Table 2.5. The running buffer was composed of 25 mM Tris-HCl pH 8.3, 192 mM glycine and 0.1% (w/v) SDS. TruPAGE™ Precast (4-20%) gels (Sigma) were also used with the running buffer supplied with the gels. Protein sample loading buffer 4X (40% Glycerol, 240 mM Tris-HCl pH 6.8, 8% SDS, 0.04% bromophenol blue, 5% beta-mercaptoethanol) was mixed with the samples at a final concentration of 1X and the samples were boiled for 5 min before loading onto the gel. Gels were run at 200 V for 40 min, then 50 ml of staining solution (50% ultra-pure water (v/v), 40% (v/v) methanol, 10% (v/v) acetic acid, 0.1% (w/v) Coomassie blue R250) was used to stain the gels on a shaker for 1 h. Gels were

destained using a destain solution (20% (v/v) methanol, 10% (v/v) acetic acid, 70% ultra-pure water (v/v)) until the background colour was sufficiently removed. Prestained protein marker 3 µl (NEB) was used on each gel to allow estimation of the molecular weights of protein samples.

Table 2.5 Composition of separating and stacking gels

Materials	separating gel	stacking gel
Tris-HCl [a]	0.375 mM	65 mM
Acrylamide [b]	12.5% (w/v)	5% (w/v)
SDS	0.1 % (w/v)	0.2 % (w/v)
APS	0.05 % (w/v)	0.1 % (w/v)
TEMED	0.02 % (w/v)	0.02 % (w/v)

[a] pH 8.8 and 6.8 solutions were used for separating gel and stack gel respectively.

[b] Acrylamide/bis Acrylamide 37.5:1 40 % solution

2.4.2 Determination of protein concentration

Bio-Rad DC protein assay was used for protein quantification as described in manufacturer's instructions using the microplate protocol. A 5 µl volume of protein standard and samples were added in triplicate into a 96 well plate (Nunc, Thermo-Fisher) followed by DC assay reagents as described in the manufacturer's instructions before incubation at 25˚C for 15 min. The absorbance was then measured at 750 nm using a microplate reader (CLARIOstar, BMG LABTECH, UK). A standard curve of bovine serum albumin (BSA) with a range of concentrations (0.1 -1 mg/ml) was used to determine the protein concentration of Vip1, Vip2 and TEV enzyme, while a standard curve of WT Venus with a range of concentrations (0.1 -1 mg/ml) was used to determine the protein concentration of Venus variants and sfGFP variants.

2.4.3 Removal of the fusion partner (Cyt b_{562}) from Vip1Ac1 protein and His-tag

TEV protease (containing C-terminal His-tag), which was prepared as described in Section 2.3.3 and purified as described in Sections 2.3.8.1 and 2.3.8.3 was used to cleavage the fusion partner (Cyt b_{562}) from the N-terminus of Vip1Ac1 protein. A TEV protease recognition site (ENLYFQ/G) had been engineered between Cyt b_{562} and Vip1Ac1 for this purpose. Different ratios of TEV protease and fusion protein were used (1:1, 2:1, 3:1 4:1 and 5:1) for 24 h at room temperature. Different buffer conditions were also used Tris-HCl pH 8 buffer containing 0.5 mM EDTA and 1 mM DTT (optimum conditions for TEV protease). Samples were analysed by SDS–PAGE (Section 2.4.1). To remove the His-tag from the protein sample, it was incubated with TEV protease at ratio 100:1 for 1 h at room temperature before purification from cleavage the His-tag and TEV protease (containing its own C-terminal His-tag) as described in Section 2.3.8.1 except that the flow through, which contains the protein sample, was collected.

2.4.4 UV-visible absorption spectroscopy and calculation of extinction coefficients

Venus and sfGFP absorbance spectra (UV-visible) were recorded using Cary spectrophotometer in 1 cm pathlength cuvettes (Hellma, Müllhein, Germany). A maximum of 500 µl of protein sample was pipetted into a cuvette and the spectrum was recorded from 200-800 nm. The extinction coefficients of Venus variants were calculated by recording absorbance spectra, as above, of protein samples with known concentrations (10-20 mM) (Section 2.4.2). The values of concentration and absorbance at λ_{max} were then substituted into the Beer-Lambert law equation (Equation 2.1).

Equation 2.1 $$A = \varepsilon.c.\ l$$

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

2.4.5 Fluorescence spectroscopy

Excitation and emission spectra of fluorescent proteins were recorded on a Cary Eclipse Fluorimeter (Varian) using 5 mm x 5 mm QS quartz cuvette (Hellma). A maximum of 400 µl (0.5 µM except for Venus ^{Y66azF} variant, which was used at 10 µM) of protein sample was used for measurement. A fixed scan rate of 120 nm/min with a 5 nm slit width and medium scan control was used to record the spectra. First, the maximum excitation wavelength (λ_{ex}) for the fluorescent protein variant was determined from its absorbance spectrum as described in Section 2.4.4 and used as a fixed excitation wavelength to record the emission spectrum up to 650 nm. Excitation spectra were recorded by monitoring excitation at a fixed maximum emission wavelength (λ_{em}) for each variant over a range of wavelength down to 350 nm.

2.4.6 Fluorescence quantum yield determination

Quantum yield (QY) is defined as a ratio of emitted photons to absorbed photons. A comparative method was used to determine the QY of fluorescent proteins using known QY molecules (fluorescein with QY 0.95, λ_{max} 496 nm) as the reference. In brief, several dilutions of fluorescein in 0.1 M NaOH were prepared, having a range of absorbances (0.012, 0.29, 0.053, 0.069 and 0.094) at the excitation wavelength 496 nm. Dilutions of protein samples were prepared in 50 mM Tris-HCl buffer pH 8 to produce the same absorbances as the reference ± 0.001 (0.012,0.29, 0.053, 0.069 and 0.094) at the specific λ_{max} of each protein sample. Then the emission spectrum for each dilution was recorded as described in Section 2.4.5 and the integrated fluorescence intensity for each dilution was calculated using a|e - UV-Vis-IR spectral analysis software (www.fluortools.com). The integrated fluorescence intensity was plotted against the absorbance as shown in Figure 2.2 and the slope of the linear fit line was determined using excel software. Then the values of slopes were substituted into Equation 2.2 to calculate the QY of protein samples.

Equation 2.2 $$QY = QY_R \left(\frac{M}{M_R}\right)\left(\frac{N}{N_R}\right)$$

Here, M is the slope, N is the refractive index of solvent and the subscript R corresponds to the reference (fluorescein). As the refractive index of aqueous solvent and 0.1 M NaOH differ by <1% (Arpino, 2011), the refractive index was ignored.

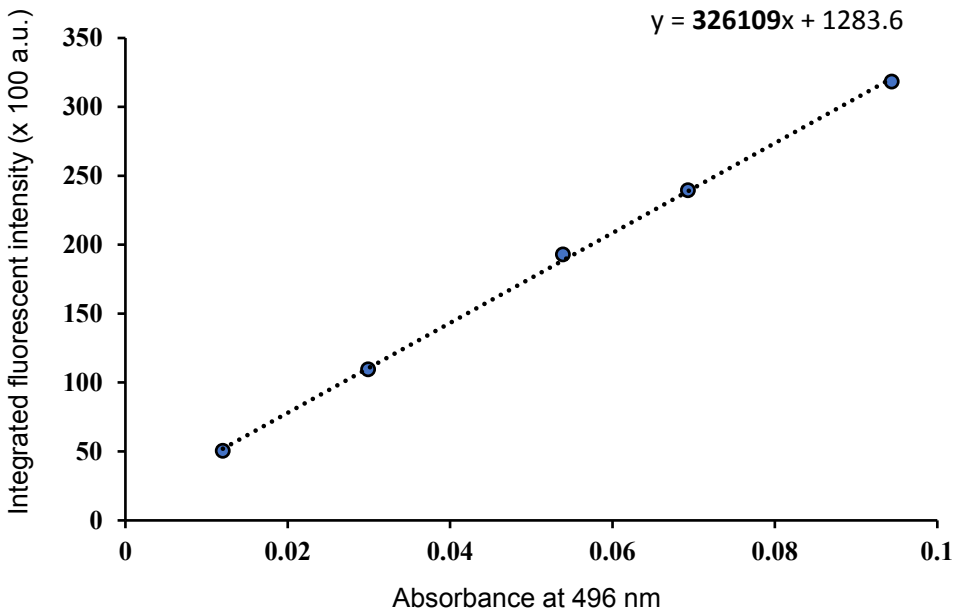

Figure 2.2. Integrated fluorescence intensity curve of the fluorescein.
The integrated fluorescence intensity of the 5 different dilutions of the reference (fluorescein) absorbance at 490 nm.

2.4.7 Relative FRET efficiency calculations

Relative förster resonance energy transfer (FRET) efficiency between the monomeric units (sfGFP; donor and Venus; acceptor) in the heterodimer (GFVen^{204}) was measured experimentally and calculated using Equation 2.3 after spectral decomposition for the GFVen^{204} dimer.

$$E_{rel} \quad \frac{I_A}{I_A + I_D}$$

Here, E_{rel} is relative FRET efficiency, I_A is the integrated fluorescence of acceptor (Venus^{Q204azF}), I_D is the integrated fluorescence of donor (sfGFP$^{Q204SCO-K}$), when excited at the donor's wavelength (450 nm).

2.4.8 Photolysis of Venus^{WT} and Venus ^{AzF} variants

Photolysis experiments of Venus and its variants were carried out using a UVM-57 Handheld UV lamp (6 W; 302 nm UV, UVP Cambridge, UK) and 1 cm pathlength quartz cuvette (Hellma). A maximum of 500 µl of protein sample (10 µM) was pipetted into a cuvette and exposed to the UV (302 nm) for the indicated periods of time at a distance of 1 cm. The absorbance spectra, excitation and emission spectra were recorded immediately afterwards, as described in Sections 2.4.4 and 2.4.5 respectively.

2.4.9 Trypsin activation of Vip1Ac1 protein

Vip1Ac1 full length protein (~82 kDa) was activated with trypsin to allow the activated form (~60 kDa) to generate an oligomer. Trypsin was added in ratio of 1/50 to Vip1Ac1 and incubated on ice. To monitor the activation process, 10 µl aliquots was taken at different periods of time (0, 10, 15, 20, 30, 45, 60 and 75 min) before the reaction was stopped by adding PMSF at a final concentration 1 mM and SDS loading buffer then boiling for 10 min. All aliquots were then analysed by SDS–PAGE (Section 2.4.1). The activated form (~60 kDa) was purified from the full-length form and the smaller cleavage product, as described in Section 2.3.8.3. and Section 2.3.8.2. Purified protein was kept at 4˚C for further analysis.

2.4.10 Determination of molecular weight by gel filtration chromatography

The oligomeric state of activated Vip1Ac1 and the estimated molecular weight of Vip1Ac1-Vip2Ae3 complex were determined by gel filtration chromatography as well as by SDS-PAGE as described in Section 2.4.1. First, A Hiload^{TM} 16/60 Superdex^{TM} S200 pg column was calibrated with gel filtration standard proteins (Bio-Rad) of molecular weight (670, 158, 44, 17 and 1.35 kDa) at a flow rate of 0.5 ml/min with 50 mM Tris-HCl buffer pH 8. Protein sample was run on the column under the same conditions. The partition coefficient (K_{av}) for each protein (standard and sample) was calculated using Equation 2.4. Then the estimated molecular mass of the test protein sample was calculated using molecular weight standard curve as shown in Figure 2.3.

Equation 2.4 $$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

Here, the void volume (V_0) of the column was determined using blue dextran (2000 kDa), total volume V_t is 120 ml (column volume) and the elution volume V_e for each sample was monitored by absorption at 280 nm (λ_{max} of elution peak) as shown in Figure 2.3.

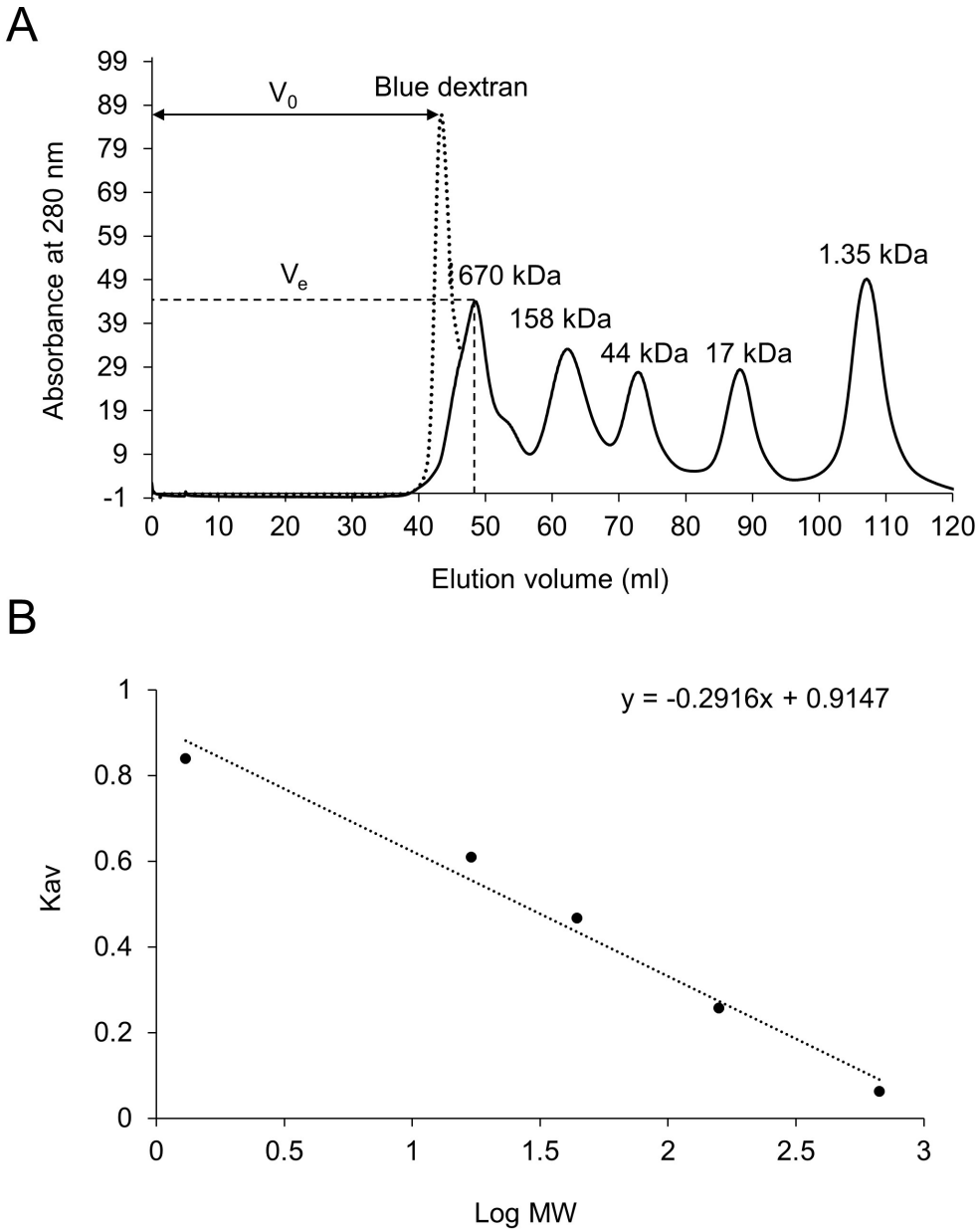

Figure 2.3. Determination of molecular weight by gel filtration chromatography.
A. Elution profile of standard proteins of molecular weight 670, 158, 44, 17 and 1.35 kDa run in a Hiload™ 16/60 Superdex™ S200 pg column at a flow rate of 0.5 ml/min with 50 mM Tris-HCl buffer pH 8. **B**. The plot of the log of molecular weight (log MW) against the partition coefficient (K_{av}).

2.4.11 Mass spectrometry

Experimental analysis of the molecular mass of target proteins (10 μM) was performed at Cardiff University School of Chemistry's Mass Spectrometry suite. Theoretical molecular mass was calculated using ExPASy ProtParam tool (Gasteiger *et al*. 2005). The theoretical mass for the ncAA-containing fluorescent protein was calculated as follows. First for the azF variants, the mutation site was replaced with F and for the SCO-K variants the mutation site was replaced with K to get the theoretical mass based on this amino acid sequence. Then the theoretical mass was altered to account for chromophore maturation (-20 Da) and the addition of the functional group of the ncAA (+41 Da for the azide group and +150 Da for the cyclooctyne group)(Worthy 2018). To generate a theoretical heterodimer mass, the two altered monomer masses were added together.

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

2.5.1 Detection of ncAAs incorporation via Click chemistry

Click chemistry via a SPAAC reaction was used to confirm incorporation of (azF) in Venus at positions 148 and 204 and Vip1Ac1 at positions 358 and 536. Purified protein containing (azF) (10 μM in 50 mM Tris-HCl buffer pH 8) was mixed with dibenzocyclooctyne tetramethyl rhodamine fluorescent dye (DBCO-TAMRA) at a final concentration of 25 μM and incubated in the dark with gentle mixing at room temperature for 24 h. After incubation, samples were analysed by SDS-PAGE (Section 2.4.1) and before staining, fluorescent protein bands were visualised and imaged using a GelDoc-It (Bio-Imaging Systems).

2.5.2 Creation of artificial heterodimer

Equal volume of purified Venus^{H148azF} (150 μM) and purified sfGFP$^{148SCO\text{-}K}$ (150 μM), or, in separate reactions, Venus^{Q204azF} and sfGFP$^{Q204SCO\text{-}K}$, were mixed together and incubated in the dark with gentle mixing at room temperature for 24 h. After incubation, samples were analysed by SDS-PAGE (Section 2.4.1) to confirm the conjugation of the two proteins. Successfully formed dimer was purified as described in Section 2.3.8.3. The purified dimer was kept at 4˚C for further analysis.

2.5.3 Labelling Vip1 protein containing ncAA with florescent dye

The dibenzocyclooctyne tetramethyl rhodamine (DBCO-TAMRA) fluorescent dye was used for labelling Vip1Ac1$^{358\text{-}536azF}$ at two positions. Fluorescent dye was added to purified Vip1Ac1$^{358\text{-}536azF}$ at a ratio of 2.1 and incubated in the dark with gentle mixing at room temperature for 24 h. After incubation, the sample was analysed by SDS-PAGE (Section 2.4.1) and before staining, labelled protein was visualised using a GelDoc-It (Bio-Imaging Systems). Successfully labelled protein was purified as described in Section 2.3.8.3 to remove the excess of fluorescent dye. Purified labelled protein was kept at 4˚C for further used.

2.6 Lipid vesicle preparation

Lipid vesicles (composed of one or more lipid bilayer) were used as membrane models to assess binding affinity and pore formation of Vip1Ac1 (the binding component of binary toxin). A stock solution 50 mg/ml of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) lipid (Avanti® polar lipids, INC, USA) was prepared in chloroform and kept at -20˚C.

2.6.1 Preparation of giant lipid vesicles (GLVs)

The mixture of giant vesicles, composed of one or multiple lipid bilayers, were prepared by a heat and sonication method (Walde *et al.* 2010). Lipid (100 µl) from stock (Section 2.6) was transferred into a glass tube and chloroform was evaporated using a dry nitrogen stream. Once the lipid was dry, 5 ml of 50 mM Tris-HCl buffer pH 8 was added and mixed by vortexing and was then placed in a 40˚C water bath (Grant, England) for 1 h. After that, the mixture was sonicated for 5 min using an ultrasonic bath (Grant XB2), before 50 µl of the solution was examined under a microscope to assess lipid vesicle formation. Successfully produced lipid vesicles were used directly to assess the binding of Vip1Ac1 protein or kept at room temperature for maximum 3 days.

2.6.2 Preparation of DNA-liposomes

Liposomes are small spherical lipid vesicles (composed of at least one lipid bilayer) and were used to encapsulate DNA for use in a YO-PRO-1 uptake assay as described in Section 2.8. Lipid (200 µl) from stock (Section 2.6) was transferred into a glass tube and the chloroform was evaporated using a dry nitrogen stream. When the lipid was

completely dry, 1 ml of 50 mM Tris-HCl buffer pH 8 was added and mixed by vortexing for 10 s and was then extruded 10 times through a polycarbonate filter (200 nm pore size) using an Avanti Mini-Extruder. DNA at a final concentration of 40 µM was incorporated into liposomes by freezing in liquid nitrogen and thawing at room temperature for three cycles. After that, the lipid was re-extruded 10 times. To digest unencapsulated DNA, 0.5 µl of DNase l (Sigma-Aldrich, 2 U/µl) was added and incubated for 1 h at room temperature. The DNA- liposomes were washed three times with 50 mM Tris-HCl buffer pH 8 by repeating centrifugation at 50,000 rpm for 45 min at 4˚C (Optima TLX Ultracentrifuge, Beckman). DNA-liposomes were directly used in the YO-PRO-1 uptake assay.

2.7 Binding of Vip1Ac1 to lipid vesicles

To evaluate the ability of Vip1Ac1 protein to bind to the membrane models (lipid vesicles), Vip1Ac1 protein was labelled with a fluorescent DBCO-TAMRA dye (Section 2.5.3). The fluorescent labelled full-length Vip1Ac1 and activated Vip1Ac1 at a final concentration 100 µM with respect to the monomeric form (Section 2.4.9), were added to the lipid vesicles 1 mg/ml (Section 2.6.1) and incubated for 1 h at room temperature. Samples were imaged by fluorescence microscopy (Section 2.9.2).

2.8 YO-PRO-1 uptake assay

The green-fluorescent YO-PRO-1 Iodide (491/509) stain (Invitrogen^{TM}, fisher scientific, UK) was used as described by Karasawa *et al.* (Karasawa *et al.* 2017) with slight modification to assess the ability of Vip1Ac1 to form a pore in an artificial lipid bilayer membrane as illustrated in Figure 2.4. YO-PRO-1 is a green fluorescent dye known as a nucleic acid marker owing to its ability to bind to, and become fluorescent after, binding to nucleic acids. Briefly, 10 µl of DNA-liposomes (10 mg/ml) (Section 2.6.2) was mixed with 80 µl of YO-PRO-1 (5 µM final concentration in 50 mM Tris-HCl buffer pH 8 or phosphate buffer pH 5.5), then the mixture was placed in a 96 well plate (Nunc). Target protein (10 µl of 50 µM protein in 50 mM Tris-HCl buffer pH 8) was added and incubated in the dark for 5 min at room temperature before fluorescence intensity was recorded using a microplate reader (CLARIOstar). A positive control was used in which 10 µl of 1% Triton X100 was added instead of the protein, to assess the DNA-liposome formation and lysis. Fluorescence intensity was normalized to the

fluorescence intensity of the negative control (10 µl of DNA-liposomes, 80 µl of YO-PRO-1 and 10 µl of Tris-HCl buffer pH 8 or phosphate buffer pH 5.5).

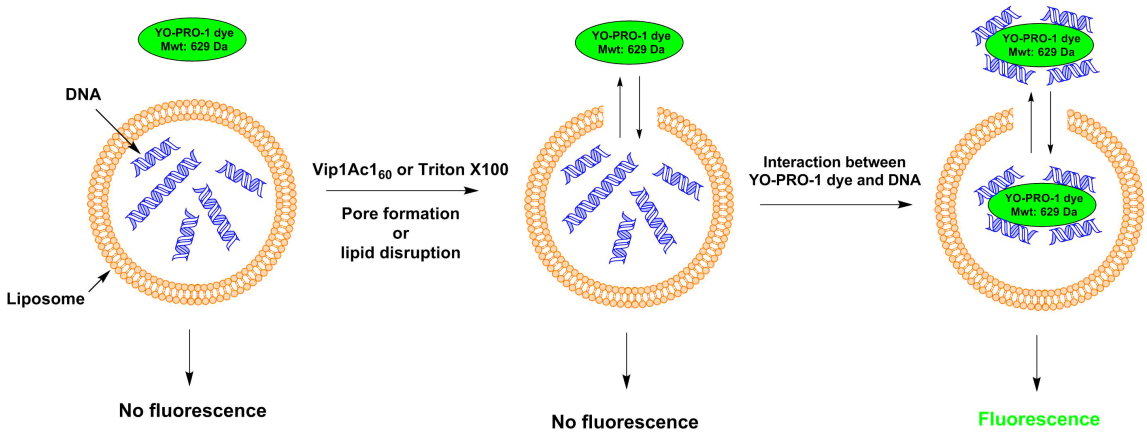

Figure 2.4. Schematic representation of proposes interaction between YO-PRO-1 dye and entrapped DNA in liposomes following pore formation.
After the liposome membrane is disrupted by either Vip1Ac1$_{60}$ or Triton X100 the interaction occurred between the dye and DNA causing the fluorescence of the YO-PRO-1 dye.

2.9 Fluorescence microscopy

2.9.1 Bacterial live cell imaging

Widefield fluorescence microscopy was used for bacterial live cell imaging. For photoactivation of Venus^{Y66azF}, *E. coli* Top10 cells were first induced to express Venus^{Y66azF} (Section 2.3.4.2). A Coverwell^{TM} imaging chamber (Sigma-Aldrich) was fixed on a glass slide and 0.2 ml of 1% agarose was applied before the induced bacterial cells (0.2 ml) were added and covered with cover slide. A sample of uninduced bacterial cells was prepared in the same manner as a control. Slides for induced and uninduced bacterial cells were exposed to UV-light for 1 min at a distance of about 1 cm. Transmitted and widefield fluorescence images visualised using an inverted Olympus IX73 widefield fluorescence microscope. Images were collected with a Hamamatsu Orca flash 4.0 camera with x100 objective lens using HCImaging software and a Prior Lumen200Pro light source. Fluorescence emission was separated by a multiband dichroic emission filter set 69002 (Chroma), a wavelength of 450 nm was used for the excitation.

2.9.2 Lipid vesicle imaging with fluorescent modified Vip1Ac1

To visualise the fluorescent modified Vip1Ac1 after interaction and binding with lipid vesicles (Section 2.7), samples were prepared by dispersing 0.4 ml of the mixture (Vip1Ac1 with lipid) on the Coverwell^{TM} imaging chambers. Transmitted and widefield fluorescence images were visualised using a Spinning Disc Confocal Microscope (Olympus IX71 motorised, inverted microscope) at the Bioimaging research hub/ Biosciences/ Cardiff university. Images were collected with a Hamamatsu ORCA Flash 4.0 CMOS monochrome digital camera using x40 objective lens controlled through Molecular Devices MetaMorph microscopy automation and image analysis software. This microscope has a tri-line laser bank (405 nm, 488 nm, 561 nm), and the 488 nm laser line suitable for DBCO-TAMRA dye was used for the excitation.

2.10 Transmission electron microscopy (TEM)

Negative stain for transmission electron microscopy was prepared as described by Jiang *et al*. (Jiang *et al.* 2015). In briefly, 5 µl of 0.1% polylysine solution was added to the surface of parafilm; a glow-discharged grid (Carbon-coated 400-mesh copper grids (Agar Scientific; S160-4)) was touched onto the polylysine droplet (carbon film facing the polylysine droplet) and incubated for 1 min at room temperature, before the solution was removed by blotting with filter paper. Treatment the grid with polylysine solution allows arrangement of the particles in different orientations on grids (Ortega *et al.* 2000). Test protein sample droplets (5 µl, 0.1 mg/ml) were added to the surface of parafilm. A glow-discharged, polylysine grid was added on the protein droplet (carbon film faced the protein droplet) and incubated for 1 min at room temperature. The excess of protein sample was removed by blotting the grid with filter paper. Then the grid was washed twice with ultrapure water droplets and blotted with filter paper to remove excess liquid before staining the grid with 2% (w/v) of Uranyl acetate (Sigma) for 1 min. Imaging of the grids was performed in the Transmission Electron Microscopy (TEM) Suite, Cardiff university using JEOL JEM-2100 TEM.

2.11 Bioinformatic analysis and computer software

2.11.1 Molecular modelling of Vip1Ac1

To engineer Vip1Ac1, preliminary structural information on Vip1Ac1 was required to direct the choice of surface resides to incorporate ncAAs and label with fluorescent dye. *In silico* modelling was performed using three different fully automated protein structure homology-modeling services; Swiss Model (Kiefer *et al.* 2009), Geno3D (Combet *et al.* 2002) and I-TASSER (Zhang 2008). Subsequently, the models were subjected to validation to select the best model using Ramachandran plots analysis (Ramachandran *et al*. 1963) and ERRAT validation (Colovos and Yeates 1993).

2.11.2 Prediction of potential interface regions of artificial heterodimers

To successfully generate an artificial dimer, residues for dimerisation need to be selected at the interface region. The ClusPro protein-protein docking server (Kozakov *et al.* 2017) was used to predict the potential interface region between sfGFP and Venus by generation of *in silico* heterodimer models. In this server, one of the monomers is kept constant as a receptor (sfGFP^{WT}; PDB 2B3P) and the other is allowed to move as a ligand (Venus ^{WT}; PDB 1MYW). The 10 best models of the interface interaction prediction were ranked based on balanced energy coefficients by the server as part of the output. The positions of candidate residues (H148 and Q204) for dimerisation were then tested for their availability in the potential interface region of the ClusPro heterodimer models.

2.11.3 Computer software

For multiple sequence alignments the CLC Sequence Viewer 8.0 program (https://www.qiagenbioinformatics.com) was used. The CCP4 package (Potterton *et al*. 2003) was used for structure determination (Section 2.12.4). PyMOL (Delano 2002) and CCP4MG software (McNicholas *el al.* 2011) were used for visualisation and graphical representations of protein structure. Structural alignments were performed using PyMOL with defaults option for backbone/Cα alignment. ImageJ software (Schneider *et al.* 2012) was used to analyse the images. ChemDraw Professional software (https://www.perkinelmer.com/category/chemdraw) was used to present vector maps and chemical component diagrams.

2.12 X-Ray crystallography

2.12.1 Preparation of protein crystals

For Vip1Ac1 protein, a concentrated purified protein sample, ~10 mg/ml, was used for crystal preparation. Vip1Ac1 protein was screened for crystal production using two types of commercially available screens: PACT *premier*™ HT-96 and JCSG-plus™ HT-96 (Molecular Dimensions, Suffolk, UK). Each screen contains 96 different conditions for crystal formation. The sitting drop vapour diffusion method was used to set up crystal trials in laboratories at the Heath Hospital, using the automated liquid handling systems Phoenix, supplied by Alpha Biotech UK (with Dr Pierre Rizkallah, School of Medicine, Cardiff University). The sitting drop vapour diffusion method is a popular technique used in protein crystallization, where the principle is to incubate a liquid reservoir of concentrated reagents, 60 µl, against a drop of 0.2 µl of reagent mixed with another 0.2 µl of protein sample, in a sealed tray, and kept at a steady temperature of ~20˚C. Typically, the liquid reservoir of reagent contains a higher reagent concentration than the drop. Water vapour will move from the drop into the reservoir to achieve equilibrium. This leads to an increase in relative saturation of the protein sample in the drop and may allow crystals to grow under favourable conditions.

For Venus^{YazF66} sample (dark state), concentrated purified sample (~20 mg/ml) was used for screening crystals formation using the PACT *premier*™ HT-96 screen and JCSG-plus™ HT-96 screen (Molecular Dimensions). Crystal trials were set up at School of Chemistry, Cardiff University (with Dr Alan Scott, School of Chemistry, Cardiff University) using the automated liquid handling systems Oryx (Douglas instruments, UK). The sitting drop vapour diffusion method was used, as above, with a slight change in volumes (1 µl of protein sample mixed with 1 µl of reagent against 50 µl of reagent) in a sealed tray and kept at 20˚C. All the steps of crystal preparation were performed in the dark.

The growth of crystals was monitored under a light microscope, and grown crystals were harvested by picking individual crystals in a mounted litholoop (Molecular Dimensions) and plunging them into liquid nitrogen. Crystals were kept in liquid nitrogen until the X-ray diffraction data were collected. For the Venus^{YazF66} sample,

crystals of dark state were harvested first, and the remaining crystals were irradiated for 5 min using UVM-57 mid-range UV lamp (UVP) to produce light state.

2.12.2 Optimization of crystal conditions

Different strategies were used for optimization of crystallization conditions, which included changing; crystallization method -using hanging drop instead of sitting drop vapour diffusion method using Easy Xtal 15-well tools (Qiagen), drop ratio (4/500); pH (varied over the range 5, 5.5 and 6); protein concentration (~8 mg/ml); and precipitant type. Protein sample (2 µl) was added into the screw lid and 2 µl of reservoir solution was mixed with the protein sample and closed over the well containing 500 µl of reservoir solution. Crystal growth was monitored under a light microscope at days 0, 1, 2 and 3. Grown crystals were harvested as described in Section 2.12.1.

2.12.3 Preparation of heavy atom derivatives

Heavy atom derivatives of Vip1Ac1 crystals were prepared by a soaking method in which 1 µl of heavy atom solution was added to the crystal drop (4 µl) and sealed again. The heavy atom solution was prepared in crystallisation buffer. Different final concentrations of heavy atom were used for both sodium bromide and sodium iodide (10, 20 and 50 mM) (Garman and Murray 2003). Crystals were left with heavy atoms for different incubation times (short times 1, 5 mins and long times 24 h). Vip1Ac1 protein containing ncAA 4-Iodo-L-phenylalanine as a heavy atom and SeMet-containing Vip1Ac1 protein were prepared as described in Sections 2.3.4.1 and 2.3.5 respectively, before crystals were prepared as described in Section 2.12.2.

2.12.4 Data collection and structural determination

Diffraction data were collected at the Diamond Light Source, Harwell, UK, at beamline I04, I24, I03 (with Dr Pierre Rizkallah, School of Medicine, Cardiff University). Data were reduced with the XIA2 package (Winter 2010), POINTLESS was used for space group assignment , scaling and merging were completed with Aimless (Evans 2006) and TRUNCATE (Bailey 1994). Molecular replacement was used to solve the structure of Venus^{Y66azF} dark state using sfGFP^{Y66azF} dark state (PDB accession 4J88) as a model using PHASER (McCoy *et al.* 2007). The COOT program was used to adjust the structure manually for several repeated cycles, TLS restrained refinement was used

to refine the structure using RefMac. The CCP4 package was used for the above routines (Bailey,1994). For Vip1Ac1 native crystals, molecular replacement was not successful (more details can be found in chapter 4).

For derivative Vip1Ac1 crystals (native crystals soaked with sodium bromide and sodium iodide solution) and crystals containing ncAA 4-Iodo-L-phenylalanine, single wavelength anomalous diffraction (SAD) data were collected at the Diamond Light Source on beamline I03 for iodide and I04 for bromide. Data were processed as previously described using the CCP4 package. CRANK (automated EP pipeline) and CRANK 2 in the CCP4 package were used for SAD data analysis as an attempt to solve the phasing problem. For SeMet containing-crystals, multiple-wavelength anomalous diffraction (MAD) were collected and analysed by Dr. Neil Paterson (Diamond Light Source, Harwell Science and Innovation Campus, Didcot, UK) at the Diamond Light Source using beamline I24. First, a fluorescence scan across the selenium edge was performed to identify accurate peak and inflection energies for data collection. A gentle MAD dataset with 3 sweeps of 540 degrees at peak (0.9793 Å), inflection (0.9795 Å) and high remote (0.9686 Å) were collected, then a further 1500 degrees at the peak energy were collected to help the substructure determination. All datasets were autoprocessed by the xia2 pipelines and used as input to the CRANK2 pipeline. The CRANK2 pipeline resulted in a final model that was then used as a search model for MOLREP with a high resolution data set from native crystals (1.47 Å). The resulting docked model was used as a starting model for PHENIX autobuild to build the structure. The COOT program was used to adjust the structure manually for several repeated cycles, TLS restrained refinement was used to refine the structure using RefMac. All the structures were validated by validation tools in COOT.

Chapter Three

Production and purification of recombinant proteins (Vip1 and Vip2)

3. Production and purification of recombinant proteins (Vip1 and Vip2)

3.1 Introduction

Vip toxins are usually secreted as soluble proteins to their surrounding environment by Bt during the vegetative growth phase due to the fact that they have a putative N-terminal signal peptide (Chakroun *et al.* 2016). Some Bt strains have the ability to produce more than one type of Vip toxin such as the strain used in this study (*Bacillus thuringiensis* strain H29.3), which produces Vip1, Vip2 and Vip3. To study each toxin individually without contamination of other proteins, high amounts of pure toxin are required, especially for crystallographic studies.

This chapter focuses on production of Vip1 and Vip2 recombinant proteins using *E. coli* as an expression host and overcoming solubility issues using protein engineering and strategies including changing: expression conditions, expression vector and fusion with a soluble partner. Purification methods for protein under native conditions for subsequent use will also be described in this chapter.

3.2 Results and Discussion

3.2.1 Expression of Vip1Ac1 protein in *E. coli* using pET-28b vector

The original clone (pET28b vector encoding Vip1Ac1 obtained as a kind gift from Professor Primitivo Caballero, Universidad Publica de Navarra, Spain) comprised the *Vip1Ac1* gene with an N-terminal His-tag preceding its own signal peptide. The construct was expressed in *E. coli* BL21 (DE3) by incubating at 37˚C until an OD_{600} of 0.7 (± 0.1) was reached prior to induction for 3 h with 1 mM IPTG. The Vip1Ac1 protein (MW ~87.6 kDa) was found to be produced in insoluble form (Figure 3.1). A previous study by Shi *et al.*(2004) also found that Vip1Ac protein expressed with the signal peptide accumulated as insoluble inclusion bodies in the *E. coli* BL21 (DE3) expression system and soluble protein was hardly detected (Shi *et al.* 2004). Often recombinant proteins form insoluble inclusion bodies because the *E. coli* expression system lacks advanced machinery to perform posttranslational modifications, leading to poor solubility of the recombinant protein and formation of inclusion bodies (which is still the major issue in recombinant protein expression (Costa *et al.* 2014)). Recombinant proteins expressed in the microenvironment of the expression host, which is different from the original source

in terms of osmolarity, cofactors, pH and folding mechanisms, as well as expression in high levels, are factors that may result in protein aggregation and instability (Rosano and Ceccarelli 2014).

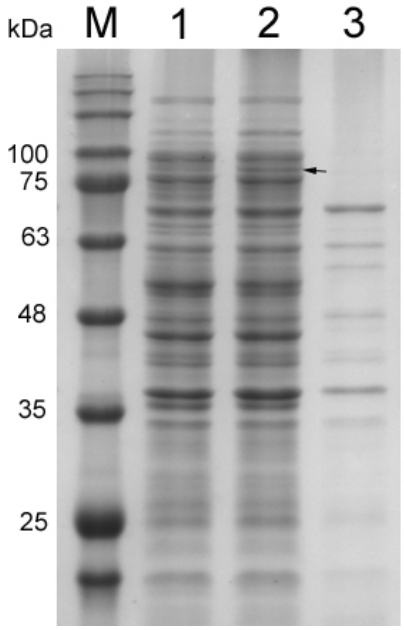

Figure 3.1. Vip1Ac1 protein expressed from original clone.
SDS-PAGE (12.5%) analysis of Vip1Ac1 protein. Lane M corresponds to protein molecular weight standards, lane1 corresponds to total cell proteins of preinduced cells, lane 2 corresponds to total cell proteins of induced cells, lane 3 corresponds to soluble cell proteins of induced cells. Arrow shows the Vip1Ac1 protein (~ 87.6 kDa).

3.2.2 Optimization of expression system

Expression systems and vectors play a vital role in production of a high level of soluble recombinant protein (Joseph *et al.* 2015). Different strategies have been used to enhance the solubility of recombinant protein and increase the level of expression such as lowering incubation temperature, reducing IPTG concentration and using different expression vectors. These different factors were explored in an attempt to improve the production of soluble Vip1Ac1 protein.

3.2.2.1 Low temperature and IPTG concentration expression conditions

Lowering the cell culture incubation temperature during expression and the IPTG concentration can lead to a decrease in the rate of protein synthesis and folding kinetics,

as well as reduced hydrophobic interactions, which are involved in protein self-aggregation (Costa *et al.* 2014). Expressing at 15˚C using three different IPTG concentrations (0.05, 0.1 and 0.2 mM) showed no significant difference in solubility (Figure 3.2). The presence of the His-tag at the N-terminus could block the signal peptide from binding with the signal recognition particle on the membrane and prevent protein secretion. In 2013, Singh *et al.* found that the presence of signal sequences in recombinant protein (thioredoxin) leads to reduction in protein activity and increases the aggregation propensity due to thermodynamic destabilization (Singh *et al.* 2013). In addition, the signal peptide of Vip1Ac1 may not be compatible with the expression system, due to the fact that signal peptides in *Bacillus* species differ from signal peptides in *E. coli* in terms of their length (five to seven amino acids longer in the former (Van Wely *et al.* 2001)). This may lead to accumulation of the Vip1Ac1 in the cytoplasm.

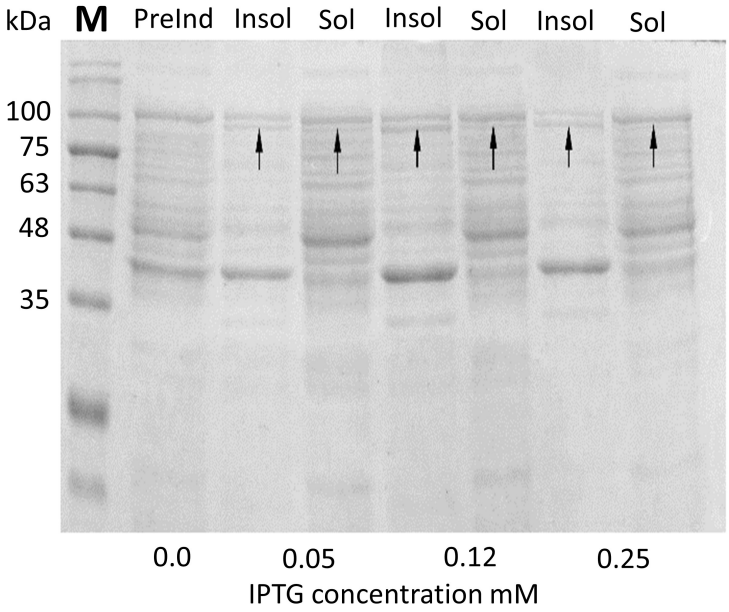

Figure 3.2. Vip1Ac1 protein expression at different IPTG concentrations.
SDS-PAGE (12.5%) analysis of Vip1Ac1 protein expression at 15 °C using different IPTG concentration. Lane M corresponds to protein molecular weight standards; PreInd corresponds to total cell proteins of preinduced cells; Insol corresponds to insoluble cell proteins of induced cells; Sol corresponds to soluble cell proteins of induced cells. Bold arrows show the Vip1Ac1 protein (~87.6 kDa).

A new pET28b construct was produced that removes the N-terminal His-tag and Vip1Ac1 signal peptide (Section 2.2.9.1) as shown diagrammatically in Figure 3.3. The construct was transformed into *E. coli* BL21 and induction was carried out as before. Removing the N-terminal His-tag and signal peptide (first 28 amino acids) encoded by

the original pET28b construct made no difference to Vip1Ac1 protein solubility when expressed at 15˚C for 3 h using 0.1 mM IPTG, although protein yields did increase, as shown in Figure 3.4 compared to the original constructs (Figure 3.1).

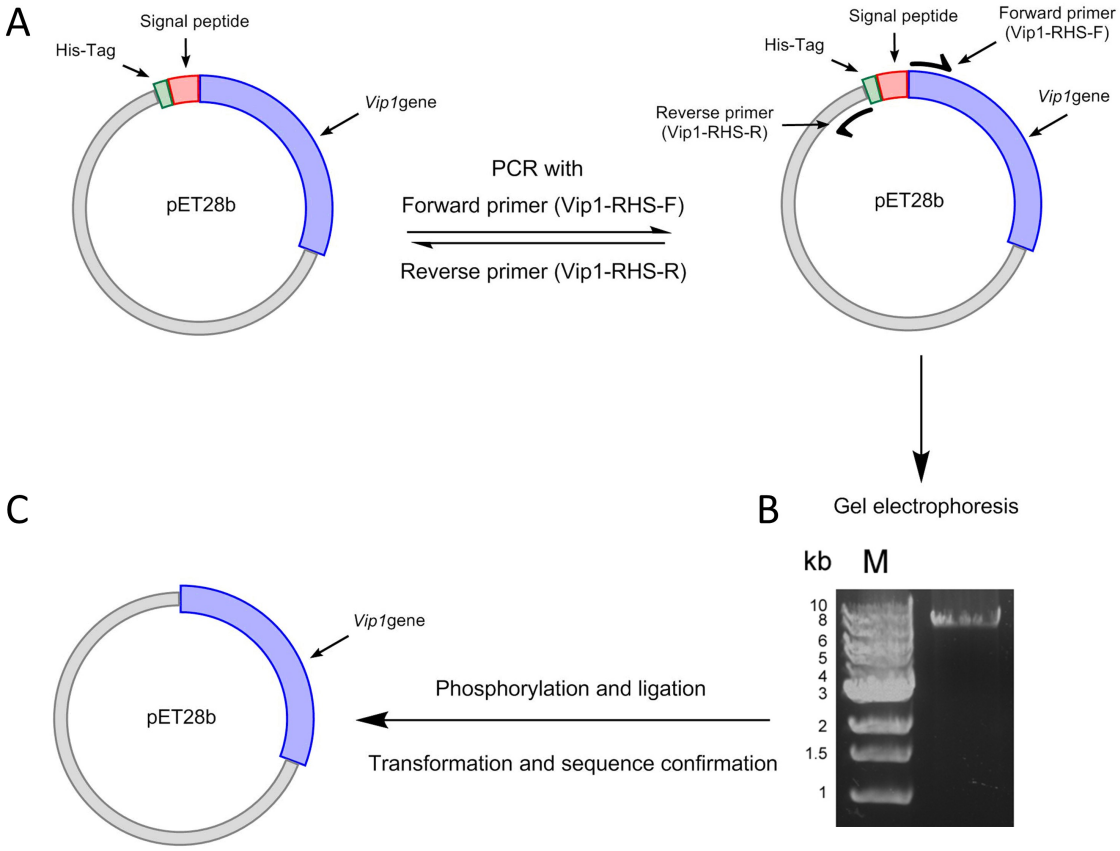

Figure 3.3. Schematic map of constructed pET28b recombinant plasmids.
A) Original pET28b construct map (~ 7.7 kb) containing *Vip1Ac1* gene with N-terminal 6xHis (green) and signal peptide (red) using whole plasmid PCR with mutagenic primers to remove both N-terminal 6xHis and signal peptide. **B)** Agarose (1%) gel electrophoretic analysis of PCR product showing linear DNA. **C)** Final map of pET28b construct containing *Vip1Ac1* gene without N-terminal 6xHis and signal peptide.

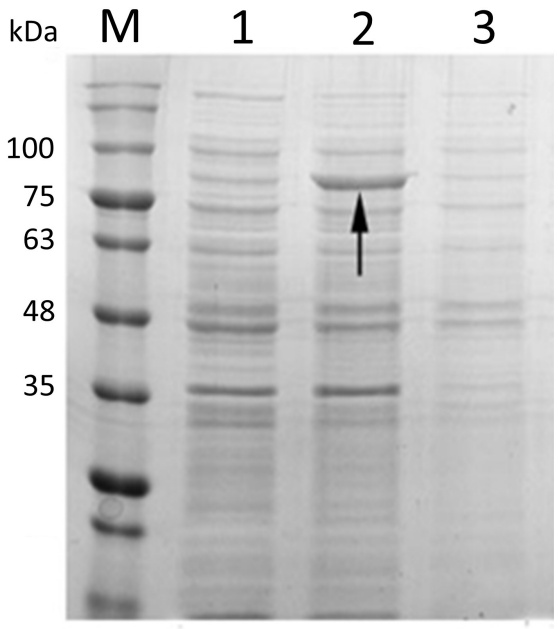

Figure 3.4. Vip1Ac1 protein expression from new pET28b constructs.
SDS-PAGE (12.5%) analysis of Vip1Ac1 protein without N-terminal 6xHis and signal peptide. Lane M corresponds to protein molecular weight standards, 1 corresponds to total cell proteins of preinduced cells, 2 corresponds to total cell proteins of induced cells, 3 corresponds to soluble cell proteins of induced cells. Protein expressed at 15˚C for 3 h using 0.1 mM IPTG as inducer. Bold arrow shows the Vip1Ac1 protein (~ 82 kDa).

3.2.2.2 Construction of pGEX-4T-2 vector plus *Vip1Ac1* gene

One approach to aid the production of soluble protein is to attach a protein of interest to a highly soluble fusion partner such as Glutathione S-transferase (GST) (eg it has been successfully used to increase the solubility of PA, which expressed as inclusion bodies in *E. coli* (Wu *et al.* 2010)). Using GST also facilitates purification of protein by an affinity-chromatography approach. The *Vip1Ac1* gene was cloned into the pGEX-4T-2 vector as outlined in Section 2.2.9.1 and illustrated in Figure 3.5. A successful construct containing the *GST-Vip1Ac1* fusion gene was selected using colony PCR (Section 2.2.2) as described in Figure 3.6 and validated by sequencing of the inserted DNA. The GST-Vip1Ac1 fusion (MW 108 kDa) was induced for 30, 60,120 and 180 min using 0.1 mM IPTG at 15˚C.

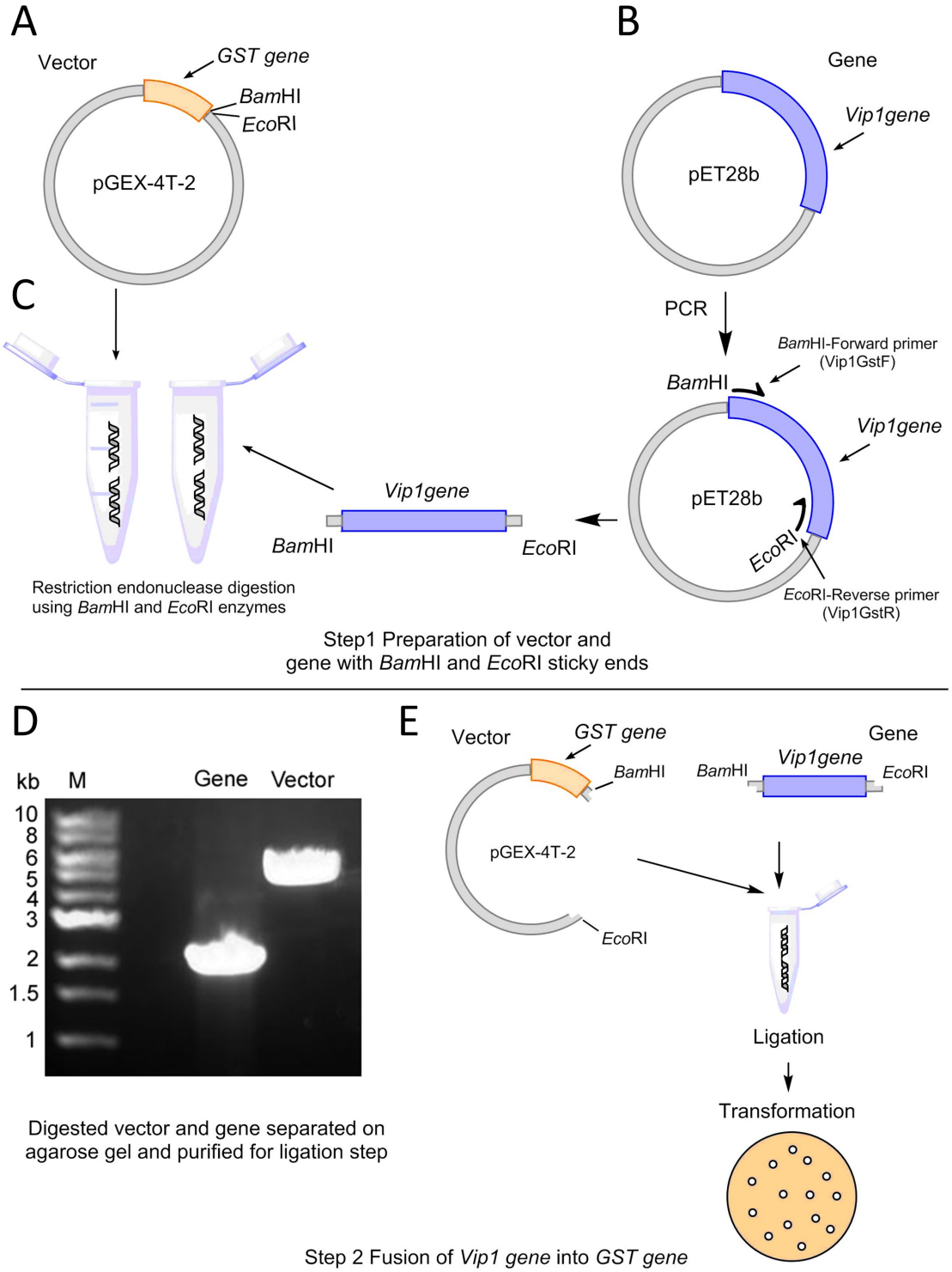

Figure 3.5. Schematic map of constructed pGEX-4T-2 vector containing *GST-Vip1Ac1* fusion. **A.** pGEX-4T-2 vector (~ 5 kb) containing *GST* gene (orange) with *Bam*HI and *Eco*RI restriction sites at the 3'-end of *GST* gene. **B.** pET28b vector map containing the *Vip1Ac1* gene (blue) used in whole plasmid PCR with mutagenic primers to amplify the *Vip1Ac1* gene (~ 2.2 kb) with *Bam*HI and *Eco*RI restriction sites. **C.** Restriction endonuclease digestion for both vector and gene using *Bam*HI and *Eco*RI. **D** and **E** showing the second step for fusing the *Vip1Ac1* and *GST* genes. **D.** Agarose (1%) gel electrophoretic analysis of digested vector (~ 5 kb) and *Vip1Ac1*gene (~ 2.2 kb). **E**. *Vip1Ac1* gene inserted into the pGEX-4T-2 vector between the *Bam*HI and *Eco*RI sites by ligation and the mixture transformed into competent cells.

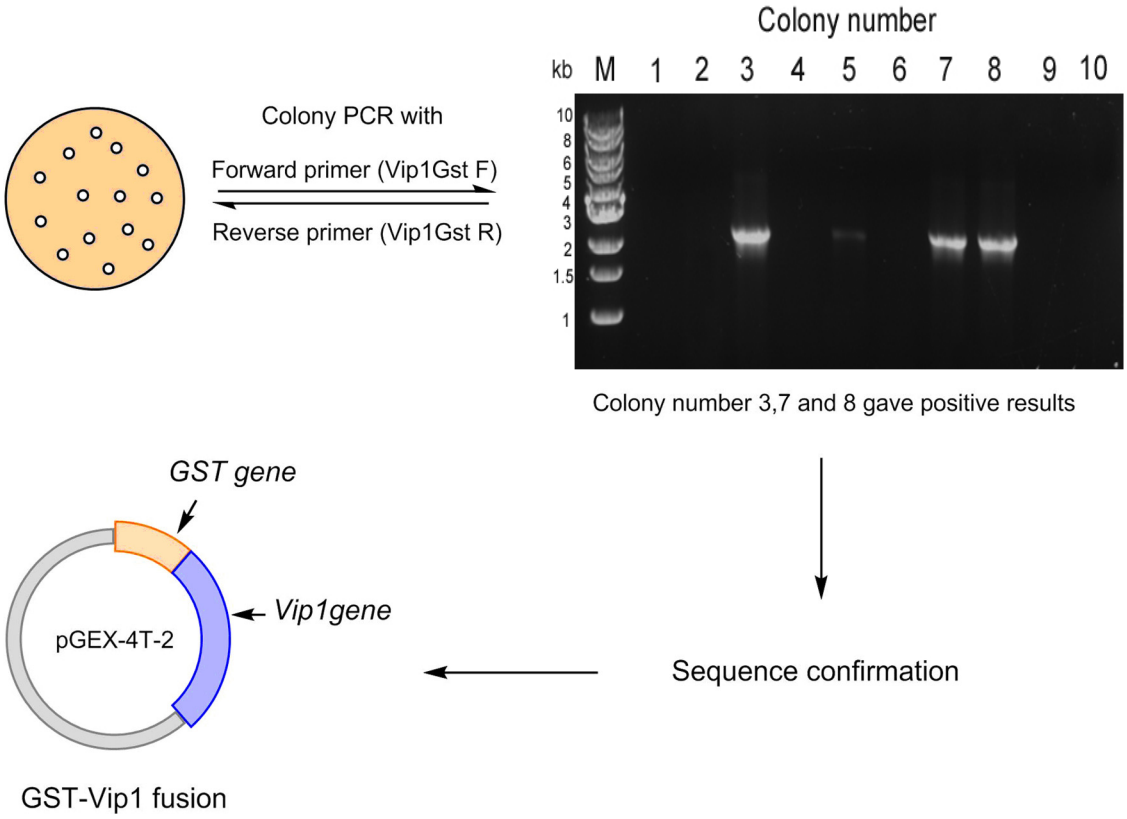

Figure 3.6. Colony PCR for *GST-Vip1Ac1* fusion gene selection.
A) Colonies transformed with ligation mixture (pGEX-4T-2 vector and *Vip1Ac1* gene), 10 colonies were selected for colony PCR using Vip1Ac1 primers (Vip1GstF and Vip1GstR). **B)** Agarose (1%) gel electrophoresis of PCR products showing colonies number 3,7 and 8 contain the *Vip1Ac1* gene (~ 2.2 kb). **C)** Final map of pGEX-4T-2 construct containing *GST-Vip1Ac1* fusion gene.

Results from expression experiments indicate that the GST fusion partner did not enhance the solubility of the Vip1Ac1 protein, which still accumulated in insoluble form as shown in (Figure 3.7). Insolubility may be due to expression in the cytoplasm of a protein that is normally secreted. However, the protein yield of fusion protein (GST-Vip1Ac1) as shown in Figure 3.7 was increased compared with the yield of Vip1Ac1 protein (Figure 3.1).

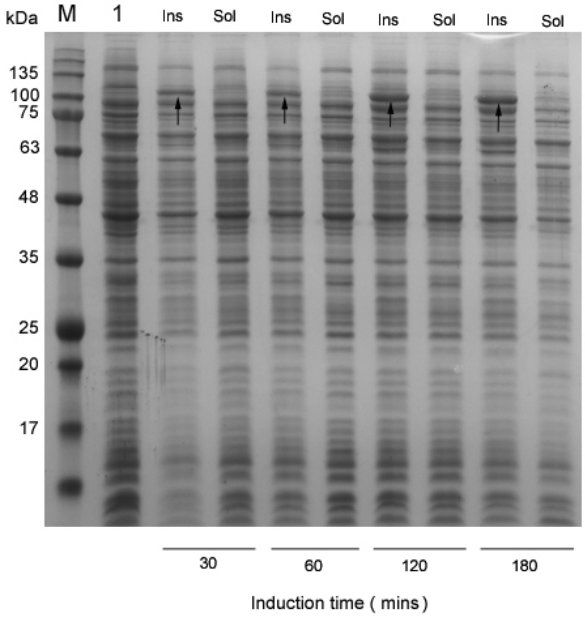

Figure 3.7. GST-Vip1Ac1 fusion protein expression at different times.
SDS-PAGE (12.5%) analysis of GST-Vip1Ac1 fusion protein**.** Lane M corresponds to protein molecular weight standards. Lane 1 corresponds to total cell proteins of preinduced cells. Ins corresponds to total cell proteins of induced cells. Sol corresponds to soluble cell proteins of induced cells. Bold arrows show the GST-Vip1Ac1 fusion protein (~108 kDa).

3.2.2.3 Construction of pET22b vector containing cytochrome b_{562} - *Vip1Ac1* fusion gene.

Cytochrome b_{562} (Cyt b_{562}) is a red coloured haem binding protein that is expressed as highly soluble protein in the periplasmic space of *E. coli* (Nikkila *et al.* 1991). Cyt b_{562} was chosen as a soluble fusion partner in an attempt to enhance the solubility of the Vip1Ac1 protein and allow it to secrete into the periplasmic space. The reasons for this choice were that this protein is highly soluble, small (about 12 kDa), contains a compatible signal peptide that directs it into the periplasmic space (naturally expressed in *E. coli* as a periplasmic protein (Mathews *et al.*1979)) and coloured protein (red colour), which can be used as expression marker. The pET22b vector containing the *Cyt* b_{562} gene was previously constructed in the laboratory and was used to construct a new vector. During construction using site directed mutagenesis (Section 2.2.2), a tobacco etch virus (TEV) protease cleavage site (ENLYFQ/G) and restriction sites (*Bam*HI and *Eco*RI) were added at the 3'-end of the *Cyt* b_{562} gene, before the *Vip1Ac1* gene was inserted into this new vector between the *Bam*HI and *Eco*RI sites to form pET22b vector

containing *Cytb*$_{562}$-*Vip1Ac1* gene as outlined in Section 2.2.2 and shown in Figure 3.8. A construct containing the fusion gene was selected using colony PCR as described in Figure 3.9 and was verified by sequencing.

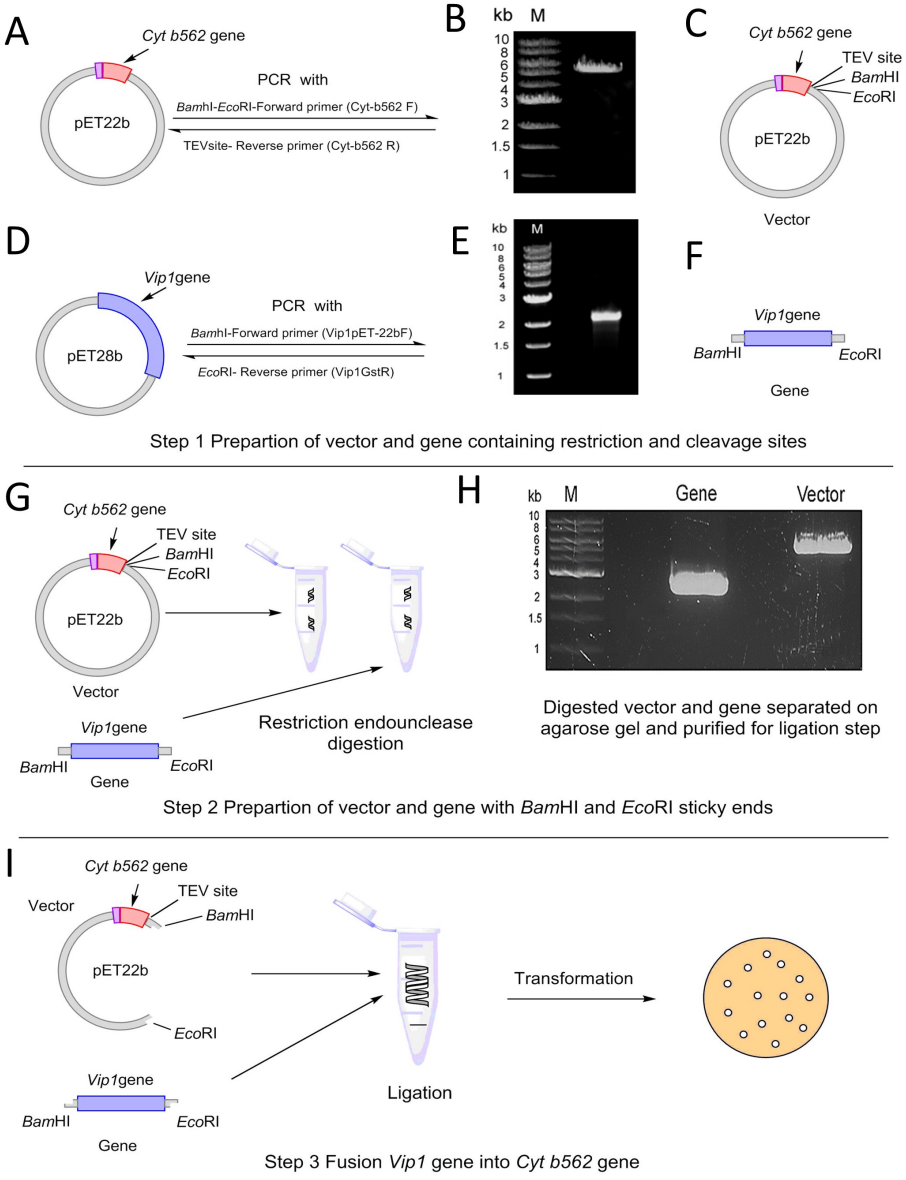

Figure 3.8. Construction of pET22b vector containing *Cytb*$_{562}$-*Vip1Ac1* gene.
A. pET22b construct map (~ 5.75 kb) containing *Cyt b*$_{562}$ gene (red) with periplasmic signal peptide (purple) used in whole plasmid PCR with mutagenic primers to add TEV cleavage site, *Bam*HI and *Eco*RI restriction sites. **B**. Agarose (1%) gel electrophoresis of PCR product showing linear pET22b vector (~ 5.75 kb). **C.** Final pET22b *Cyt b*$_{562}$ construct map. **D**. pET28b construct map containing *Vip1Ac1* gene (blue) used in whole plasmid PCR with mutagenic primers to amplify *Vip1Ac1* gene (~ 2.2 kb) with *Bam*HI and *Eco*RI restriction sites. **E.** Agarose (1%) gel electrophoresis of PCR product showing *Vip1Ac1* gene (~ 2.2 kb). **F.** *Vip1Ac1* gene with *Bam*HI and *Eco*RI restriction sites. **G.** Restriction endonuclease digestion for both vector and gene using *Bam*HI and *Eco*RI enzymes. **H.** Agarose (1%) gel electrophoresis of digested construct and gene. **I**. *Vip1Ac1* gene inserted into the pET22b construct between the *Bam*HI and *Eco*RI sites by ligation and the mixture transformed into competent cells.

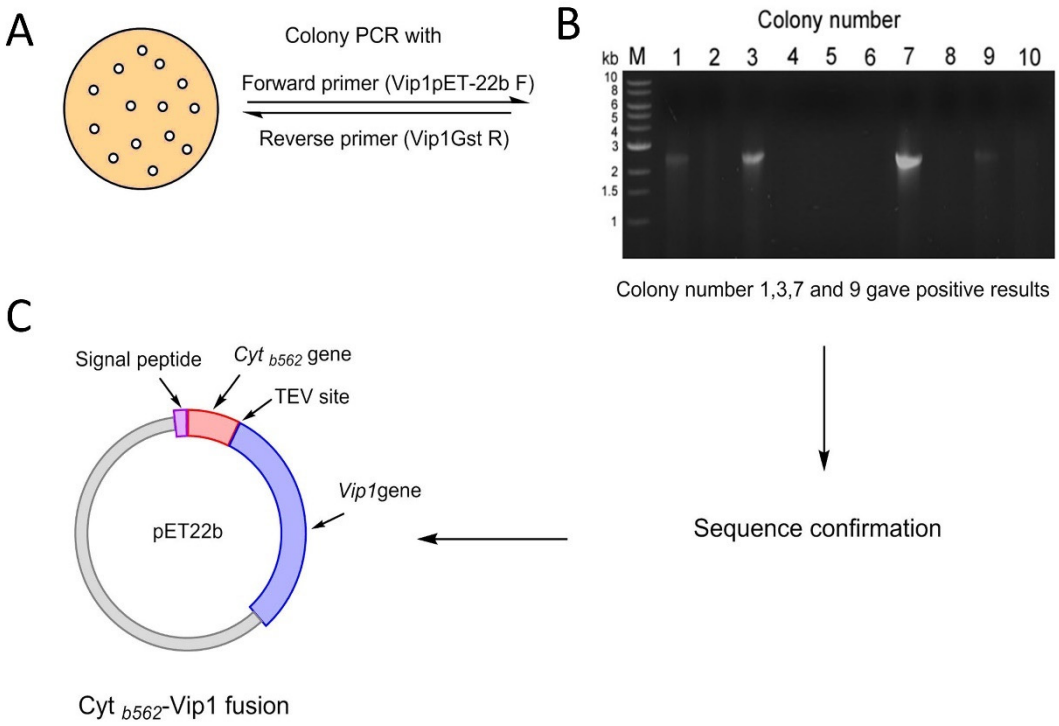

Figure 3.9. Colony PCR for *Cytb$_{562}$-Vip1Ac1* fusion gene selection.
A) Colonies transformed with ligation mix were selected for colony PCR using Vip1Ac1 primers (Vip1pET-22bF and Vip1GstR). **B)** Agarose (1%) gel electrophoresis of PCR products showing colonies number 1, 3, 7 and 9 contain the *Vip1Ac1* gene (~ 2.2 kb). **C)** Final map of pET22b construct containing *Cytb$_{562}$-Vip1Ac1* fusion gene.

Following transformation of the construct into *E. coli* BL21 cells, induction of protein expression was carried out at 37˚C for 3 h using 1 mM IPTG. The fact that induced cells exhibited a distinct red colour (due to haem binding to Cyt b_{562}) highlighted successful expression of the fusion protein (Figure 3.10A). However, the fusion protein (predicted MW 95.77 kDa) appeared in the insoluble fraction as shown in Figure 3.10B.

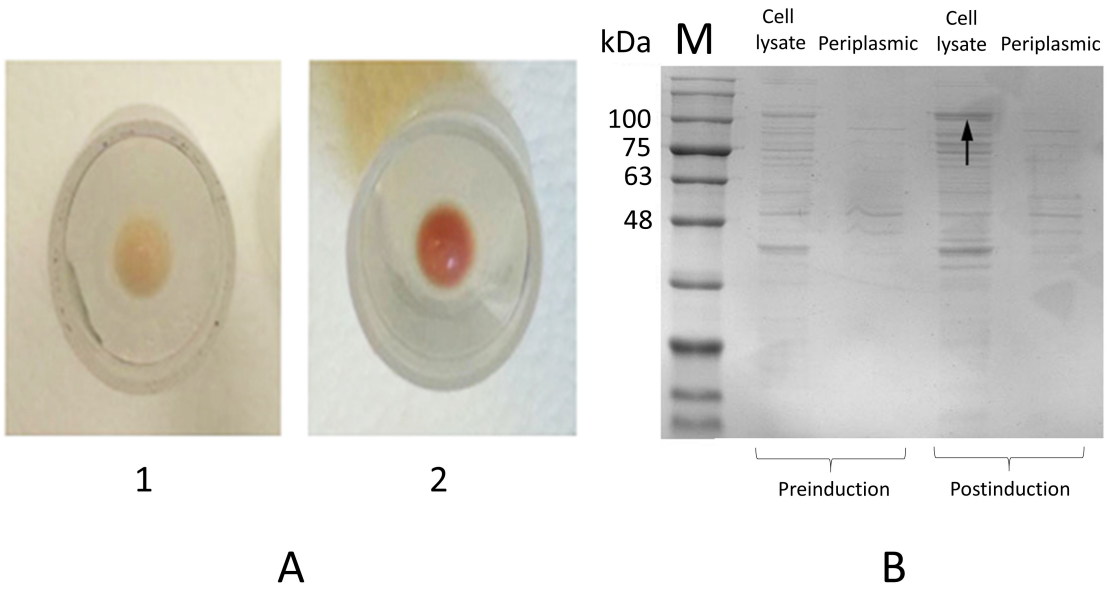

Figure 3.10. Expression of the Cyt b_{562}-Vip1Ac1 fusion protein in *E. coli* BL21 cells.
A. Image showing cell pellet. **1**.Pellet from 10 ml culture of pre-induction cells. **2**. Pellet from 10 ml culture of Post-induction cells. **B**. SDS-PAGE analysis of Cyt b_{562}-Vip1Ac1 expression at 37˚C for 3 h using 1 mM IPTG. Lane M corresponds to protein molecular weight standards, Cell lysate corresponds to total cell proteins, Periplasmic corresponds to total periplasmic proteins. Bold arrow shows the Cyt b_{562}- Vip1Ac1 protein (95.77 kDa).

Expression of the fusion protein (Cyt b_{562} -Vip1Ac1) at 15˚C for 3 h using 0.1 mM IPTG was found to increase the solubility of Vip1Ac1 fusion protein dramatically (Figure 3.11 A). Optimum temperature, IPTG concentrations and induction time for expression the fusion protein (Cyt b_{562}-Vip1Ac1) were investigated over a range of IPTG concentrations of (0.05, 0.12, 0.25 and 0.5 mM), temperatures of (15, 20, 25 and 30˚C) and induction times (3 h and 24 h). Results showed that over the range of IPTG concentrations and different incubation temperatures, the fusion protein expressed at the same level based on the intensity of the protein band on SDS-PAGE as shown in Figure 3.11 B and C. Increased induction time led to an increase the expression level as shown in Figure 3.11 D as a results of biomass increased. Based on these results, large scale expression was carried out using the optimum conditions (25˚C, 0.1 mM IPTG for 24 h), as described in Section 2.3.2. Fusion protein was extracted from induced cells as described in Section 2.3.6.2 and purified as shown below.

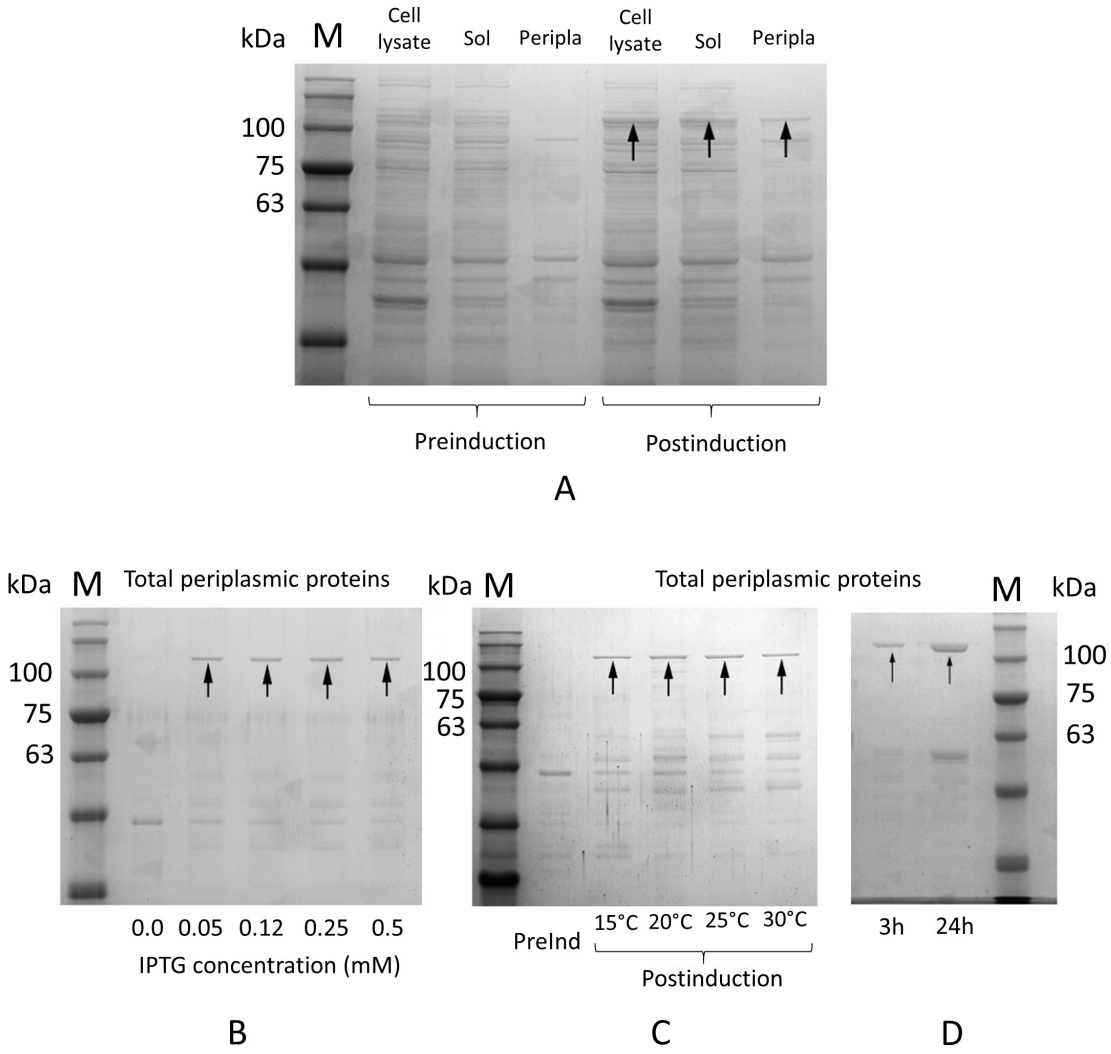

Figure 3.11. Cyt b_{562}-Vip1Ac1 fusion protein expression and optimisation.
A. Expression of the fusion protein (Cyt b_{562}-Vip1Ac1) at 15˚C for 3 h using 0.1 mM IPTG. Lane M corresponds to protein molecular weight standards, Cell lysate corresponds to total cell proteins, Sol corresponds to soluble cell proteins, Peripla corresponds to total periplasmic proteins. **B**. Expression of the fusion protein (Cyt b_{562}-Vip1Ac1) at 15˚C for 3 h using a range of IPTG concentrations. **C**. Expression of the fusion protein (Cyt b_{562}-Vip1Ac1) at different incubation temperatures using 0.1 mM IPTG. **D.** Expression of the fusion protein (Cyt b_{562}-Vip1Ac1) at 25˚C for different incubation times using 0.1 mM IPTG. Bold arrows show the Cyt b_{562}-Vip1Ac1 fusion protein (95.77 kDa).

3.2.3 Purification of Cyt b_{562}-Vip1Ac1 fusion protein

Cyt b_{562}-Vip1Ac1 fusion protein was purified in two steps: ion exchange chromatography (IEX) followed size exclusion chromatography (SEC).

3.2.3.1 Ion Exchange Chromatography (IEX)

The net charge of fusion Cyt b_{562}-Vip1Ac1 protein was predicted using the Bachem website (http://www.bachem.com/service-support/peptide-calculator/): the net charge at pH 7 is -13 and at pH 8 is -15 so an anion exchanger was chosen. Periplasmic extracts from induced cells (Section 2.3.2) were loaded onto a HiTrap Q HP 5 ml column connected to an ÄKTA purifier. Proteins were eluted with a linear gradient of NaCl 0-500 mM buffer B (50 mM Tris-HCl, 500 mM NaCl pH 8). The elution of protein was monitored by absorption at 280 nm. The six observed elution peaks, Figure 3.12A were analysed by SDS-PAGE as shown in Figure 3.12 B and showed that the Cyt b_{562}-Vip1Ac1 fusion protein was eluted in the second peak. Fractions containing partially purified Cyt b_{562}-Vip1Ac1 fusion protein were pooled and subjected to the second step of purification.

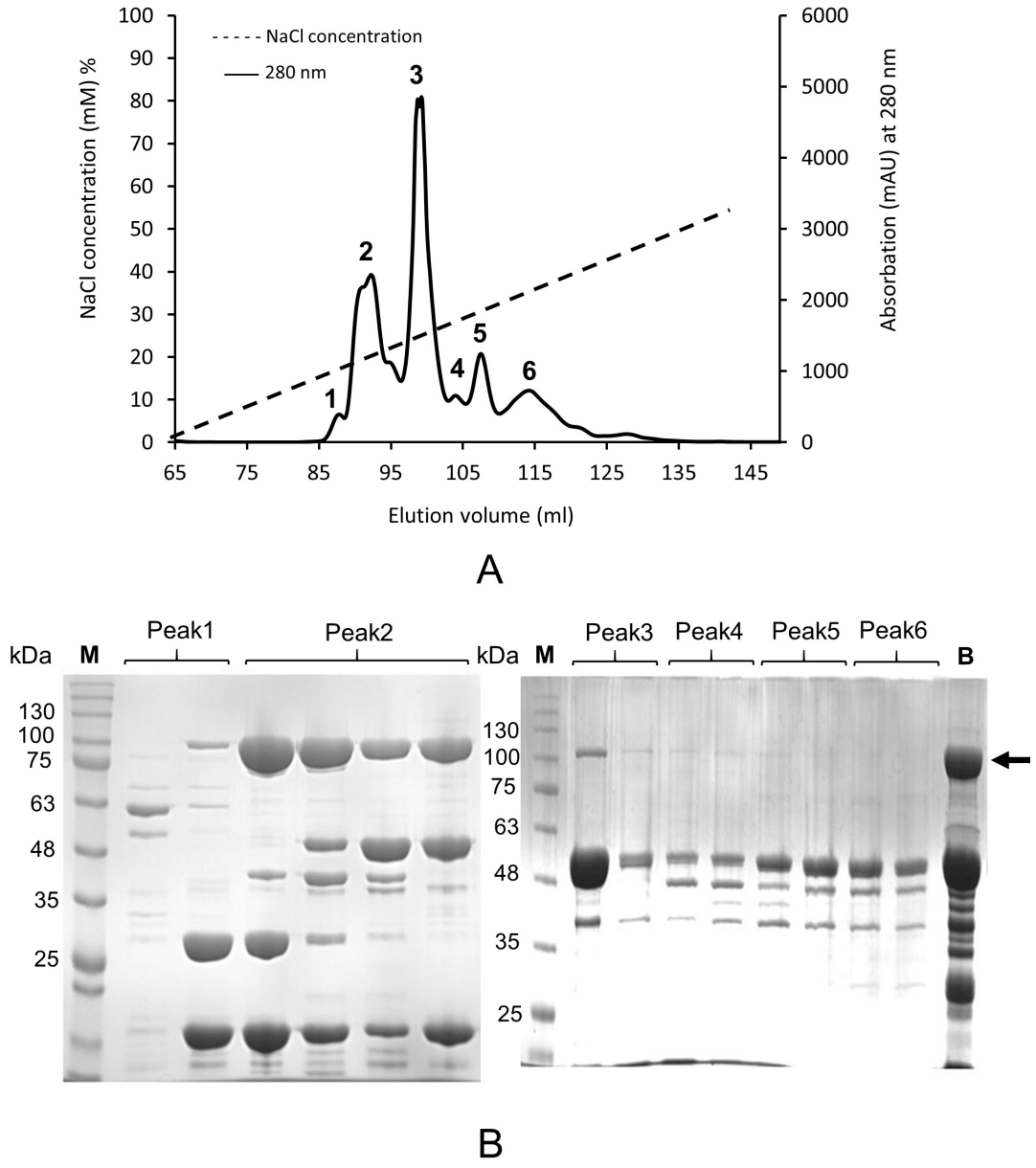

Figure 3.12. Ion exchange chromatography of Cyt b_{562}-Vip1Ac1 fusion protein.
A. Anion exchange chromatography elution profile. Absorbance at 280 nm monitoring Cyt b_{562}-Vip1Ac1 fusion protein (solid line) on application of a NaCl gradient (dotted line). **B.** SDS-PAGE (12.5) analysis of protein elution peaks. Lane M corresponds to protein molecular weight standards. Lane B corresponds to the periplasmic extraction before loaded into the column. The remaining lanes are a selection of fractions from each peak collected during the NaCl gradient. Bold arrow shows the Cyt b_{562}-Vip1Ac1 fusion protein (95.77 kDa).

3.2.3.2 Size Exclusion Chromatography (SEC)

Fractions from IEX containing the Cyt b_{562}-Vip1Ac1 fusion protein were pooled and concentrated to 2 ml and loaded on a Hiload^{TM} 16/60 Superdex^{TM}200 prep grade gel filtration column as described in Section 2.3.8.3. The Cyt b_{562}-Vip1Ac1 fusion protein has a predicted molecular weight 95.77 kDa and its elution was monitored by absorption at 280 nm. Three protein elution peaks were observed, Figure 3.13A and fractions corresponding to these peaks were analysed via SDS-PAGE (Figure 3.13B). This showed that the Cyt b_{562}-Vip1Ac1 fusion protein was in the first peak. The identity of proteins in the other peaks (other proteins or possible degradation products) was not investigated. Fractions containing the purified Cyt b_{562}-Vip1Ac1 fusion protein were pooled, concentrated and used for further study.

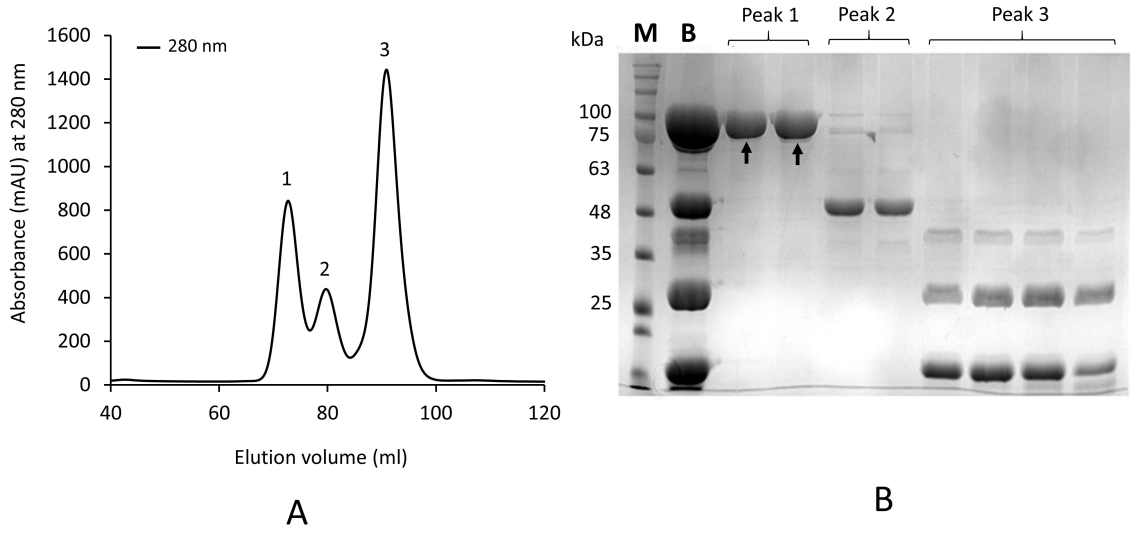

Figure 3.13. Size exclusion chromatography of Cyt b_{562}-Vip1Ac1 fusion protein.
A. Size exclusion chromatography elution profile. Absorbance at 280 nm monitoring Cyt b_{562}-Vip1Ac1 fusion protein (solid line). **B.** SDS-PAGE (12.5) analysis of protein elution peaks. Lane M corresponds to protein molecular weight standards. Lane B corresponds to sample before loaded into the SEC column. The remaining lanes are a selection of fractions from peaks collected during the elution. Bold arrows show the Cyt b_{562}-Vip1Ac1 fusion protein (95.77 kDa).

3.2.4 Removal of Cyt b_{562} from Vip1Ac1 by TEV protease

After purification of Cyt b_{562}-Vip1Ac1, the fusion partner Cyt b_{562} should be removed from Vip1Ac1 before subsequent use of the Vip1Ac1 for further analysis. The TEV cleavage site between the fusion partner Cyt b_{562} and Vip1Ac1 protein allowed

cleavage of the using TEV protease to produce two proteins, Vip1Ac1 (82.73 kDa) and Cyt b_{562} (12k Da). In-house produced TEV protease was used for cleavage experiments as described below.

3.2.4.1 Expression and purification of TEV protease

TEV protease is a 27 kDa enzyme that recognizes a specific seven amino acid sequence EXXYXQ*G/S where X corresponds to any amino acid and * is the cleavage point. Kostallas *et al.* (2011) found that the best recognition and cleavage sequence was ENLYFQ*G/S, which is the identical sequence to the natural substrate of TEV protease (Kostallas *et al.* 2011). This specificity makes it a valuable reagent for cleaving fusion proteins or removing N-terminal His-tags from recombinant proteins and leaves only one extra amino acid at the N-terminus of the cleaved protein. The plasmid pRK793 encoding the GST-TEV fusion was donated from the Jones lab and was used for TEV protease expression (Section 2.3.3). Cell lysis and protein extraction as described in Section 2.3.6.3. C-terminally His-tagged TEV protease was expressed in soluble form as a fusion protein with GST, itself containing a TEV recognition site between GST and TEV protease. After expression, the TEV protease cleaved itself from the fusion protein and was purified from cell lysate by two purification steps. The first used Protino^{R} Ni-TED 2000 Packed Columns (Machery-Nagel, Germany) nickel affinity chromatography as described in Section 2.3.8.1. Partially purified protein from this step was then subjected to the second step of purification by size exclusion chromatography as described in Section 2.3.8.3. After two steps of purification TEV protease showed high purity judged by SDS-PAGE (Figure 3.14). Purified protein was concentrated to 1 mg/ml and aliquoted into 100 µl samples and stored at -80˚C until use.

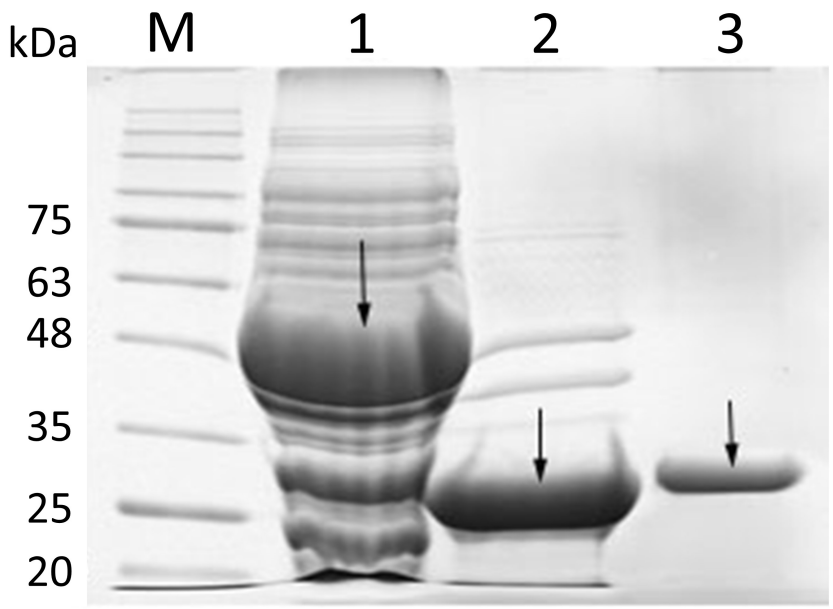

Figure 3.14. TEV protease purification.
SDS-PAGE (12.5%) analysis of TEV protease. Lane M corresponds to protein molecular weight standards. Lane 1 corresponds to whole cell lysate; bold arrow shows the GST-TEV fusion protein (~53 kDa). Lane 2 corresponds to elution fraction from nickel affinity chromatography, bold arrows show the TEV protease (27 kDa). Lane 3 corresponds to elution fraction from size exclusion chromatography, bold arrows show the pure TEV protease (~27 kDa).

3.2.4.2 Cleavage of the Cyt b_{562}-Vip1Ac1 fusion protein

Purified TEV protease (3.2.4.1) was incubated with the Cyt b_{562}-Vip1Ac1 protein at a range of fusion protein to TEV ratios (1:1, 1:2, 1:3, 1:4 and 1:5) for 24 h at room temperature but did not show any cleavage, Figure 3.15A. Commercial TEV-Express enzyme was used as a comparison to our in-house produced TEV protease at a ratio 1:1 in the optimum conditions for TEV protease (50 mM Tris-HCl, pH 8 containing 0.5 mM EDTA and 1 mM DTT) and incubated overnight at room temperature but, again, did not cleave the fusion protein (Figure 3.15B). This may indicate that the folding of the Cyt b_{562}-Vip1Ac1 fusion protein may bury the TEV cleavage site and prevent access of the enzyme.

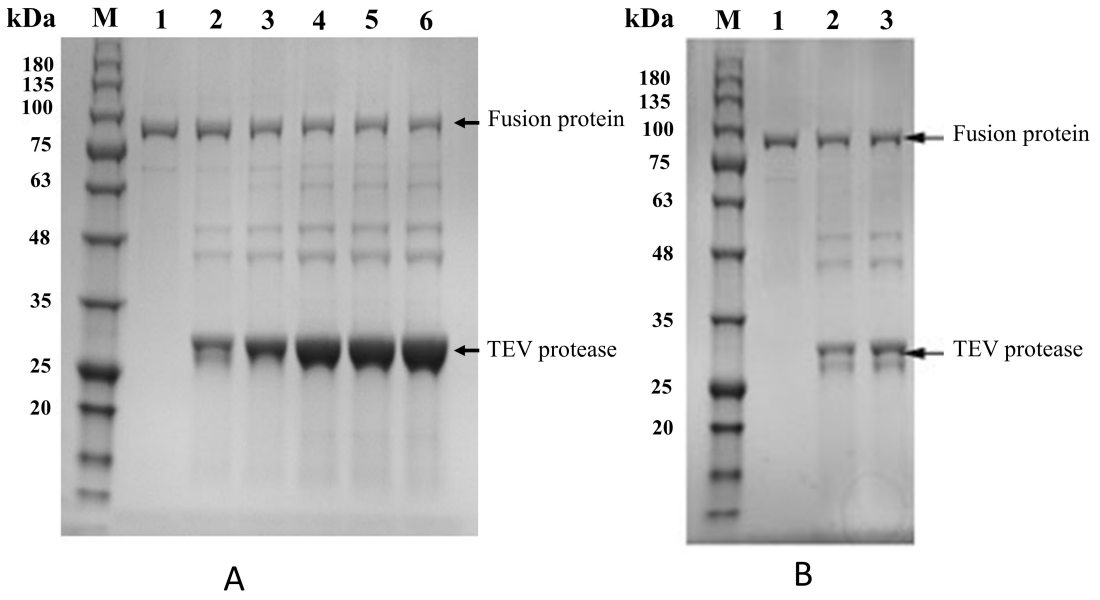

Figure 3.15. SDS-PAGE (TruPAGE™ Precast gels (4-20%)) analysis of Cytb_{562}-Vip1Ac1 fusion protein (95.77 kDa) after treated with TEV protease.
A. Incubation of fusion protein with different ratio of TEV protease at R.T for 24 h**,** lane 1 corresponds to Cytb_{562}-Vip1Ac1 fusion protein**,** lanes 2-6 correspond to Cytb_{562}-Vip1Ac1 fusion protein with TEV protease at ratios of 1:1, 1:2, 1:3, 1:4, and 1:5. **B**. Incubation of fusion protein with commercial TEV-Express enzyme at ratio 1:1 at R.T for 24 h , lane 1 corresponds to Cytb_{562}-Vip1Ac1 fusion protein, lane 2 corresponds to Cytb_{562}-Vip1Ac1 fusion protein with TEV protease, lane 3 corresponds to Cytb_{562}-Vip1Ac1 fusion protein with commercial TEV-Express enzyme.

3.2.5 Reconstruction of pET22b vector encoding Vip1Ac1 without Cyt b_{562}

To overcome the problem that the fusion protein could not be cleaved, the pET22b vector was reconstructed (Section 2.2.9.1) to remove cytochrome *b* 562 and the TEV cleavage site, leaving only the signal sequence from cytochrome *b* 562 linked with the Vip1Ac1 protein (Figure 3.16). This new construct should allow Vip1Ac1 protein secretion into the periplasmic space and avoid the need for TEV, thereby diminishing the steps of production and purification for Vip1Ac1 protein.

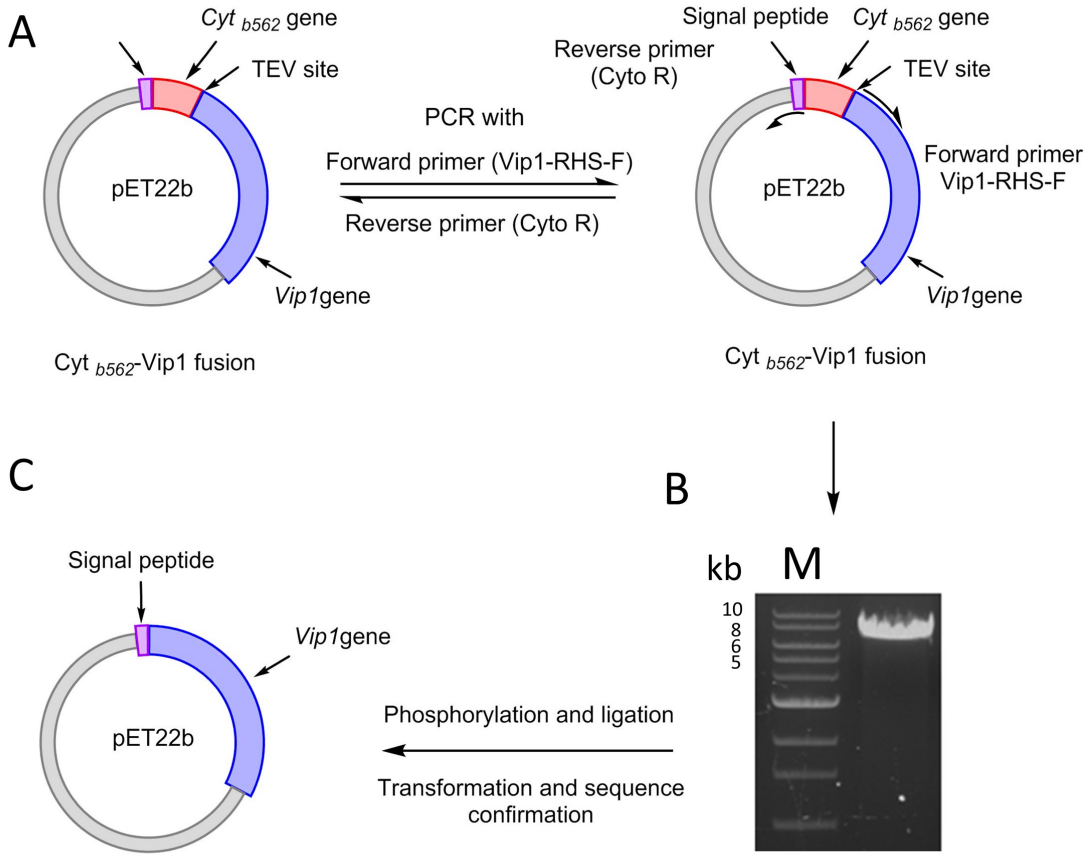

Figure 3.16. Construction of pET22b plasmid encoding Vip1Ac1 protein only.
A. pET22b vector with *Cyt* b_{562} gene (red) plus TEV site and *Vip1Ac1* gene (blue) used in whole plasmid PCR with mutagenic primers to remove *Cyt* b_{562} gene and TEV site. **B.** Agarose (1%) gel electrophoresis of PCR product showing linear pET22b vector (~ 7.9 kb). **C.** Final map of pET22b construct with signal sequence of cytochrome b_{562} (purple) plus *Vip1Ac1* gene (blue).

The construct was transformed into *E. coli* BL21 cells and protein expression was induced using the previous optimum conditions for fusion expression (0.1 mM IPTG and 24 h induction period) at different temperatures (20, 25, 30 and 37˚C). The removal cytochrome b_{562} did not affect the solubility of the Vip1Ac1 protein and the presence of only the signal sequence from cytochrome b_{562} led to successful secretion of Vip1Ac1 protein (~82 kDa) into the periplasmic space. The conditions for optimum production were the same as for the fusion protein except that the best temperature was 20˚C, Figure 3.17.

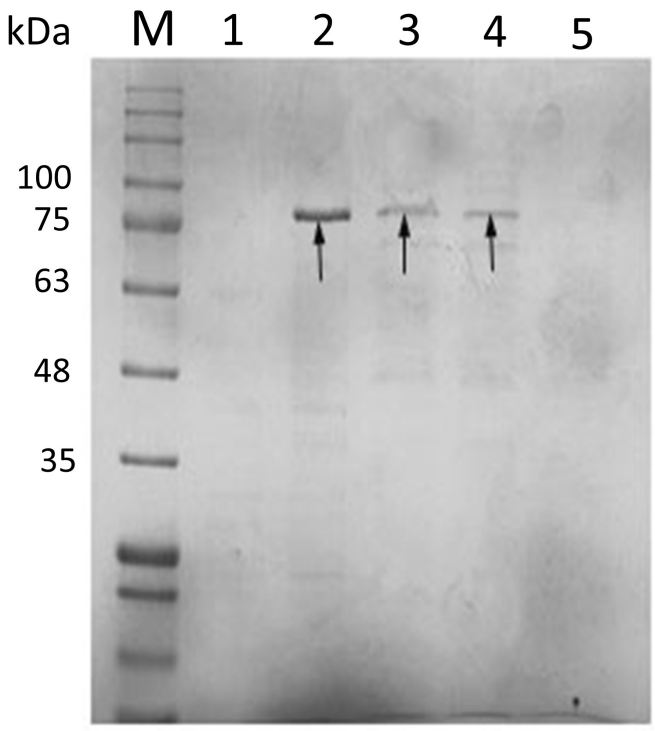

Figure 3.17. Vip1Ac1 protein expression in periplasmic space.
SDS-PAGE (12.5%) analysis of Vip1Ac1 protein expression in periplasmic space at different temperatures. Lane M corresponds to protein molecular weight standards. Lane 1 corresponds to total periplasmic proteins of preinduced cells. Lane 2 corresponds to total periplasmic proteins of induced cells at 20˚C. Lane 3 corresponds to total periplasmic proteins of induced cells at 25℃. Lane 4 corresponds to total periplasmic proteins of induced cells at 30℃. Lane 5 corresponds to total periplasmic proteins of induced cells at 37˚C. Expression was induced using 0.1 mM IPTG as inducer for 24 h. Bold arrows show the Vip1Ac1 protein (~ 82 kDa).

Overall, the above results show that expression of Vip1Ac1 in the cytoplasm affects the solubility of the protein leading to accumulation in insoluble form, while expression of the protein in the periplasmic space dramatically enhances its solubility. To allow efficient purification of Vip1Ac1 from crude lysates a further modification was made to the construct to add an N-terminal His-tag and TEV cleavage site between the signal peptide and Vip1Ac1 protein as described in Section 2.2.9.1 and shown in Figure 3.18.

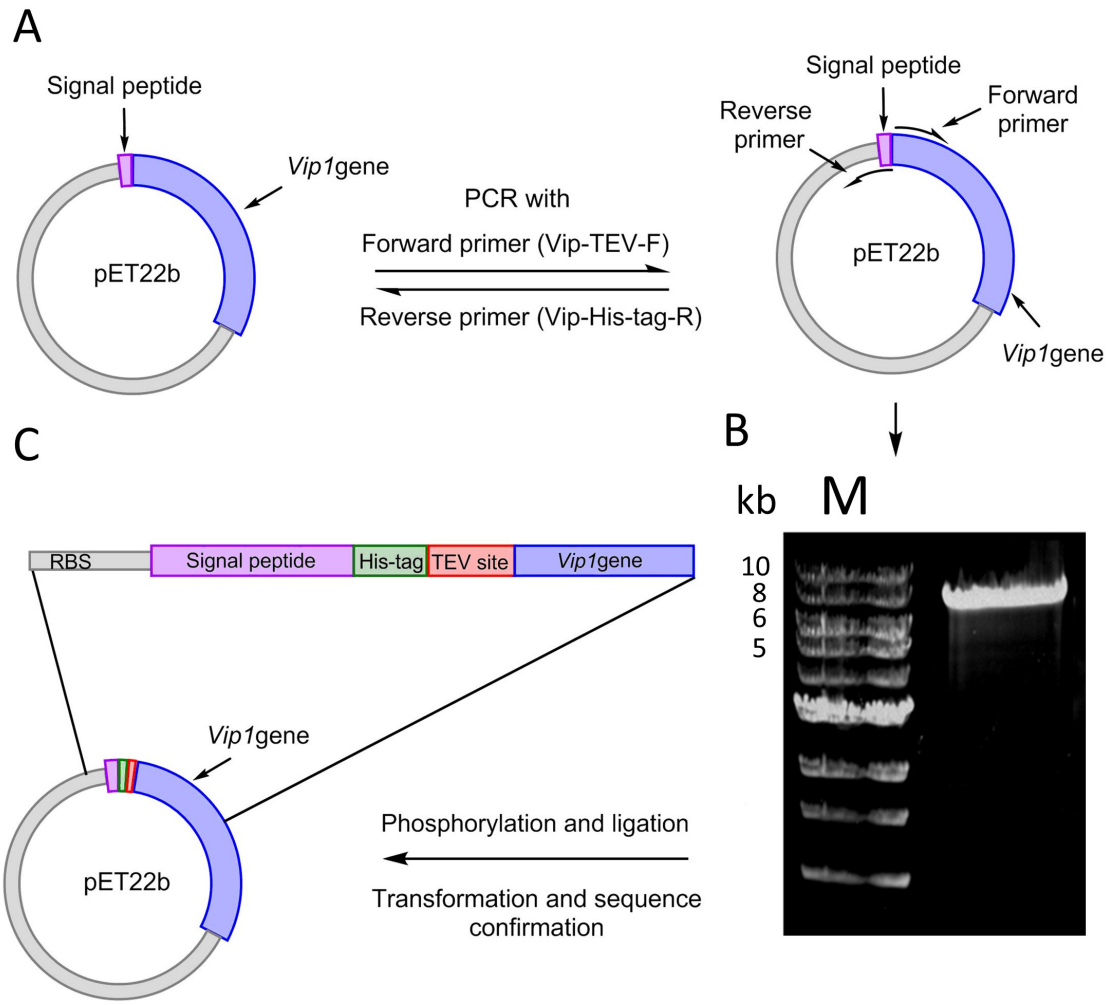

Figure 3.18. Construction of pET22b plasmid encoding Vip1Ac1 protein only.
A. pET22b vector with *Vip1Ac1* gene (blue) plus periplasmic signal sequence (purple) used in whole plasmid PCR with mutagenic primers to add TEV site and His-tag. **B.** Agarose (1%) gel electrophoresis of PCR product showing linear pET22b vector (~ 7.9 kb) **C.** Final map of pET22b construct with periplasmic signal sequence (purple), His-tag (green), TEV site (red) and *Vip1Ac1* gene (blue).

Following transformation of the construct into *E. coli* BL21 cells and induction of protein expression using the previous conditions (Section 3.2.5) (0.1 mM IPTG for 24 h induction period) it is clear that, addition of the His-tag and cleavage site at the N-terminus did not affect the solubility of the protein and the Vip1Ac1 expressed as a soluble protein in the periplasmic space and purified. The His-tag was successfully used to purify the Vip1Ac1 protein and was then removed by TEV enzyme as shown in Figure 3.19. This construct was used for large scale Vip1Ac1 production for all further studies.

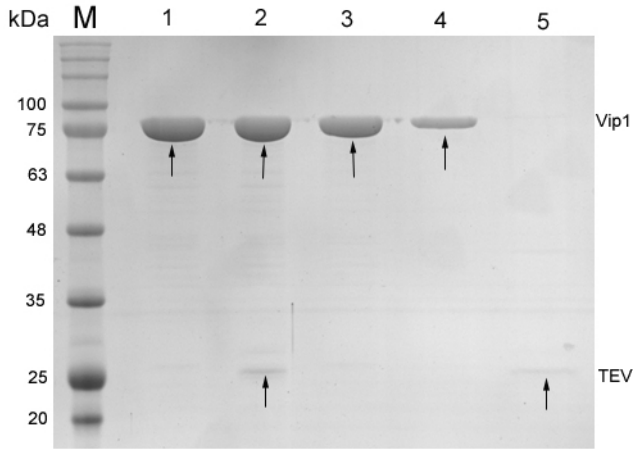

Figure 3.19. His-tag removing from Vip1Ac1 protein by TEV protease.
SDS-PAGE analysis (12.5%) of removing His-tag from Vip1Ac1 protein. Lane M corresponds to protein molecular weight standards. Lane 1 corresponds to pure Vip1Ac1 protein from Ni-NTA affinity purification. Lane 2 corresponds to the pure Vip1Ac1 protein plus pure TEV enzyme before Ni-NTA affinity purification. Lane 3 corresponds to the flow-through from the Ni-NTA column. Lane 4 corresponds to the wash fraction. Lane 5 corresponds to elution fraction which shows only TEV enzyme.

3.2.6 Construction of pET22b vector containing *Vip2Ae3* gene

The partner protein to Vip1, Vip2 was required to study the interaction between two proteins during pore formation in artificial membranes. Vip2 protein has been found in both the soluble and insoluble fractions when expressed in BL21(DE3) using the pET28b vector as reported by Shi *et al.* (Shi *et al.* 2004). To overcome the solubility issue of Vip2, the same strategy that was used to express Vip1Ac1 in the periplasmic space as a soluble protein, was used for Vip2Ae3. The *Vip2Ae3* gene without signal sequence was isolated from the original Bt strain H29.3 (Section 2.2.2 using bacterial culture as a DNA template and using forward and reverse primers with *Bam*HI and *Eco*RI sites respectively; Table 2.3) and inserted between the *Bam*HI and *Eco*RI sites of the pET22b vector containing the *Cyt* b_{562} gene, which was previously constructed as described in Section 3.2.2.3 (Section 2.2.6 and Section 2.2.8) to form pET22b vector containing *Cytb*$_{562}$*-Vip2Ae3* gene as outlined in Section 2.2.9.2 and shown in Figure 3.20. *Cytb*$_{562}$*-Vip2Ae3* fusions were selected using colony PCR as shown in Figure 3.21.

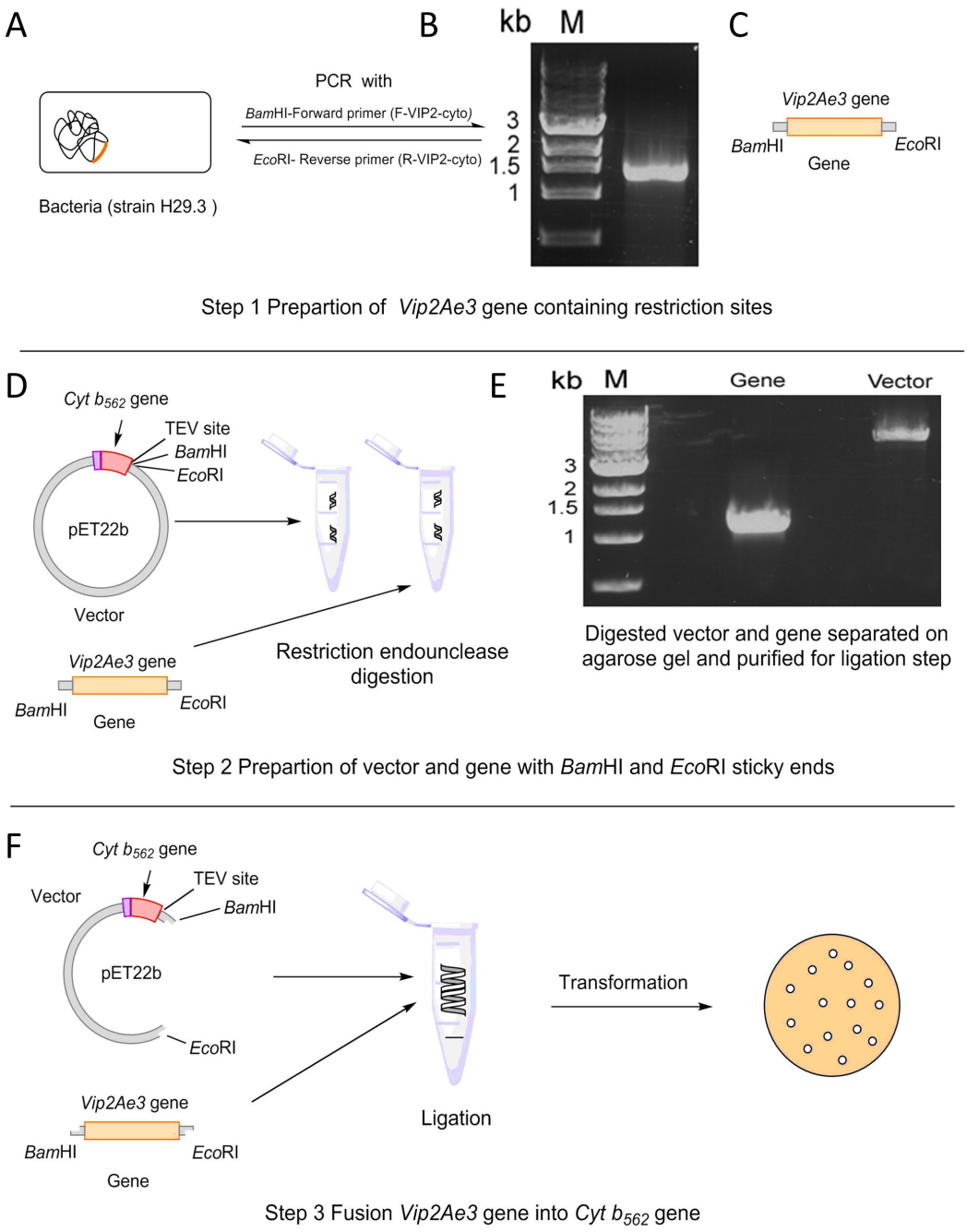

Figure 3.20. Construction of pET22b vector containing *Cytb$_{562}$-Vip2Ae3* gene.
A. Bacterial strain H29.3 used in PCR as a template to mutagenic primers to amplify *Vip2Ae3* gene with *Bam*HI and *Eco*RI restriction sites. **B**. Agarose (1%) gel electrophoresis of PCR product showing *Vip2Ae3* gene (~ 1.26 kb). **C.** *Vip2Ae3* gene with *Bam*HI and *Eco*RI restriction sites. **D**. Restriction endonuclease digestion for both vector and gene using *Bam*HI and *Eco*RI enzymes. **E.** Agarose (1%) gel electrophoresis of PCR product showing *Vip2Ae3* gene (~ 1.26 kb) and linear pET22b vector (~ 5.75 kb). **F**. *Vip2Ae3* gene inserted into the pET22b vector between the *Bam*HI and *Eco*RI sites by ligation and the mixture transformed into competent cells.

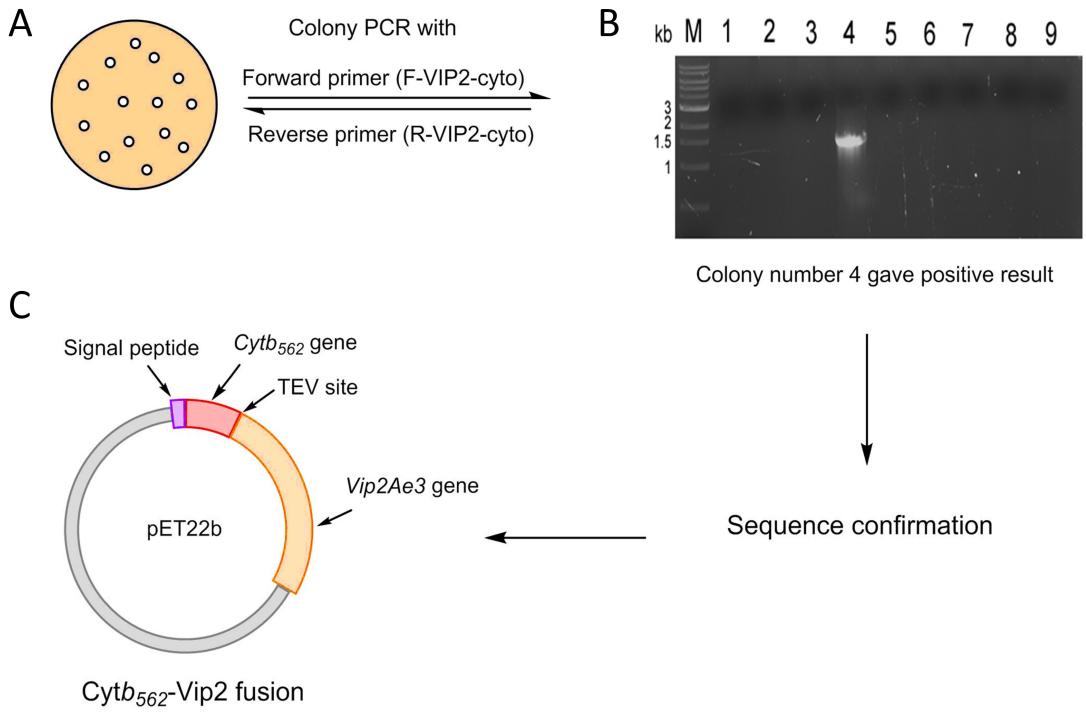

Figure 3.21. Colony PCR for *Cytb$_{562}$-Vip2Ae3* fusion gene selection.
A) Colonies transformed with ligation mixture (pET22b vector containing *Cyt b$_{562}$* gene and *Vip2Ae3* gene, were selected for colony PCR using Vip2Ae3 primers (F-Vip2-cyto and R-Vip2-cyto). **B)** Agarose (1%) gel electrophoresis of PCR products from selected colonies showing colonies number 4 contain the *Vip2Ae3* gene (~ 1.26 kb). **C)** Final map of pET22b construct containing *Cytb$_{562}$-Vip2Ae3* fusion gene.

Expression of fusion protein (Cyt*b*$_{562}$-Vip2Ae3) was carried out using the expression conditions for Vip1Ac1 (Section 3.2.5). The fact that induced cells exhibited a distinct red colour highlighted successful expression of the fusion protein as shown in Figure 3.22A. Fusion protein was expressed as a soluble protein in the periplasmic space as shown in Figure 3.22B.

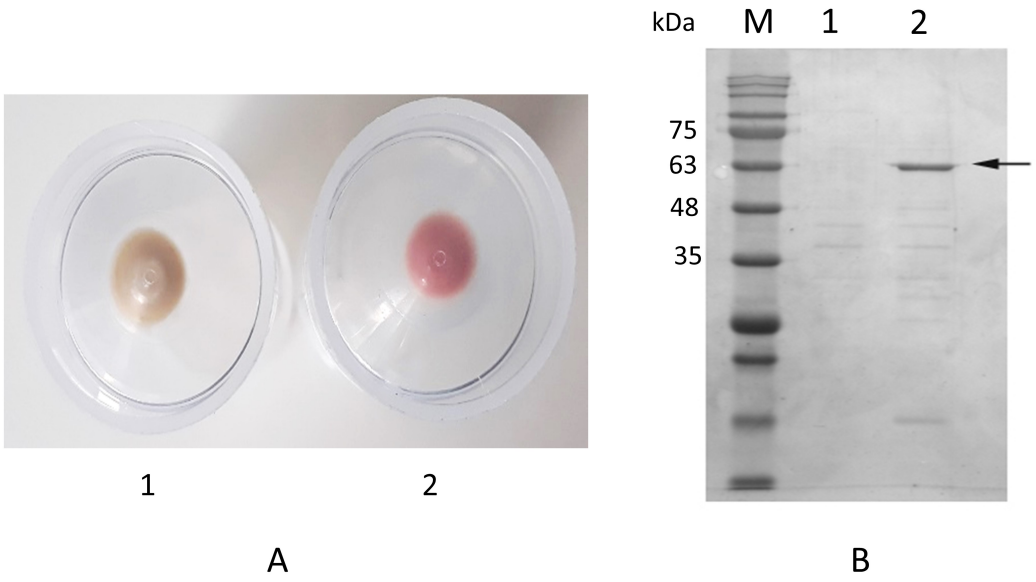

Figure 3.22. Expression of the Cyt b_{562}-Vip2Ae3 fusion protein in *E. coli* BL21 cells.
A. Image showing cell pellet. **1**.Pellet from 10 ml culture of pre-induction cells. **2**. Pellet from 10 ml culture of post-induction cells. **B**. SDS-PAGE (12.5%) analysis of Cyt b_{562}-Vip2Ae3 fusion protein expression at 20˚C for 3 h using 0.1 mM IPTG. Lane M corresponds to protein molecular weight standards. Lane number 1 corresponds to total periplasmic proteins of pre-induction cells. Lane number 2 corresponds to total periplasmic proteins of post-induction cells. Bold arrow shows the Cyt b_{562}-Vip2Ae3 protein (61.5 kDa).

The vector containing the fusion gene was reconstructed to remove the *Cyt* b_{562} gene leaving only the signal sequence of cytochrome b_{562} linked with Vip2Ae3 protein to direct the protein into the periplasmic space (Section 2.2.9.2). During the construction an N-terminal His-tag and TEV site were added between the signal peptide and the Vip2Ae3 sequence (Figure 3.23). Vip2Ae3 protein from this construct was successfully expressed in soluble form in the periplasmic space. This construct was used for large scale Vip2Ae3 production in further studies.

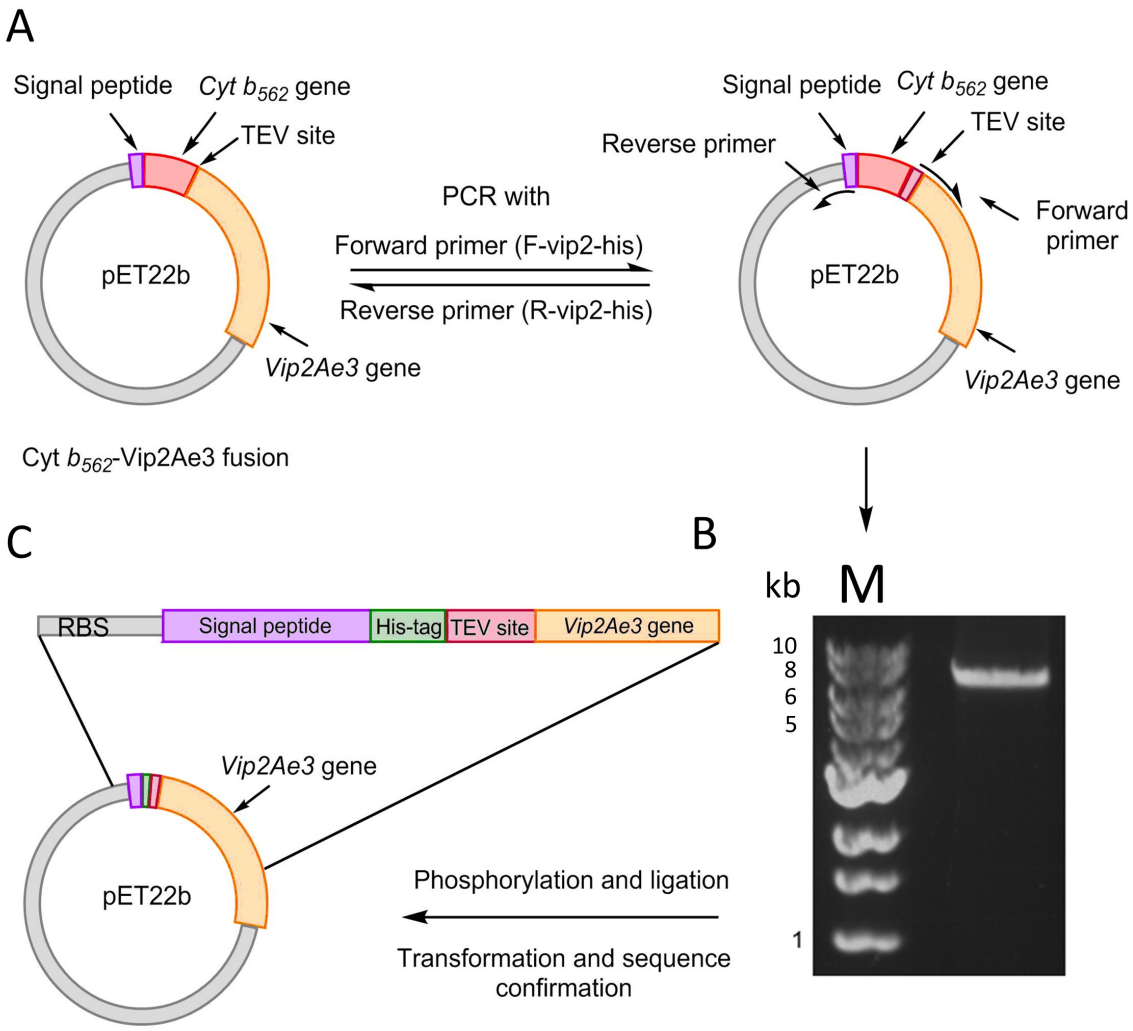

Figure 3.23. Schematic maps of pET22b plasmid encoding Vip2Ae3 protein only.
A. pET22b construct containing *Cytb*$_{562}$ gene (red) plus periplasmic signal sequence (purple) and *Vip2Ae3*gene (orange) used in whole plasmid PCR with mutagenic primers to remove *Cytb*$_{562}$ gene and add His-tag and TEV site. **B.** Agarose (1%) gel electrophoresis of PCR product showing linear pET22b vector (~ 7 kb) **C.** Final map of pET22b construct with periplasmic signal sequence (purple), His-tag (green), TEV site (red) and *Vip2Ae3* gene (orange).

3.2.7 Production and purification of Vip1Ac1 and Vip2Ae3 proteins

The final constructs carrying either the *Vip1Ac1* gene or *Vip2Ae3* gene as described in Section 3.2.5 and Section 3.2.6 respectively were used to express the Vips in the periplasm as described in Section 2.3.2 and Section 2.3.6.2.

During Ni^{2+}-affinity chromatography (Section 2.3.8.1), three elution fractions were collected. More than ~ 90% of the protein eluted in the first elution fraction with a high

level of purity (Figure 3.24A and Figure 3.25A). Fractions containing each Vip protein were pooled and subjected to the second step of purification.

Samples were concentrated into 2 ml and loaded into a SEC column as described in Section 2.3.8.3. Protein elution was monitored by absorption at 280 nm and showed a single major peak as shown in Figure 3.24B and Figure 3.25B. Fractions corresponding to these peaks were analysed via SDS-PAGE (Figure 3.24 C and Figure 3.25C). Fractions containing the purified Vip proteins were pooled and concentrated into 1 ml, then the His-tag was removed by incubation of the purified protein with TEV protease as described in Section 2.4.3. Purified Vip proteins were either used directly or stored at 4˚C until used.

Expression of both proteins (Vip1Ac1 and Vip2Ae3) in the periplasmic space with His-tags allows simple purification of both proteins with high levels of purity using a single purification step (affinity purification) as shown in Figure 3.24 A and Figure 3.25A compared with the fusion (Cyt b_{562}-Vip1) without His-tag (Figure 3.12 B). The fact that there are fewer contaminating proteins present in the periplasm compared with the cytoplasm, facilitating purification. High level of purity has been achieved from the first purification step (affinity purification; Figure 3.24A and Figure 3.25A). SEC was then used as a second purification step followed by buffer exchange to remove all other minor contamination and imidazole while transferring the protein to Tris-HCl buffer for further study.

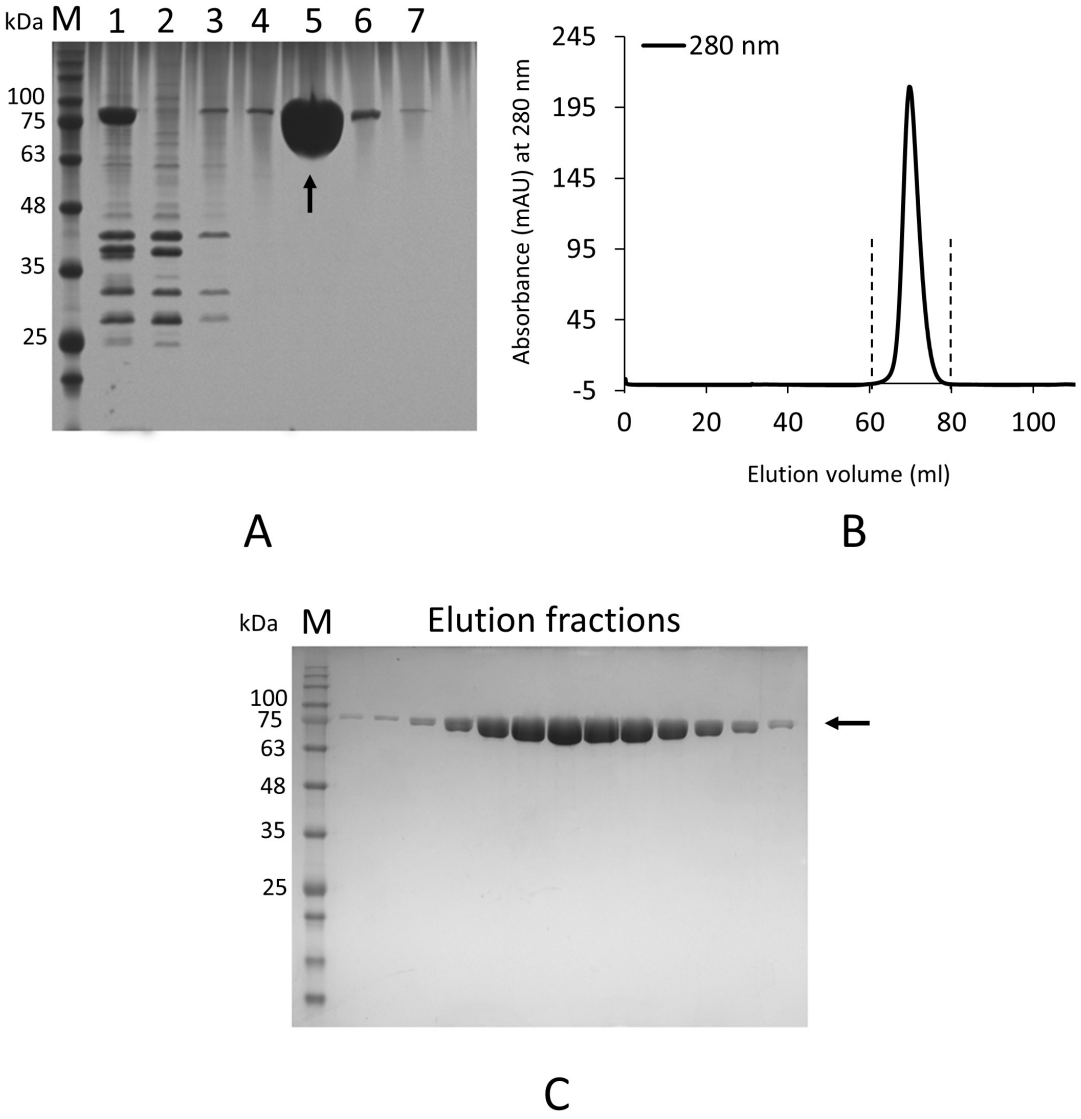

Figure 3.24. Purification profile of Vip1Ac1 protein.
A) SDS-PAGE (12.5%) analysis of nickel affinity purification. Lane M corresponds to the protein marker of known molecular weights. Lane 1 corresponds to whole cell lysate, lane 2 corresponds to flow-through fraction, lanes 3 and 4 correspond to wash 1 and wash 2 fractions respectively and lanes 5, 6 and 7 correspond to elution 1, elution 2 and elution 3 fractions respectively. **B)** Chromatogram of SEC purification showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/60 Superdex™ S200 pg column. The peak represents Vip1Ac1 protein and dashed line corresponds to the elution fraction area **C)** SDS-PAGE (12.5%) analysis of elution fractions from SEC. Bold arrows show the Vip1Ac1 protein (~ 82 kDa).

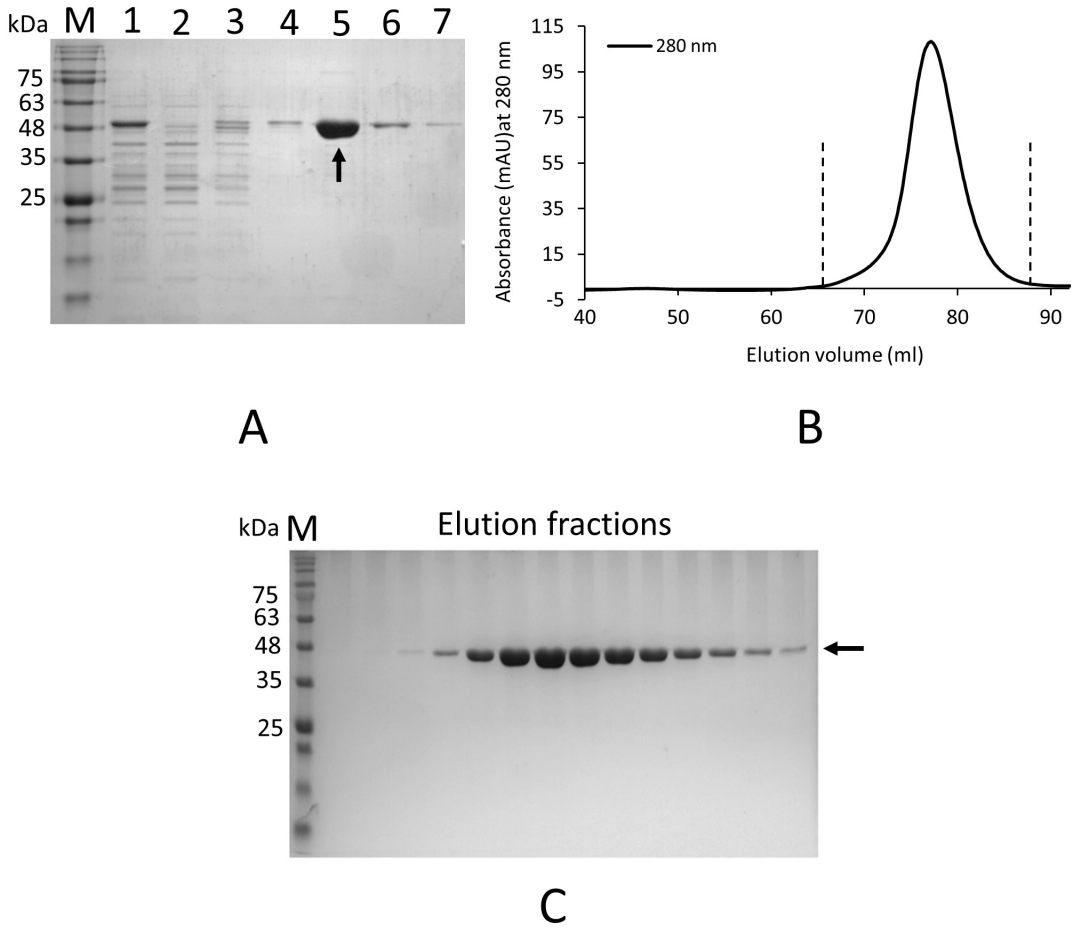

Figure 3.25. Purification profile of Vip2Ae3 protein.
A) SDS-PAGE (12.5%) analysis of nickel affinity purification. Lane M corresponds to the protein marker of known molecular weights. Lane 1 corresponds to whole cell lysate, lane 2 corresponds to flow-through fraction, lanes 3 and 4 correspond to wash 1 and wash 2 fractions respectively and lanes 5, 6 and 7 correspond to elution 1, elution 2 and elution 3 fractions respectively. **B)** Chromatogram of SEC purification showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/60 Superdex™ S200 pg column. The peak represents Vip2Ae3 protein and dashed line corresponds to the elution fraction area. **C)** SDS-PAGE (12.5%) analysis of elution fractions from SEC. Bold arrows show the Vip2Ae3 protein (~47.9 kDa).

3.2.8 Incorporation of ncAAs into Vip1Ac1 protein

As mentioned in chapter one, incorporation of ncAAs (amino acid analogues with a functional group) into target proteins allows for the engineering of proteins for different purposes by introduction of new and useful chemistry that does not exist in nature. In this study, Vip1Ac1 protein was engineered by genetically incorporating ncAAs for two main purposes: incorporation of the ncAA 4-Iodo-L-phenylalanine containing iodine as

a heavy atom into Vip1Ac1 protein to solve the phasing problem of Vip1Ac1 crystals experimentally using a heavy atom approach as described in chapter 4. Secondly, incorporation of the ncAA *p*-azido-L-phenylalanine (azF) containing an azide group as reactive handle for bioorthogonal reaction and labelling Vip1Ac1 protein with fluorescent dye via SPAAC reaction (Section 2.5.3) and used for visualization of the Vip1Ac1 protein interacting with lipid membrane as described in chapter 5.

3.2.8.1 Generation of amber stop codon (TAG) mutations in the *Vip1Ac1* gene

As previously demonstrated in chapter one, the incorporation of ncAAs requires reprogramming the genetic code (amber stop codon; TAG). Two tyrosine residues at positions Y358 and Y536 were chosen as surface residues in the Vip1Ac1 model for ncAA incorporation. TAG codons were generated in the *Vip1Ac1* gene at Y358TAG and Y536TAG by whole vector site-directed mutagenesis as described in Section 2.2.9.3. Introduction of TAG mutant in *Vip1Ac1* gene are illustrated in Figure 3.26. The vector containing the *Vip1Ac1* gene was amplified by PCR using a pair of oligonucleotide primers containing the mutation for Y358TAG (Table 2.3). PCR results showed a DNA band of ~ 7.9 kb that corresponded to the original size of pET22b construct. DNA was extracted and purified from the gel as described in Section 2.2.4.3, phosphorylated and ligated to re-circularise the vector as described in Sections 2.2.7 and 2.2.8 respectively. Re-circularised construct was transformed into Top10^{TM} electrocompetent *E. coli* (Section 2.2.10.2) and subsequently plated onto LB agar supplemented with ampicillin for clone selection. Vectors were extracted (Section 2.2.4.1) and sent for sequencing to confirm the Y358TAG mutation. A successful mutant was used as a template to introduce the second mutation, Y536TAG, in the same manner.

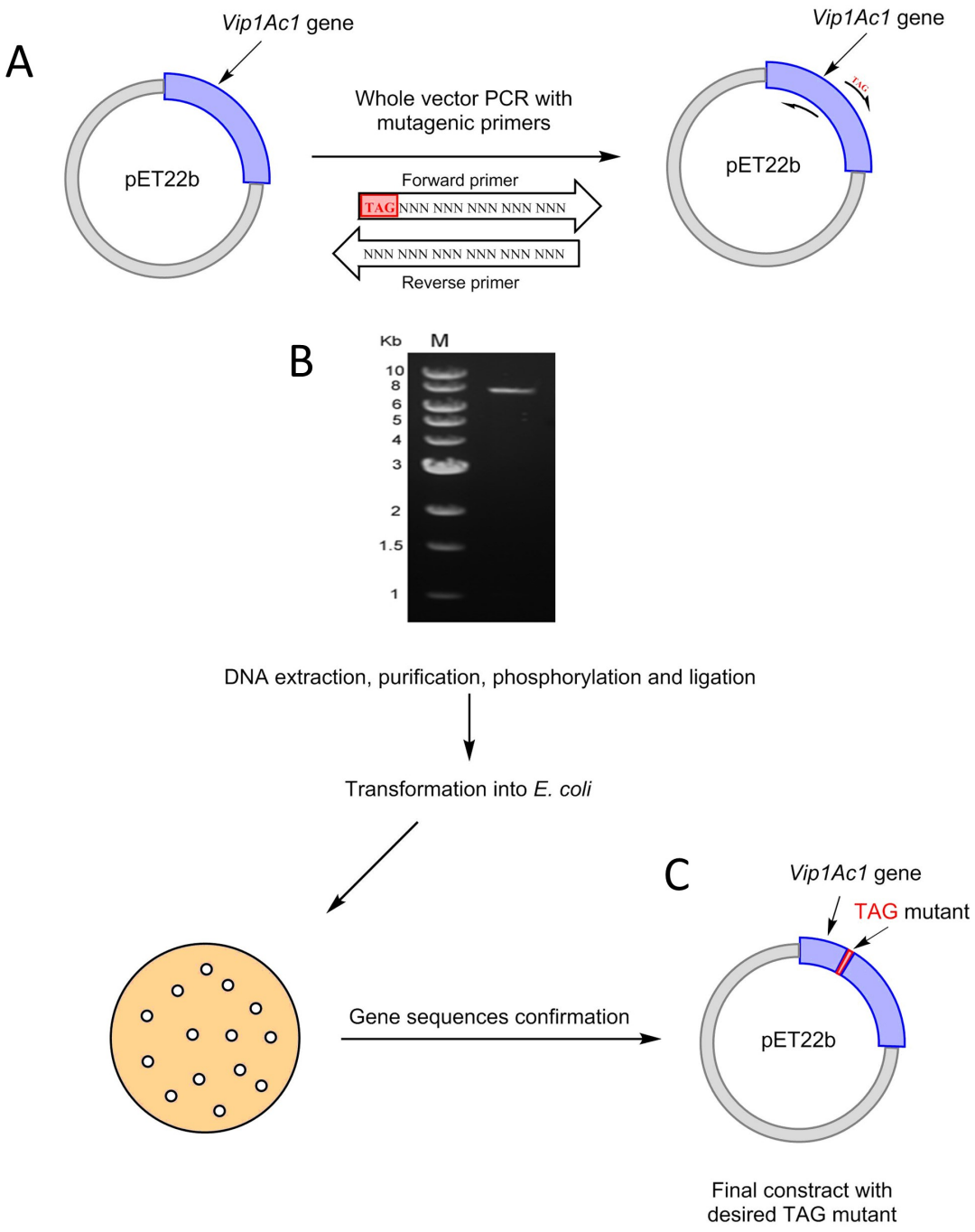

Figure 3.26. Schematic maps of introduce TAG mutation in *Vip1Ac1* gene.
A. pET22b construct with *Vip1Ac1* gene (blue) using in whole plasmid PCR with mutagenic primers to introduce TAG mutation. **B.** Agarose (1%) gel electrophoresis of PCR product showing linear pET22b construct (~ 7.9 kb) **C.** Final map of pET22b construct with *Vip1Ac1* gene (blue) plus desired TAG mutation (red).

3.2.8.2 Expression and purification of Vip1Ac1 containing ncAA.

Incorporation of ncAA was achieved by transforming BL21 electrocompetent *E. coli* with the pET22b construct with the *Vip1Ac1* gene containing two TAG mutations (Y358TAG and Y536TAG) and the pDULE vector (Table 2.2; Figure 2.1) carrying the genes for the tRNA/aaRS pair for 4-Iodo-L-phenylalanine and *p*-azido-L-phenylalanine (azF) incorporation (tyrosine derivative) as described in Section 2.2.10.2.

Vip1Ac1 containing ncAA was expressed as described in Section 2.3.4.1, assessed in the presence and absence of ncAA by SDS-PAGE as described in Section 2.3.4. As expected, in the absence of ncAA and IPTG no protein was produced and in the presence of IPTG without ncAA, truncated protein was produced. Only a small amount of full-length protein containing ncAA was successfully produced in the presence of ncAA and IPTG Figure 3.27A and B. This indicates successful incorporation of ncAA into full-length Vip1Ac1 at low efficiency.

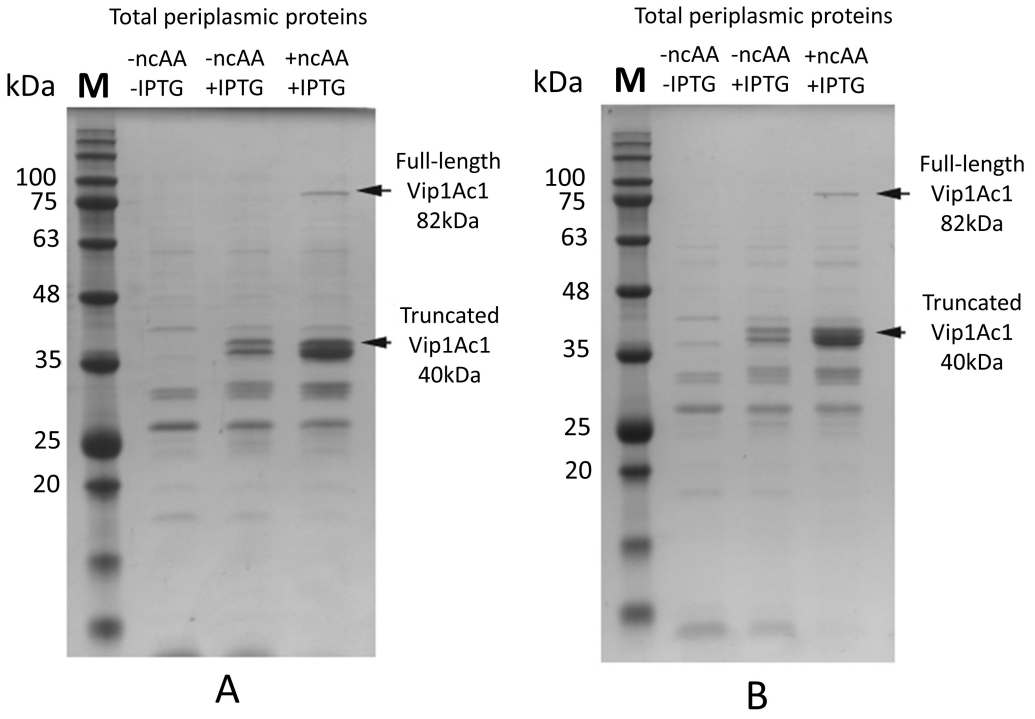

Figure 3.27. Expression of Vip1Ac1 containing ncAA.
A. SDS-PAGE (12.5%) analysis of Vip1Ac1 containing 4-Iodo-L-phenylalanine. **B.** SDS-PAGE (12.5%) analysis of Vip1Ac1 containing azF. Lane M corresponds to protein molecular weight standards. Lane - ncAA and - IPTG corresponds to total periplasmic extract of preinduced cells without ncAA. Lane - ncAA and + IPTG corresponds to total periplasmic extract of induced cells without ncAA. Lane + ncAA and + IPTG corresponds to total periplasmic extract of induced cells in the presence of ncAA. Total volume of periplasmic extract was 10 ml from 100 ml batch culture and 5 µl was loaded in the gel.

Vip1Ac1$^{358\text{-}536\ ncAA}$ was extracted and purified as for wildtype Vip1Ac1 as described in Section 3.2.7. The yield of Vip1Ac1$^{358\text{-}536\ ncAA}$ using both ncAA were low compared with wild type Vip1Ac1 yield as can be observed in Figure 3.28B and Figure 3.29B. The majority of synthesised protein stopped at the first stop codon (~40 kDa) with a minor band at ~60 kDa corresponding to truncation at the second stop codon Figure 3.28A and Figure 3.29A. However, the small amount of full-length protein with 2 ncAA was enough to perform the experiments (Crystallography studies and labelling).

The first step of purification by His-tag affinity purification did not achieve a high level of purity since both full-length and truncated forms possess the N-terminal His-tag. SDS-PAGE analysis of elution fractions from nickel affinity chromatography showed a number of protein bands which correspond to full-length Vip1Ac1 with 2 ncAA (~82 kDa), truncated protein with 1 ncAA at position 358 (~60 kDa), truncated protein with no ncAA (~40 kDa) and other bands, which may represent digested truncated proteins as shown in Figure 3.28A (4-Iodo-L-phenylalanine) and Figure 3.29A (azF). Fractions containing full-length Vip1Ac1$^{358\text{-}536\ ncAA}$ (~ 82 kDa) were pooled and purified by SEC as shown in Figure 3.28B (4-Iodo-L-phenylalanine) and Figure 3.29B (azF).

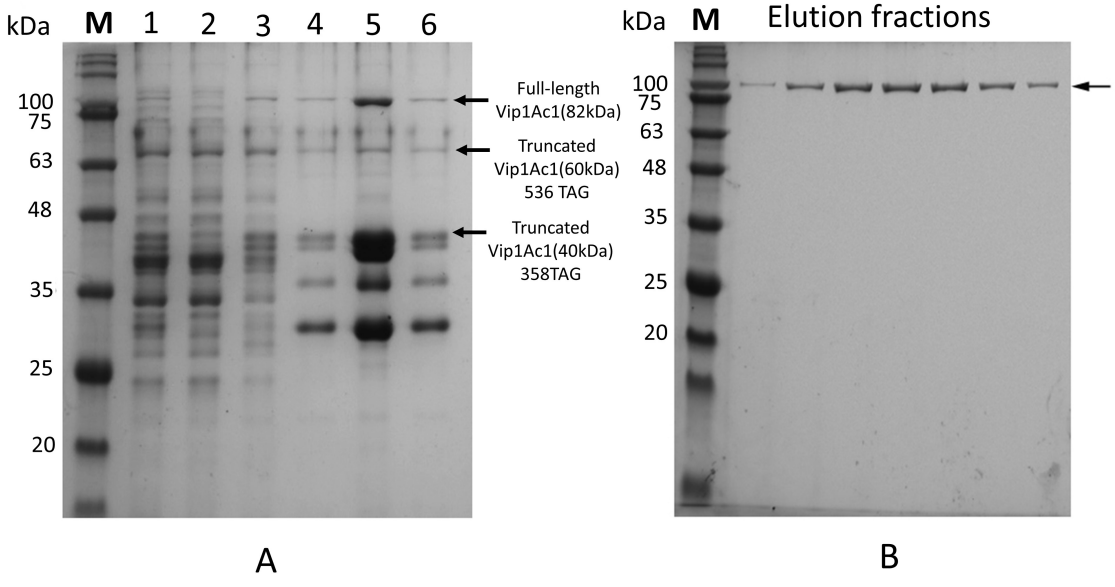

Figure 3.28. Purification profile of Vip1Ac1 containing 4-Iodo-L-phenylalanine.
A. SDS-PAGE (12.5%) analysis of His-tag purification. Lane 1 corresponds to periplasmic extract (60 ml), lane 2 corresponds to flow-through, lanes 3 and 4 correspond to washing fractions (4 ml), lanes 5 and 6 correspond to elution fractions (4 ml). **B**. SDS-PAGE (12.5%) analysis of SEC purification. Lane M corresponds to protein molecular weight standards and other lanes correspond to the elution fractions. Bold arrow shows the Vip1Ac1$^{358\text{-}536\ ncAA}$ protein (~ 82 kDa).

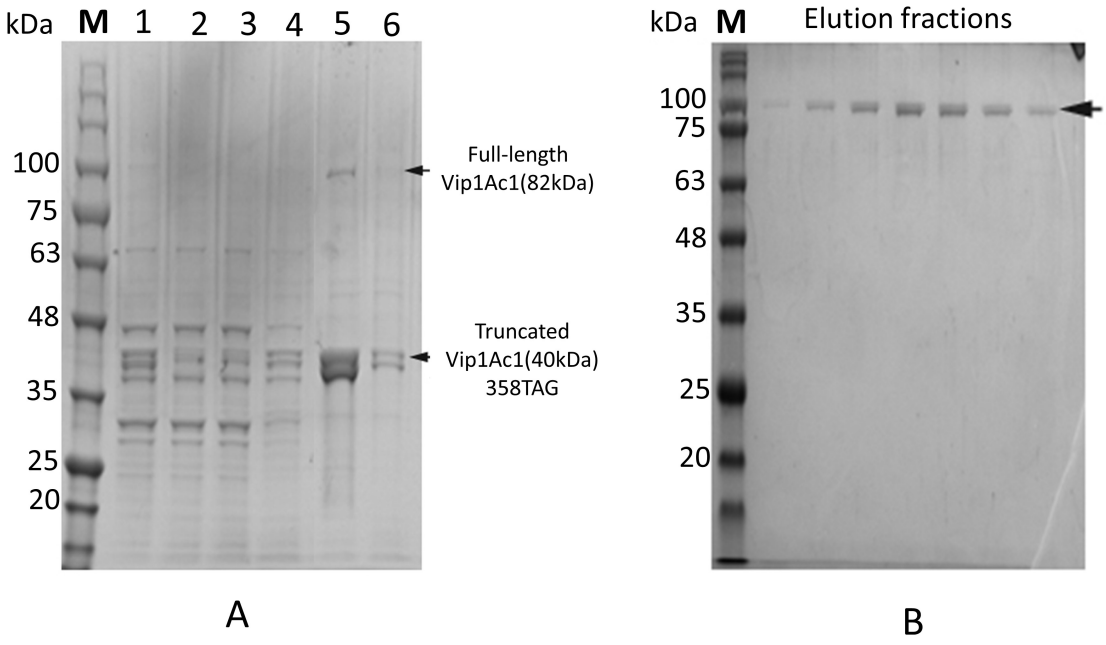

Figure 3.29. Purification profile of Vip1Ac1 containing azF.
A. SDS-PAGE analysis (TruPAGE^{TM} Precast gels (4-20%)) of His-tag purification. Lane 1 corresponds to periplasmic extract (60 ml), lane 2 corresponds to flow-through, lanes 3 and 4 correspond to washing fractions (4 ml), lanes 5 and 6 correspond to elution fractions (4 ml). **B**. SDS-PAGE (12.5%) analysis of SEC purification. Lane M corresponds to protein molecular weight standards and other lanes correspond to the elution fractions. Bold arrow shows the Vip1Ac1$^{358\text{-}536\ ncAA}$ protein (~ 82 kDa).

3.2.9 Expression of Vip1Ac1 containing selenomethionine (SeMet)

Substitution of methionine with SeMet in recombinant proteins to phase molecular structures experimentally is a common technique to solve the molecular structure by multiwavelength anomalous dispersion (MAD; Strub *et al.* 2003). There are different methods for SeMet substitution in recombinant proteins. The first uses methionine auxotrophic *E. coli* as an expression host to achieve 100 % substitution (Hendrickson, *et al.* 1990) but this method has disadvantages of long log time, low cell density and poor protein yield (Walden 2010). Another uses non auxotrophic *E. coli* as an expression host. In this method, the methionine biosynthesis pathway must be blocked, and this can be achieved by adding isoleucine, lysine, and threonine in high concentrations to inhibit the enzyme (aspartokinase) responsible for methionine biosynthesis in *E. coli*. This method has become most common and routinely used for SeMet substitution (Doublié 2007) and was used to substitute SeMet in Vip1Ac1 as described in Section 2.3.5. Protein extraction was performed as described in Section 2.3.6.2 and purified by nickel affinity

chromatography (Section 2.3.8.1) followed by SEC (Section 2.3.8.3) as shown in Figure 3.30. As SeMet is prone to oxidation, which might affect the crystal diffraction (Bateman *et al.* 2002), all the purification buffers were degassed and contained reducing agent (5 mM DTT). The yield of Vip1Ac1 protein containing SeMet was low compared with the yield of wild type Vip1Ac1 Figure 3.24 A-C. This yield of protein is expected as the nutrient restricted, defined M9 salts medium was used for expression. However, this amount of protein was enough to establish crystal trials using the same or very similar conditions as for native protein.

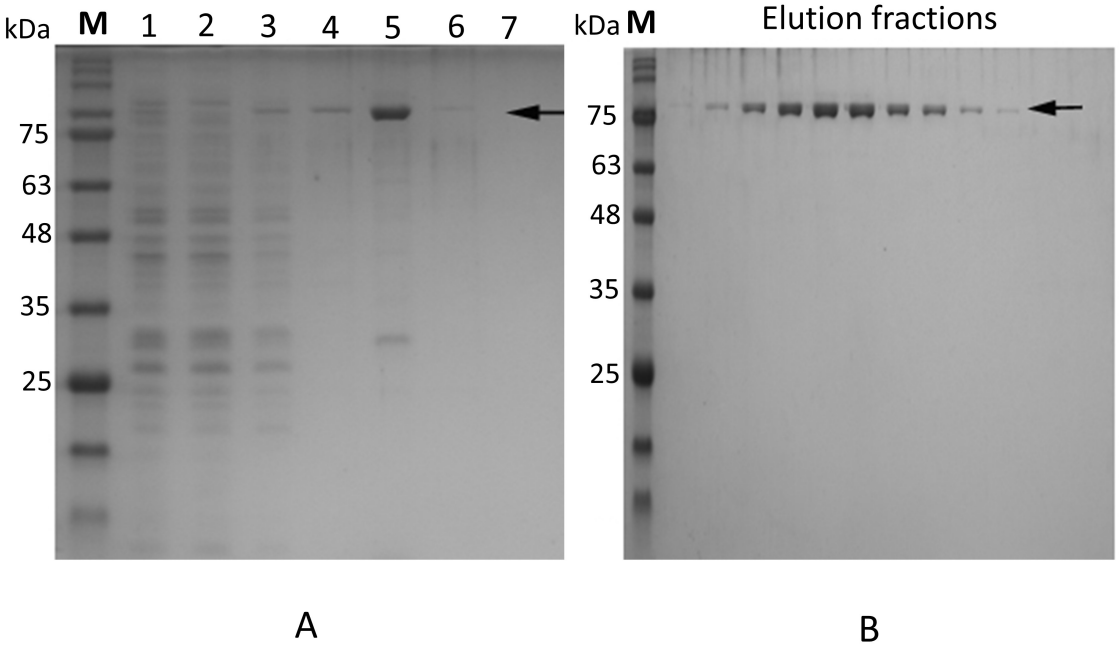

Figure 3.30. Purification profile of Vip1Ac1 containing SeMet.
A. SDS-PAGE (12.5%) analysis of His-tag purification. Lane 1 corresponds to periplasmic extract (60 ml), lane 2 corresponds to follow-through, lanes 3 and 4 correspond to washing fractions (4 ml), lanes 5,6 and 7 correspond to elution fractions (4 ml). **B**. SDS-PAGE (12.5%) analysis of SEC purification. Lane M corresponds to protein molecular weight standards and other lanes correspond to the elution fraction. Bold arrows show the Vip1Ac1 protein (~ 82.51 kDa).

3.3 Conclusion

As mentioned at the start of this chapter, high level of pure protein is required for crystallography studies, and this can be achieved by recombinant protein technology and protein engineering. These technologies allow proteins to be produced in their native forms, in high amounts, using an expression host while also producing protein with improved properties such the presence of a purification tag or protein fusion to enhance solubility and stability. Additionally, by these technologies a new chemistry can be added to the protein at a specific position by incorporation ncAA. Successful recombinant protein production needs a good correlation between three elements; expression of the gene; solubility of protein; and its ability to be purified to homogeneity (Costa *et al.* 2014). Some recombinant proteins express as inclusion bodies (insoluble protein) or unstable protein, which undergoes precipitation or degradation after expression. These problems of recombinant protein production arise from the fact that they are expressed heterologously in a different environment and at a different expression level from their natural one (Rosano and Ceccarelli 2014).

Vip1Ac1 as a secreted protein produced by *Bacillus sp*. forms insoluble inclusion bodies when expressed as a recombinant protein in the cytoplasm of *E. coli*. To overcome the solubility issue of Vip1Ac1 protein, different strategies have been applied, which include changing: expression conditions, expression vector and fusion with a soluble partner (Gopal and Kumar 2013). A range of the above strategies was unsuccessful in enhancing the solubility of Vip1Ac1, suggesting that the expression of Vip1Ac1 in the cytoplasm of *E. coli* was the issue and that expressing the protein in the periplasmic space could be the solution for the solubility. The periplasmic signal sequence from cytochrome b_{562} was successfully linked to the N-terminus of Vip1Ac1 and allows the protein to be expressed as a soluble protein in the periplasmic space. The same strategy was successfully applied to Vip2Ae3 expression as a soluble protein. The periplasmic signal sequence from cytochrome b_{562} seems to be a useful tool to express a soluble recombinant protein in the periplasm. Expression of Vips in the periplasmic space with N-terminal His-tags allows simple purification of both proteins with a high level of purity using a single purification step (affinity purification) as shown in Figure 3.24A and 3.25A. With a second SEC purification step, high levels of purity were achieved for the proteins in preparation for crystallography and further studies.

Vip1Ac1 has been successfully expressed as a soluble protein with ncAA incorporated at two different positions (Y358 and Y536) using amber codon suppression technology as mentioned in chapter one. However, the yield of protein was low compared with the wildtype. The reduction in protein yield could be explained by the presence of two stop codons in a long protein, which will lead to production of truncated protein. Despite the lower yield of Vip1Ac1 containing ncAA, it was sufficient to carry out the experiments (Crystallography studies in chapter 4 and labelling with fluorescent dye in chapter 5).

Chapter Four

Bioinformatic analysis and structural determination of Vip protein

4. Bioinformatic analysis and structural determination of VIP protein

4.1 Introduction

Bioinformatics is a general term that refers to the use of computational techniques for management and analysis of biological information during biological studies, it is also known as computational biology (Jiang *et al.* 2013). Bioinformatic tools have been widely used to analyse DNA sequences and amino acid sequences in terms of searching for similarity targets, sequence alignments, prediction of secondary structure, building three-dimensional structures (models and actual structure), structure validation and to predict protein functions (Edwards and Cottage, 2003). In this chapter, different bioinformatic tools have been used to analyse Vip1 and Vip2 proteins.

The three-dimensional structure of a protein is a unique feature for each protein and in the majority of cases defines the protein's function. This molecular structure is a result of the final arrangement of different structural levels of amino acids: primary structure, which refers to the amino acid sequence that forms the polypeptide chain; secondary structure, which refers to the folding of segments of the polypeptide chain due to the interaction between amino acid backbone groups via hydrogen bonds (to form ordered local structures such as α-helices and β-sheets); tertiary structure, which results from the interaction between side chain groups and the backbone of amino acids from different regions, giving the protein its final structure. The secondary and tertiary structural levels depend mainly on the primary structure and any change at this level can lead to change in higher levels of folding and, as a result, change the protein's properties and functions.

A knowledge of a protein's molecular structure is a key part of understanding protein function, including how they interact with other molecules. Knowing the molecular structure of a protein is of extraordinary value for the rational engineering of a protein. The molecular structure of a protein can be predicted via molecular modelling using a known experimental structure of a homologous protein (the template). Nowadays, a large number of bioinformatic tools have been developed for prediction of protein three-dimensional structure. Predicted structures can be very useful tools for probing protein structure-function relationships but sometimes miss the important details gained from determining the actual structure.

The molecular structure of a protein can be experimentally determined using X-ray crystallography techniques. Briefly, in this technique, the crystals of the target molecule are exposed to an X-ray beam, the waves of X-rays diffract differently depending on the electron density of the atoms in the crystal. This diffraction, resulting from detailed features of the atoms in the structure, is recorded as a pattern of spots on a film or detector and is known as diffraction pattern. The amplitude of the X-ray waves can be calculated from the diffraction pattern but vital information (phase) is missing and this gives rise to the phasing problem in X-ray crystallography and this phasing needs to be determined separately using different approaches. Subsequently, the electron density can be reconstructed using the information from the diffraction pattern and phase information using computer software to build a structural model. The model generated is then refined against the experimental data and finally an accurate molecular structure of a target protein can be determined.

X-ray crystallography may be possible following a number of steps including gene isolation, recombinant protein expression, purification of the protein, crystal preparation, crystal harvesting, data collection (diffraction pattern), solving the phasing problem, structure building and refinement using bioinformatic tools. Some of these stages can represent big challenges and may be time consuming. In this chapter, a model structure of the Vip1Ac1 protein has been predicted by homology modelling. Also, in this chapter, the actual molecular structure of monomeric Vip1Ac1 has been determined experimentally at a resolution of 1.47 Å. Vip1Ac is a potential biopesticide and it will be valuable to understand its molecular structure and functions, and how this might contribute to specificity.

4.2 Results and discussion

4.2.1 Bioinformatic analysis of Vip2

4.2.1.1 Vip2 structure and function

The Vip2 structure was solved by Han and co-workers in 1999 at a resolution 1.5 Å. In general, Vip2 consists of two domains: the N-domain (from amino acids 60 to 265) and the C-domain (from amino acids 266 to 461). Despite the fact that both domains share low sequence identity (22%) they are structurally homologous (Han *et al.* 1999). Vip2 has ADP-ribosyltransferase (ART) activity as indicated by Han *et al.* when they solved the structure of Vip2 in complex with nicotinamide adenine dinucleotide (NAD) (PDB 1QS2), showing that the C-domain is the catalytic domain, which contains the cleft for NAD binding (Han *et al.* 1999) (more information can be found in chapter one). However, the target protein that Vip2 modifies is unknown. It is reported by many researchers that G-actin is the target protein for Vip2 (Warren, 1997; Han *et al.* 1999; Chakroun *et al.* 2016) but there are no published accounts of experimental data on G-actin. The identification of Vip2 as an actin-ADP-ribosylating toxin appears to be based on the fact that it has similar overall folding with component A of several *Clostridium* AB binary toxins that are known to target G-actin for ADP-ribosylation (Figure 4.1). Also, the common important motifs are conserved in Vip2, including; aroma-R motif, STS motif (which is an important motif for NAD binding and is conserved in most ART family proteins (Sun *et al.* 2004)) and ARTT motif (ADP-ribosyltransferase turn turn) (E-X-E), which is a motif believed to have a role in recognition and interaction with a target proteins (Han *et al.* 2001)) are present in component A of several *Clostridium* AB binary toxins and in other ADP- ribosyltransferases (ART) that catalyse modification of G-actin (Laing *et al.* 2011) as illustrated in Figure 4.2.

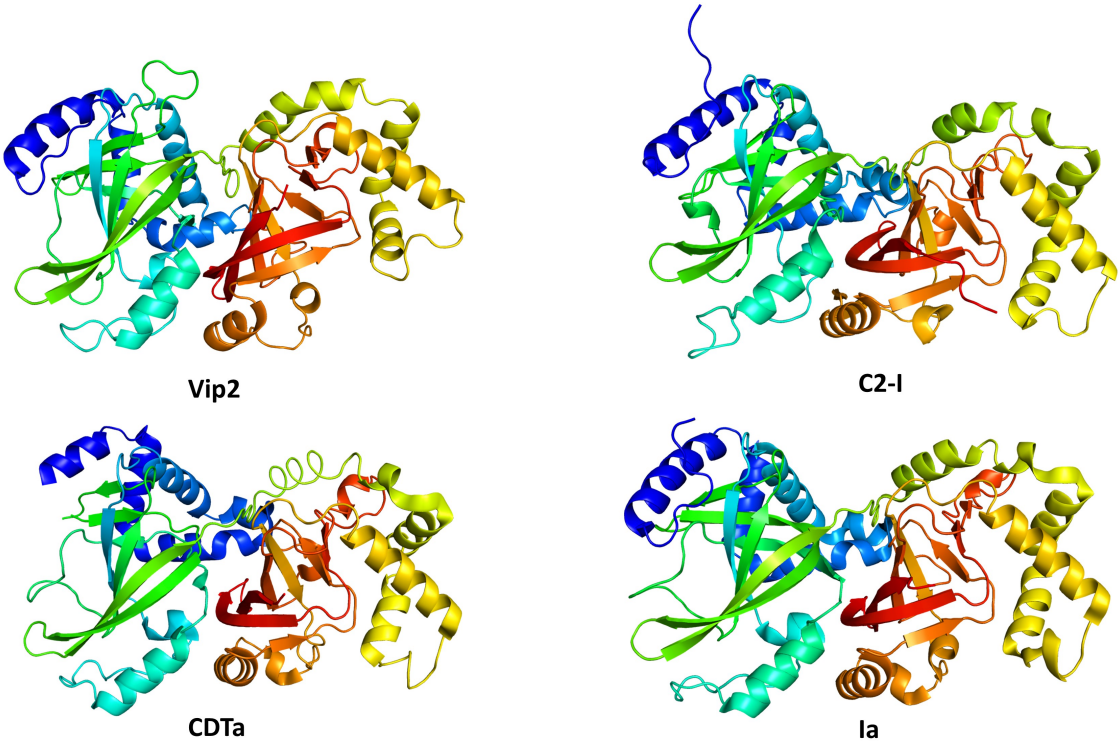

Figure 4.1. Comparison between Vip2 structure and other A components of binary toxins. Vip2 corresponds to the *Bacillus cereus* Vip2 toxin (PDB code 1QS1). C2-I corresponds to *Clostridium botulinum* C2 toxin (PDB code 2J3X). CDTa corresponds to *Clostridium difficile* toxin (PDB code 2WN4). Ia corresponds to *Clostridium perfringens* iota toxin (PDB code 1GIQ). The N-terminus is coloured blue and the C- terminus coloured red.

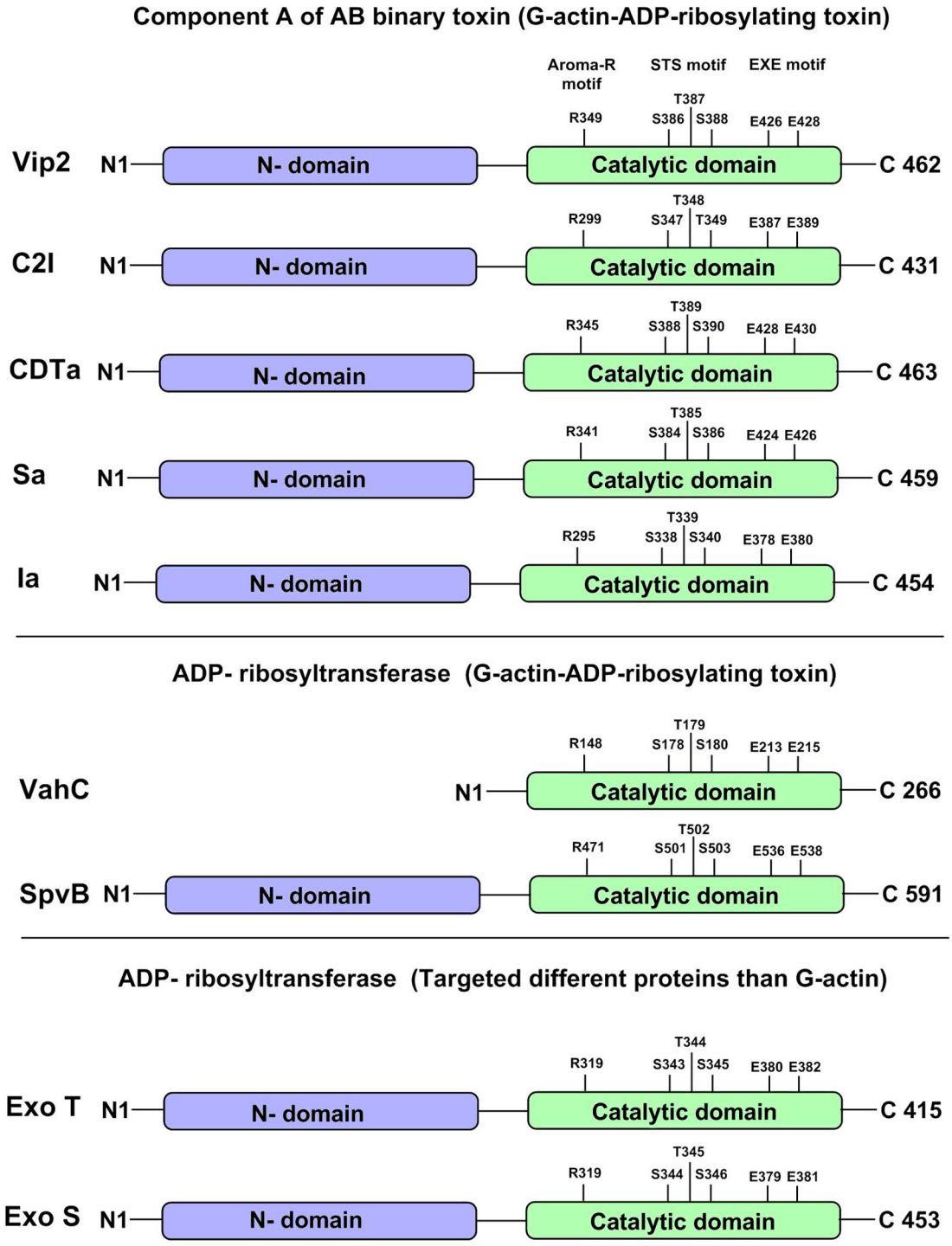

Figure 4.2. Conserved motifs in ADP-ribosylating toxins.
Vip2 corresponds to the *Bacillus cereus* Vip2 sequence. C2-I corresponds to *Clostridium botulinum* C2 toxin. CDTa corresponds to *Clostridium difficile* toxin. Sa corresponds to *Clostridium spiroforme* toxin. Ia corresponds to *Clostridium perfringens* iota toxin. VahC corresponds to *Aeromonas hydrophila* toxin. SpvB corresponds to *Salmonella enterica* toxin. Exo T and Exo S correspond to *Pseudomonas aeruginosa* toxins.

Interestingly, sequence comparisons of Vip2 with other ART that target different proteins rather than G-actin, showed that it shared the same important motifs (aroma-R motif, STS and ARTT motifs) as shown in Figure 4.2. For instance, *Pseudomonas aeruginosa* exoenzymes T and S that target different proteins such as Rho factor, moesin and Crk protein (Deng and Barbieri 2008) shared the same motifs (Radke *et al.* 1999) and have the same overall structure for the catalytic domain (C-domain in Vip2), as shown in Figure 4.3. These results suggested that Vip2 might target G-actin, Rho factor or other proteins. ART proteins can target more than one protein as reported by Rocha *et al*. (Rocha *et al.* 2003) and without any experimental data the protein target of Vip2 still unknown. The N-domain of ART proteins have different functions: in some ART the N-domain is known as targeting domain and plays a vital role in target recognition (Cohen and Chang 2018). While the N-domain of components A of AB binary toxins have a role in interaction with the B components during translocation into the cytosol of the target cell (Barth *et al.* 2004). The N-domain of Vip2 has unknown functions but compared with the N-domain of other A components of AB binary toxins in which the N-domains have a vital role in interaction with component B, the non-catalytic N-domain of Vip2 may play a similar role in interaction with Vip1 during translocation into the cytosol of the target cell (more information about interaction of Vip2 with Vip1 can be found in chapter 5).

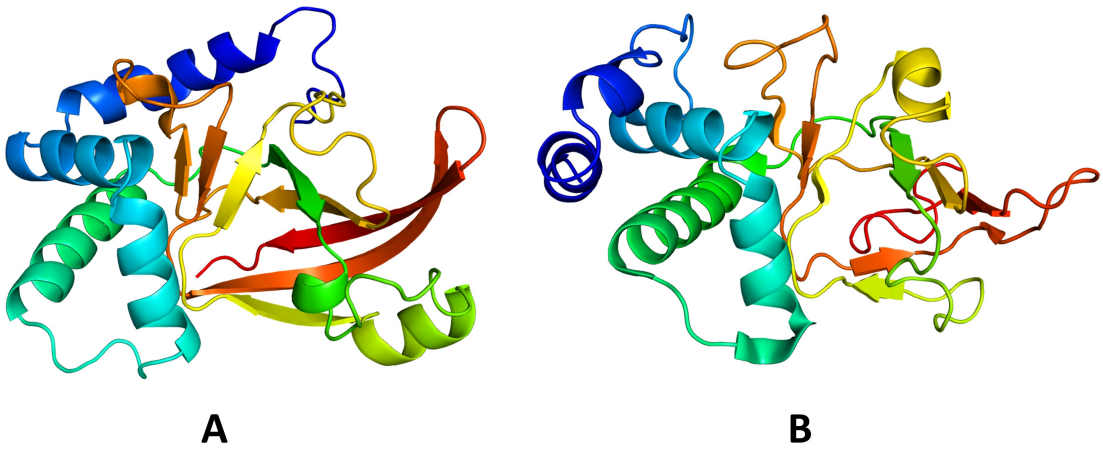

Figure 4.3. Comparison between catalytic C- domain of Vip2 and C- domain of Exo toxin.
A. Structure of catalytic C- domain of Vip2 (PDB code 1QS1). **B.** Structure of catalytic C- domain of Exo toxin from *Pseudomonas aeruginosa* (PDB code 6GNN). The N-terminus is coloured blue and the C- terminus is coloured red.

4.2.1.2 Multiple sequence alignments of Vip2 proteins

The results obtained from the sequence alignment of 16 Vip2 variants available in sequence databases are presented in Figure 4.4. It is apparent from this Figure that Vip2 proteins are highly conserved in both the N-terminal and C-terminal domains. The cleft for NAD binding in the C-terminal (enzymatic) domain, which is composed of six residues C351, M385, L389, R393, R400 and the catalytic residue E428 (numbered according to the sequence of PDB code 1QS1). Four residues are conserved (C351, L389, R400 and the important catalytic residue E428). The STS motif and ARTT motifs are also conserved among all Vip2 variants. This conservation in Vip2 variants indicates that Vip2 variants have the same function, which is ADP-ribosylation and are unlikely to have a role in toxin specificity. This proposal is supported by the findings of another study, showing that using Vip1Aa in combination with Vip2Aa (from the same operon) produced comparable toxic activity against *D. virgifera virgifera* to Vip1Aa in combination with Vip2Ab (from a different operon) and, thus, changing Vip2 variants did not affect the toxicity (Warren 1997).

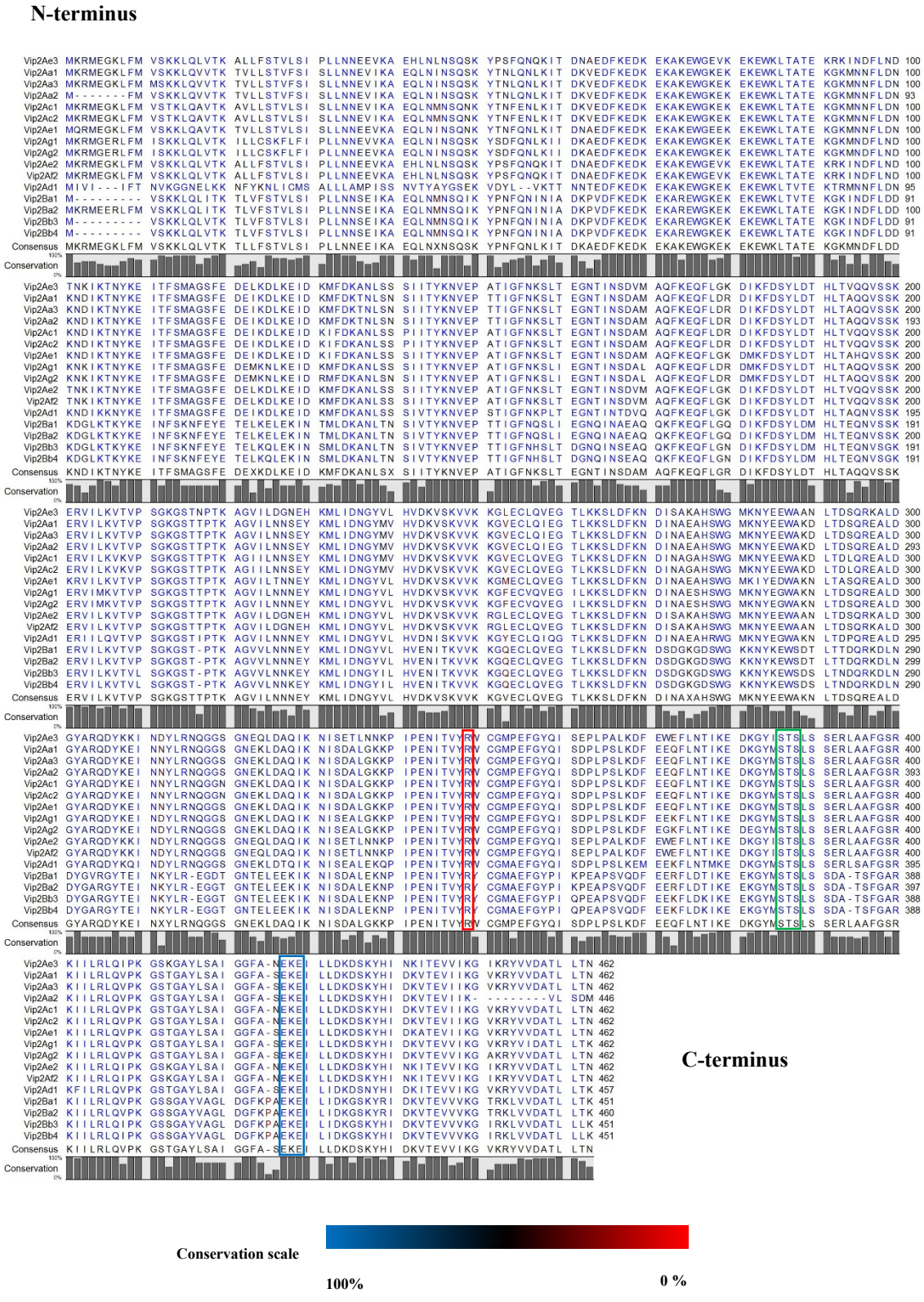

Figure 4.4. Multiple sequence alignments of Vip2 variants.
Multiple sequence alignment of Vip2 variants generated by the CLC Sequence Viewer 8.0 program. Blue colour corresponds to the conserved residues. Red marker represents Aroma-R motif, green marker represents STS motif and blue marker represents EXE motif.

4.2.2 Bioinformatic analysis of Vip1

4.2.2.1 Molecular modelling of Vip1Ac1 protein

To try and generate some preliminary structural information on Vip1Ac1, *in silico* modelling was performed using the closest available homologues in the Protein databank (PDB). Molecular modelling is based on the fact that naturally-occurring proteins that share sequence identity, usually have similar three-dimensional structures (Kaczanowski and Zielenkiewicz 2010). The first step in molecular modelling is to find a similar protein of known molecular structure in the PDB. The quality and accuracy of the modelling depends on the degree of identity between the target protein and the homologous template proteins, with sequence identities ~20% or less likely to have very different structures (Chothia and Lesk 1986) and not suitable for homology modelling. Analysis using Swiss model (Kiefer *et al.* 2009), Geno3D (Combet *et al.* 2002) and I-TASSER (Zhang 2008) showed that the Vip1Ac1 protein has sequence identity with *Bacillus anthracis* protective antigen (PA, template PDB accession 4H2A) calculated as 36.12% by Swiss model, and by I-TASSER as 31.3% and Geno3D (using template PDB accession 3TEW) as 34%. Three models were built by Swiss model, Geno3D and I-TASSER web services (using default parameters) as shown in Figure 4.5. The models have overall the same general folding with the greatest difference in conformation at the C-terminal region (red in Figure 4.5). These results suggested that the Vip1Ac1 and PA have similar overall folding at the N-terminus and different at the C-terminal region, which explained why the prediction servers generated different conformations for the C-terminal region.

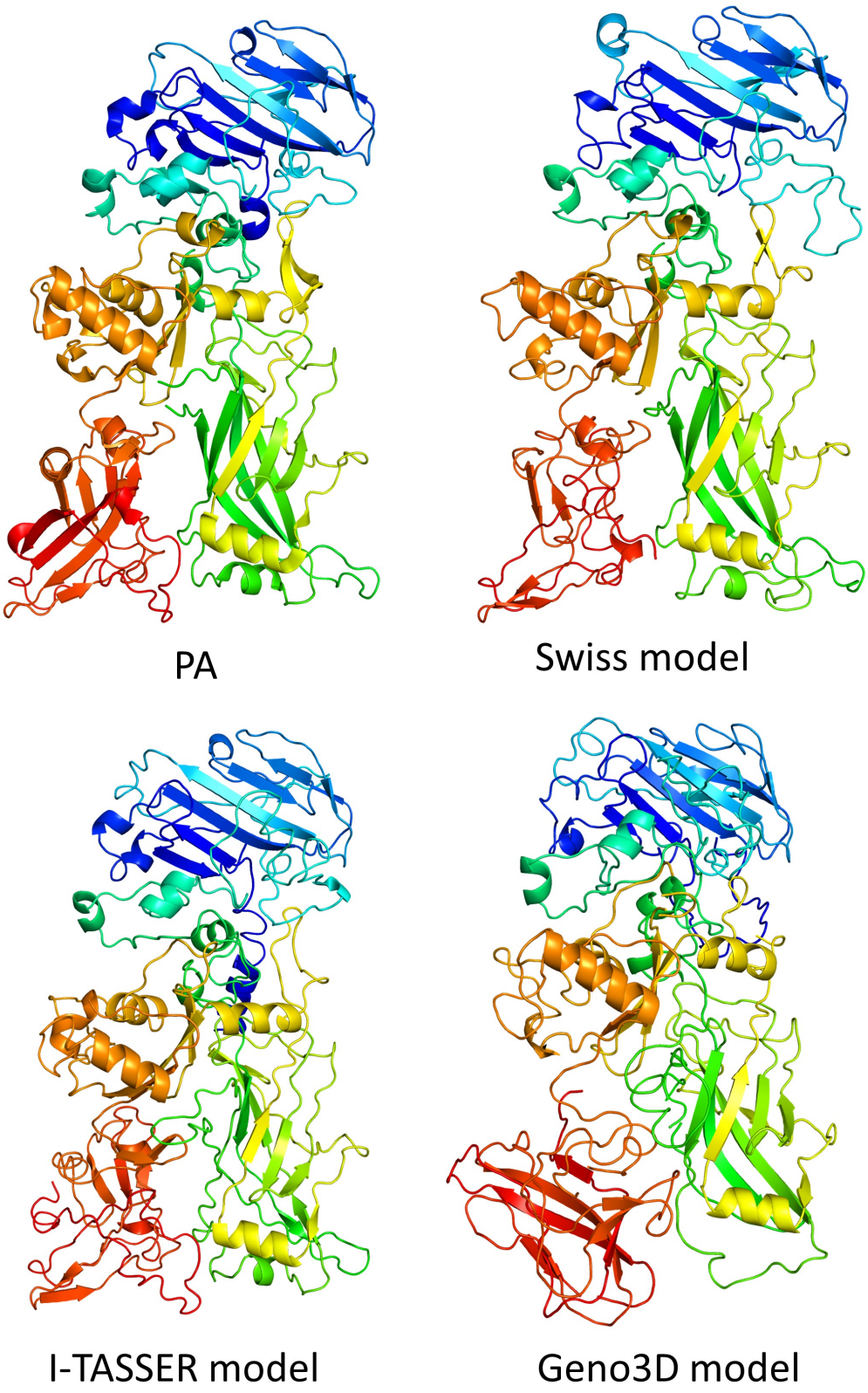

Figure 4.5. Models of Vip1Ac1.
PA corresponds to the molecular structure of anthrax toxin protective antigen (PDB code 4H2A). The proteins are shown in rainbow colouring with the N-terminus coloured blue and the C- terminus coloured red.

4.2.2.2 Model validation

Every molecular model will have a number of errors, these errors depend on the percentage of sequence identity between target and template. As mentioned above, if the sequence identity between target and template is around 20% or less, the target and template have very different structures (Chothia and Lesk 1986) and cannot be used for modelling as the number of errors will be high and the model will be inaccurate. High sequence identity(> 90%), results in a model with a precision that can be compared with an experimentally determined structure with a few individual side chain exceptions (Walker 2010). Models having sequence identity between 90% and 20%, have a decreasing level of accuracy. Vip1Ac1 has sequence identity with PA calculated by Swiss model, Geno3D and I-TASSER web services as 36.12%, 32.3% and 34% respectively. With this percentage of sequence identity, the Vip1Ac1 models are expected to have several errors. However, validation of the models depends on the purpose for which they will be used: if the model is built for docking studies with a ligand, more assessment is required, especially for the active site, and reliable models should have a level of accuracy around the active site. In this study, Vip1Ac1 models were built to generate some preliminary structural information and select surface residues for labelling with fluorescent dye (chapter 5) and adding a heavy atom for phasing (see Section 4.2.3.3.2). To select the most reliable model, Vip1Ac1 models were subjected to general validation using Ramachandran plot validation and ERRAT validation. Results of Ramachandran plot validation are presented in Table 4.1. It can be seen that the Vip1Ac1 model generated by Swiss model has a lower error (outlier residues) than other models in terms of the positions of amino acids and 90.4% of amino acids were in the favoured region.

Table 4.1 Ramachandran plot validation for template and Vip1Ac1 models

PDB	No of residues in favoured region	No of residues in allowed region	No of residues in outlier region
4H2A (template)	686 (98.0 %)	14 (2.0 %)	0 (0.0 %)
Swiss model	624 (90.4 %)	45 (6.5 %)	21 (3.0 %)
Geno3D models	500 (70.9 %)	155 (22.0 %)	50 (7.1 %)
I-TASSER model	531 (72.0 %)	127 (17.2 %)	79 (10.7 %)

ERRAT validation is based on the statistical analysis of atomic interactions and distinguishes between correct distributions and incorrect distributions of atoms in protein structures. Atoms in experimentally determined protein structures are distributed nonrandomly, while different types of atoms are distributed randomly in the model due to errors in model building (Colovos and Yeates 1993). Results from ERRAT calculate an overall score for the model and the ERRAT validation results are set out in Figure 4.6. The Vip1Ac1 model generated by the Swiss model has the highest overall quality factor (82.65) compared with other models.

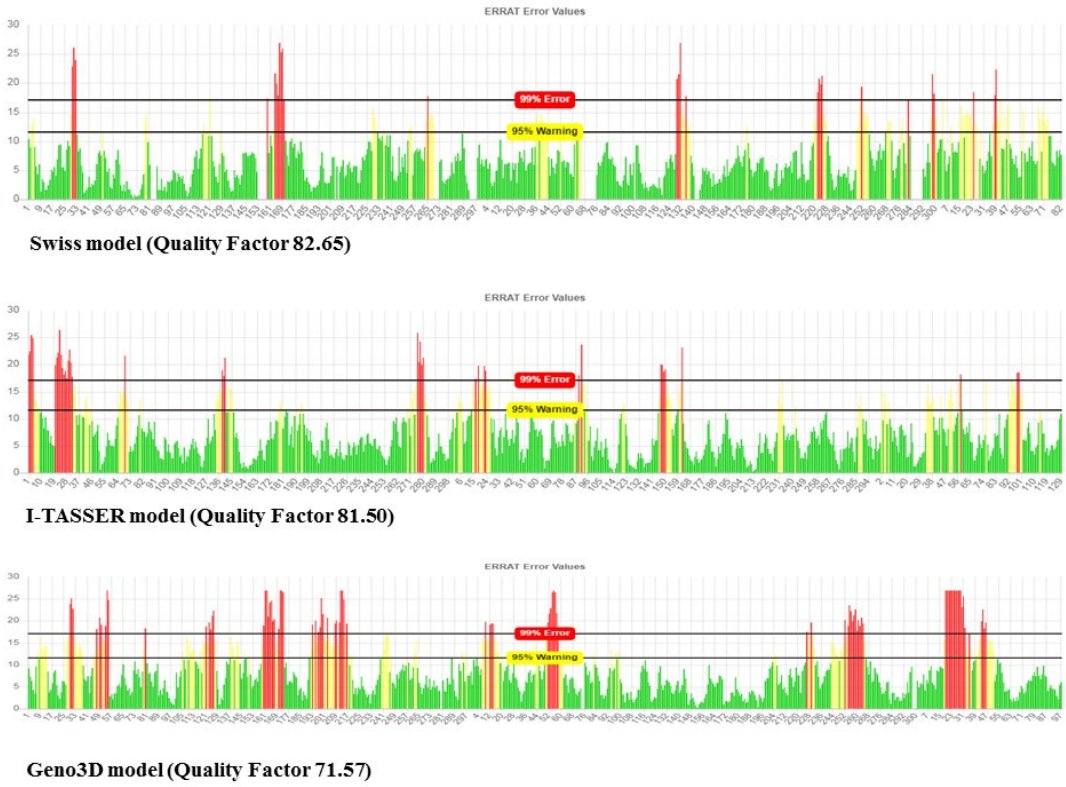

Figure 4.6. ERRAT validation for Vip1Ac1 models.
Red colour indicates 99% error, yellow colour indicates 95% warning and green colour indicates no error.

Results from both validations indicate that the Vip1Ac1 model generated by the Swiss model service is more reliable than the other models. Two tyrosine residues have been chosen based on this model, at positions 358 and 536, for incorporating non-canonical amino acids (ncAAs) (chapter 3 Section 3.2.8.2), as both residues appear to be

surface exposed and located in favoured regions in the protein model according to Ramachandran plot validation.

4.2.2.3 Comparison between Vip1Ac1 model and PA in terms of structure

PA, which is the receptor binding subunit of anthrax toxin, has a molecular weight of 83 kDa and is composed of four domains. Domain 1 (D1 amino acids 1 to 258) contains a furin cleavage site that allows removal of a ~20 kDa fragment from full-length PA (83 kDa; PA_{83}) to produce activated PA (63 kDa; PA_{63}), which allows the assembly of the heptamer (Petosa *et al.* 1997) and also binds to the enzymatic subunits of anthrax toxin (lethal factor and edema factor) (Brossier *et al.* 2004; Alisaraie and Rouiller 2016). Domain 2 (D2 amino acids 259 to 487) contains the 2β2-2β3 loops, which are responsible for channel formation (Jiang *et al.* 2015). Domain 3 (D3 amino acids 488 to 595) contains a number of key residues for oligomerisation (Mogridge *et al.* 2001), and domain 4 (D4 amino acids 596 to 735) binds to receptors on target cells (Alisaraie and Rouiller 2016).

Alignment of the Vip1Ac1 model constructed using Swiss model (Figure 4.5, above) with PA, as shown in Figure 4.7, indicates that the Vip1Ac1 protein has the same overall fold as PA. This structural comparison (along with comparison of primary structures, Figure 4.8) may give insight into regions involved in the Vip1Ac1mechanism of action through similarity between their principle functions: receptor binding subunits and oligomerisation. Thus, D1 may have a specific cleavage site for Vip1Ac1 protein activation or binding site with Vip2 (the enzymatic subunit). D2 could be responsible for channel formation. D3 may play a role in oligomerisation, D4 may be involved in interaction between Vip1Ac1 protein and receptors on target cells. These issues will be explored further later in this thesis.

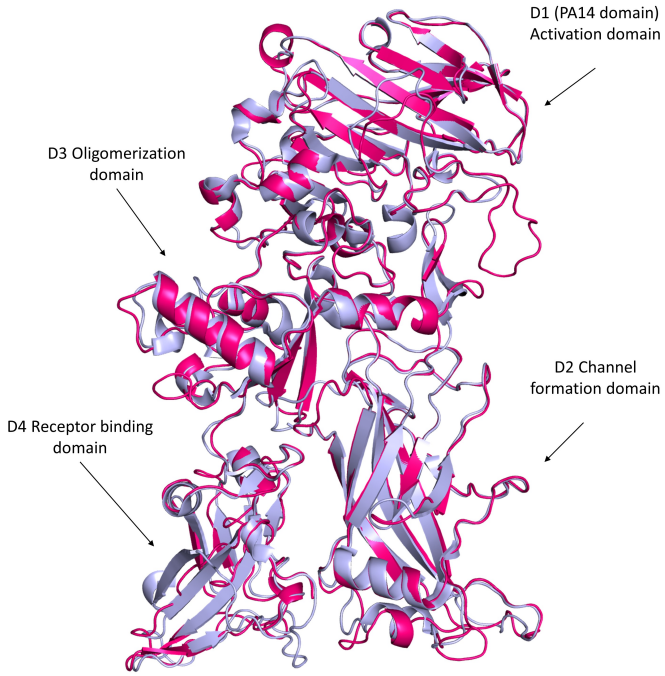

Figure 4.7. Structural overlay between PA and Vip1 model.

Light blue colour corresponds to the molecular structure of PA (PDB code 4H2A) with domains marked. Pink colour corresponds to the Swiss model of Vip1Ac1.

```
                              20                                  40                                  60                                  80
Vip1Ac1 MKNMKKKLAS VVTCTLLAPM FFNGNVNTAY ADNKINQDSS KQEYQ---QK EMDRKGLLGY YFKDKDF-SN LTMFSPTRYN 76
4H2A    ---------- ---------- ---------- --------EV KQENRLLNES ESSSQGLLGY YFSDLNFQAP MVVTSSTTGD 42
                             100                                 120                                 140                                 160
Vip1Ac1 TLIYDQQTAN KLLDKKQQEY QSIRWIGLIQ SNKTGDFTFE LSDDECAIIE MDGKVISNKG KEKQVVHLEK GKLVPIKIEY 156
4H2A    LSIPSSELEN --IPSENQYF QSAIWSGFIK VKKSDEYTFA TSADNHVTMW VDDQEVINKA SNSNKIRLEK GRLYQIKIQY 120
                             180                                 200                                 220                                 240
Vip1Ac1 QLDEPLNIDD EKFKGFKLLK VDNQKQLHQV QQDELRNPEF NKKESQEFLA KASKINLFTK KIKRDIDEGT ---DTDGDSI 233
4H2A    QRENPT---- EKGLDFKLYW TDSQNKKEVI SSDNLQLPEL KQKSSNS--- ---------- RKKRSTSAGP TVPDRDNDGI 183
                             260                                 280                                 300                                 320
Vip1Ac1 PDMWEENGYT I--QNR--IA VKWNDSL-AS KGYTKFVSNP LDSHTVGDPY TDYEKASRDL DLSNAKETFN PLVAAFPSVN 308
4H2A    PDSLEVEGYT VDVKNKRTFL SPWISNIHEK KGLTKYKSSP EKWSTASDPY SDFEKVTGRI DKNVSPEARH PLVAAYPIVH 263
                             340                                 360                                 380                                 400
Vip1Ac1 VSMEKVILSP NKNLSN-SVE SHS-----ST NWSYTNTEGA SVEAGIGPKG FSFG--VSAN YQHSE--TVA QEWGASIG-- 376
4H2A    VDMENIILSK NEDQSTQNTD SQTRTISKNT STSRTHTSEV HGNAEVHASF FDIGGSVSAG FSNSNSSTVA IDHSLSLAGE 343
                             420                                 440                                 460                                 480
Vip1Ac1 ----DTTQLN TASAGYLNAN VRYNNVGTGA IYDVKPTTSF VLEKN-TIAT ITAKSNSTAL SISPGESYPK KGQNGIAITS 451
4H2A    RTWAETMGLN TADTARLNAN IRYVNTGTAP IYNVLPTTSL VLGKNQTLAT IKAKENQLSQ ILAPNNYYPS KNLAPIALNA 423
                             500                                 520                                 540                                 560
Vip1Ac1 MDDFNSHPIT LNKKQLDQVL TNNPIMLETD QTDG---IYK IKDTHGNIVT GGTWNGVTQQ IKAKTASIIV DDGKQV--AE 526
4H2A    QDDFSSTPIT MNYNQFLELE KTKQLRLDTD QVYGNIATYN FENGRVRVDT GSNWSEVLPQ IQETTARIIF N-GKDLNLVE 502
                             580                                 600                                 620                                 640
Vip1Ac1 KRVAAKDYAY PEDKT-PSLT LKDALKLSFP EEIKETDGLL YYNNKPIYES SVMTYLDGNT AKEVKKQIND KTGEFKDVQH 605
4H2A    RRIAAVNPSD PLETTKPDMT LKEALKIAFG --FNEPNGNL QYQGKDITEF DFN--FDQQT SQNIKNQLAE L-----NATN 573
                             660                                 680                                 700                                 720
Vip1Ac1 LYAV----KL TPKMNFTIKV P-VAYDTAKQ AVNLGGDNPW ---------- -GAKGLLGTW VNAMVVDNSG DKAYKRVEPG 669
4H2A    IYTVLDKIKL NAKMNILIRD KRFHYDRNNI AV--GADESV VKEAHREVIN SSTEGLL--- LNI------- DKDIRKILSG 641
                             740                                 760                                 780                                 800
Vip1Ac1 YLLSPTLEFS EGSLDNLKKN YSFY-VSMYV KSDKPFTLRI NAGPYSTKRT IEASNDFKRV DIPAFYIEGF PIDTIRLEGS 748
4H2A    YIVE--IEDT EGLKEVINDR YDMLNISSLR QDGKTF---I DFKKYNDKLP LYISNPNYKV NVYAVTKENT IINPSE-NGD 715
                             820
Vip1Ac1 DYPSAIWWKD VSITEVSAVK -K 769
4H2A    TSTNGI--KK ILIFSKKGYE IG 735
```

Figure 4.8. Alignment of Vip1Ac1 with PA.

Sequence alignment between Vip1Ac1 and PA generated by the CLC Sequence Viewer 8.0 program. Red letters correspond to conserved residues. Sequence identity 36.12%.

4.2.2.4 Multiple sequence alignments of Vip1 proteins

Based on the results of comparison between 11 Vip1 variants (Figure 4.9) available in sequence databases, Vip1 proteins are highly conserved in regions corresponding to domains 1-3, which may indicate that these regions have the same functions, including containing the proteolytic cleavage site, oligomerisation, binding with the enzymatic subunit and channel formation. The region towards the C-terminus (domain 4) may be involved in receptor binding, which may explain specificity of Vip1 variants to different target insects. These findings further support the previous work by Warren 1997 who reported that the products of the *vip1Aa/vip2Aa* operon are toxic to *D. virgifera virgifera* while the products from the *vip1Ab/vip2Ab* operon are non-toxic to this insect (Warren 1997).

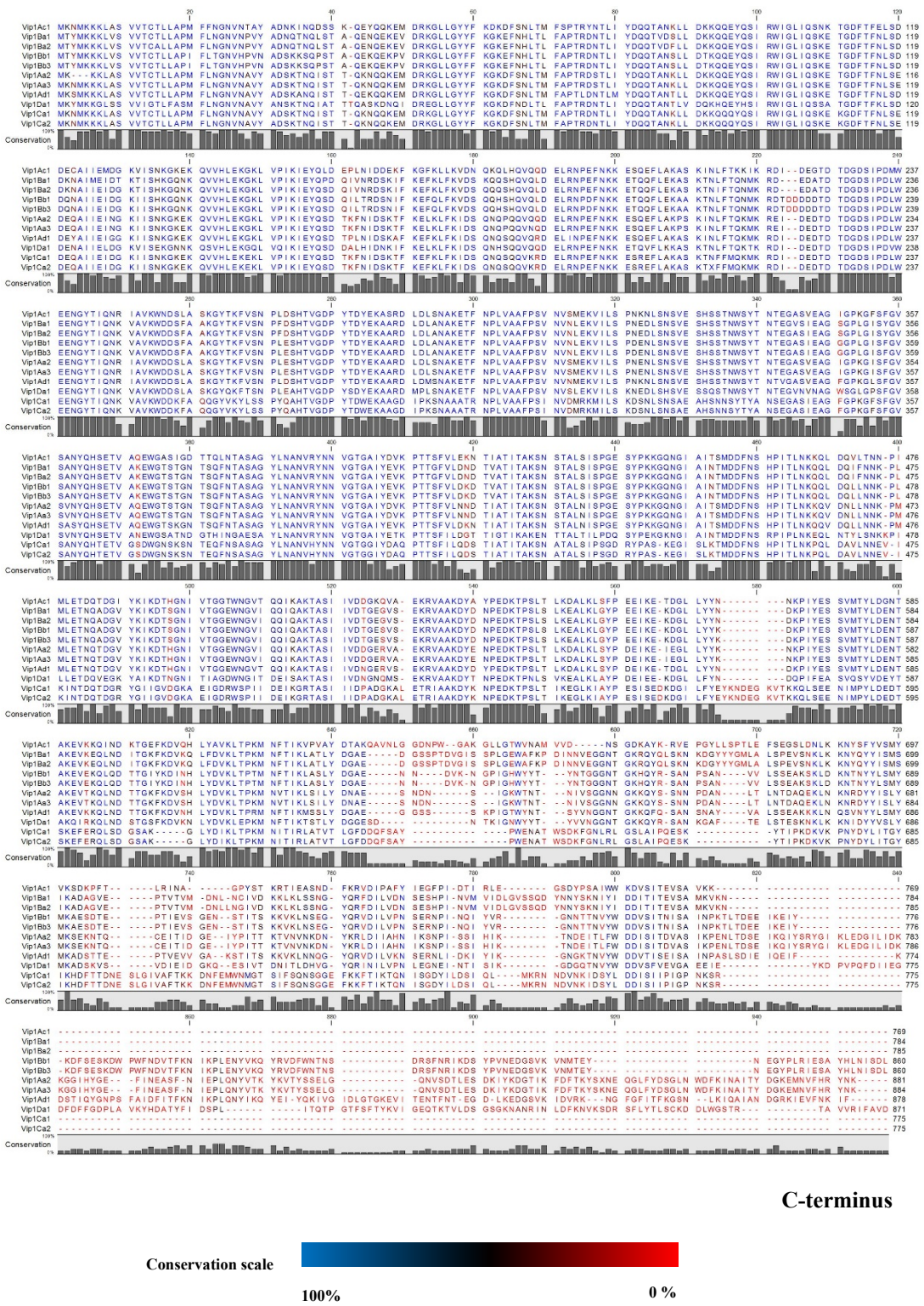

Figure 4.9. Multiple sequence alignments of Vip1 variants.
Multiple sequence alignment of Vip1 variants generated by the CLC Sequence Viewer 8.0 program. Blue colour corresponds to conserved residues.

4.2.2.5 Protein sequence analysis and classification of Vip1Ac1

InterPro services is a powerful bioinformatic tool, which analyses the sequence of target proteins in terms of function, predicts sequence motif (conserved sequence) and domains, and classifies the proteins into families (Mitchell *et al.* 2018). Interestingly, results from InterPro (Figure 4.10) classified Vip1Ac1 as component B of a binary toxin family and supports the interpretation for the domain functions of the Vip1Ac1 model. As can be seen from Figure 4.10, Vip1Ac1 protein has 4 predicted domains. The first domain from amino acids 19 to160, is a PA14 domain (a β-barrel architecture domain named PA14 as it resembles domain one of PA and a conserved domain in a wide variety of eukaryotic and bacterial proteins with proposed carbohydrate binding function (Rigden *et al.* 2004)). The second predicted InterPro domain, from amino acids 197 to 272, is a Ca^{2+}-binding domain. The first and second InterPro domains correspond to the D1 structural domain in the Vip1Ac1 model. The third InterPro domain, from amino acids 277 to 461, is predicted to be a heptamerisation domain and corresponds to the D2 in the Vip1Ac1 structural model. The fourth InterPro domain, from amino acids 477 to 567, has features similar to domain 3 in both PA and Clostridial binary toxin B and corresponds to D3 in the Vip1Ac1 structural model. The rest of amino acids from 568 to 739 did not match any InterPro protein domain and this is consistent with this more variable region being involved in receptor binding and toxin specificity.

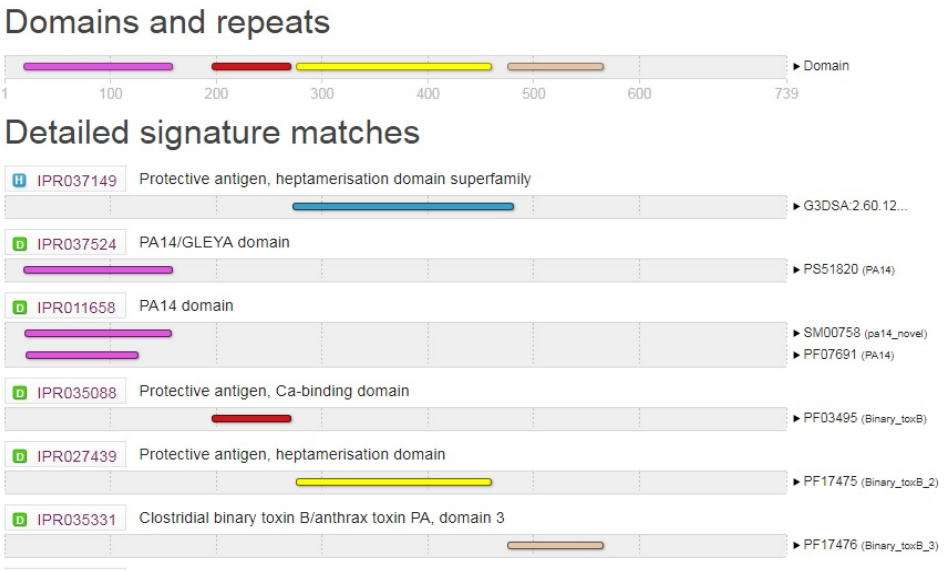

Figure 4.10. Vip1Ac1 sequence analysis by InterPro services.

Results from sequence comparison of Vip1Ac1 with other B components of binary toxins (*Bacillus* and *Clostridium* binary toxins) showed that the calcium binding motif as described by Sakurai *et al.* (Sakurai, *et al.* 2009) was conserved in Vip1Ac1 (as shown in Figure 4.11) in addition to other residues around this motif, which may play a role in calcium binding. This also supports the interpretation for the domain 2 function in the Vip1Ac1 model.

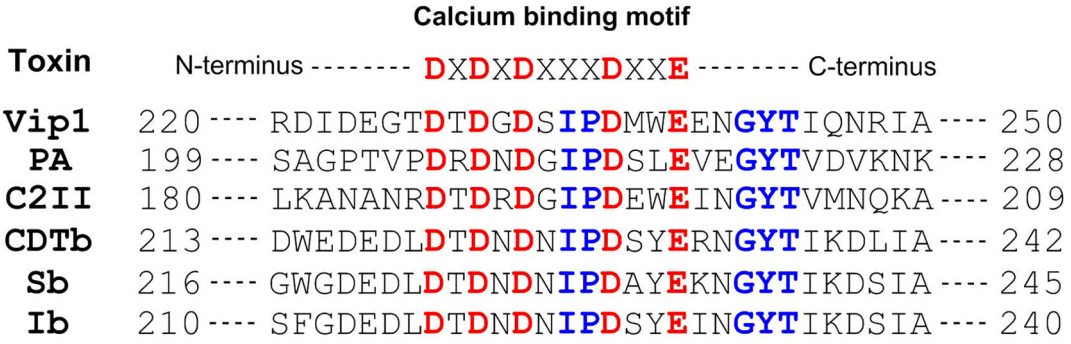

Figure 4.11. Calcium binding motif shared between B components of AB binary toxins.
Red coloured letters refer to the residues involved in calcium chelating (calcium binding motif). Blue coloured letters correspond to other conserved residues in the calcium binding region. Vip1 corresponds to Vip1Ac1 toxin, PA corresponds to protective antigen toxin, C2II corresponds to *Clostridium botulinum* toxin, CDTb corresponds to *Clostridium difficile* toxin, Sb corresponds to *Clostridium spiroforme* toxin, Ib corresponds to *Clostridium perfringens* toxin.

4.2.3 X-ray crystallography and Vip1Ac1 structure determination

X-ray crystallography is a powerful technique to elucidate the precise positions of atoms within the molecular structure of proteins with the aid of computational techniques. First, a high concentration of a pure protein sample is required to set up crystal trials using a range of buffers and precipitants in a trial and error technique to attempt to find appropriate conditions for forming usable protein crystals. In these experiments, the protein sample becomes slowly concentrated and may precipitate from the solution during the extended incubation time and, under favourable conditions, accumulate as crystals (more information can be found in chapter 2; Section 2.12.1). Protein samples with high levels of purity have a higher chance of forming crystals. Vip1Ac1 was

expressed and purified to a high level of purity as described in chapter 3 and used for crystal trial experiments.

4.2.3.1 Initial crystal screening

Growing of protein crystals is a critical stage in X-ray crystallography. High quality protein crystals result in good diffraction data at a high resolution. It is difficult to predict favourable conditions for growing high quality protein crystals as each protein has specific conditions and some proteins may fail to form any crystals under a wide range of conditions. Therefore, the optimal conditions must be determined for each protein individually, often using commercial crystal screens containing a range of common crystallisation buffers.

A concentrated, purified Vip1Ac1, ~10 mg/ml, was used for initial crystal preparation using two types of screens: PACT *premier*™ HT-96 and JCSG-plus™ HT-96. The sitting drop vapour diffusion method was used as described in chapter two (Section 2.12.1). After three weeks of monitoring, some protein crystals appeared in different conditions as illustrated in Table 4.2. All crystals were harvested and tested for X-ray diffraction at the Diamond Light Source (with Dr Pierre Rizkallah, School of Medicine, Cardiff University) as described in chapter two (Section 2.12.1). Crystals grown in conditions C3 (PACT *premier*™), A2, A10 and G1 (JCSG-plus™) did not diffract, but crystals grown in condition B10 (PACT *premier*™) diffracted to lower resolution (3.5˚A). New crystal trials were set up using the same screens above, but no crystals grew. An optimization around the condition B10 (PACT *premier*™) was undertaken in order to improve the quality of crystals for high resolution data collection.

Table 4.2 Crystallization conditions of Vip1Ac1 protein

Screen type	Well NO.	Salt and conc.	Buffer	pH	Precipitant
PACT *premier*™	B10	Magnesium chloride 0.2 M	MES 0.1 M	6	20 % w/v PEG 6000
PACT *premier*™	C3	None	PCTP 0.1 M	6	25 % w/v PEG 1500
JCSG-plus™	A2	None	Sodium citrate 0.1 M	5.5	20 % w/v PEG 3000
JCSG-plus™	A10	Potassium formate 0.2 M	None	--	20 % w/v PEG 3500
JCSG-plus™	G1	None	HEPES 0.1 M	7	30% v/v Jeffamine® ED-2003

MES= 2-(N-morpholino) ethanesulphonic acid. PEG= Polyethylene glycol. PCTP buffer= Propionic acid, Cacodylate, Bis-tris propane. Jeffamine® ED-2003= Polyvinylpyrolidine

4.2.3.2 Optimization of crystals conditions

Successful X-ray diffraction data collection needs good quality crystals. Some crystals from the initial screen were microcrystals, which have unsuitable morphology, yield poor diffraction data or do not diffract at all. An improvement in crystallization conditions around the initial condition could improve the quality of crystals for better diffraction data. The optimisation was carried out using the hanging drop instead of sitting drop vapour diffusion method using Easy Xtal 15-well tools as described in chapter two (Section 2.12.2). The drop ratio (the ratio between protein drop volume to the reservoir volume changed from 0.4/60 µl to 4/500 µl) and protein concentration (changed from ~10 mg/ml to ~8 mg/ml). Depending on the initial screening and the conditions that gave the crystals, two types of available PEG and pH range (5.0, 5.5 and 6.0) were used. The optimisation conditions around B10 conditions (0.2 M $MgCl_2$, 0.1 M MES, pH 6 and 20% w/v PEG 6000) are illustrated in Table 4.3.

Table 4.3 Optimisation conditions around the initial condition (B10)

Precipitant type 20 % w/v	pH 5	pH 5.5	pH 6
PEG 6000	A1	A2	A3
PEG 3350	B1	B2	B3

Interestingly, after 2 days many medium size crystals appeared in B3, where the PEG precipitant was changed (0.2 M $MgCl_2$, 0.1 M MES, pH 6 and 20 % w/v PEG 3350) as shown in Figure 4.12 and a few in B2 (0.2 M $MgCl_2$, 0.1 M MES, pH 5.5 and 20 % w/v PEG 3350). Crystals were harvested from the B3 condition (Figure 4.12A) and tested through X-ray diffraction at the Diamond Light Source, Harwell, UK, using beamline I04. Three crystals diffracted to high resolution (1.47 Å, 1.5 Å and 1.55 Å) (diffraction pattern shown in Figure 4.13). It can be seen that in the optimisation trial, B3 was the most favourable condition for growing good quality Vip1Ac1 crystals in a short time (2-3 days).

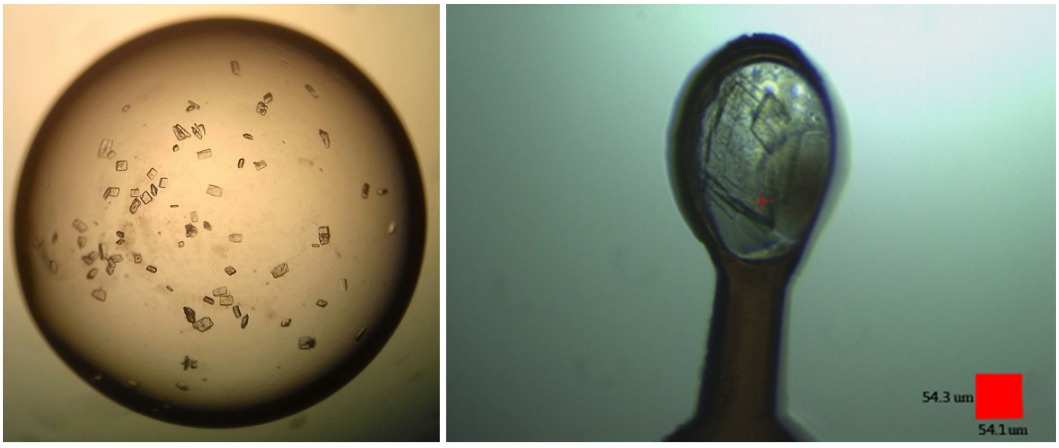

Figure 4.12. Vip1Ac1 crystals.
A. Crystals growing in B3 optimisation conditions (40X). **B.** Crystal in a loop prior to X-ray analysis.

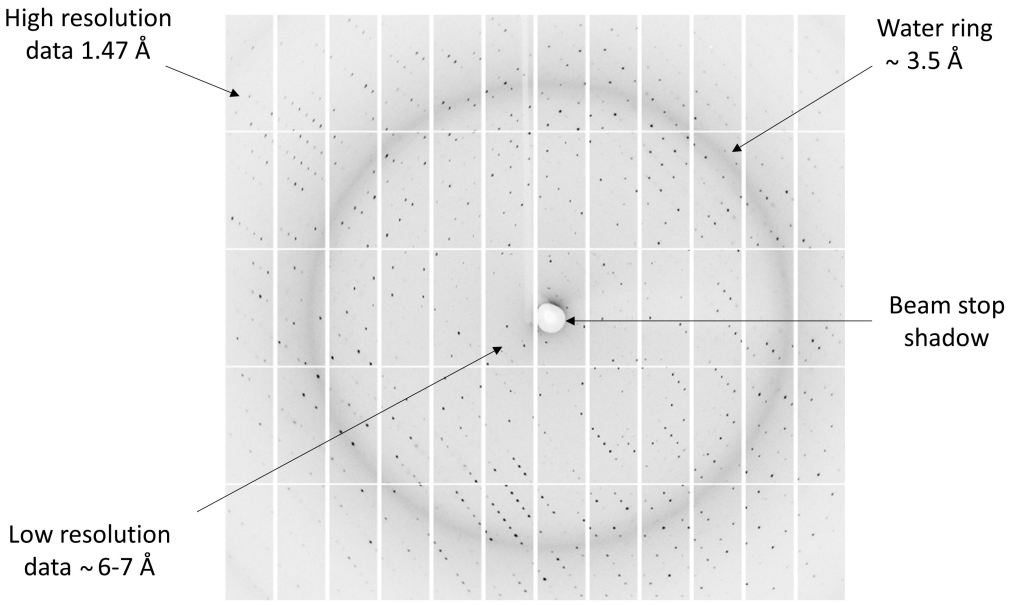

Figure 4.13. X-ray diffraction pattern of Vip1Ac1.
Diffraction pattern generated from Vip1Ac1 crystals produced under condition B3.

4.2.3.3 Solving the phasing problem of Vip1Ac1 crystals

As mentioned before, phase information is missing during data collection in X-ray crystallography. Without this information, it is difficult to solve the structure as this information has details about the atom positions because the phase determines the position of the peak of electron density for each atom in the crystal. Based on the known crystal structure of a homologue, phases can be determined using a molecular replacement approach but if no homologue structure is available, the phase needs to be determined separately using a different approach (Taylor 2003). To determine the phase information of Vip1Ac1 crystals, different approaches were applied.

4.2.3.3.1 Molecular replacement (MR)

Molecular replacement is a powerful approach to solving the phasing problem for unknown crystal structures when structures of similar proteins are available. This

approach is the most popular for solving the phasing problem and does not need further experiments and is quick, inexpensive and automated. As detailed previously, proteins that share less than 20% sequence identity are likely to have different structures; above this percentage of identity MR can take advantage of similarity to solve phasing for unknown crystal structures by comparison with a known structure. The first step and main criterion in MR is to find a similar structure(s) and chose the best one. Often MR is successful when the template shares more than 30% sequence identity and fails below this level (Dimaio *et al.* 2011).

As shown above during the modelling (Section 4.2.2.1), anthrax protective antigen (PA, PDB accession 4H2A, resolution 1.62 Å; the best structure amongst other PA structures available in the protein databank) has a known structure potentially similar to Vip1Ac1, with a 34% sequence identity between the proteins. In addition, BLAST search showed another structural homology with Vip1Ac1, the transport component C2-II of the C2-toxin from *Clostridium botulinum,* with 32% sequence identity (PDB accession 2J42, resolution 3.13 Å). Comparison of the validation metrics for the above known structures (Table 4.5) indicates that the PA structure (PDB accession 4H2A) is the best refined and, therefore, was the best model for MR to attempt to solve the structure of the Vip1Ac1 protein using PHASER (McCoy *et al.* 2007). However, the PHASER output from this attempt indicated that it was unable to find a MR solution from the data. This could be because the template and Vip1Ac1 have quite a low percentage sequence identity and may have significant differences between Vip1Ac1 and PA structures. When MR fails to solve the phasing, this indicates significant differences between the model and the target structure (Evans and McCoy 2007). In the absence of any other model in the protein databank that could be used, it became necessary to determine phases experimentally with heavy atom derivatives.

Table 4.5 Structure validation of PA and transport component C2-II

Metric	Value*	
	PA (1.62 Å)	C2-II (3.13 Å)
R_{free}	0.223	0.472
Clashscore	2	80
Ramachandran outliers	0	14.3%
Sidechain outliers	0.3%	26.6%
RSRZ outliers	5.4%	22.4%

* lower value better

4.2.3.3.2 Experimental phasing methods with heavy atom derivatives

Phase information for a new structure can be determined experimentally using a heavy atom approach. In this standard approach, heavy atoms need to be incorporated into target protein; by mixing the crystals with a solution containing heavy atoms (Garman and Murray 2003), by substituting methionine with selenomethionine (Hendrickson *et al.* 1990), or by genetically incorporating unnatural amino acids containing heavy atoms (Xie *et al.* 2004).

Two different common heavy atoms were used in attempts to solve the phasing (Br and I). Iodine and bromine have been widely used to produce crystals derivatives with heavy atom for the phasing problem (Pike *et al.* 2016). Vip1Ac1 crystals with heavy atom derivatives were prepared by addition of NaI / NaBr to the crystal drop as described in Section 2.12.3. The anomalous signal from crystals soaked with the two heavy atoms was not sufficient to solve the phasing. There are many factors that can affect the binding of heavy atoms to protein crystals that lead to weak signals or no signal at all. The binding sites for the heavy atoms in the protein (imidazole and thiol groups) may be blocked by buffer or present in non-surface residues. The pH of the buffer in which the crystals grew has a strong effect on the state of both the heavy atom and protein side chains (Pike *et al.* 2016). Prior to data collection, it is not possible to assess whether soaking with heavy atoms was successful or not, nor to know the number and positions of any heavy atoms incorporated. This is a major problem in the heavy atom derivatives approach as it is time-consuming and may not be successful (Garman and Murray 2003).

To overcome the issues of the binding of heavy atoms into protein crystals, genetic incorporation of the ncAA 4-Iodo-L-phenylalanine (tyrosine derivative), containing iodine as a heavy atom, was attempted. Two tyrosine residues at positions 358 and 536 were chosen as surface residues in the Vip1Ac1 model. Vip1Ac1 protein containing 4-Iodo-L-phenylalanine was prepared as described in chapter 3 (Section 3.2.8). Protein was concentrated to ~8 mg/ml and subjected to the optimum conditions for native crystal preparation (described in Section 4.2.3.2) for 6 days. The ncAA produced many crystals (Figure 4.14) that were smaller than those produced from native toxin.

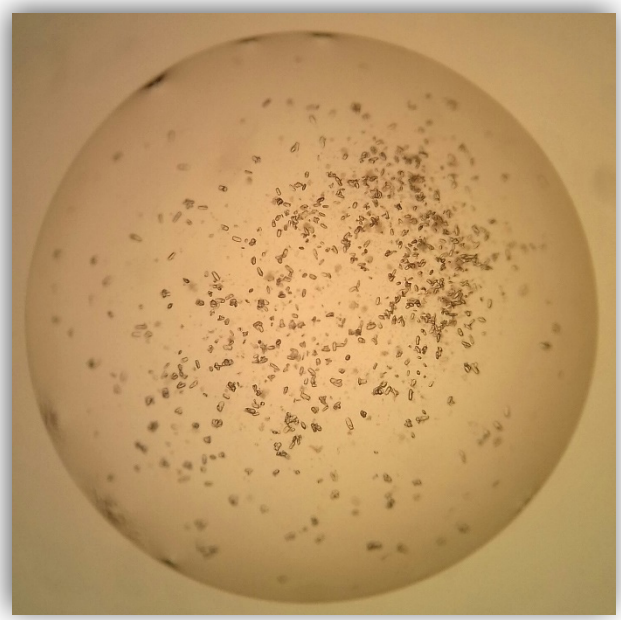

Figure 4.14. Vip1Ac1 crystals containing ncAA 4-Iodo-L-phenylalanine (40X).

Crystals were harvested and tested for X-ray diffraction and anomalous signal at the Diamond Light Source, Harwell, UK, using beamline I03. Results showed that the anomalous signal was weak and not sufficient to solve the phases. This is could be because there are only 2 heavy atoms in the protein and that they were unable to produce enough signal for phasing. Xie *el al.*(2004) used the same approach to solve the structure of bacteriophage T4 lysozyme by incorporating 4-Iodo-L-phenylalanine at position 153 and single incorporation of 4-Iodo-L-phenylalanine was enough to solve the structure of this protein of 164 amino acids (Xie *et al.* 2004). For the Vip1Ac1 protein (739 amino

acids), which is more than 4 times bigger than T4 lysozyme, incorporation of two 4-Iodo-L-phenylalanine residues might be not enough to generate the signal for phasing. It would be difficult to add more ncAA to Vip1Ac1 protein, since the expression results as described in chapter 3 (Section 3.2.8.2), from incorporation of only 2 ncAA clearly showed significant early termination and only small amounts of full-length protein produced. The addition of 3 or 4 ncAA would be likely to reduce the amount of protein to non-viable levels. Vip1Ac1 crystals containing 4-Iodo-L-phenylalanine were soaked in NaI solution as an attempt to increase the signal but the results showed no improvement in the signal.

As ncAA incorporation and heavy atom soaking methods had proved unsuccessful, replacement of methionine by selenomethionine (SeMet) was attempted. SeMet has been widely used as a phasing tool for solving protein structure using multiwavelength anomalous diffraction (MAD) (Walden 2010). It has been found that one SeMet per 100 amino acids is requited for phasing (Hendrickson and Ogata 1997). Vip1Ac1 protein has 11 methionine residues and this should provide enough incorporation sites if ~70 % substitution is achieved. There are different methods for SeMet substitution in recombinant proteins. One method uses methionine auxotrophic *E. coli* as an expression host to achieve 100 % substitution (Hendrickson, *et al.* 1990) but this method has disadvantages of long lag time, low cell densities and poor protein yield (Walden 2010). An alternative method uses non-auxotrophic *E. coli* as an expression host but blocks the methionine biosynthesis pathway by adding isoleucine, lysine, and threonine at high concentrations to inhibit the enzyme (aspartokinase) responsible for methionine biosynthesis in *E. coli*. This method has become the most common and routinely used for SeMet substitution (Doublié 2007) and was use here to produce Vip1Ac1 as described in Section chapter 3 (Section 3.2.9). The yield of Vip1Ac1 protein containing SeMet was not high compared with wild type yield as mentioned in chapter 3. This yield of protein is, however, expected since defined M9 salts medium was used for expression and yield was sufficient to produce crystals under the same conditions as native protein.

It is necessary to check the SeMet substitution before crystal trials and data collection using intact mass analysis (Walden 2010). The molecular weight of SeMet is 196.106 Da and methionine is 149.21 Da, so the difference between them is 46.896 Da. An increase of the molecular weight of Vip1Ac1 protein due to SeMet substitution can

be measured by mass spectrometry. Theoretically, for 11 SeMet substitutions in Vip1Ac1 protein, the mass difference should be 515.856 Da. Results of mass spectrometry, Figure 4.15A and B, show an experimental difference (520 Da), very close to this theoretical value and consistent with substitution of all methionine residues in Vip1Ac1 protein with SeMet. Zhao *et al.* (2007) found that incorporation of 11 SeMet in Monster Green® fluorescent protein increased the experimental molecular weight of the protein by 543 Da (Zhao *et al.* 2007).

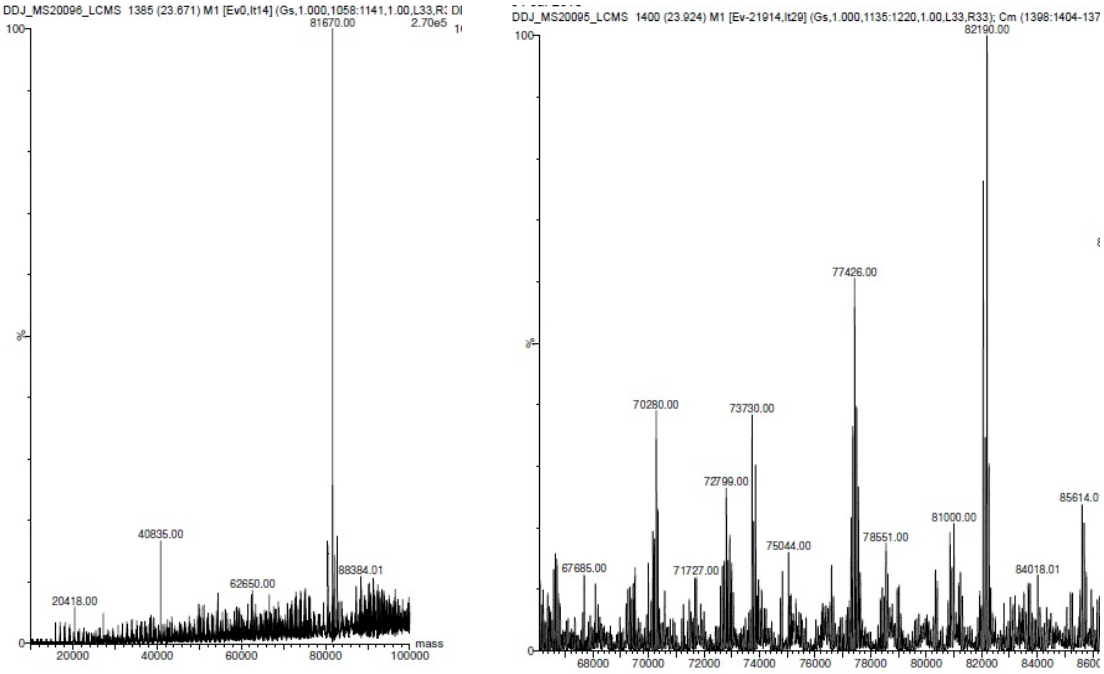

Figure 4.15. Mass spectrometry of Vip1Ac1.
A. Mass spectrometry of native Vip1Ac1. **B**. Mass spectrometry of Vip1Ac1 containing SeMet. The major peak in A at 81670 is consistent with the sequence of the Vip1Ac1 protein from amino acids KQEYQQKE-XXXX-EVSAVKK (figure 4.8; 81667 MW) and has shifted to 82190 in B.

Vip1Ac1 protein containing SeMet was concentrated to the optimum concentration for the native crystal growth (~8 mg/ml) and crystallised using the conditions native crystals as described above (4.2.3.2). After 2-3 days medium sized crystals grew as can be seen in Figure 4.16.

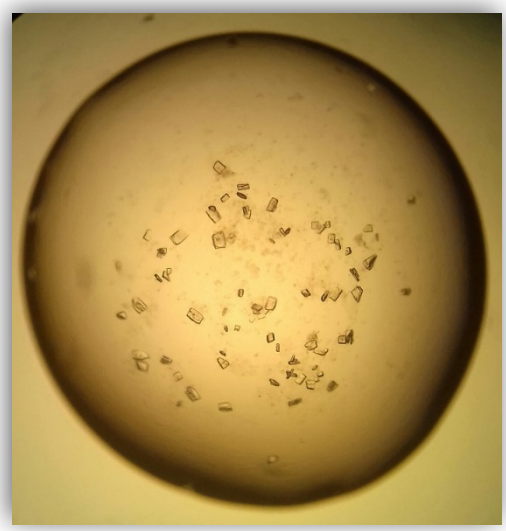

Figure 4.16. Vip1Ac1 crystals containing SeMet (40X).

Crystals were harvested as before (Section 2.12.1) and tested for X-ay diffraction and anomalous signal by Dr Neil Paterson at the Diamond Light Source, Harwell, UK, using beamline I03. The results confirmed the presence of 11 SeMet residues and the signal was sufficient to solve the phasing problem, after identification of accurate peak and inflection energies for data collection by a fluorescence scan across the selenium edge. MAD dataset with 3 sweeps of 540 degrees at peak (0.9793 Å), inflection (0.9795 Å) and high remote (0.9686 Å) were collected, then a further 1500 degrees at the peak energy were collected to help the substructure determination. All datasets were autoprocessed by the xia2 pipelines in $P2_12_12$ and used as input to the CRANK2 pipeline. Two thousand trials yielded two clear solutions and two partials. The top solution after density modification had an overall figure of merit (FOM) of 0.365 with 22 sites (occupancy > 0.2) and 2 non-crystallographic symmetry (NCS) operators detected, these agree with there being 2 copies of Vip1Ac1 per asymmetric unit (11 SeMet per one copy). The CRANK2 pipeline resulted in a final model of 1446 residues for the SeMet protein (for two monomers in the symmetrical unit) with R/R_{free} of 30.1/36.9%. This model was used as a search model for MOLREP with the high-resolution data set from native

crystals (1.47 Å) in space group $P2_12_12$. The resulting docked model was used as a starting model for PHENIX autobuild which built a model with R/R_{free} 23.0/25.3%. The COOT program was used to adjust the structure manually for several repeated cycles, TLS restrained refinement was used to refine the structure using RefMac. The final Vip1Ac1 structure (724 residues from Q16 to the C-terminal K739) was determined at a resolution of 1.47 Å.

Table 4.6 Crystal diffraction and refinement statistics of Vip1Ac1

PDB Code	6SMS
Data collection/reduction statistics	
Diamond Beamline	I04-1
Wavelength	0.9795
a, b, c (Å)	105.17, 176.24, 48.78
Space group	P 2_1 2_1 2
Resolution (Å)	1.47 – 51.29
Outer shell	1.47 – 1.51
R-merge (%)	5.4 (164.2)
CC1/2	0.999 (0.605)
I / σ	16.0 (1.1)
Completeness (%)	100 (100)
Total Measurements	1,126,168 (79,928)
Unique Reflections	154,836 (11,339)
Wilson B-factor(Å^2)	19.5
Refinement Statistics	
Non-H Atoms	6,780
R-work reflections	147,014
R-free reflections	7,752
R-work/R-free	15.6 / 19.8
rms deviations	
Bond lengths (Å)	0.012 (0.013)
Bond Angles (°)	1.707 (1.648)
[1]Coordinate error	0.053
Mean B value (Å^2)	26.9
Ramachandran Statistics	
Favoured/allowed/Outliers	635 / 21 / 1
%	96.7 / 3.2 / 0.2

* Numbers in brackets refer to outer resolution shell, where applicable.
[1] Coordinate estimated standard uncertainty in (Å), calculated based on maximum likelihood statistics.

4.2.3.4 General features of Vip1Ac1 structure

The structure of the Vip1Ac1 protein shows a flat long molecule composed of four domains as can be seen in Figure 4.17. Domain 1 (D1, from amino acid 1 to 260), contains 12 beta strands which form Greek-key topology with 7 α-helices of varied lengths and several loops connecting the β-strands and α-helices (Figure 4.18). The Greek-key topology represents the PA14 domain (predicted by modelling, above Section 4.2.2.5). D1 also contains a pair of structurally close calcium ions maintained in a buried calcium binding motif (Kobayashi *et al.* 2008) as shown in Figure 4.19. There are 8 residues (D197, D199, D201, I203, E208, N237, D240 and D250) involved in calcium binding as shown in Figure 4.20. This domain has a cleavage site in a surface loop where trypsin is able to cut to remove an N-terminal 22 kDa fragment (from amino acid 1 to 190; called Vip1Ac1$_{22}$). The remaining part of domain 1 (from amino acid 191 to 260) with the 2 calcium ions, 1 β sheet (1β$_{11}$,1β$_{12}$) and two α-helices (Figure 4.18) forms the N-terminus of the activated 60-kDa Vip1Ac1 (Vip1Ac1$_{60}$). More information about the activation and oligomerisation process can be found in chapter 5.

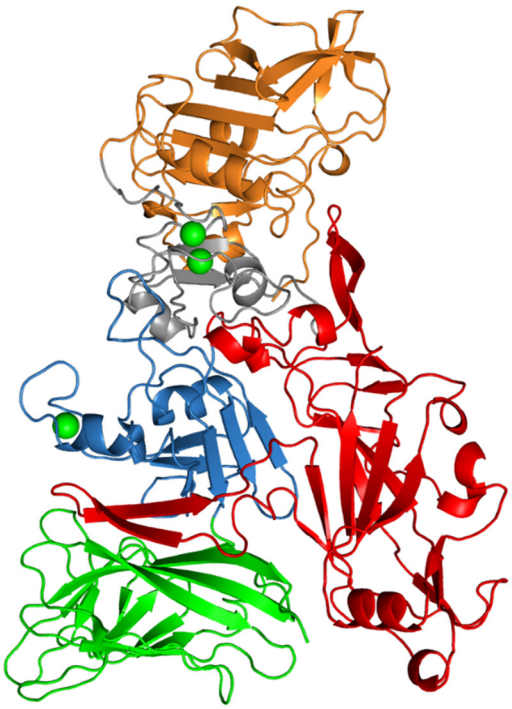

Figure 4.17. Orthogonal view of Vip1Ac1, coloured by domain.
Domain 1 consists of Vip1Ac1$_{22}$ (orange) plus the remaining part of domain 1 (grey) and two Ca^{2+} ions (green spheres). Domain two is coloured red, domain three is coloured sky-blue and contains one Ca^{2+} ion (green sphere) and domain four is coloured green. PDB code 6SMS.

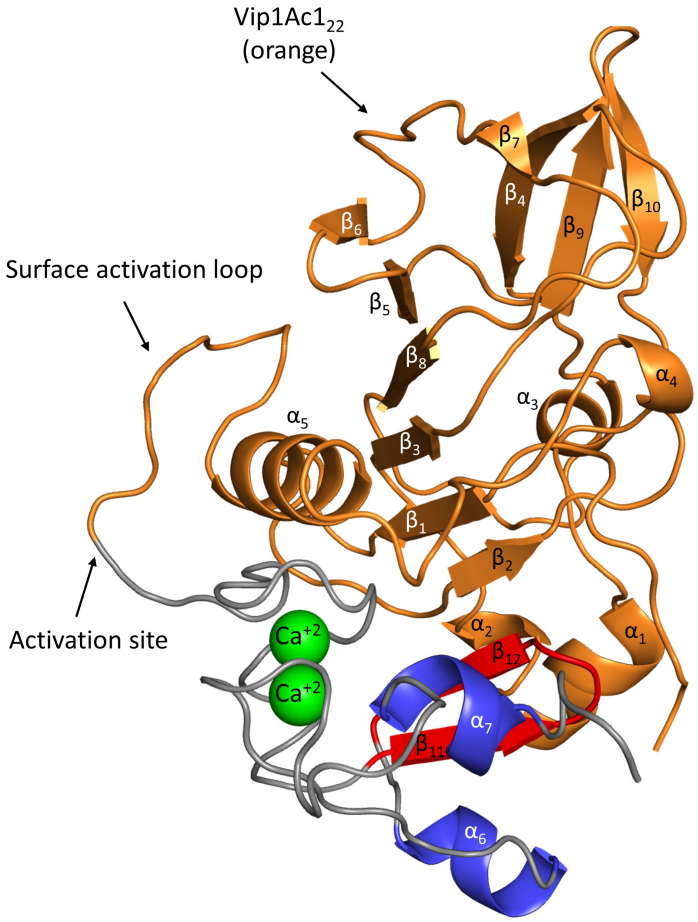

Figure 4.18. Structure of domain 1 of Vip1Ac1.
Domain 1 consists of Vip1Ac1$_{22}$ (orange) plus the remaining part of domain 1 (grey, red and blue) and two Ca^{2+} ions (green spheres). β sheet ($1\beta_{11}$,$1\beta_{12}$) is coloured red and α helices are coloured blue of the remaining part of domain 1. PDB code 6SMS.

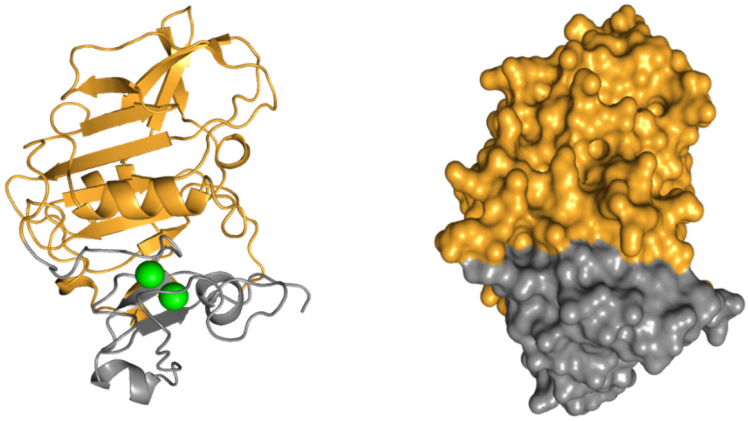

Figure 4.19. Domain 1 of Vip1Ac1 with structural calcium ions.
A. Cartoon representation of domain 1 consisting of Vip1Ac1$_{22}$ (orange) plus the remaining part of domain 1 (grey) and two structural Ca^{2+} ions (green spheres). B. Surface representation of domain 1 showing that the 2 Ca^{2+} ions in the remaining part of domain 1 (grey) are buried. PDB code 6SMS.

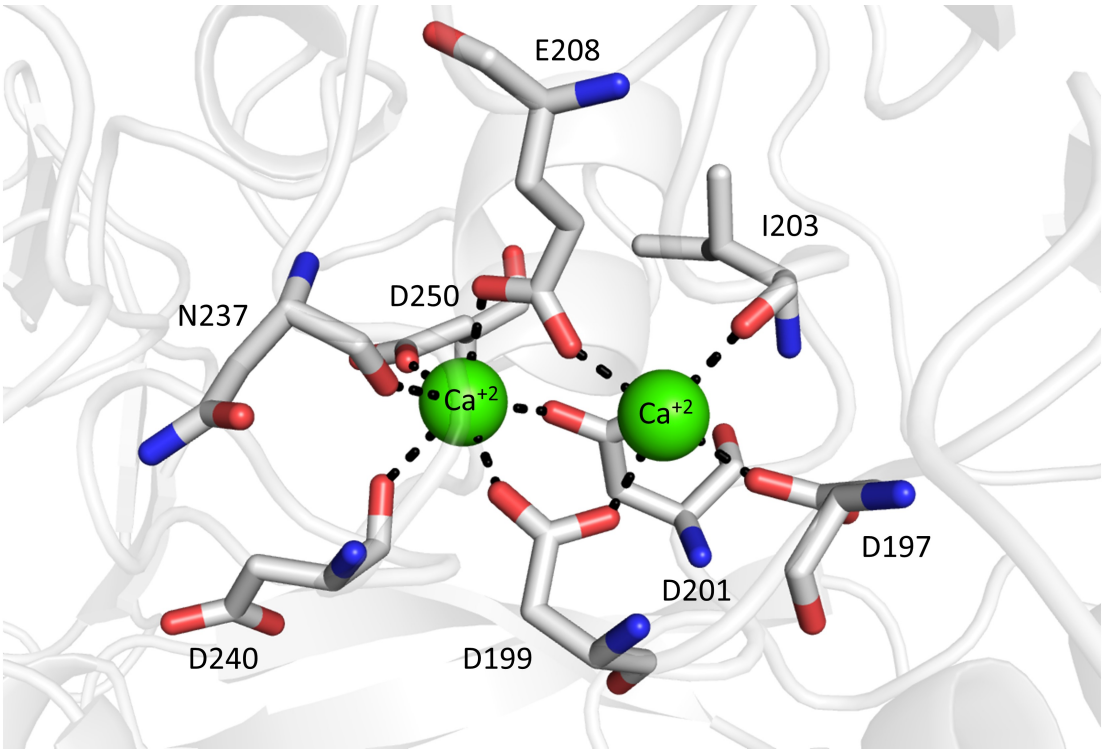

Figure 4.20. Residues involved in calcium binding.
Residues (D197, D199, D201, I203, E208, N237, D240 and D250) carbon coloured grey involved in calcium binding in full-length Vip1Ac1$_{83}$. Green spheres correspond to the calcium ions.

Domain 2 is shown in Figure 4.21 (D2, from amino acid 261 to 482). This domain has 17 β-strands, 13 of them form the β-sheet core of a Greek-key topology and the rest form 2 pairs of β-sheets. The first β-sheet (2β$_4$, 2β$_5$) is clearly separated from the rest of the domain by long loops of 9 amino acids (Figure 4.21). Interestingly, the amino acid sequence of this β-sheet has been proposed as a putative channel formation region in Vip1 (Petosa *et al.*1997; Leuber *et al.* 2006), that may form a long β-hairpin in each monomer after oligomerisation of activated Vip1Ac1$_{60}$. The second β-sheet (2β$_{16}$, 2β$_{17}$) is in the main body of the domain. This domain also has 8 short α-helices and several long and small loops connecting the β-strands and α-helices.

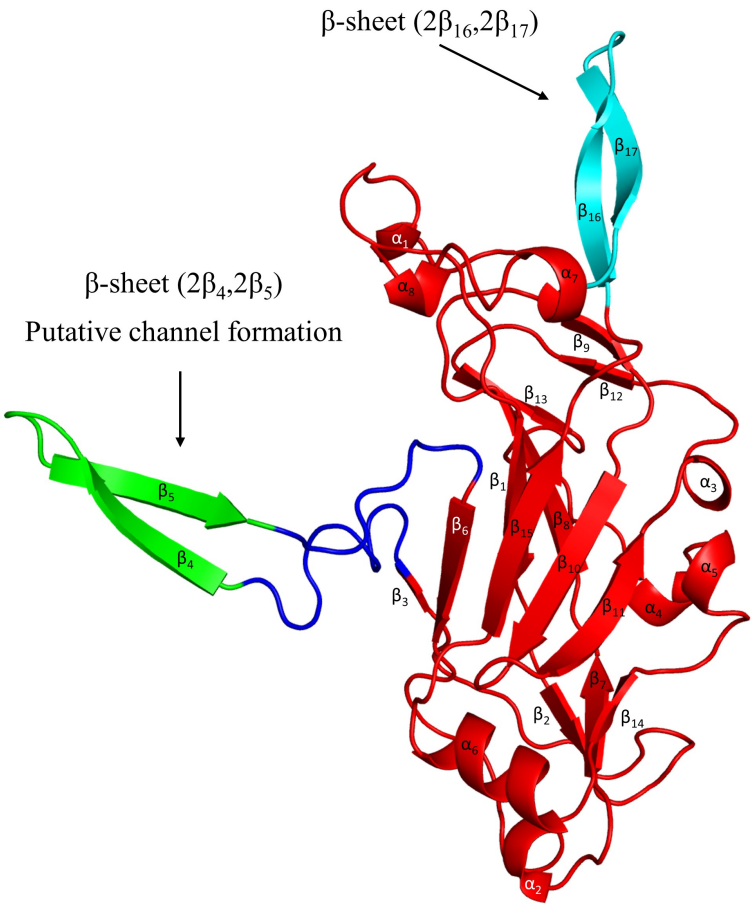

Figure 4.21. Structure of domain 2 of Vip1Ac1.
Cartoon representation of domain 2 showing the β-sheet core of a Greek-key topology and the 2 pairs of β-sheets. The first β-sheet ($2\beta_4$, $2\beta_5$) is coloured green (putative channel formation) and separated from the core of the domain by long loops of 9 amino acids (blue). The second β-sheet ($2\beta_{16}$, $2\beta_{17}$) is coloured cyan. PDB code 6SMS.

Domain 3 (D3, amino acids 483-591), is the shortest domain in the Vip1Ac1 structure, and comprises four β-strands forming a β-sheet, one separate β-sheet and three α-helices with several long and small loops connecting β-strands and α-helices as shown in Figure 4.22. This domain also has a loosely bound calcium ion bound at the end of the third α-helix.

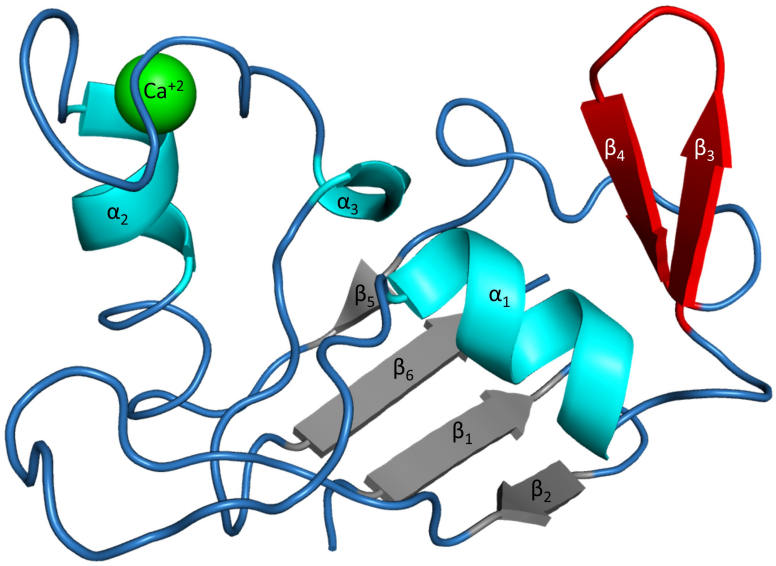

Figure 4.22. Structure of domain 3 of Vip1Ac1.
Cartoon representation of domain 3 showing the four β-strands forming a β-sheet (grey), one separate β-sheet (red) and three α-helices (cyan) plus a loosely bound calcium ion (green sphere) located at the end of the third α-helix. PDB code 6SMS.

Domain 4 (D4, from amino acid 592 to 739), comprises 13 β-strands connected by loops and generates a β-barrel like shape as shown in Figure 4.23. In general, the first three domains (D1, D2 and D3) are intimately connected as can be seen in Figure 4.24, while D4 is a more independent domain with limited interactions with the other domains. As mentioned above, based on sequence alignment of Vip1 variants, Vip1 proteins are highly conserved at the N-terminus (D1, D2 and D3) and more variable at the C-terminus (D4) and this may indicate that the conserved domains have the same functions. Domain 4 is proposed to be involved in receptor binding, based on the result of sequence alignment, as explained in Section 4.2.2.4. The structural analysis of this domain plus the results of sequence alignment (Figure 4.9), showed that part of domain 4 is highly conserved (from amino acid 592 to 644) while its C-terminal part (from amino acid 644 to 739) is more variable and, therefore, we can speculate that the more variable region (Figure 4.23) may contain important residues that interact with receptors in target cells.

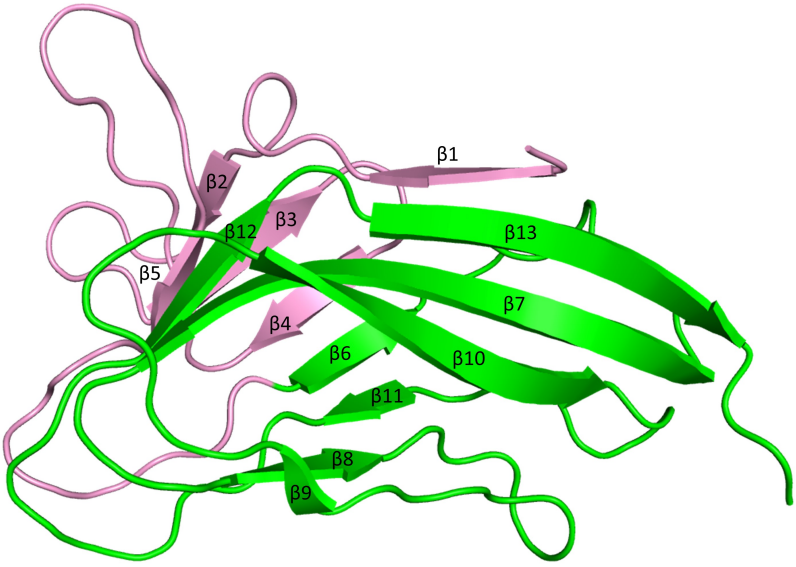

Figure 4.23. Structure of domain 4 of Vip1Ac1.
Cartoon representation of domain 4 showing the13 β-strands connected by loops and generating a β-barrel like shape. The part of domain 4 of Vip1Ac1 that is conserved with other Vip1 variants is coloured pink and the variable region amongst Vip1 variants is coloured green. PDB code 6SMS.

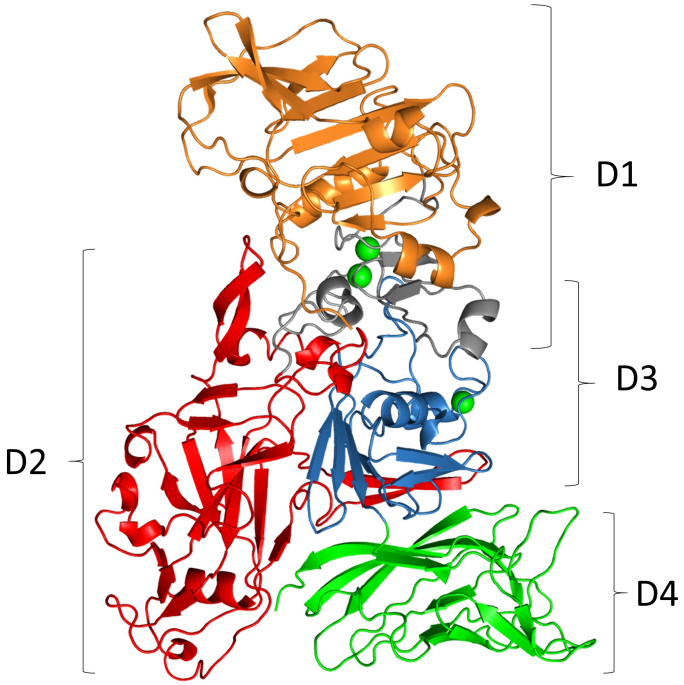

Figure 4.24. Vip1Ac1, coloured by domain, showing the intimate connection of domains 1-3
Domain 1 consists of Vip1Ac1$_{22}$ (orange) plus the remaining part of domain 1 (grey) and two Ca^{2+} ions (green spheres). Domain 2 is coloured red, domain 3 is coloured sky-blue and contains one Ca^{2+} ion (green sphere). D1, D2 and D3 are intimately connected via loops, α-helices and β-strands. Domain 4 is coloured green with limited contact with the other domains. PDB code 6SMS.

4.2.3.5 Comparison of Vip1Ac1 structure with related toxins

Homologues of Vip1 protein are known in both *Bacillus anthracis* (PA toxin) and *Clostridium* species, where they are known as iota-like binary toxins that include; *Clostridium perfringens* iota toxin (Ib), *Clostridium botulinum* C2 toxin (C2II), *Clostridium spiroforme* toxin (CSTb also known as Sb) and *Clostridium difficile* toxin (CDTb). Despite the fact that there are multiple similar toxins produced by *Clostridium* species, the molecular structure of only one (C2II) has been solved and only at a low resolution of 3.1 Å for the first three domains, without the C-terminal, receptor-binding domain 4, as shown in Figure 4.25 (Schleberger *et al.* 2006). In contrast, the structure of PA has been extensively studied and the wild-type structure has been solved at different resolutions (1.62 Å highest resolution, PDB code 4H2A). The structure of C2II and PA were used to compare with the structure of Vip1Ac1 in general as shown in Figure 4.25. The structure of PA was then used to for more detailed comparison of each domain.

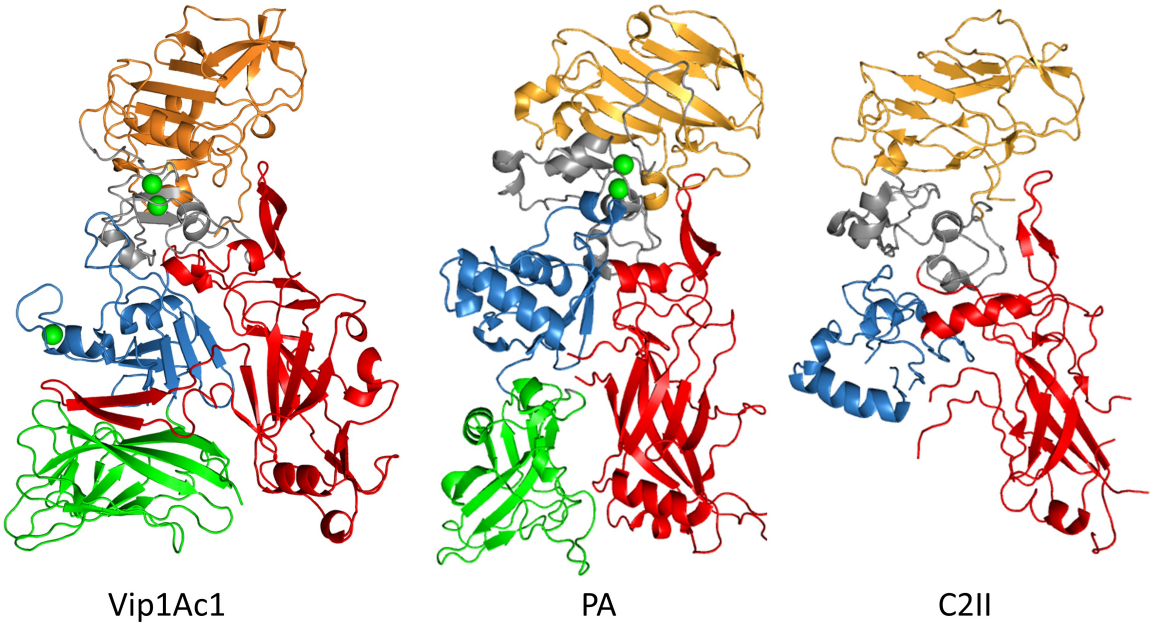

Figure 4.25. Comparison of Vip1Ac1 structure with related toxins.
Cartoon representation of toxins. Domain 1 consists of the cleavage fragment (orange) plus the remaining part of domain 1 (grey), domain 2 is coloured red, domain 3 is coloured sky-blue, domain 4 is coloured green and green spheres representing Ca^{2+} ions. PDB codes: Vip1Ac1(6SMS), PA (4H2A) and C2II (2J42).

Comparison between the three structures showed that all the toxins share the same overall folding for the first three domains (D1, D2 and D3) and are different for the domain 4 for PA and Vip1Ac1. In 1997, Petosa *et al.* suggested that the D1,D2 and D3 of PA, Ib and Vip1 have the same folds based on the high sequence similarity between them (Petosa *et al.* 1997). The proteins also share other properties such as a pfam PA14 domains, calcium binding motifs and cleavage sites for activation (orange part in the Figure 4.19 representing the cleavage fragment) in domain 1 and intimate association of D1, D2 and D3. This similarity in the first three domains in terms of structure and properties indicate the functional similarity between them (activation, oligomerisation and pore formation). The results showed another important feature, which is the difference of domain 4 amongst the three toxins (domain 4 of C2II is absent from the structure because the authors could not solve the structure by MR using PA as a model and this indicates the difference between domain 4 of C2II and PA). These differences are logical and not surprising as this is the receptor binding domain, which recognises specific receptors in the target cells. In PA, domain 4 recognises and interacts with two receptors known as ANTXR1 and ANTXR2 (Scobie *et al.* 2006). Iota-like binary toxins share high sequence identity and recognise a common receptor, lipolysis-stimulated lipoprotein receptor (LSR) (Aktories *et al.* 2018). The receptor for Vip1 is unknown.

Comparison of each domain of Vip1Ac1 with PA showed significant similarity between them, which may indicate they have the same function and mode of action during toxicity.

Domain 1: In domain 1 of both PA and Vip1Ac1, two calcium ions are strongly associated with the structure via specific residues known as calcium binding motif. In PA, the calcium ions have a stabilisation role and maintain the conformation of PA_{63} after the activation process (Gao-Sheridan *et al.* 2003). In domain one, full length PA loses a 20 kDa fragment (PA_{20}) from the N-terminus after cleavage by a furin-like enzyme or by trypsin (Benson *et al.* 1998). The activated PA_{63} then forms a soluble pre-pore oligomer (Alisaraie and Rouiller 2016). The cleavage site is separated from the calcium binding motif by 9 amino acids that form a loop that represents the N-terminus of the activated PA_{63} (Figure 4.26D). In Vip1Ac1, activation by trypsin occurs at a cleavage site separated from the calcium binding motif by 7 amino acids to remove a 22 kDa fragment from the N-terminus (chapter 5, Section 5.2.1) as shown in Figure 4.26B. The activated

Vip1Ac1$_{60}$, like PA$_{63}$, also forms a soluble pre-pore oligomer (see chapter 5). The cleaved parts of both PA$_{20}$ and Vip1Ac1$_{22}$ share a common motif, which is the pfam PA14 domain that blocks oligomerisation and represents the majority of the cleaved region as shown in Figure 4.26A and C (coloured cyan). This motif (PA14 domain) is believed to have carbohydrate binding function (Rigden *et al.* 2004).

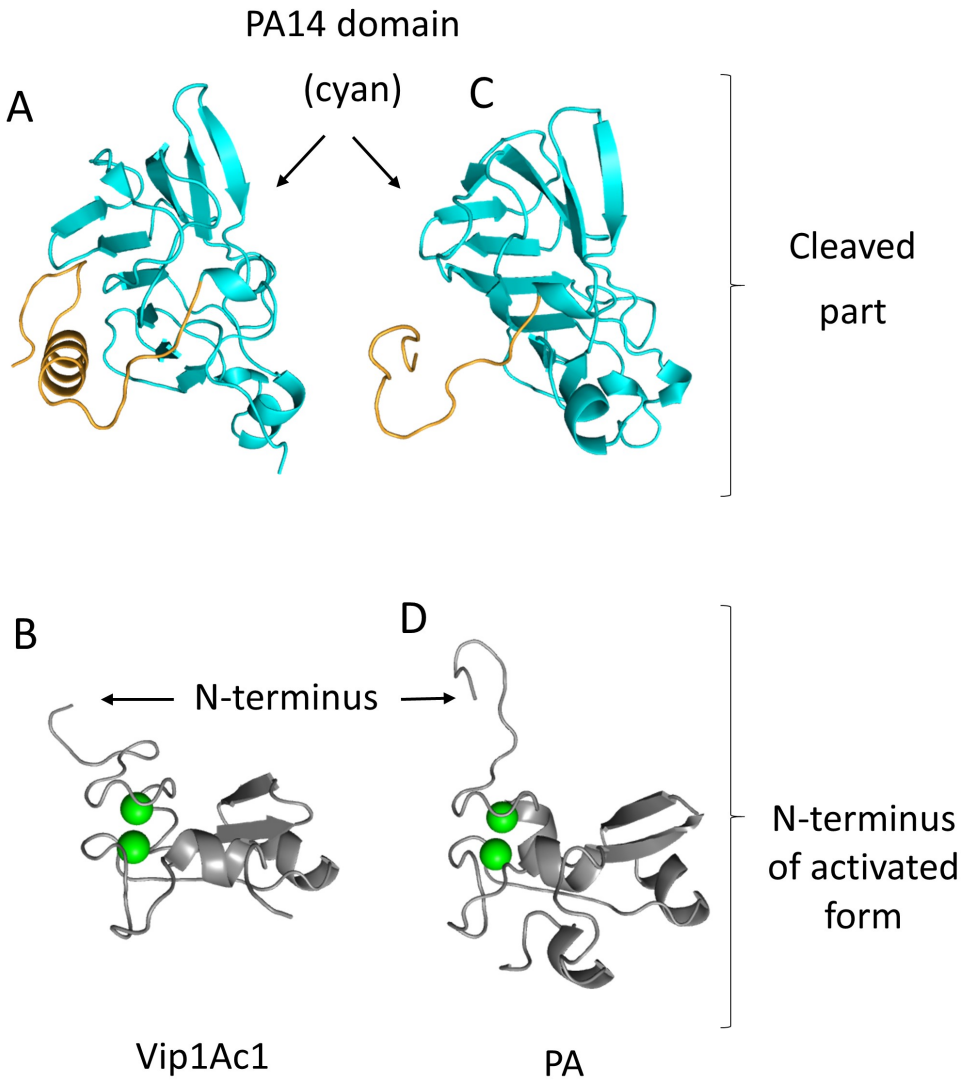

Figure 4.26. Comparison of domain 1 between Vip1Ac1 and PA.
A and **B** correspond to the Vip1Ac1 toxin, **C** and **D** correspond to the PA toxin (PDB code 4H2A). The regions removed from domain 1 by trypsin activation, Vip1Ac1$_{22}$, and furin activation (PA$_{20}$) are shown in **A** and **C** respectively with parts corresponding to the PA14 motif coloured cyan. The remaining regions of domain 1 that represent the N-termini of the activated forms are shown in **B** and **D**.

Domain 2: The domain 2 regions of both toxins are very similar in terms of structure, folding and placing of amphipathic residues as well as the residues that are involved in oligomerisation, indicating that they are likely to have the same function. Residue Q483 in PA has been found to be involved in the oligomerisation process (Mogridge *et al.* 2001) and is conserved in $2\alpha_6$ of Vip1Ac1 (Q479) (Figure 4.27).

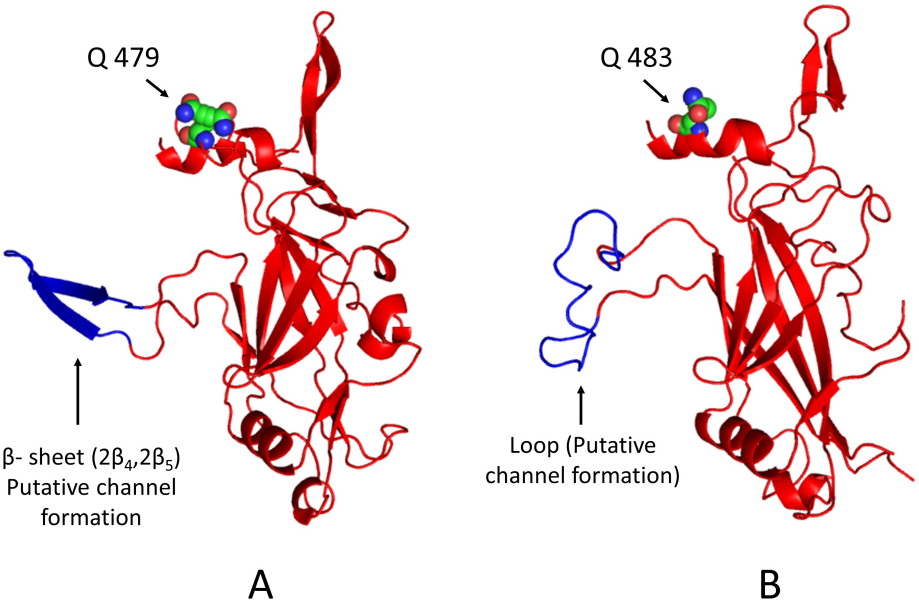

Figure 4.27. Comparison of domain 2 between Vip1Ac1 and PA.
A. Domain 2 of Vip1Ac1 showing β-sheet ($2\beta_4$, $2\beta_5$) coloured blue (putative channel formation), which separated from the core of domain by long loops. **B.** Domain 2 of PA showing the long loop coloured blue (putative channel formation). Coloured spheres correspond to the Q479 and Q483 residue of Vip1Ac1 and PA respectively.

The important amphipathic loop in PA (amino acids 302-325) is responsible for putative channel formation (producing a 14 stranded β-barrel derived from 7 subunits in the heptameric pore) after conformational rearrangement of the pre-pore oligomer (Benson *et al.* 1998). In Vip1Ac1, the structurally equivalent region is the β-sheet ($2\beta_4$ and $2\beta_5$) (Figure 4.27) as previously suggested by Leuber *et al.* (Leuber *et al.* 2006). The structure of the oligomeric pore of PA (Jiang *et al.* 2015) shows that the long β-sheet of each subunit that forms the β-barrel channel was generated after conformational rearrangement of residues from 275 to 352. During the rearrangement, the $2\beta_2/2\beta_3$ antiparallel β-sheet in the oligomer (Figures 4.27 and 4.28) becomes elongated as surrounding regions extend each of the β-strands (the loop N-terminal to $2\beta_2$ and part of the long, amphipathic loop connecting $2\beta_2$ and $2\beta_3$ (amino acids 302-325) extend $2\beta_2$;

while other residues in this loop 302-325 and the $2\alpha_1$ helix extend $2\beta_3$) as illustrated in Figure 4.28. In Vip1Ac1, the putative channel formation region already contains a β-sheet composed of $2\beta_4$ and $2\beta_5$ as shown in Figure 4.27. It is clearly separated from the rest of domain 2 by two long loops which make this β-sheet moveable. The two loops are likely to undergo conformational rearrangement and change to extend the β-strands to form an equivalent putative channel to that formed by PA although the Vip1Ac1 loop will be shorter than that for PA.

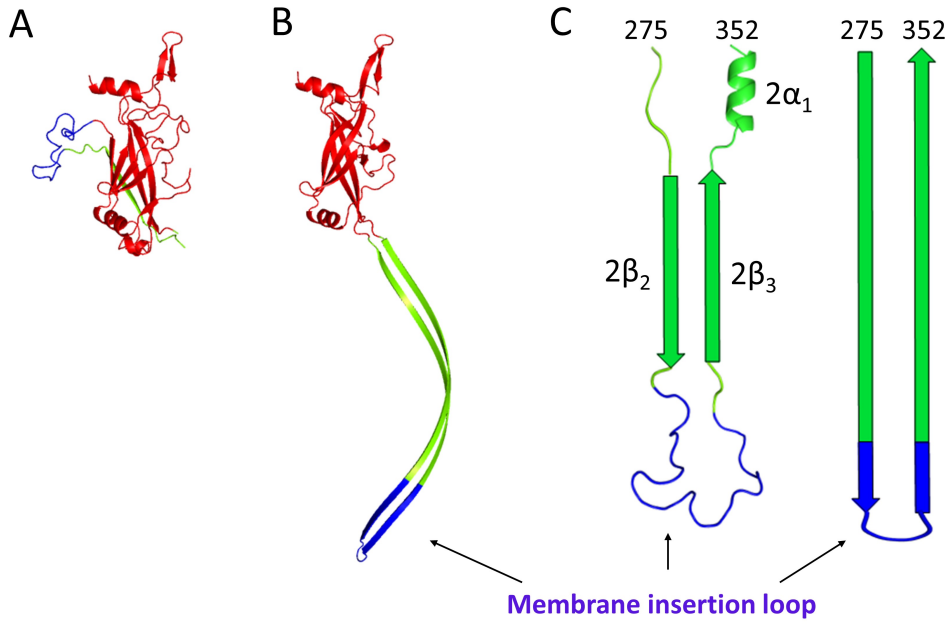

Figure 4.28. β sandwich conversion during channel formation in PA.
A. Domain 2 of Prepore PA. **B.** Domain 2 of pore PA showing the long β-sheet that involved in β barrel channel formation after conversion of $2\beta_2$,$2\beta_3$,$2\alpha_1$ (green) and the membrane insertion loop (blue). **C.** Schematic of the formation the long β-sheet after conversion of $2\beta_2$,$2\beta_3$,$2\alpha_1$ (green) and the membrane insertion loop (blue) into a long β-sheet. (PDB code: 1TZO).

Domain 3: Domain 3 of Vip1Ac1 is similar in structure and folding with domain 3 of PA, which is involved in oligomerisation and stabilisation of PA as described by Mogridge *et al.* (Mogridge *et al.* 2001). In PA, residue D512 has been found to play a critical role in oligomerisation, with the mutation to alanine abolishing the oligomerisation (Mogridge *et al.* 2001). This residue is conserved in domain 3 of Vip1Ac1 (D509) suggesting a similar role in oligomerisation. The only major difference between Vip1Ac1 and PA in this domain is the loosely bound calcium ion in Vip1Ac1 (Figure 4.29).

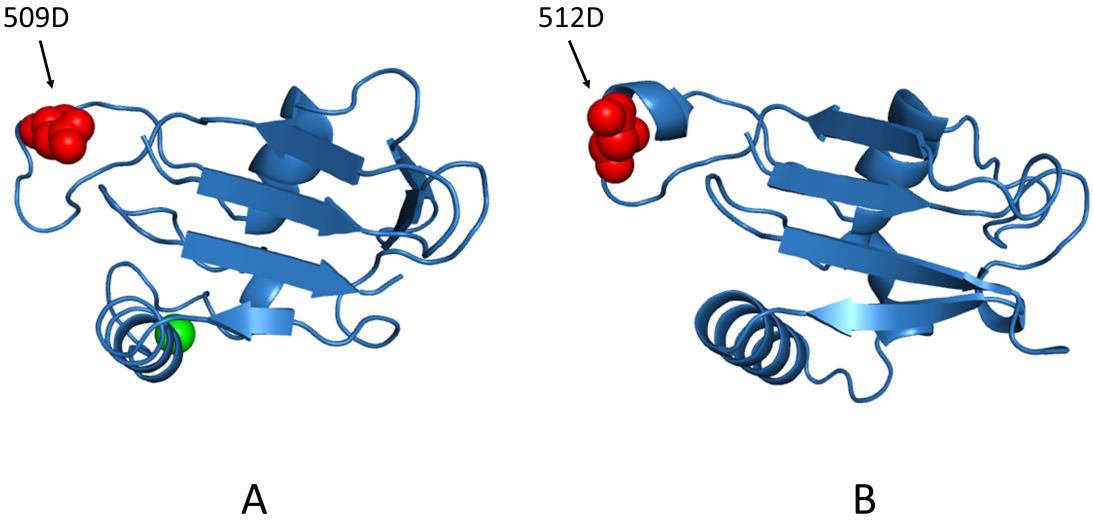

Figure 4.29. Comparison of domain 3 between Vip1Ac1 and PA.
A. Domain 3 of Vip1Ac1: green sphere corresponds to the calcium ion. **B**. Domain 3 of PA. Both domains have common features including β-sheet, 2 α-helices and the important residue for oligomerisation Asp512 of PA, conserved as Asp509 in Vip1Ac1 (red spheres).

Domain 4: Domain 4, as mentioned in the general comparison between Vip1Ac1 and related toxins, above, is the most different among these toxins. The difference between domain 4 of Vip1Ac1 and PA can be clearly seen in the alignment between the two full structures as shown in Figure 4.30. The β-strands in this domain are arranged in a different orientation and this explains why MR was unsuccessful in solving the structure. In Vip1Ac1, the β-strands are arranged perpendicular to the main axis of the protein while in PA they are arranged parallel to the main axis, which makes the PA look longer than Vip1Ac1 (Figure 4.30). It can be also speculated that this arrangement allows the pore-forming loop in domain 2 to be brought into closer proximity to the membrane on binding. This, in turn, may explain why Vip1Aa is able to function with a shorter membrane insertion loop than PA.

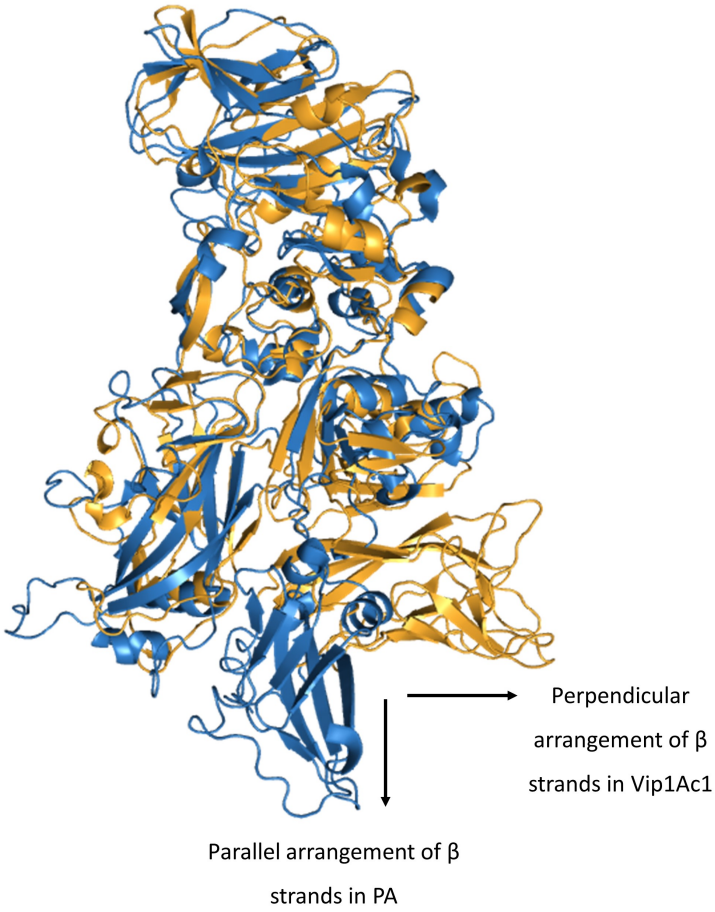

Figure 4.30. Alignment of Vip1Ac1 with PA.
Domain 1, 2 and 3 of both Vip1Ac1 (orange) and PA (blue) have similar folds. D4 in both toxins has a different orientation relative to the major axis of the protein.

4.2.4 Conclusion

The combination of different bioinformatic tools provides a powerful approach for protein studies. Analysis of Vip2 in terms of sequences and structural comparison using bioinformatic tools revealed the similarity of Vip2 with actin-ADP-ribosylating toxins in terms of structure, overall folding and important motifs (aroma-R motif, STS motif and ARTT motif). These similarities have been reported previously, leading to the prediction that Vip2 is an actin-ADP-ribosylating toxin. In this chapter, the bioinformatic analysis revealed an interesting result, which is the similarity of Vip2 in terms of important motifs mentioned above and overall folding with other ADP-ribosylating toxins that target different proteins rather than G-actin (such as Rho factor, moesin and Crk protein). These results suggested that while Vip2 might target G-actin it may also target other proteins as the ADP-ribosylating toxins can have more than one target protein (Rocha *et al.* 2003).

Sequence analysis of Vip2 variants showed that all are highly conserved, which indicates they are likely to have the same function and be able to interact with any variant of Vip1 and not be restricted to one variant. Also, these results suggest that the Vip2 has no role in toxin specificity.

Sequence analysis of Vip1 variants showed that they are highly conserved in regions corresponding to first three predicted domains and more variable in the C-terminal regions (domain 4). It can be predicted that the first 3 predicted domains have the same function amongst Vip1 variants, and this conservation explains the ability of Vip2 to interact with any variant of Vip1 and suggests that the interaction involves the N-terminal domains of Vip1 and not the C-terminal region which may be involved in receptor binding. Warren (1997) reported that two distinct variants, Vip2Ab and Vip2Aa, have a toxic effect against the Western corn rootworm (*Diabrotica virgifera virgifera*) in combination with Vip1Aa. While replacing the Vip1 component Vip1Ab with Vip1Aa led to loss of toxicity against the same insect (Warren 1997). The model prediction web servers provide powerful tools to predict and build models of a target protein when a homologous molecular structure, with high sequence identity is available.

Determination of protein molecular structure by X-ray crystallography may present several challenges starting from cloning the gene to solving the final structure. Cloning, expression and purification of Vip1Ac1 has been previously addressed in chapter 3. Protein crystals can grow in a period ranging from days to months depending on the favourable conditions for crystallisation. Vip1Ac1 produced low quality crystals diffracting to low resolution (3.5 Å) after 3 months using commercial crystal screens while optimisation around the initial conditions led to production of a good quality crystals in 2 days that diffracted at a high resolution (1.47 Å). Obtaining good diffraction data is the first step in solving the structure; the phasing problem for the crystal data need the needs to be resolved. Phasing of the Vip1Ac1 crystals to allow elucidation of the structure proved to be a major problem and various strategies (molecular replacement approach, soaking the crystals with different heavy atom solutions and incorporating a non-natural amino acid with a heavy atom into the protein) were followed unsuccessfully. Ultimately, production of crystals with methionine residues replaced selenomethionine allowed the solving of the phases and the molecular structure of Vip1Ac1 was solved at 1.47 Å resolution.

As predicted from the modelling, the crystal structure showed that Vip1Ac1 has the same overall folding as PA, with significant similarities in the first three domains, which include: a pfam PA14 domain; calcium binding motif; cleavage site for activation; the placing of amphipathic residues for putative channel formation; and conserved residues involved in oligomerisation (residues Q483/Q479 and D512/D509; numbered according to Vip1Ac1/PA respectively). Based on these similarities, it can be suggested that the first three domains in both Vip1Ac1 and PA have the same functions (activation, oligomerisation and pore formation).

Chapter five

Analysis of Vip1 and Vip2 proteins

5. Analysis of Vip1 and Vip2 proteins

5.1 Introduction

As described in chapter one, the Vip1 and Vip2 proteins constitute an AB binary toxin with activity against a number of insects. Vip2 corresponds to the A subunit (enzymatic subunit) and Vip1 corresponds to the B subunit (receptor binding/translocation subunit). Both of the subunits in AB binary toxins are required for full toxicity and the individual proteins have no toxic effect. The B subunit plays a vital role as it is responsible for toxin specificity through interactions between the its C-terminal domain and specific receptors on a target cell. It is also thought to be responsible for translocating the A subunit into the cytosol of target cell. B subunits produced by *Bacillus* and *Clostridium* species as components of the anthrax and *Clostridium* toxins listed in Section 1.3.5.3, are monomeric pro-toxins which undergo proteolytic activation in the target host by cleavage of part of the N-domain (removing about 16 to 22 kDa depending on the toxin type) to generate an active form. This active form then assembles to a homo-oligomeric pore and possibly translocates the A subunit into the cytosol (Barth *et al.* 2004).

The mechanisms of homo-oligomer formation and A subunit translocation into the cytosol of the target cell seems to be similar among AB binary toxins. In *Bacillus anthracis* anthrax toxin, protective antigen (83 kDa, PA_{83}), corresponds to the B subunit, while lethal factor (LF, 90 kDa) and edema factor (EF, 89 kDa) correspond to alternative A subunits. The monomer PA_{83} first binds to the specific cell-surface receptor in the target host (Bradley *et al.* 2001) and is activated by a furin-like protease which recognises the sequence ($_{164}$RKKR$_{167}$; Molloy *et al.* 1992). It has also been found that PA_{83} can be activated by trypsin in solution (Cunningham *et al.* 2002) by removal of 20 kDa from N-terminus generating a 63 kDa protein termed PA_{63}.This activation allows PA_{63} to oligomerise (heptamer) and to bind with a maximum of three A subunits (LF and/or EF) (Cunningham *et al.* 2002) to generate a complex. Via receptor-mediated endocytosis this complex is moved to the acidic endosomal compartment (Gordon *et al.* 1988) where the pH causes the heptameric pre-pore PA_{63} to undergo conformational change and form a membrane-channel pore and then translocate bound A subunit across the membrane to the cytosol (Benson *et al.* 1998; Cunningham *et al.* 2002).

Clostridium species produce different AB binary toxins known as iota-like binary toxins that include: *Clostridium perfringens* iota toxin (A subunit (Ia) and B subunit (Ib)), *Clostridium botulinum* C2 toxin (A subunit (C2-I) and B subunit (C2-II)), *Clostridium spiroforme* toxin (CST) (A subunit (CSTa also known Sa) and B subunit (CSTb also known Sb)) and *Clostridium difficile* toxin (CDT) (A subunit (CDTa) and B subunit (CDTb)). The different iota-like binary toxins share high sequence identity and the A subunit targets the same protein (actin) while the B subunit recognises the same receptor (lipolysis stimulated lipoprotein receptor (LSR)) (Aktories *et al.* 2018). The homo-oligomer formation and A subunit translocation into the cytosol of target cells has been proposed by many researchers. In general, the iota-like binary toxins have a parallel mode of action to anthrax toxins. First the B subunit is activated proteolytically by removing part from N-terminal domain to produce an activated heptameric form that translocates the A subunit. The activation may occur in solution or after binding of the B subunit to LSR. However, in either scenario the activation is essential for oligomerisation and interaction with the A subunit. The complex (oligomer, A subunit and LSR) is endocytosed and the low pH of endosomes induces the oligomer to form a pore that translocates the enzymatic component into the cytosol (Gerding *et al.* 2013; Aktories *et al.* 2018).

Vip1 protein has also been found to form an oligomer in solution, which could be found in the supernatants when expressed by *B. thuringiensis* strain HD201 (Leuber *et al.* 2006). The activation of Vip1 could occur in the supernatants through the action of an, as yet unidentified, extracellular protease. The same study showed the ability of the oligomer to form channels in an artificial lipid bilayer. This chapter will focus on activation, oligomerisation of Vip1Ac1 and the study the ability of Vip1Ac1 to bind and form a pore in an artificial membrane.

5.2 Results and discussion

5.2.1 Activation of Vip1Ac1 by trypsin

Some toxins are produced as pro-toxins that show no activity until they are activated by removal of part of their structure that blocks activity. The B subunit of AB binary toxins are produced as pro-toxins by both by *Bacillus* and *Clostridium* species before proteolytic activation results in the loss of about 16 to 22 kDa from N-terminus (see above). Both PA_{83} and the B subunit of iota-like binary toxins can be activated *in vitro* by trypsin. Incubation of pure, full-length Vip1Ac1 (82 kDa, $Vip1Ac1_{82}$) with trypsin at a 100/1 ratio on ice, showed that after 45 min, most of the $Vip1Ac1_{82}$ had been cleaved into two products: ~22 kDa N-terminal ($Vip1Ac1_{22}$) and ~60 kDa C-terminal part ($Vip1Ac1_{60}$) (based on the band intensity and known protein marker comparison on the SDS-PAGE gel) as shown in Figure 5.1 A. It is believed that the $Vip1Ac1_{60}$ is the activated form of the toxin, similar with other B subunits of AB binary toxins. Results from mass spectrometric analysis of purified $Vip1Ac1_{60}$ (Section 5.2.2) showed that the pure $Vip1Ac1_{60}$ has a molecular weight of 60337 Da (Figure 5.2). This matches the predicted molecular weight of the Vip1Ac1 sequence from amino acid D191 to the final residue of the protein, K739 (60336 Da).

Interestingly, this cleavage site is located in a surface loop in domain 1 (Figure 5.3) and 7 amino acids from the calcium binding motif, which is similar to the activation site of PA_{83} that is located 9 amino acids away from its calcium binding motif. Based on the results above, the trypsin cleavage site for Vip1Ac1 activation was determined to be between amino acids R190 and D191.

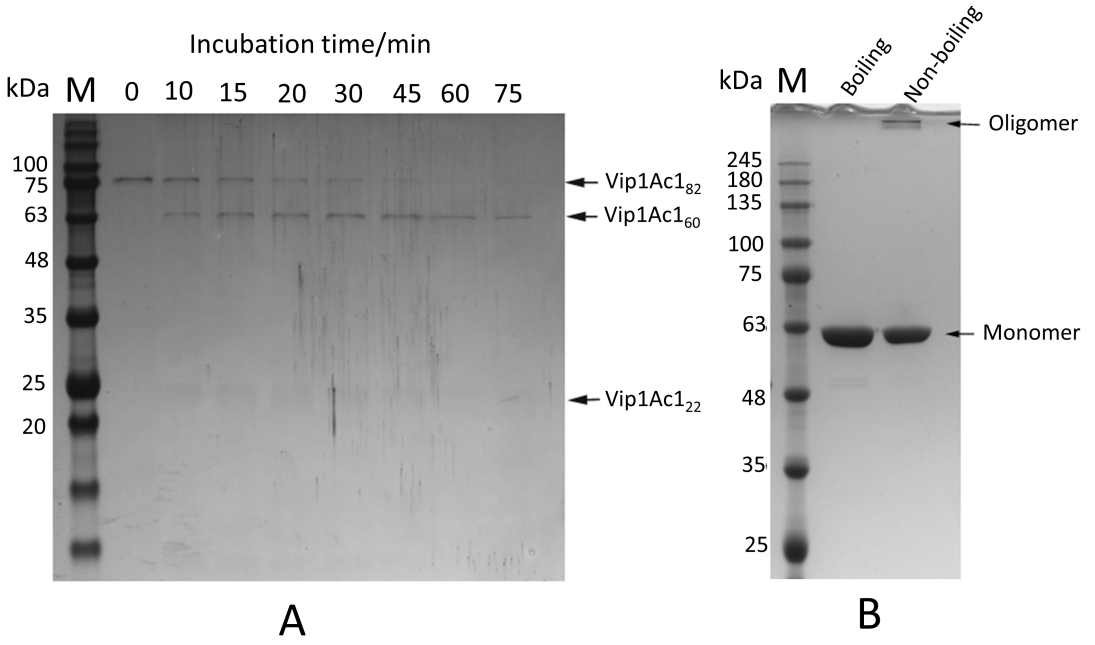

Figure 5.1. Vip1Ac1 activation with trypsin.
A. SDS-PAGE (12.5%) analysis of Vip1Ac1$_{82}$ activated with trypsin at a 100:1 ratio over a 75 min time course in ice. After 75 min all the Vip1Ac1$_{82}$ was activated to Vip1Ac1$_{60}$ **B**. SDS-PAGE (12.5%) analysis of Vip1Ac1$_{60}$. The boiling lane corresponds to the sample boiled for 5 min, which shows the monomer only; the non-boiling lane corresponds to the un-boiled sample, which shows both the oligomeric and monomeric forms of Vip1Ac1$_{60}$.

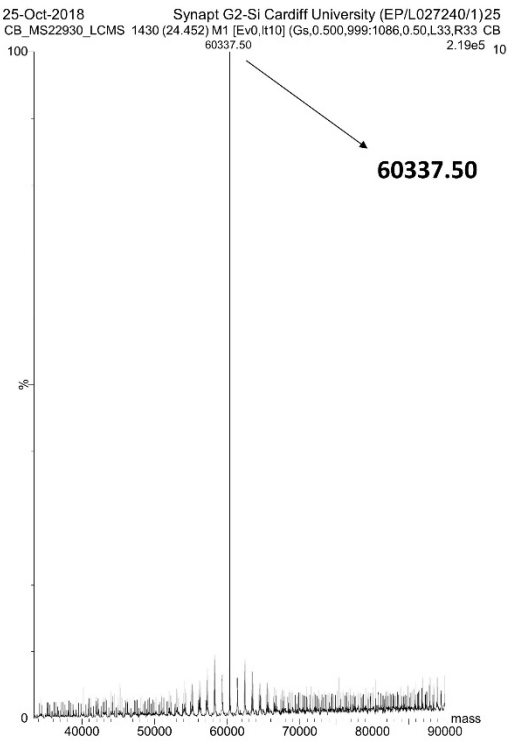

Figure 5.2. Mass spectrum analysis of Vip1Ac1$_{60}$.

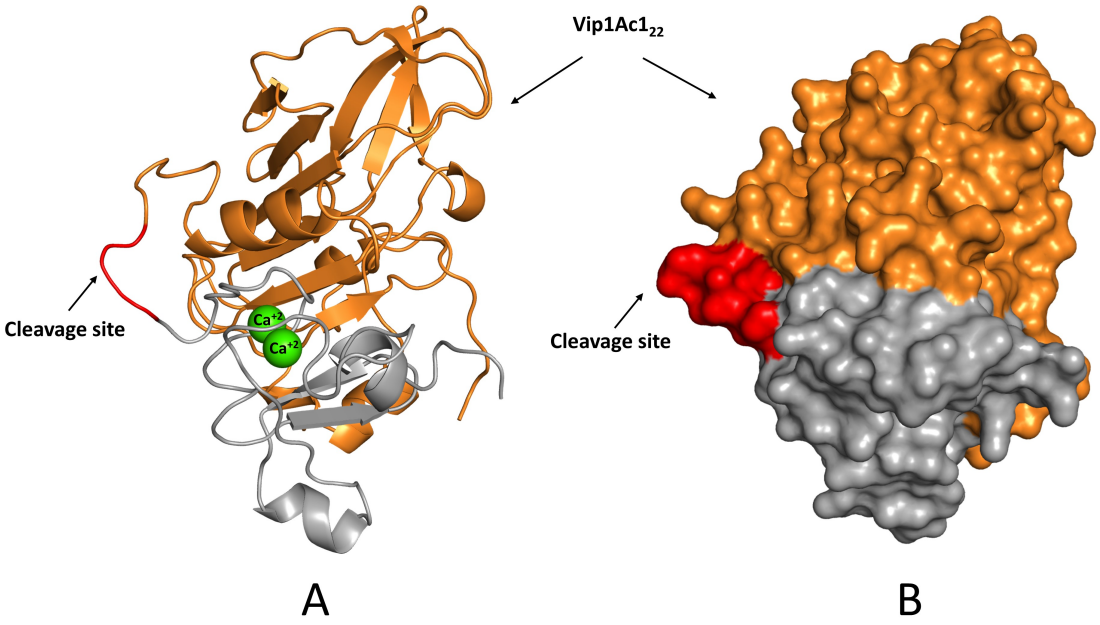

Figure 5.3. Domain 1 of Vip1Ac1 showing activation site.
A. Cartoon representation of domain 1 showing the surface loop with activation site (coloured red). **B**. Surface representation of domain 1 showing surface exposure of the cleavage site (coloured red). The region coloured orange represents Vip1Ac1$_{22}$ that is removed on activation to produce Vip1Ac1$_{60}$, the grey coloured part represents the N-terminal region of the Vip1Ac1$_{60}$. Green spheres correspond to the structural calcium ions.

5.2.2 Oligomerisation and purification of Vip1Ac1$_{60}$

Vip1Ac1$_{60}$ self-associated into an oligomer in solution after the removal of Vip1Ac1$_{22}$ by trypsin as shown by SDS PAGE in Section 5.2.1. Vip1Ac1$_{60}$ is present as an oligomer in solution and this oligomer appears to be resistant to SDS to an extent but not to boiling (Figure 5.1B). Purification of Vip1Ac1$_{60}$ by SEC after trypsin treatment (Section 5.2.1), showed two peaks (Figure 5.4A). The smaller first peak (Figure 5.4A and B) corresponds to a high molecular weight oligomer of Vip1Ac1$_{60}$ based on the comparison with SEC molecular weight standards (Section 2.4.10) and the larger second peak corresponds to the mixture of (Vip1Ac1$_{60}$ and Vip1Ac1$_{22}$) as shown in Figure 5.4 C, which have the same elution volume as inactivated Vip1Ac1$_{82}$ that does not form oligomers (chapter 3; Figure 3.24 B). Some of cleaved fragment, Vip1Ac1$_{22,}$ is still associated with Vip1Ac1$_{60}$ as seen in SDS PAGE analysis (Figure 5.4C). As mentioned in chapter 4, the first three domains (D1, D2 and D3) of Vip1Ac1$_{82}$ are intimately connected. Only small amount of

Vip1Ac1$_{60}$ was purified by SEC (Figure 5.4B) and most of the protein still associated with the cleavage part (Figure 5.4 C).

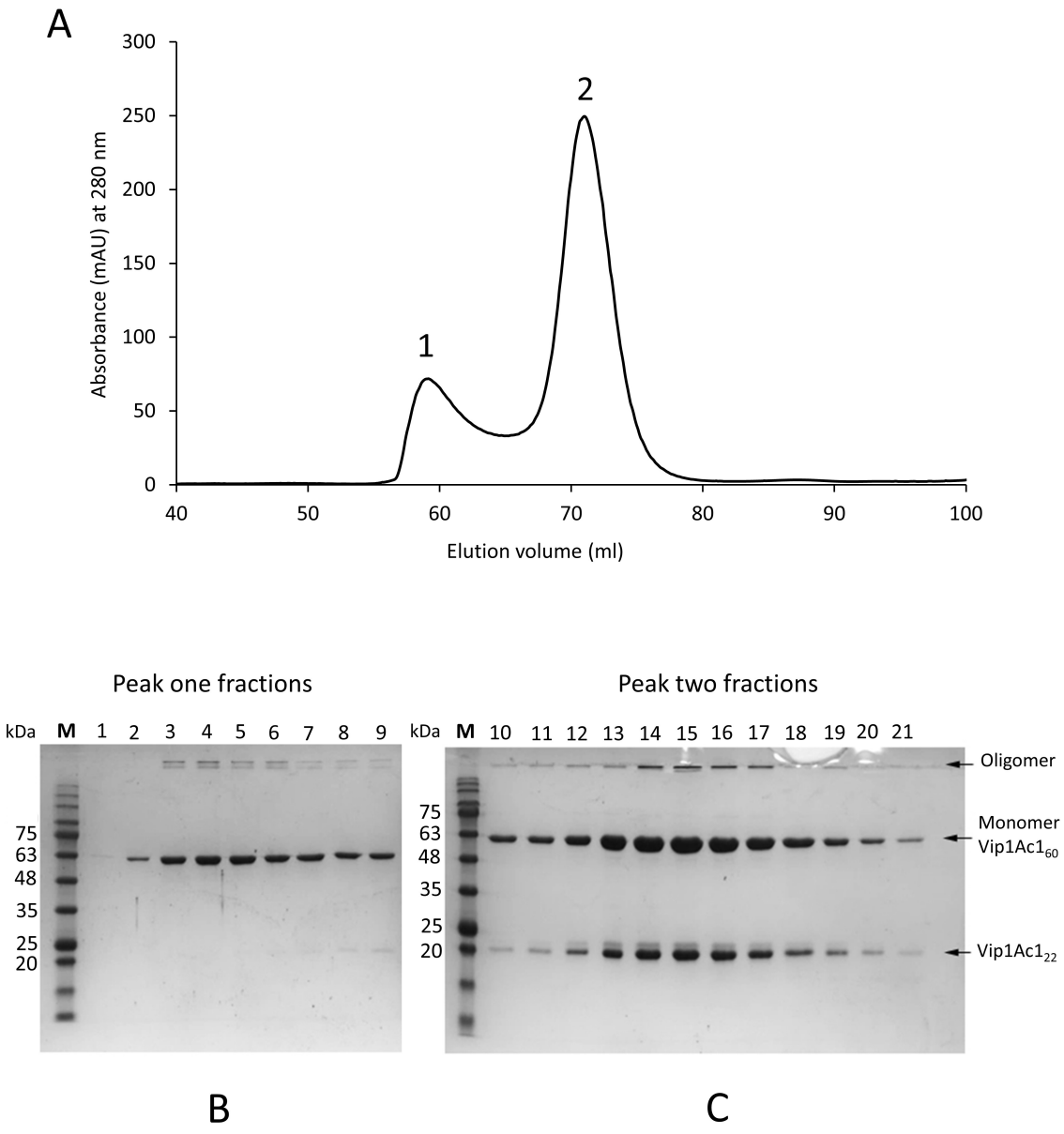

Figure 5.4. Purification profile of Vip1Ac1$_{60}$ (oligomeric state) by SEC.
A. Chromatogram of SEC purification showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/600 Superdex™ S200 pg column. **B.** SDS-PAGE (12.5%) analysis of elution fractions from SEC (peak one) showing both oligomeric and monomeric form of Vip1Ac1$_{60}$. **C**. SDS-PAGE (12.5%) analysis of elution fractions from SEC (peak two) showing oligomer, monomer (Vip1Ac1$_{60}$) and cleavage product, Vip1Ac1$_{22}$. Lane M corresponds to the protein marker of known molecular weights. Non-boiled samples were loaded into the gel.

As a result of the difference in isoelectric point (IP) between $Vip1Ac1_{22}$ (IP = 8.4) and $Vip1Ac1_{60}$ (IP = 5.1) anion ion-exchange was successfully used to purify $Vip1Ac1_{60}$ as shown in Figure 5.5A-B. Three protein elution peaks were observed on application of a NaCl gradient (0-500 mM): the first two elution peaks of low binding affinity proteins contain the mixture of $Vip1Ac1_{60}$ and $Vip1Ac1_{22}$ and the last elution peak (number 3) contains pure $Vip1Ac1_{60}$ (Figure 5.5A-B). The fractions containing the mixture of $Vip1Ac1_{60}$ and $Vip1Ac1_{22}$ were desalted and purified again by ion exchange until all the $Vip1Ac1_{60}$ was separated from $Vip1Ac1_{22}$. The fractions of purified $Vip1Ac1_{60}$ from ion exchange were pooled and subjected to SEC to remove the NaCl and achieve a high level of purity.

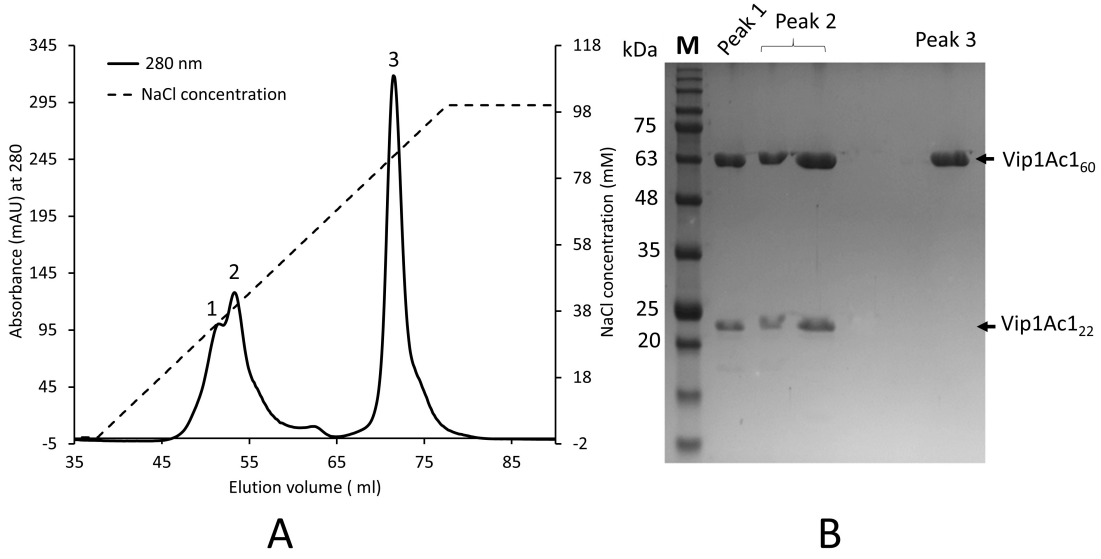

Figure 5.5. Purification profile of $Vip1Ac1_{60}$ (oligomeric state) by ion exchange.
A. Chromatogram of ion exchange purification showing absorbance (280 nm, black line) and NaCl concentration (black dashed line). **B**. SDS-PAGE (12.5%) analysis of elution fractions from ion exchange (peak 1,2 and 3). Lane M corresponds to the protein marker of known molecular weights.

Similar results have been found with PA after trypsin treatment when the cleaved part (PA_{20}) noncovalently bound to PA_{63} and cannot be separated by SEC, necessitating the use of ion-exchange chromatography to purify PA_{63} from the mixture (Collier and Young 2003).

5.2.3 Interaction of Vip2Ae3 with Vip1Ac1

The structure of the A subunit of the binary toxin such as in anthrax toxin (LF and EF) and iota-like binary toxins consist of two domains: the C-terminal domain with enzymatic activity and the N-terminal domain, which is responsible for interaction with the B subunit for translocation into the cytosol of the target cell (Barth *et al.* 2004). The structure of a Vip2 component of the Vip1/Vip2 binary toxin has been elucidated (Han *et al.*1999) and this A subunit also consists of two domains. C-terminus domain of Vip2 has been confirmed as an enzymatic domain (Han *et al.*1999). The N-domain, as discussed in the previous chapter, is predicted to interact with Vip1. To study the ability of Vip2Ae3 to interact with Vip1Ac1, Vip2Ae3 was mixed with Vip1Ac1$_{83}$ and pre-pore oligomer (Vip1Ac1$_{60}$). In initial experiments Vip1Ac (in either the Vip1Ac1$_{83}$ or Vip1Ac1$_{60}$ form) was added to Vip2Ae3 at a molar ratio of approximately 3:1 and incubated for 1 h at 37˚C before separation using SEC (Section 2.3.8.3) and analysis by SDS-PAGE.

For Vip2Ae3 and Vip1Ac1$_{83}$, the SEC chromatogram shows two distinct peaks (Figure 5.6A) corresponding to Vip1Ac1$_{83}$ and Vip2Ae3 respectively, as confirmed by SDS-PAGE analysis (Figure 5.6B). This indicates that Vip2Ae3 and Vip1Ac1$_{83}$ did not interact, at least in the conditions used. Vip2Ae3 did show binding affinity toward the pre-pore oligomer (Vip1Ac1$_{60}$). In this case, the SEC chromatogram shows 3 peaks (Figure 5.7A). The first peak eluted before the elution volume of pre-pore oligomer (Vip1Ac1$_{60}$), and corresponds to the complex of Vip1Ac1$_{60}$ and Vip2Ae3 as confirmed by SDS-PAGE analysis as shown in Figure 5.7 B. As can be seen, all Vip2Ae3 protein interacted with Vip1Ac1$_{60}$, no single species of Vip2Ae3 was observed in either the SEC chromatogram or SDS-PAGE analysis. The two other, overlapping peaks (peak 2 and 3), have the same elution profile (ranging from 60-75 ml) as Vip1Ac1$_{60}$ in the absence of Vip2Ae3, as can be seen in Figure 5.7A (solid line) and showed only Vip1Ac1$_{60}$ in SDS-PAGE analysis (Figure 5.7B), indicating that these peaks correspond to the excess of Vip1Ac1$_{60}$ in this experiment. Recently, similar results have been reported by Geng *et al.* who found that Vip2Ag binds and interacts with the trypsin treated form of Vip1Ad only and did not bind to the precursor Vip1Ad (Geng *et al.* 2019). These results indicate that the activation of Vip1Ac1$_{82}$ is not only essential for oligomerisation but also for interaction with Vip2Ae3 to allow complex formation before translocation of Vip2Ae3

into the cytosol of the target cell. It is believed that the residues involved in interaction are buried in the Vip1Ac1$_{82}$ structure and removal of the ~ 22 kDa N-terminal region (Vip1Ac1$_{22}$) result in them becoming surface exposed. Similar results have been reported for anthrax and *Clostridium* toxins. The CII protein (subunit B of *Clostridium botulinum* C2 binary toxin) required activation to allow oligomerisation and cellular uptake of subunit A (C2) (Barth *et al.* 2000). Mogridge *et al.* (2002) reported similar findings with PA: the A subunits of anthrax toxin (lethal and edema factors) bind only to the oligomeric state of PA (Mogridge *et al.* 2002).

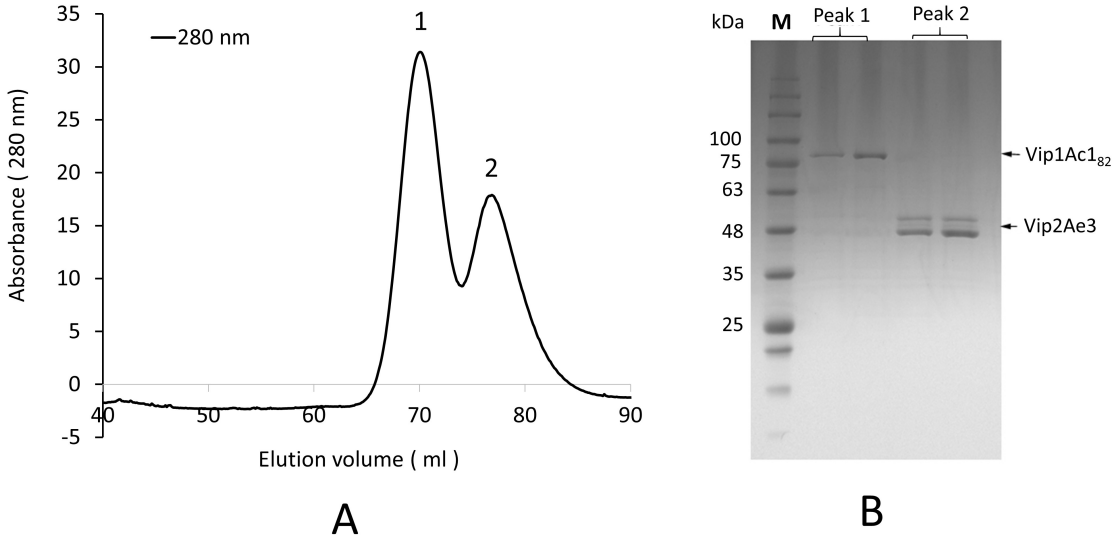

Figure 5.6. Interaction of Vip2Ae3 with Vip1Ac1$_{83}$.
A. Chromatogram of SEC purification showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/600 Superdex™ S200 pg column. **B**. SDS-PAGE (12.5%) analysis of elution fractions from SEC, peak 1 shows monomeric Vip1Ac1$_{82}$ and peak 2 shows Vip2Ae3. Lane M corresponds to the protein marker of known molecular weights.

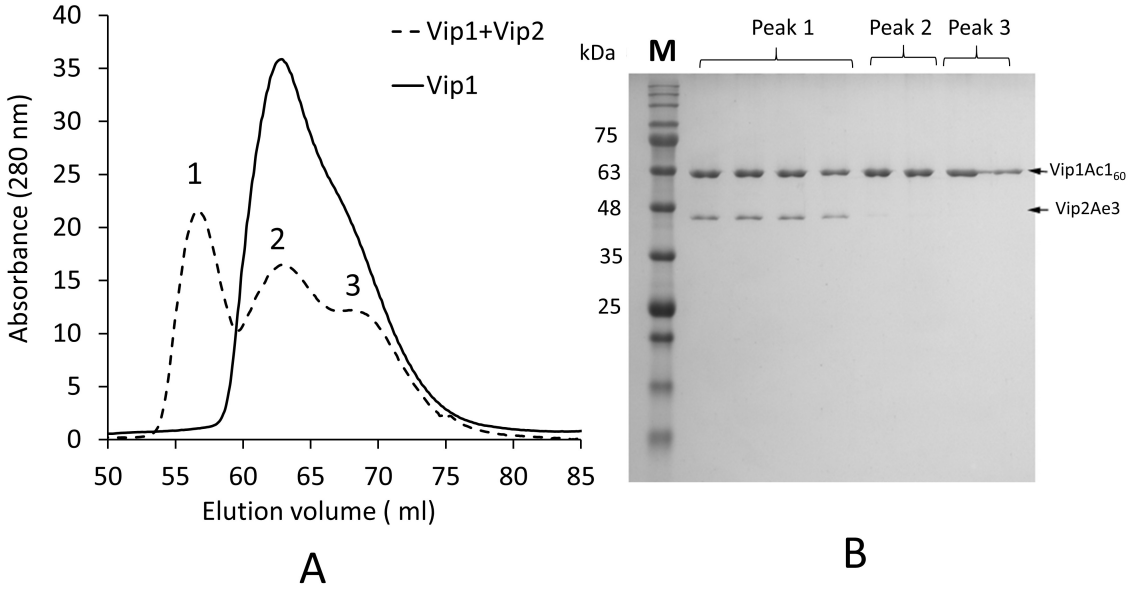

Figure 5.7. Interaction of Vip2Ae3 with Vip1Ac1$_{60}$ (oligomeric state).
A. Chromatogram SEC purification showing absorbance (280 nm) for Vip1Ac1$_{60}$ (black line) and mixture of Vip1Ac1$_{60}$ and Vip2Ae3 (black dashed line) over total volume run through the Hiload™ 16/600 Superdex™ S200 pg column. **B.** SDS-PAGE (12.5%) analysis of elution fractions from SEC, peak 1 showing both Vip1Ac1$_{60}$ and Vip2Ae3, peak 2 and peak 3 showing Vip1Ac1$_{60}$. Lane M corresponds to the protein marker of known molecular weights.

Based on the calibrated SEC column and molecular weight estimation as described in Section 2.4.10, the complex of (Vip1Ac1$_{60}$ and Vip2Ae3) eluted as part of the first peak (Figure 5.7A) has a molecular weight *crica* 131 kDa higher compared to the molecular weight of the oligomeric state of Vip1Ac1$_{60}$ which eluted in the second peak (Figure 5.7A). This increase in molecular weight is approximately equal to 2.7 times the molecular weight of Vip2Ae3 (~48.5 kDa). The SDS-PAGE (Figure 5.7B) for lanes from peak 1 was scanned and analysed using ImageJ software (Schneider *et al.* 2012), which indicated an average intensity ratio of 2.88:1(0.1SD). Adjusting for the molecular weights of the two proteins (60.3 and 48.5 kDa, respectively), this approximates to a molar ratio of 2.25:1. If we assume that all of the Vip1Ac1$_{60}$ is present as a heptamer, this equates to 3.1 Vip2Ae3 monomers bound to each Vip1Ac1$_{60}$ heptamer. The above results are similar to findings for anthrax toxin in which the oligomeric state of PA has been shown to be capable of binding three molecules of A subunits (lethal factor or edema factor ; Mogridge *et al.* 2002).

Interestingly, incubation of Vip2Ae3 with an N-terminal His-tag with the pre-pore oligomer (Vip1Ac1$_{60}$) showed no binding affinity and two distinct peaks were observed in the SEC chromatogram for the mixture as shown in Figure 5.8A. These two peaks correspond to Vip1Ac1$_{60}$ and Vip2Ae3 respectively as confirmed by SDS-PAGE analysis (Figure 5.8B). This result suggests that N-terminal His-tag in Vip2Ae3 prevents the interaction with Vip1Ac1$_{60}$ and this may provide evidence that the function of the N-domain of Vip2Ae3 mirrors that of other A subunits and is responsible for interaction with the B subunit for translocation into the cytosol of the target cell. Vip2Ae3 (with and without the His tag) seems to be unstable at 37˚C over one hour and seems to undergo some partial processing to form a dimer when it does not interact with Vip1Ac1$_{60}$ as can be observed in Figure 5.6B (without His-tag), above, and Figure 5.8B (with His-tag).

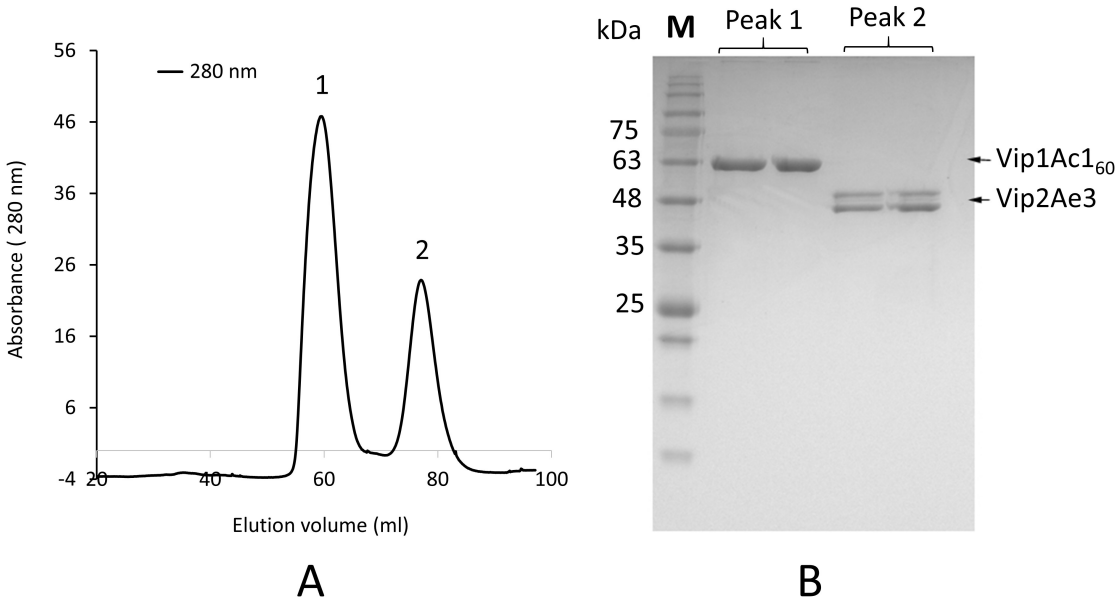

Figure 5.8. Interaction of Vip2Ae3 (with N-terminal His-tag) with Vip1Ac1$_{60}$ (oligomeric state). **A.** SEC purification chromatogram showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/600 Superdex™ S200 pg column. **B.** SDS-PAGE (12.5%) analysis of elution fractions from SEC, peak 1 shows Vip1Ac1$_{60}$ and peak 2 shows Vip2Ae3. Lane M corresponds to the protein marker of known molecular weights.

For additional evidence to support the idea that the N-domain of Vip2Ae3 is responsible for interacting with Vip1Ac1$_{60}$, both domains of Vip2Ae3 (N-domain and C-domain) were expressed separately to study the interaction of each with Vip1Ac1$_{60}$. The *Vip2Ae3* gene was reconstructed by adding a stop codon and an *Nco*I restriction site at the end of the N-domain using the whole vector site directed mutagenesis method (Section 2.2.2). This construct was used for N-domain expression. The region encoding the C-domain was removed from the above construct by digestion with two restriction enzymes (*Nco*I and *Eco*RI) and re-inserted into the pET22b vector between the *Nco*I and *Eco*RI sites as described in Sections 2.2.6 and 2.2.8. This construct provides a periplasmic signal sequence and N-terminal His-tag before the sequence of the C-domain. Reconstruction to produce these constructs is illustrated in Figure 5.9.

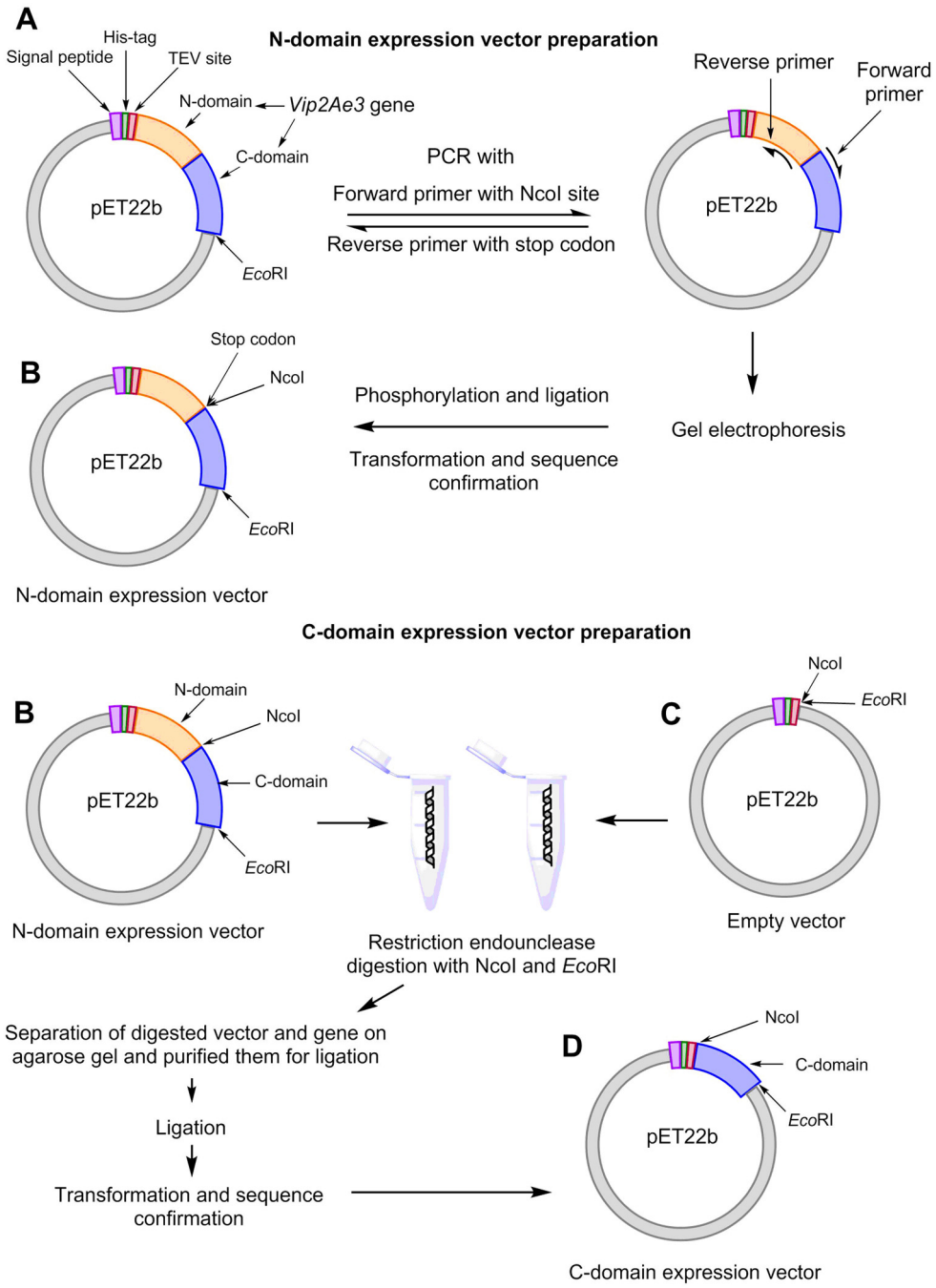

Figure 5.9. Schematic map for the construction of Vip2Ae3 N and C-domain expression vectors.

A. pET22b vector map containing *Vip2Ae3* gene (orange; N-domain and blue; C-domain) used in whole plasmid PCR with mutagenic primers to add stop codon and *Nco*I restriction site at the end of the N-domain. **B**. Final map of pET22b vector containing *Vip2Ae3* gene with stop codon and *Nco*I restriction site at the end of the N-domain. This construct was used for N-domain expression and for preparation of the C-domain expression construct. **C**. Empty pET22b vector map contains *Nco*I and *Eco*RI restriction sites. This vector was used for *C-domain* gene insertion. **D.** Final map of pET22b vector containing *C-domain* gene and used for C-domain expression.

When expression studies were carried out, neither domain was secreted into the periplasmic space and both accumulated as insoluble inclusions when expressed using the expression conditions developed for Vip2Ae3 protein (0.1 mM IPTG, 20˚C overnight incubation time). These results suggest that both domains are required for full correct folding. Due to time constraints, no further work was performed with these constructs.

5.2.4 Interaction of Vip1Ac1 with artificial membranes (lipid vesicles)

Most bacterial toxins recognise and interact with a surface receptor in the target cell as an essential step of toxicity. Most of the surface receptors are lipids or lipid derivatives, or membrane-associated proteins (Geny and Popoff 2006). Most pore forming toxins (PFTs) have been found to interact and form pores in artificial membranes (lipid vesicles) (Belmonte *et al.* 1994; Sun and Collier 2010 ; Ros *et al.* 2013; Yilmaz and Kobayashi 2016). Here the Vip1Ac1 protein was tested with lipid vesicles to assess its ability to bind and form a pore.

5.2.4.1 Lipid binding affinity

To study binding affinity, Vip1Ac1 protein was labelled with a fluorescent dye (DBCO-TAMRA) via Click chemistry to visualise the protein on the surface of an artificial membrane using fluorescent microscopy. First, an azide group (Click chemistry active group) was introduced into the Vip1Ac1 at two positions (Y358 and Y536) by incorporating ncAA *p*-azido-L-phenylalanine (azF) carrying the azide group, as described in Section 3.2.8.2, to allow the reaction between the azide group in Vip1Ac1$^{358\text{-}536azF}$ and alkyne group in the fluorescent dye via Click chemistry.

5.2.4.1.1 Labelling of Vip1Ac1$^{358\text{-}536azF}$ with fluorescent dye

The purified Vip1Ac1$^{358\text{-}536azF}$ was labelled with DBCO-TAMRA dye (contains a cyclic alkyne group for orthogonal Click chemistry reaction with the azide group) as described in Section 2.5.3. Successful labelling with fluorescent dye was confirmed as outlined in Section 2.5.1. Based on the SDS-PAGE analysis and UV-imaging of the total periplasmic protein containing Vip1Ac1$^{358\text{-}536azF}$ with DBCO-TAMRA dye, results showed that Vip1Ac1$^{358\text{-}536azF}$ had been successfully labelled (Figure 5.10A-B). The cyclic alkyne group in DBCO-TAMRA reacted specifically with the azide group present in Vip1Ac1$^{358\text{-}536azF}$ and gave a clear fluorescent band under UV (Figure 5.10A). The

lack of azide groups in other proteins present in the periplasmic space stops the reaction of DBCO-TAMRA with any other protein and no other fluorescent bands were observed under UV. Labelled Vip1Ac1$^{358\text{-}536azF}$ was purified from excess dye by SEC (Section 2.3.8.3) and was used for further studies.

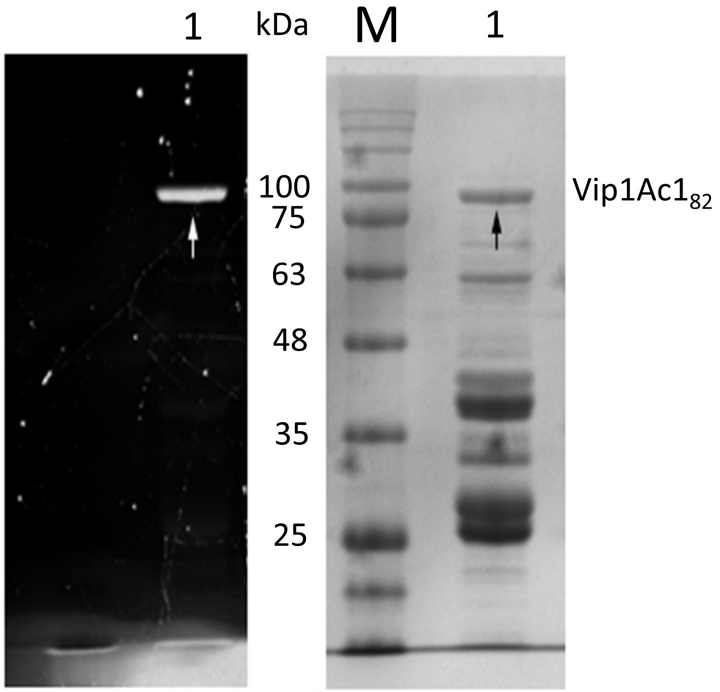

Figure 5.10. Vip1Ac1$_{82}$ labelled with fluorescent dye.

A) SDS-PAGE (12.5%) analysis visualised and imaged under UV light before staining, showing fluorescent protein bands which indicate the click reaction between the protein and compatible fluorescent DBCO-TAMRA dye. **B)** SDS-PAGE analysis after staining with Coomassie blue R250 showing the protein bands. Lane 1 corresponds to the total periplasmic protein of induced cells. Lane M corresponds to the protein marker of known molecular weights.

5.2.4.1.2 Preparation of an artificial membrane (lipid vesicles)

Lipid vesicles have been widely used as a membrane model to study the ability of toxins to interact with the membrane to facilitate pore formation (Belmonte *et al.* 1993; Vad *et al.* 2010; Karasawa *et al.* 2017). To study the interaction of Vip1Ac1 with an artificial membrane, giant lipid vesicles (GLVs) composed of one or more lipid bilayer, were prepared as outlined in Section 2.6.1. Results from examination by light microscopy (40X magnification) showed that a range of size and type of lipid vesicles were formed (Figure 5.11). Separating the lipid vesicles according to the size or type was

unnecessary for this experiment as the purpose of these lipid vesicles was to assess the ability of Vip1Ac1 to bind into the membrane surface.

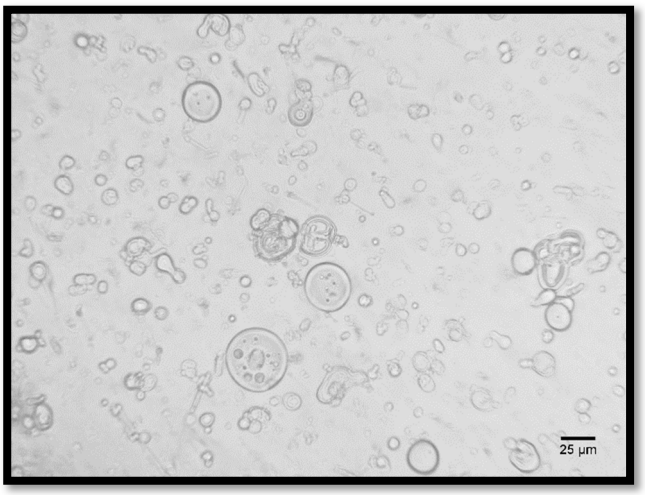

Figure 5.11. Giant lipid vesicles (GLVs) under microscopy (40X).
Giant lipid vesicles composed of one or more lipid bilayer imaging using 40X magnification.

5.2.4.1.3 Binding of Vip1Ac1 into GLVs

Fluorescently labelled full-length Vip1Ac1$_{82}$ and Vip1Ac1$_{60}$ (oligomeric state) at a final concentration of 100 µM were incubated with GLVs (1 mg/ml) for 1 h with mixing at room temperature. After the incubation period, samples were imaging (Section 2.9.2). Based on the fluorescence microscopy imaging, no fluorescent lipid vesicles were observed with full-length Vip1Ac1$_{82}$ (Figure 5.12A) suggesting that the monomeric Vip1Ac1$_{82}$ was unable to bind or has a low binding affinity. In contrast, Vip1Ac1$_{60}$ (oligomeric state) showed a number of fluorescent lipid vesicles (Figure 5.12B-C). These results indicate the ability of the oligomeric form to bind to the lipid vesicles. It is interesting to note that Vip1Ac1$_{60}$ accumulated and covered only a subset of the lipid vesicles, which may suggest that the oligomeric form has a tendency to accumulate as a population in individual lipid vesicles with a synergic affect rather than dispersing as single oligomers on separate lipid vesicles. This could occur in nature when the toxins attack the target cell.

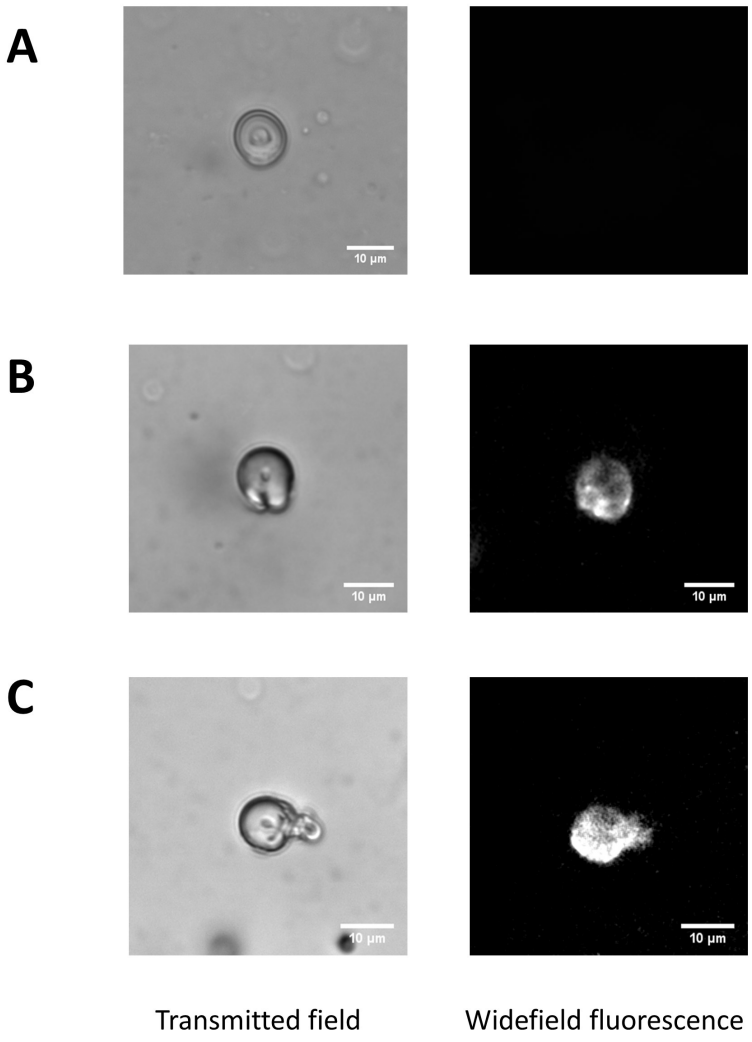

Figure 5.12. Interaction of Vip1Ac1 with artificial membrane.
A. Lipid vesicle mixed with $Vip1Ac1_{82}$, no fluorescent lipid vesicles were observed with widefield fluorescence microscopy. **B** and **C** shows two examples lipid vesicles mixed with $Vip1Ac1_{60}$. fluorescent lipid vesicles were observed with widefield fluorescence microscopy.

5.2.4.2 Pore formation and YO-PRO-1 uptake assay

5.2.4.2.1 Preparation of liposomes containing nucleic acids

Small lipid vesicles (liposomes) composed of one lipid bilayer encapsulating a DNA cargo were prepared as described in Section 2.6.2. These were used to assess the ability of Vip1Ac1 to form a pore in the artificial lipid membrane through measurement of the fluorescence intensity of YO-PRO-1 that becomes fluorescent when it interacts with entrapped DNA as a result of pore formation. As mentioned in chapter two, YO-PRO-1

is a nucleic acid stain that shows green fluorescence after binding with DNA. After preparation, the successful formation of liposomes and incorporation of DNA was tested using the YO-PRO-1 uptake assay (Section 2.8) under two different buffer conditions (50 mM Tris-HCl pH 8.0, and 50 mM phosphate buffer pH 5.5). Results from mixing YO-PRO-1 dye with DNA- liposomes in the presence (positive control) and absence (negative control) of 1 % Triton X100 indicated successful formation of DNA- liposomes as there was high fluorescence signal in the positive control compared with negative control due to interaction of YO-PRO-1 dye with entrapped DNA after lipid vesicles disruption by Triton X100 (Figure 5.13). The results also showed the ability of YO-PRO-1 to interact with DNA under both acidic and basic conditions.

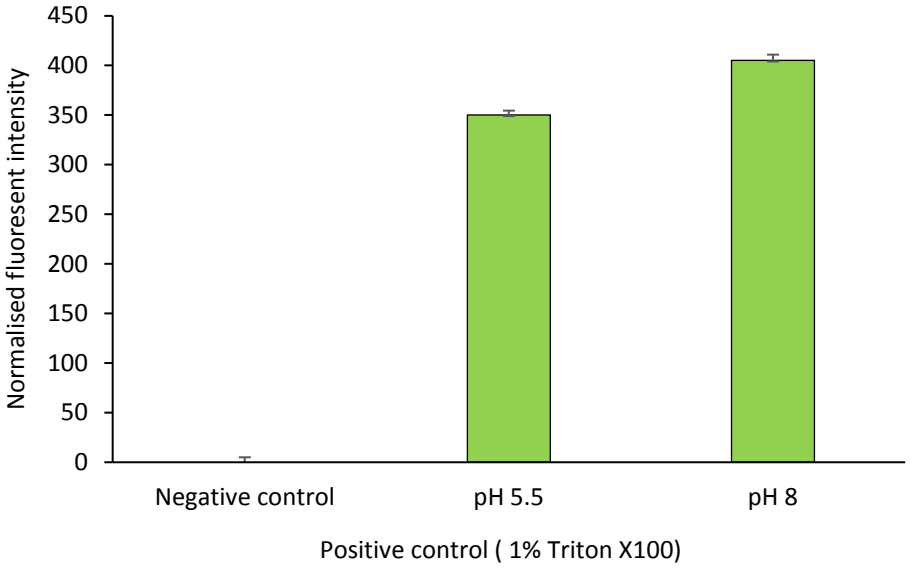

Figure 5.13. YO-PRO-1 uptake assay for DNA- liposomes.
The bars represent the means of normalised fluorescence intensity from 3 replicates and the error bars correspond to the standard deviation. Fluorescence intensity was normalised by subtractions of the fluorescence intensity of the negative control.

5.2.4.2.2 Pore formation by Vip1Ac1

Successfully formed DNA-liposomes as prepared in Section 5.2.4.2.1 were used to assess the ability of Vip1Ac1 to form a pore. As mentioned in the introduction the pore formation by oligomeric state of B subunit of AB binary toxins, especially in anthrax and iota-like binary toxins, is induced by acidic environments. The YO-PRO-1 dye was

prepared in two buffers; 50 mM Tris-HCl, pH 8.0 and 50 mM sodium phosphate buffer, pH 5.5 at a final concentration of 5 μM. The mixture of the DNA-liposomes, Vip1Ac1$_{60}$ (5 μM final concentration) and YO-PRO-1 dye was prepared as outlined in Section 2.8 and the fluorescence intensity recorded using a microplate reader.

Figure 5.14 presents the normalised fluorescence intensity obtained. It is apparent from these results that Vip1Ac1$_{60}$ forms a pore in liposomes at pH 5.5 only, as the high fluorescence intensity indicates pore formation or lipid disruption by Vip1Ac1$_{60}$, allowing interaction of the YO-PRO-1 dye with DNA. At pH 8.0, very low fluorescence intensity was recorded (95% lower than that recorded at pH 5.5), indicating little disruption of the DNA-liposomes in the presence of Vip1Ac1$_{60}$ at this pH despite accumulation of the prepore state on lipid membranes at pH 8.0 (Figure 5.12). Statistical analysis by one-way ANOVA followed by Tukey HSD's test ($p<0.01$) showed a significant difference between the sample at pH 5.5 and both samples at pH 8.0 and the negative control (without protein), there is no significant difference between the sample at pH 8.0 and the negative control. The results presented here agree with other results reported by Leuber *et al.*(2006) who found that a protein that they called Vip1Ac (currently known as Vip1Ca1 according to the *Bacillus thuringiensis* toxin nomenclature (Crickmore *et al.* 2019)) forms a channel in an artificial black lipid bilayer membranes at pH 5.5 (Leuber *et al.* 2006). The results above are consistent with the idea that at pH 5.5 the pre-pore oligomer undergoes conformational rearrangement allowing the putative channel forming region in domain 2 of Vip1Ac1$_{60}$ (as described in chapter 4;Section 4.2.3.5) to adopt a 14 stranded β barrel derived from 7 subunits to produce the heptameric pore. Electron microscopy studies of the oligomeric state of Vip1Ac1$_{60}$ confirmed the formation of the putative channel at pH 5.5 (see Section 5.2.5 later in this chapter). This pH induced pore formation is consistent with the mechanism of action of other related toxins: Schmid *et al.* (1994) demonstrated that the trypsin activated C2II and not the full-length form generates ion-permeable channels in artificial lipid bilayers at pH 6.0 (Schmid *et al.* 1994); Barth *et al.* showed that the cellular uptake of subunit A (C2) required oligomerisation of C2II and acidification of the environment (Barth *et al.* 2000).

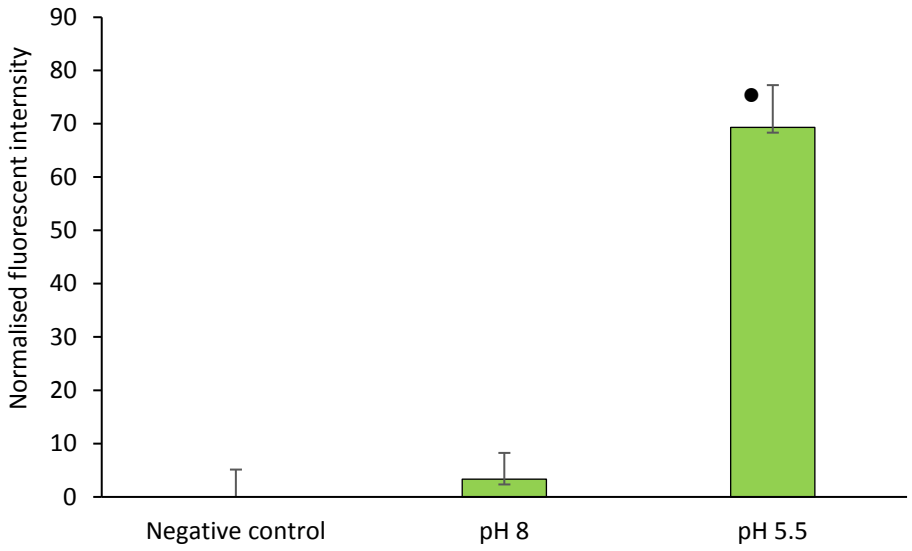

Figure 5.14. YO-PRO-1 uptake assay for Vip1Ac1$_{60}$ at two different pH.
The bars represent the means of normalised fluorescent intensity from 3 replicates and the error bars correspond to the standard deviation. Fluorescence intensity normalised to the fluorescence intensity of the negative control. Black circle indicates significant difference compared to the negative control and sample at pH 8.0 ($p<0.01$) as determined by one-way ANOVA followed by Tukey HSD's test.

As demonstrated earlier, the Vip1Ac1$_{60}$ forms pre-pore oligomers in solution. These pre-pore oligomers are slightly resistant to SDS denaturation conditions when the samples are not boiled but dissociate into monomers on boiling, as shown in Figure 5.1B. SDS resistance was used in studies of PA to assess the transition of pre-pore to pore: the pre-pore oligomeric form of PA converted to the SDS resistant oligomeric pore state under acidic condition (Milnes *et al.* 1994; Sun and Collier 2010). SDS-PAGE analysis of the Vip1Ac1 sample from the mixture at pH 8.0 shows only monomer after boiling due to disassociation of the pre-pore oligomer, Figure 5.15A. Without boiling, a minor protein band of high molecular weight corresponding to the pre-pore oligomer and a major protein band corresponding to the monomer were observed Figure 5.15B. To confirm the intact formation of pore oligomers at pH 5.5, samples from the YO-PRO-1 dye mixture at pH 5.5 were analysed. The sample from the mixture at pH 5.5 under both boiled and unboiled conditions showed a protein band of high molecular weight

corresponding to the intact oligomer and no monomers were observed (Figure 5.15A-B). This suggests the strong association of the pore oligomer after β barrel channel formation.

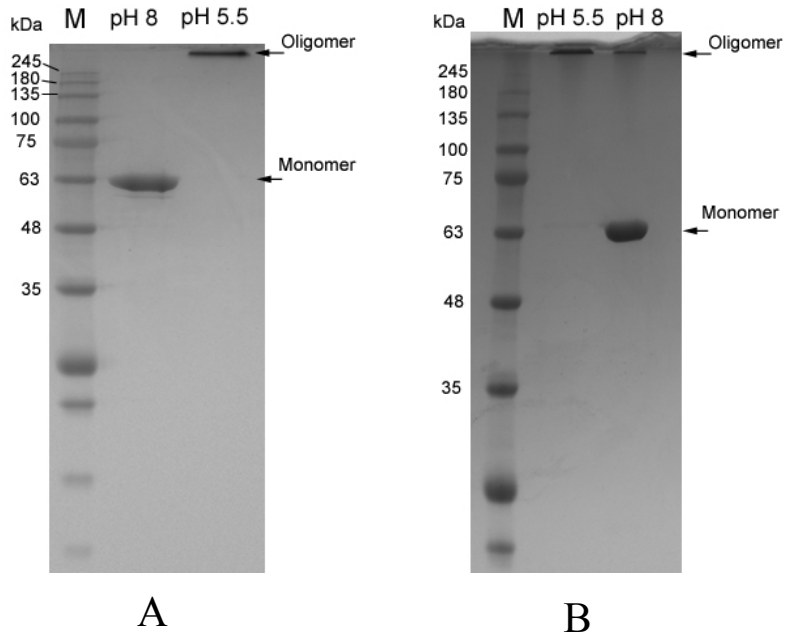

Figure 5.15. SDS and boiling resistance of Vip1Ac1$_{60}$ from lipid mixture.
A. SDS-PAGE (12.5%) analysis on boiling samples for 5 min. **B**. SDS-PAGE (12.5) analysis without boiling samples. Lane M corresponds to the protein marker of known molecular weights.

The ability of full-length Vip1Ac1$_{82}$ and Vip2Ae3 to form a pore in a lipid bilayer was investigated compared with the Vip1Ac1$_{60}$ oligomeric form. Full-length Vip1Ac1$_{82}$, Vip2Ae3 and Vip1Ac1$_{60}$ at a final concentration of 5 μM were used in YO-PRO-1 uptake assays at pH 5.5 (Section 2.8) and the fluorescence intensity was recorded. Based on the fluorescence intensity, samples treated with full-length Vip1Ac1$_{82}$ and Vip2Ae3 (Figure 5.16) showed no significant difference compared to the negative control (without protein), while the Vip1Ac1$_{60}$ showed a significant difference compared to full-length Vip1Ac1$_{82}$ and Vip2Ae3 treated samples and negative control ($p<0.01$) as determined by one-way ANOVA test followed by Tukey HSD's test.

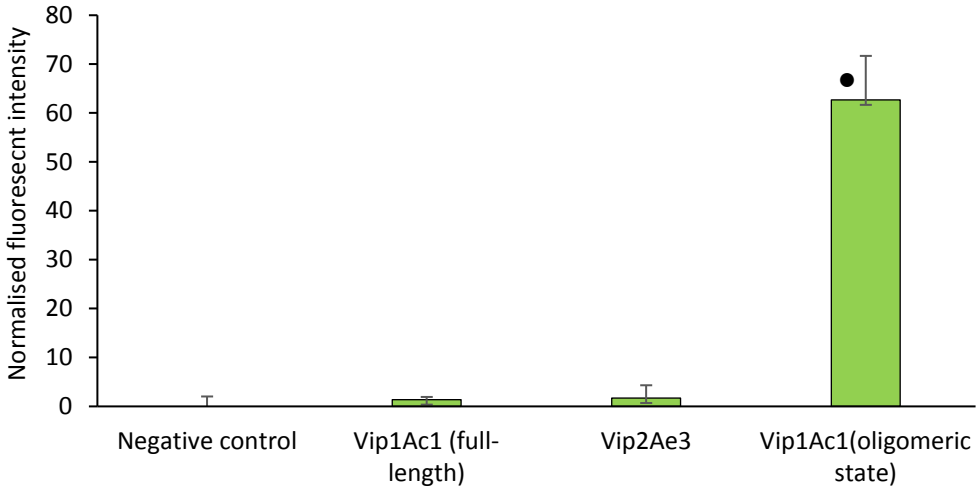

Figure 5.16. Comparison of pore formation in liposomes by Vip toxins.
The bars represent the means of normalised fluorescence intensity from 3 replicates and the error bars correspond to the standard deviation. Fluorescence intensity normalised to the fluorescence intensity of negative control (without protein). Black circle indicates significant difference compared to the negative control and sample at pH 8.0 ($p<0.01$) as determined by one-way ANOVA followed by Tukey HSD's test.

To investigate the effect of Vip2Ae3 on pore formation by the Vip1Ac1$_{60}$ protein, a complex of Vip1Ac1$_{60}$ plus Vip2Ae3, prepared and isolated as described in Section 5.2.3 was used compared to Vip1Ac1$_{60}$ alone in YO-PRO-1 uptake assay at pH 5.5 (Section 2.8). Vip1Ac1$_{60}$ was used at a concentration of approximately 5 μM and a mixture of Vip1Ac1$_{60}$ plus Vip2Ae3 was used with the same approximate concentration of Vip1Ac1$_{60}$ (based on the approximate molar ratio of Vip1Ac1$_{60}$ to Vip2Ae3 of 2.52:1 calculated above from gel scanning of the SEC peak of the complex, Section 5.2.3). The presence of Vip2Ae3 with the pre-pore oligomer increased the fluorescence signal by 20% suggesting that the oligomeric Vip1Ac1$_{60}$ had increased the capacity to form a pore (Figure 5.17A). There was a significant difference compared to oligomeric state only ($p<0.01$) as determined by one-way ANOVA test followed by Tukey HSD's test. The increase of fluorescence intensity and the pore formation activity by the complex (Vip1Ac1$_{60}$ plus Vip2Ae3) could be explained by the fact that the Vip1Ac1$_{60}$ in the complex as pre-pore oligomers may be stabilised by Vip2Ae3 and only well-formed pre-pore oligomers undergo conformational change to form viable pore-forming oligomers.

Recently, Geng *et al.*(2019) reported that the activation of Vip1Ad by trypsin led to formation of oligomers composed of trimers, tetramers, pentamers, hexamers and heptamers based on ultracentrifugation analysis (Geng *et al.* 2019). The complex of Vip1Ac1$_{60}$ plus Vip2Ae3 was not stable in SDS PAGE (without boiling) (Figure 5.17B). This indicates the interaction may be weak and reversible as might be expected for translocation of Vip2Ae3 that would require dissociation from the complex.

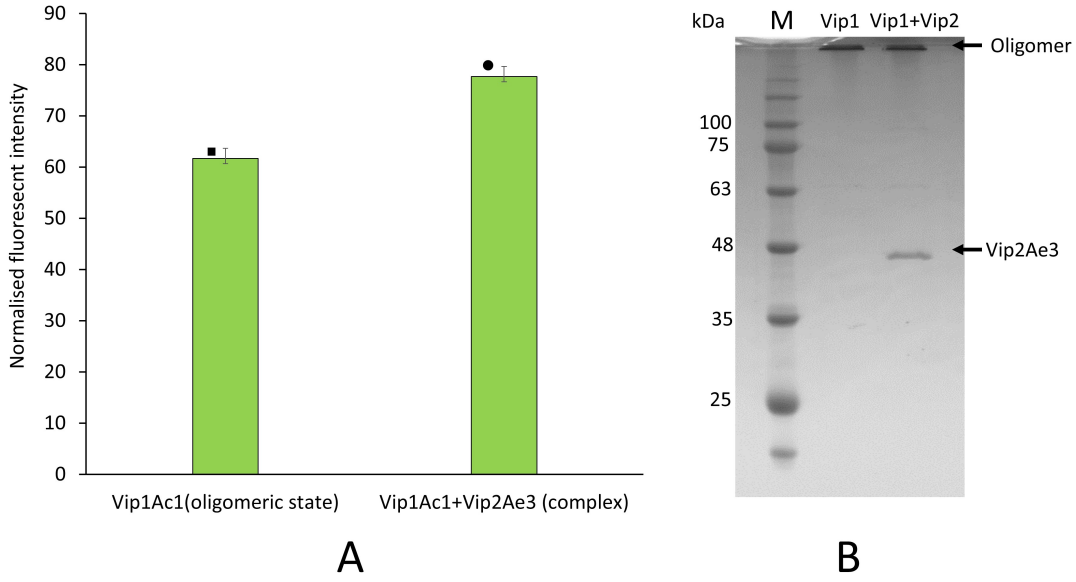

Figure 5.17. Comparison of pore formation in liposomes by Vip1Ac1$_{60}$ and complex of Vip1Ac1$_{60}$ + Vip2Ae3.

A. The bars represent the means of normalised fluorescence intensity from 3 replicates and the error bars correspond to the standard deviation. Fluorescence intensity normalised to the fluorescence intensity of negative control (without protein). Black circles indicate significant differences compared to the negative control and sample at pH 8.0 ($p<0.01$) as determined by one-way ANOVA followed by Tukey HSD's test. **B**. SDS-PAGE (12.5%) analysis of un-boiled samples. Lane Vip1 corresponds to the Vip1Ac1$_{60}$. Lane Vip1+Vip2 corresponds to the Vip1Ac1$_{60}$ plus Vip2Ae3 complex. Lane M corresponds to the protein marker of known molecular weights.

5.2.4 Involvement of domain 3 in oligomerisation

The molecular structure of Vip1Ac1 reveals the 4 domains that generate the final functional structure. Structural similarity of domain 3 between Vip1Ac1 and PA and conservation of the important residue D512 (D509 in Vip1Ac1) that has been found to play a critical role in oligomerisation in PA (Mogridge *et al.* 2001) suggests that domain 3 of Vip1Ac1 has a similar function (oligomerisation and stabilisation).

To investigate the function of domain 3 in Vip1Ac1 and the role of D509 in the oligomerisation process, a single mutation of D509 to alanine was performed by whole vector site directed mutagenesis (Section 2.2.2). The mutant Vip1Ac1^{D509A} was expressed and purified as described for wildtype Vip1Ac1 (Section 3.2.7). The activation and purification of the activated form Vip1Ac1^{D509A} (60 kDa) was carried out as described above (Section 5.2.2). The purified protein after ion exchange purification (Figure 5.18A) was concentrated and subjected to SEC to estimate the molecular weight as described in Section 2.4.10. In contrast to activated wild-type Vip1Ac1$_{60}$, which elutes in SEC as an oligomer, the Vip1Ac1^{D509A} (60 kDa) eluted as a monomer (Figure 5.18B-C) with an estimated molecular weight approximately 59.4 kDa (which is very close the measured molecular weight of the monomeric activated wild-type Vip1Ac1$_{60}$ as measured by mass spectrometry in Section 5.2.1). The Vip1Ac1^{D509A} (60 kDa) protein analysed by SDS PAGE with and without boiling to confirm the monomeric state. Results showed that no oligomer was observed under either condition and only a monomeric protein band appeared on SDS-PAGE gel (Figure 5.18D). Additionally, the sample of Vip1Ac1^{D509A} (60 kDa) was analysed by electron microscopy as described in Section 5.2.5 and no oligomer was observed (results not shown). These results suggest that domain 3 and D509 in Vip1Ac1 has the same function as in PA and are involved in oligomerisation.

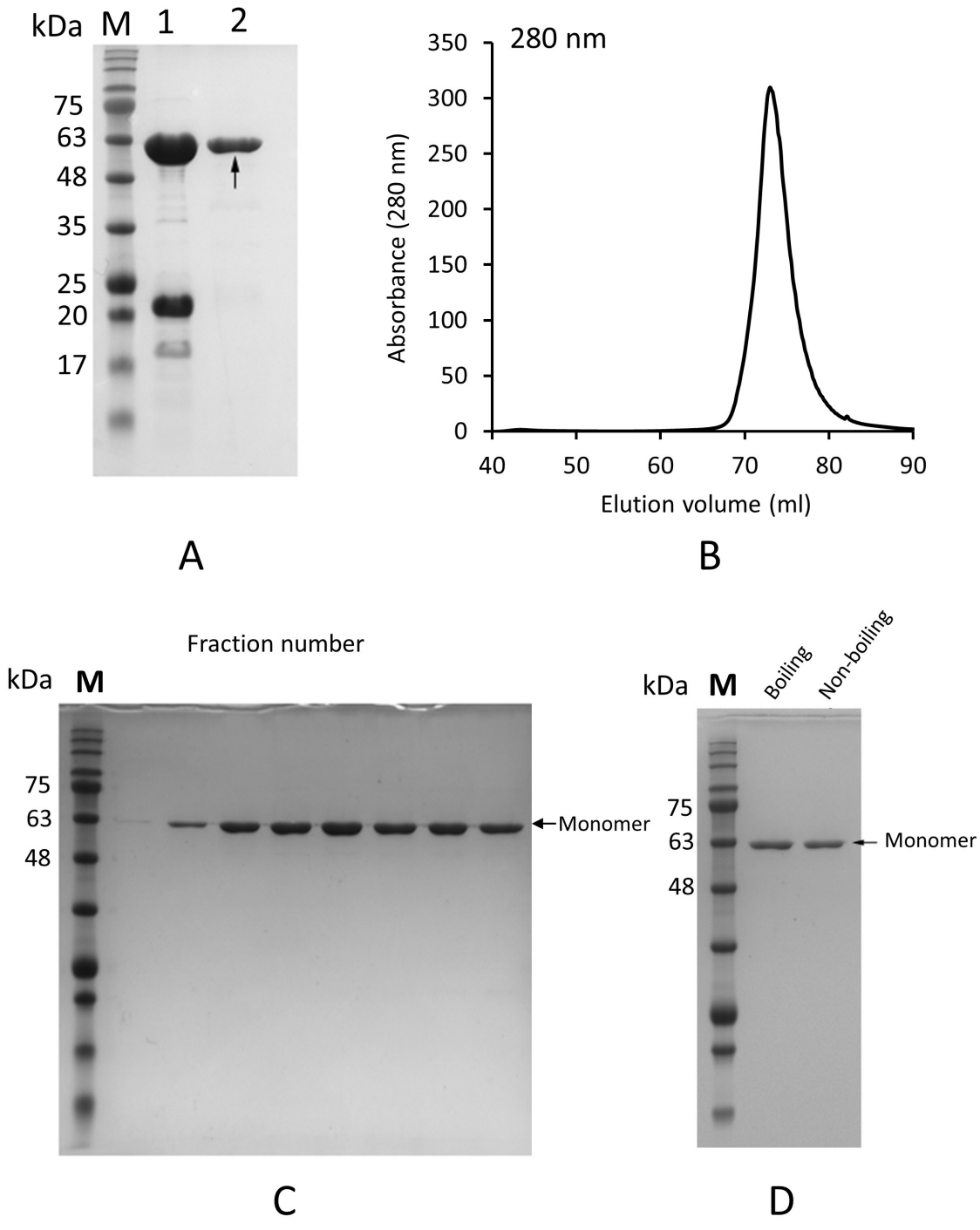

Figure 5.18. Purification and analysis of mutant Vip1Ac1^{D509A}.
A. SDS-PAGE (12.5%) analysis of: lane 1 (mixture of Vip1Ac1$_{60}^{D509A}$ and Vip1Ac1$_{22}$) before purification; and lane 2 purified Vip1Ac1$_{60}^{D509A}$ after purification by ion exchange. **B**. SEC chromatogram showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/600 Superdex™ S200 pg column. **C**.SDS-PAGE (12.5%) analysis of elution fractions from SEC showing the monomer (Vip1Ac1$_{60}^{D509A}$) only. **D**. SDS-PAGE (12.5%) analysis of boiling and un-boiled sample of purified Vip1Ac1$_{60}^{D509A}$. Lane M corresponds to the protein marker of known molecular weights.

5.2.5 Transmission electron microscopy (TEM) studies

Transmission electron microscopy (TEM) is a powerful technique to study large macromolecular complexes and viruses with image processing software (De la Rosa-Trevín *et al.* 2013). In this study, TEM was used to study the oligomeric state of Vip1Ac1 protein. Vip1Ac1$_{82}$ and Vip1Ac1$_{60}$ (0.1 mg/ml in 50 mM Tris-HCl pH 8.0) were used in grid preparation (Section 2.10). Analysis of the TEM images showed that only the trypsin activated form (Vip1Ac1$_{60}$) produced oligomers while Vip1Ac1$_{82}$ did not (Figure 5.19 parts A and B respectively). The ring-shaped oligomers along with what appeared to be partial rings, presumed to be incomplete oligomerisation states, were observed for Vip1Ac1$_{60}$ under these conditions. This indicates that the pre-pore ring shaped oligomer was formed after activation with trypsin at pH 8.0. This is consistent with the findings of Gena *et al.* (2019) who found that the activation of Vip1Ad by trypsin led to formation an range of oligomers with different oligomerisation states (trimers, tetramers, pentamers, hexamers and heptamers (Geng *et al.* 2019)). Similar findings have been reported for PA and the B subunit of *Clostridium* binary toxins. PA has been found to form ring shaped heptameric and octameric oligomers when characterised by TEM (Kintzer *et al.* 2009). The heptameric state of PA was more common and the atomic structure of the PA heptameric pre-pore was solved by X-ray crystallography (Petosa *et al.* 1997). Barth *et al.* (2000) demonstrated that the C2II toxin (subunit B of *Clostridium botulinum* C2 binary toxin) forms a ring-shaped structure after activation in solution and this was characterised by analytical ultracentrifugation showing it was composed of 7 monomeric units (Barth *et al.* 2000).

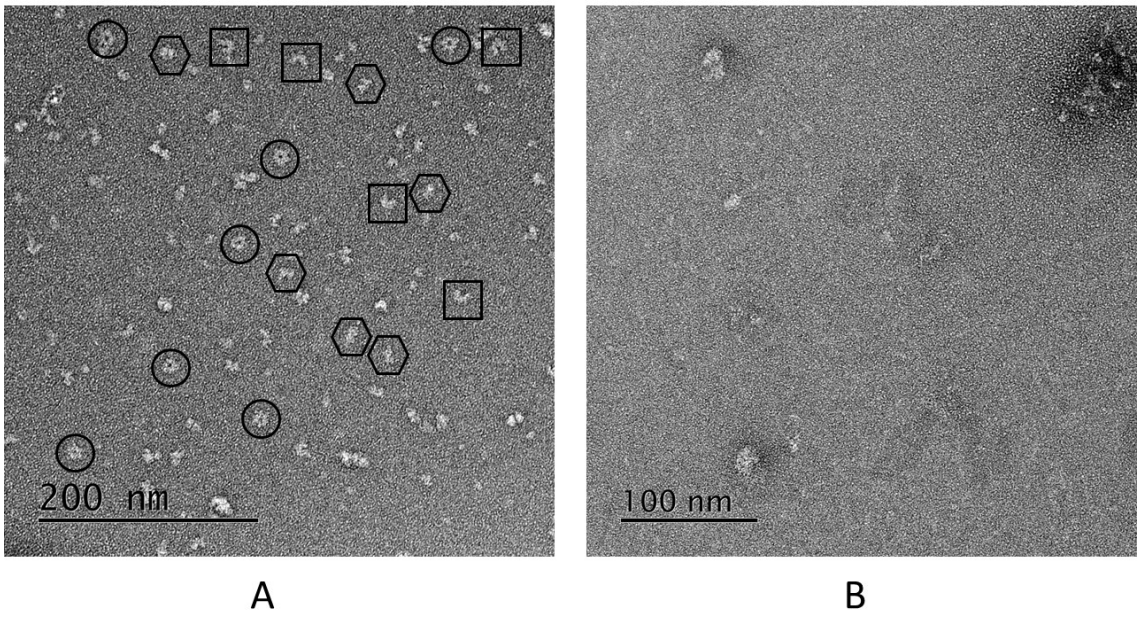

Figure 5.19. Micrograph of negative-stained TEM grid of Vip1Ac1.
A and **B** represent micrographs of negative-stained TEM grids of $Vip1Ac1_{60}$ and $Vip1Ac1_{82}$ respectively. Black circles represent top views of well-formed ring-shaped oligomers and black squares represent putative incomplete oligomerisation state. Black hexagons represent side view of ring-shaped oligomer.

To investigate the conformational change in pre-pore oligomer and β-barrel structure formation, the grids were washed with low pH buffer (based on the method used to induced β-barrel structure formation in PA as reported by Jiang *et al.* (2015), with slight modification). Briefly, after blotting the protein on the grid and before staining, the grid was washed for 1 min with low pH buffer (50 mM sodium phosphate buffer, pH 5.5). Interestingly, a well-formed ring-shaped oligomer composed of 7 units and a stem like structure (which may represent the β-barrel structure) were clearly observed (Figure 5.20A, B and C). The acidic environment appears to induce conformational rearrangement that would allow the putative channel region in domain 2 to form a long β strand in the oligomeric state and generate a 14 β stranded barrel stem in the heptamer. The PA heptameric pore state was solved by cryogenic electron microscopy (CryoEM; Jiang *et al.* 2015) and the conformational change of pre-pore to pore was described as detailed in chapter 4 (Section 4.2.3.5).

TEM images were analysed by imageJ software for measurement and comparison with previous results published for PA (Jiang *et al.* 2015). The pre-pore oligomer of Vip1Ac1 and PA seem to be similar, except the length of the insertion channel. Both pre-pore oligomers of Vip1Ac1 and PA have nearly the same average outer diameter ~14 nm (SD 0.4) and ~13.98 nm (SD 0.5) respectively. Similarly, pore sizes are the same for both: 3.4 nm (SD 0.2) and 3.38 nm (SD 0.19) for Vip1Ac1 and PA respectively. The differences in length of the channel can be seen when comparing the side views of Vip1Ac1 and PA (Figure 5.20B and E). However, the average length of the stem in Vip1Ac1 was 6.7 nm (SD 0.09) while in PA it was 10.17 nm (SD 0.5), with difference 3.45 nm in length (as predicted from the differences in the putative channel region of domain 2, which is shorter in Vip1Ac1 than PA, as outlined in chapter 4 (Section 4.2.3.5).

To study the molecular structure of $Vip1Ac1_{60}$ pre-pore oligomers, crystal trials were set up using two different types of screens: PACT *premier*™ HT-96 and JCSG-plus™ using the sitting drop vapour diffusion method as described in chapter two (2.12.1). The conditions used to grow full-length Vip1Ac1 crystals were also tested using the hanging drop method (Section 4.2.3.2). No crystals were produced under this range of conditions, using a range of protein concentrations (8 mg/l, 6 mg/ml, 4 mg/ml and 2 mg/ml). This could be a result of lack of homogeneity and the presence of $Vip1Ac1_{60}$ in different oligomerisation states.

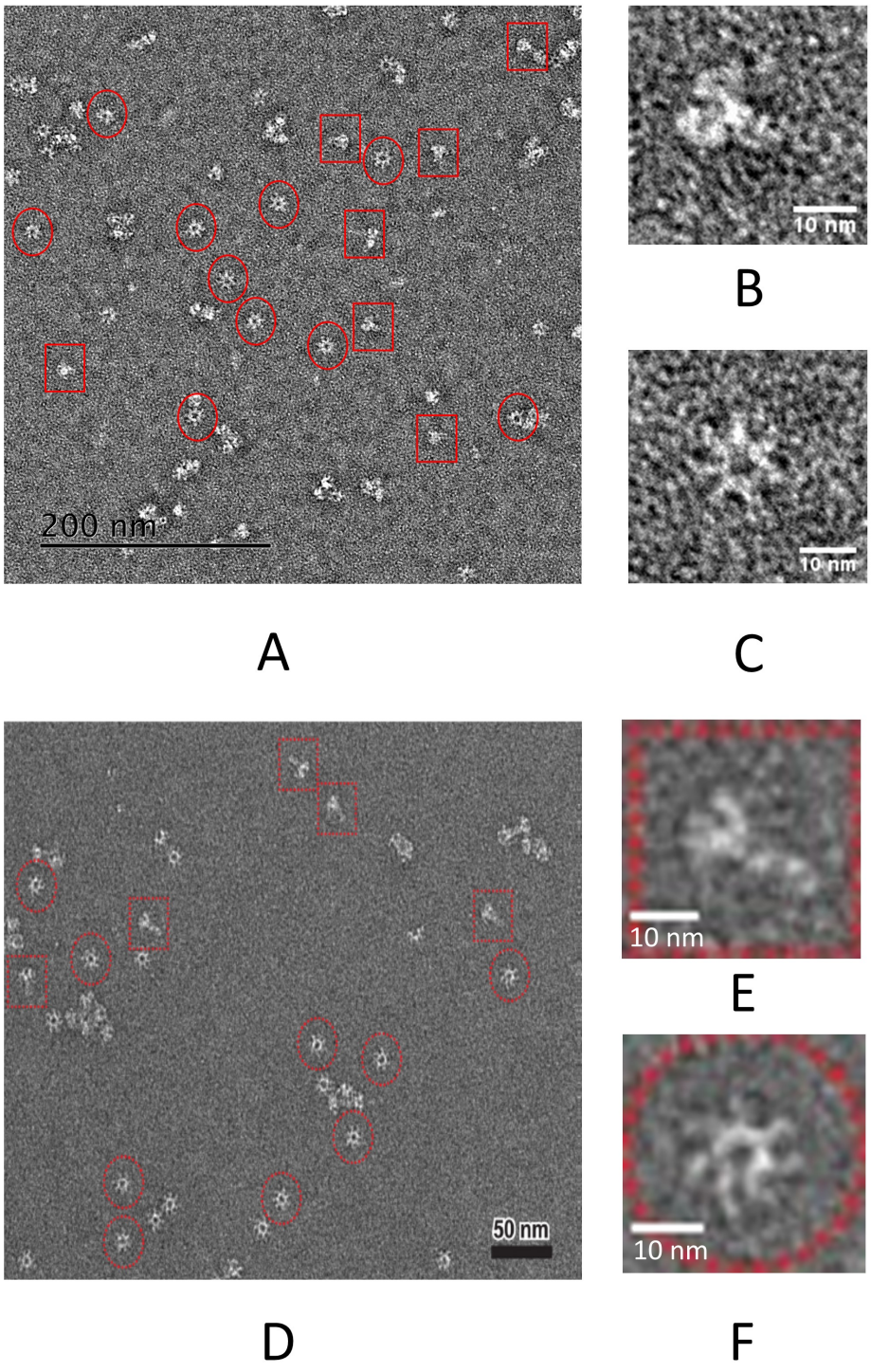

Figure 5.20. Negative stained EM grids of Vip1Ac1$_{60}$ pore compared with PA.
A, B and **C** represent Vip1Ac1$_{60}$. **D**, **E** and **F** represent PA (images taken from Jiang *et al*. 2015). **A** and **D** represent micrographs of negative-stained TEM grids for Vip1Ac1$_{60}$ and PA respectively. Red circles represent top views of particles and red squares represent side views of particles. **B** and **E** represent high magnification of areas from micrographs of TEM grids showing top views of Vip1Ac1 and PA particles respectively. **C** and **F** represent high magnification of areas from micrographs of TEM grids showing side views of Vip1Ac1 and PA particles respectively.

5.2.5 Models and proposal interaction of Vip2Ae3 with the oligomer

The results from chapter 4 and chapter 5 showed high similarity between Vip1Ac1 and PA in terms of structure and general function as pore formation toxins after oligomerisation. Additionally, modelling of Vip1Ac1 with the Swiss model server using PA as a template predicted a structure of Vip1Ac1 that was good for the first 3 domains. X-ray crystallography studies with the oligomeric state of Vip1Ac1 were unsuccessful, so models of pre-pore and pore oligomers of Vip1Ac1 were generated by the Swiss model server using a pre-pore and pore oligomer structures of PA from the PDB.

5.2.5.1 Pre-pore oligomer model of Vip1Ac1

A model of the Vip1Ac1 pre-pore oligomer was generated by the Swiss model server using a PA pre-pore oligomer structure (PDB code; 1TZO) as a template. In general, the pre-pore oligomer model has the same overall folding as the PA pre-pore oligomer, particularly for the first 3 domains (Figure 5.21) and matches the ring-shaped oligomer that was observed in TEM. Modelling of domain 4 is partial (79 residues out of 148 modelled) and this region does not reflect the beta sheet structure of domain 4 in the Vip1Ac1 monomer. This result may be an artefact caused by the greater divergence of this domain as described in chapter 4 (Section 4.2.3.5). Considering the possibility that modelling of domain 4 may have been poor, the monomer structure of Vip1Ac1 (after removal the cleavage part, $Vip1Ac1_{22}$) was aligned with the pre-pore oligomer model (Figure 5.21B).

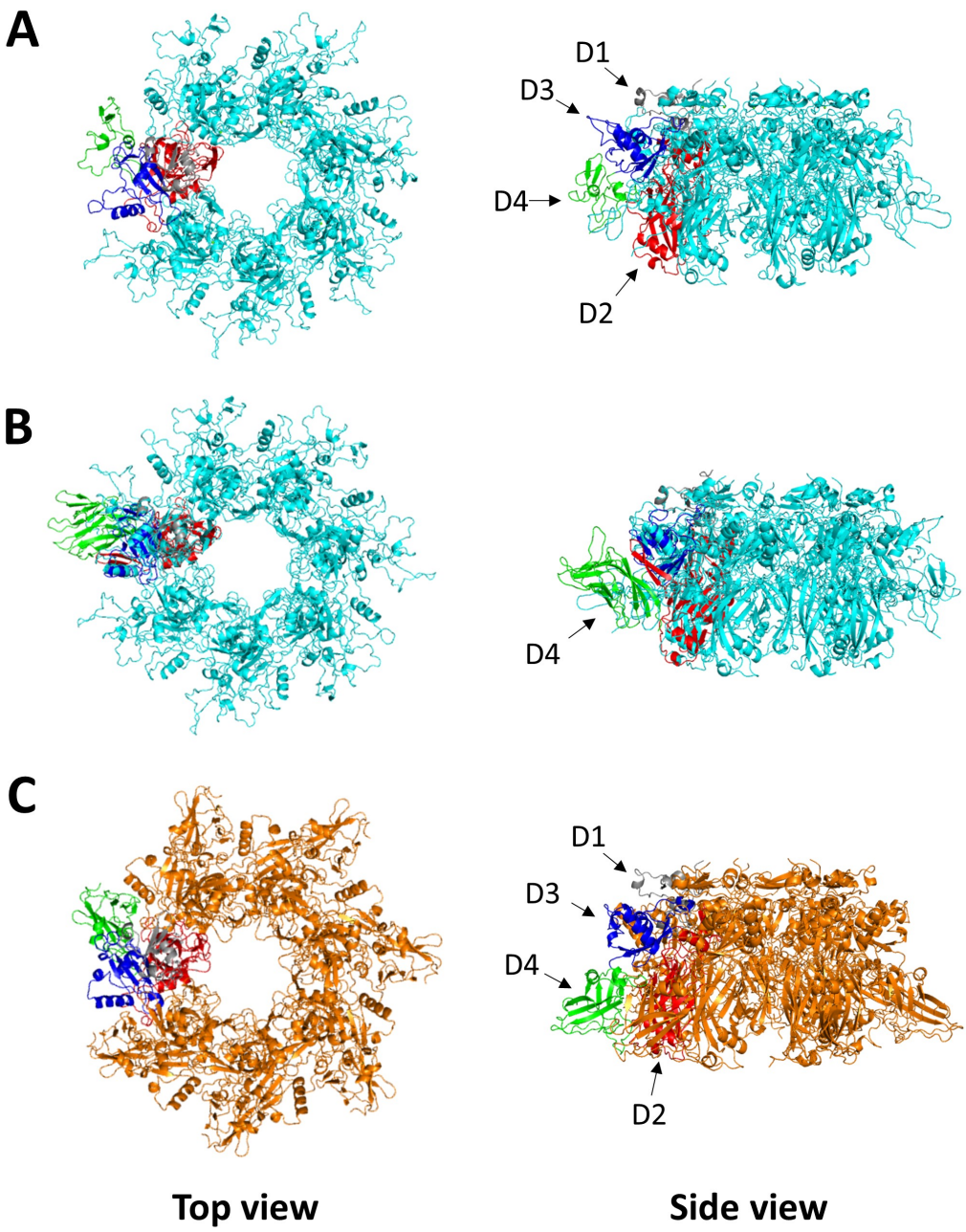

Figure 5.21. Pre-pore oligomer model of Vip1Ac1.
A and **B** are Swiss models of Vip1Ac1 pre-pore oligomer. **A**. The model with the disordered domain 4. **B.** The model with the addition of domain 4 from the monomer structure to shows its potential position. **C**. Pre-pore oligomer structure of PA used as template in Swiss model server (PDB code; 1TZO). Domains marked in the side view as following; domain 1 (D1; grey), domain 2 (D2; red), domain 3(D3; blue) and domain 4 (D4; green).

Based on the model produced by Swiss model, it can be seen that domain 2 is present in the centre of the pore, suggesting its role in pore channel formation after conformational change during transition from pre-pore to pore oligomer under an acidic conditions. Domain 1 and 3 are also present at the top of the oligomer and this suggests that one or both domains may play a role in interaction with Vip2Ae3. Domain 4 in the oligomer has few interactions with the other domains and is placed radially in the structure (Figure 5.22). The separation of domain 4 may be important in allowing it to interact with receptors independently from Vip1Ac1 oligomerisation or transition from pre-pore to pore oligomer.

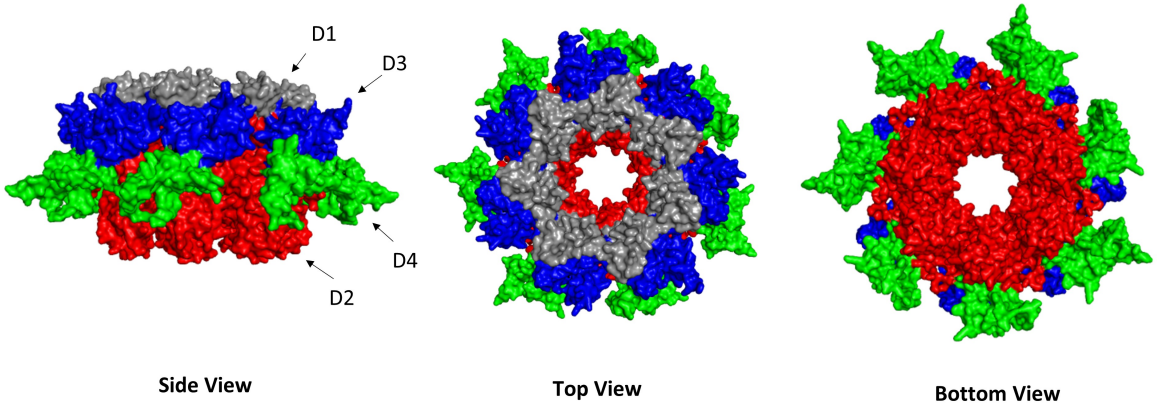

Figure 5.22. Location of Vip1Ac1 domains in the pre-pore oligomer model.
Three different orientations of the Swiss model of the pre-pore, showing the potential position of each domain. Domain 1 (D1; grey), domain 2 (D2; red), domain 3(D3; blue) and domain 4 (D4; green).

The pre-pore oligomeric structure of PA showed that no conformational changes occur during oligomerisation, as the alignment of the structure of the PA_{83} monomer with a monomer of PA_{63} in the oligomer, showed well conserved folding, with no major conformational changes as a result of oligomerisation (Figure 5.23). Based on these results, it appears likely that the active monomeric form of Vip1Ac1 ($Vip1Ac1_{60}$) will have the same general conformation as $Vip1Ac1_{82}$ after removing the cleavage part ($Vip1Ac1_{22}$) by trypsin activation, with no conformational change on oligomerisation. The lack of structural reorganisation on oligomerisation may explain the weak association of the pre-pore oligomer.

Figure 5.23. Structural alignment of the monomer and pre-pore monomeric unit of PA.
The monomer structure of full-length PA (PA_{83}) coloured grey (PDB code; 4H2A), the activated monomeric unit (PA_{63}) in the pre-pore oligomer coloured blue (PDB code; 1TZO).

5.2.5.2 Pore oligomer model of Vip1Ac1

The molecular structure of the PA pore oligomer has been solved using CryoEM by Jiang *et al.*(2015) at 2.9 Å resolution (PDB code 3J9C ; Jiang *et al.* 2015). However, the structure is missing electron density for domain 4 (due to its flexibility) and upper part of the β-barrel that runs from the corolla into the stem (forming the β-barrel channel) (Figure 5.24). As the only structure of PA pore oligomer available in the protein databank, it was used as a template to generate a model for Vip1Ac1 pore oligomer, without domain 4, using the Swiss model server with expectation that some disordering around the missing electron density may occur. However, the importance of this model is the prediction of conformational changes in domains 1-3 and formation the β-barrel as it is predicted that domain 4 has no role in the transition from pre-pore to pore oligomer.

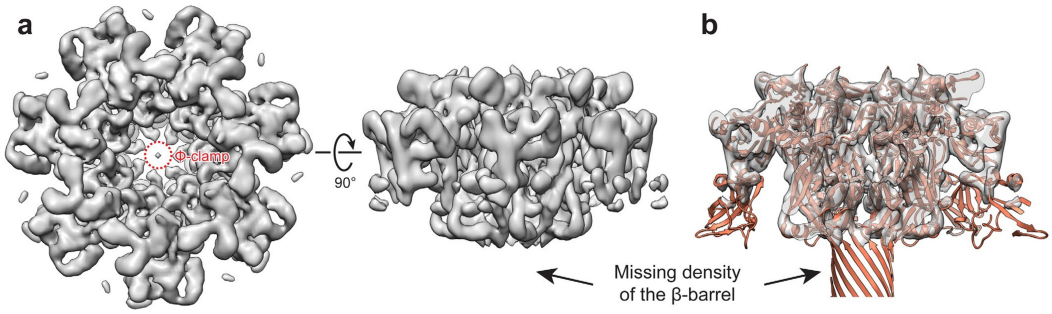

Figure 5.24. CryoEM structure of pore oligomer of PA.
The CryoEM map of PA showing the missing β-barrel and part of domain 4. **a**. Surface top view. **b** side view. Figure was taken from Jiang *et al.* (2015).

Regions defined by Jiang *et al.* (2015) as the corolla, calix and stem of the pore can be seen in the model (Figure 5.25). The corolla is composed of domain 3, part of domain 2 (and may include domain 4 but which missing from the modelling). The calyx is composed mainly of domain 2. The model shows a disordered region that consists of domain 1 and part of domain 2 (as expected, in the region corresponding to missing electron density for the PA template) and a β-barrel channel, which is composed of domain 2 only (Figure 5.25). The top view of the model has overall the same shape as the Vip1Ac1 pore form oligomer that was observed in TEM (Figure 5.25C). Based on the model, the putative channel formation region in domain 2 of the monomeric unit undergoes conformational changes and rearrangement to form the β-barrel channel with adjacent monomeric units in the oligomer (Figure 5.25A). The part of the disordered region that corresponds to the remaining part of domain 1 after activation, is separate from the rest of the structure (Figure 5.25A) but would be expected to form a part of the corolla. In the pore oligomer of PA, both domains 1 and 3 remain in the equivalent positions to those in the pre-pore, without significant conformational changes after transition to the pore oligomer (Jiang *et al.* 2015).

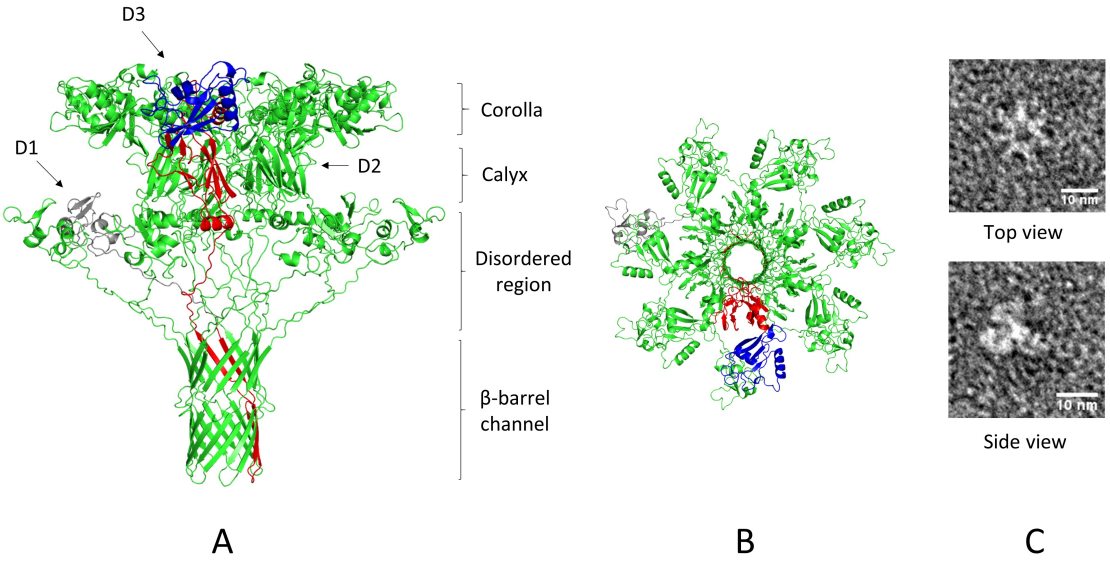

Figure 5.25. Pore oligomer model of Vip1Ac1.
A and **B** are the Swiss model of Vip1Ac1 pore oligomer side and top view respectively. A. The side view with regions marked and domains labelled; domain 1 (D1; grey), domain 2 (D2; red), domain 3 (D3; blue). **C**. The top and side view of the pore oligomer vip1Ac1 observed in TEM.

It is hard to predict the starting and ending point of the rearrangement of the putative channel region in domain 2, especially with the disorder produced by the lack of the electron density in the template but a region involved in β-barrel formation (residues from N291 to E335) can be seen. This is consistent with the prediction of Leuber *et al.* (Leuber *et al.* 2006) that residues from amino acid N308 to T333 (residues numbering after removing the putative signal sequence) could be involved in putative channel formation. The model showed that during the transition from pre-pore to pore oligomer, domain 2 undergoes rearrangements and integrates together with other subunits to generate a β-barrel channel (Figure 5.25A) and this explains the resistance of the pore oligomer to boiling and SDS denaturation conditions.

5.2.5.3 Proposed mechanism of Vip1Ac1 and Vip2Ae3 interaction.

Based on the findings in this chapter, the oligomeric form of Vip1Ac1 consists of 7 monomeric units and interacts with Vip2Ae3 at a ratio of 1:3. It is possible to propose a model for this interaction. Activation of $Vip1Ac1_{82}$ removes the $Vip1Ac1_{22}$ cleavage part to produce $Vip1Ac1_{60}$, which may expose an interaction region that can be occupied by the N-domain of Vip2Ae3. As mentioned in chapter 4, Vip2 consists of two structurally homologous domains; the C-domain (enzymatic domain) and the N-domain (the interaction domain with Vip1). Interaction of the N-domain of Vip2Ae3 with one monomer of $Vip1Ac1_{60}$ in the heptameric pore may place the C-domain in position to mask the interaction region of the neighbouring monomeric unit of $Vip1Ac1_{60}$, thus blocking this subunit from further Vip2Ae3 binding. This model would explain why the oligomeric $Vip1Ac1_{60}$ is capable of interacting with a maximum of 3 Vip2Ae3 proteins (Figure 5.26).

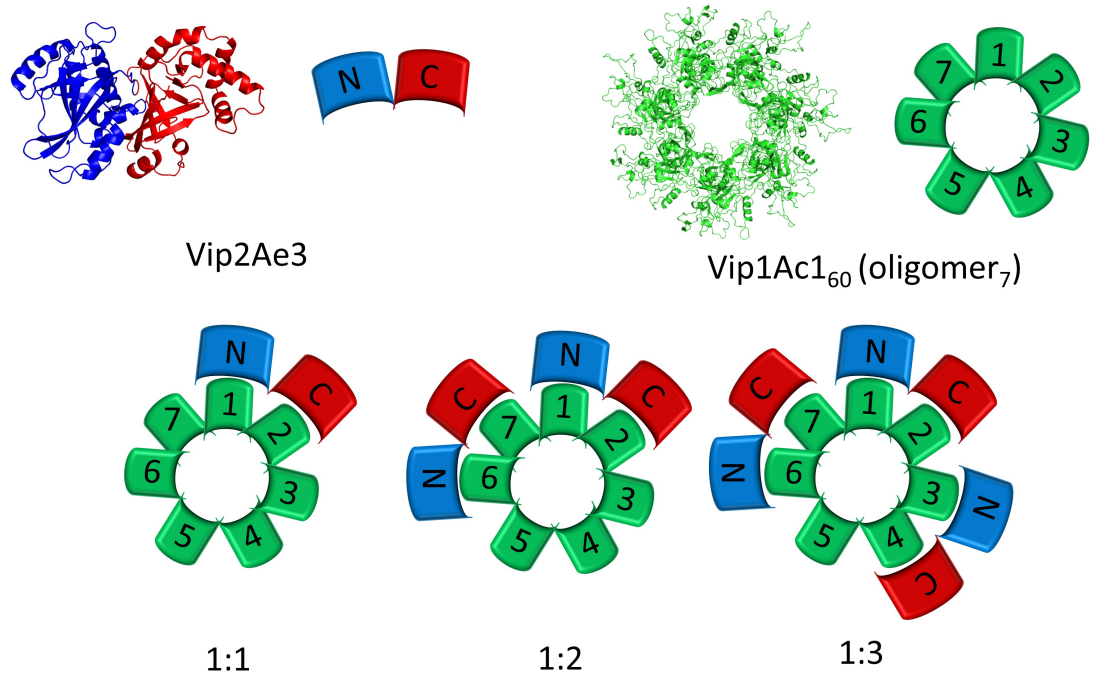

Figure 5.26. Proposed mechanism of Vip1Ac1 and Vip2Ae3 interaction.
The top panel shows Vip2Ae3 with domains coloured blue (N-domain) and red (C-domain) and a heptameric pre-pore model of Vip1Ac1. Cartoon and schematic representations are shown for both proteins. The bottom panel shows the proposed interaction of Vip2Ae3 with the $Vip1Ac1_{60}$ oligomer.

5.2.6 Conclusion

Vip1Ac1$_{82}$ and Vip2Ae3 as AB binary toxins share many properties with other AB binary toxins in terms of activation, oligomerisation, interaction between the two components, and pore formation in an acidic environment via oligomerisation of the B subunit. Vip1Ac1$_{82}$ was activated by trypsin in solution at pH 8.0 to remove an ~22 kDa N-terminal part (Vip1Ac1$_{22}$) by cleavage at a surface loop between amino acids R190 and D191. The remaining part ~60 kDa, represents the activated form (Vip1Ac1$_{60}$) of the protein. Activation is essential for self-assembly of Vip1Ac1$_{60}$ to a pre-pore oligomer, which is labile to boiling and SDS denaturation conditions due to weak interactions. The oligomer shows binding affinity to artificial lipid bilayers at pH 8.0. Moreover, pre-pore oligomer also acts as a receptor for Vip2Ae3. Vip2Ae3 appears to interact with the oligomeric state of Vip1Ac1$_{60}$ via its N-domain, as the presence of an N-terminal His-tag in Vip2Ae3 abolished its interaction with the Vip1Ac1$_{60}$ oligomer. At least 3 Vip2Ae3 monomers seems to be able to bind to the Vip1Ac1$_{60}$ oligomer.

An acidic environment induces conformational change of pre-pore state Vip1Ac1$_{60}$ to form the pore state and is predicted to generate a 14-β-barrel stem derived from 7 monomers in the heptamer, consistent with the pore formation experiments, TEM studies and models. The presence of Vip2Ae3 did not affect the conformational change of the pre-pore, which indicates the interaction of Vip2Ae3 occurs away from the putative channel formation region in domain 2 of Vip1Ac1$_{60}$. Characterisation of the oligomeric pore of Vip1Ac1$_{60}$ by TEM showed high similarities with the pore oligomer of PA in terms of the shape and size, except the length of the β barrel stem, which is shorter in Vip1Ac1$_{60}$. In general, as mentioned in the previous chapter, Vip1Ac1$_{82}$ and PA as monomers share high similarities in terms of structure. In this chapter, more similarities emerged in terms of sites of activation; self-assembly to pre-pore and function of domain 3 in oligomerisation; transition of heptameric pre-pore state to heptameric pore state in an acidic environment; ring-shaped oligomer with similar average outer diameter and pore size; and pore formation in artificial lipid bilayers. As a conclusion based on these similarities, Vip1Ac1 and PA could have the same mode of action except that they target different cells and the enzymatic components that they translocate have different activities.

Chapter six

Engineering of the yellow fluorescent protein, Venus, with non-canonical amino acids

6. Engineering of the yellow fluorescent protein, Venus, with non-canonical amino acids

6.1 Introduction

GFP and GFP-derived fluorescent proteins have been extensively engineered for different purposes ranging from improved folding properties to generating new colour variants based on the fact that the fluorescent properties of the inherent chromophore is highly sensitive to the local environment (more information about GFP engineering can be found in chapter one Section 1.5.2.4). Venus is a yellow version of the *Aequorea victoria* green fluorescent protein (avGFP) that was generated in 1996 by Ormo *et al.* via mutation of the residues S65G/V68L/S72A/T203Y, which shifted excitation/emission wavelengths from 395/509 nm to 513/527 nm, followed by further improvements by other researchers as mentioned in chapter one (Ormo *et al.* 1996; Nagai *et al.* 2002). In general, the molecular structure of Venus is similar to the original avGFP and other GFP-derived fluorescent proteins. It differs from avGFP by 9 residues (F46L, F64L, S65G, V86L, S72A, M153T, V163A, S175G and T203Y).

To incorporate new physicochemical properties into GFP and its derivatives, non-canonical amino acids (ncAAs) have been used extensively by the Jones group (Reddington *et al.*2012; Reddington *et al.*2015; Zaki *et al.* 2018) and others (Wandrey *et al.* 2016;Villa *et al.* 2017; Baumann *et al.* 2018). The use of ncAAs allows researchers to introduce new useful chemistry at defined residues (through codon reprogramming - see Section 1.7), which successfully generates new variants with useful properties. One of the most important variants in terms of changing the physicochemical properties of GFP was switching on and off fluorescence (Reddington *et al.* 2013). A single mutation of T66 (a core chromophore residue) to the ncAA *p*-azido-L-phenylalanine (azF; Figure 1.14) changed the chromophore environment by introducing a photosensitive phenyl azide group. The photochemistry of phenyl azides is based on the ability to produce a single reactive radical nitrene after activation by UV-visible light that can take various reaction pathways depending on the local environment (Figure 1.17 and shown again here in Figure 6.1 for reference). The result was that the protein was non-fluorescent until irradiated with UV-light that reduced the azide to an amine and allowed the protein to fluoresce (a ‘switch on’ approach). In contrast, a single mutation of F145 residue to azF

showed the opposite effect: the F145azF florescent protein was non-fluorescent after irradiation with UV-light (a 'switch off' approach) as a result of crosslinking of the azide group to the chromophore causing disruption of the chromophore system.

Incorporation of ncAA using stop codon suppression technology has already been addressed earlier in chapter one (Section 1.7) and allows researchers to introduce new chemistry in a specific position in a protein. This, in turn, improves our fundamental understanding a protein's molecular properties and allows generation a new variant with novel useful characteristics. This chapter focuses on using the amber codon suppression approach (reprogramming the TAG codon) to genetically encode of ncAA and engineer the yellow fluorescent protein Venus to generate new useful variants. This chapter, thus, describes the generation of TAG stop codon mutations, incorporation of ncAA, production and purification of proteins containing ncAA, characterisation of the new variants' properties and the structural analysis of one variant (Venus^{66azF}).

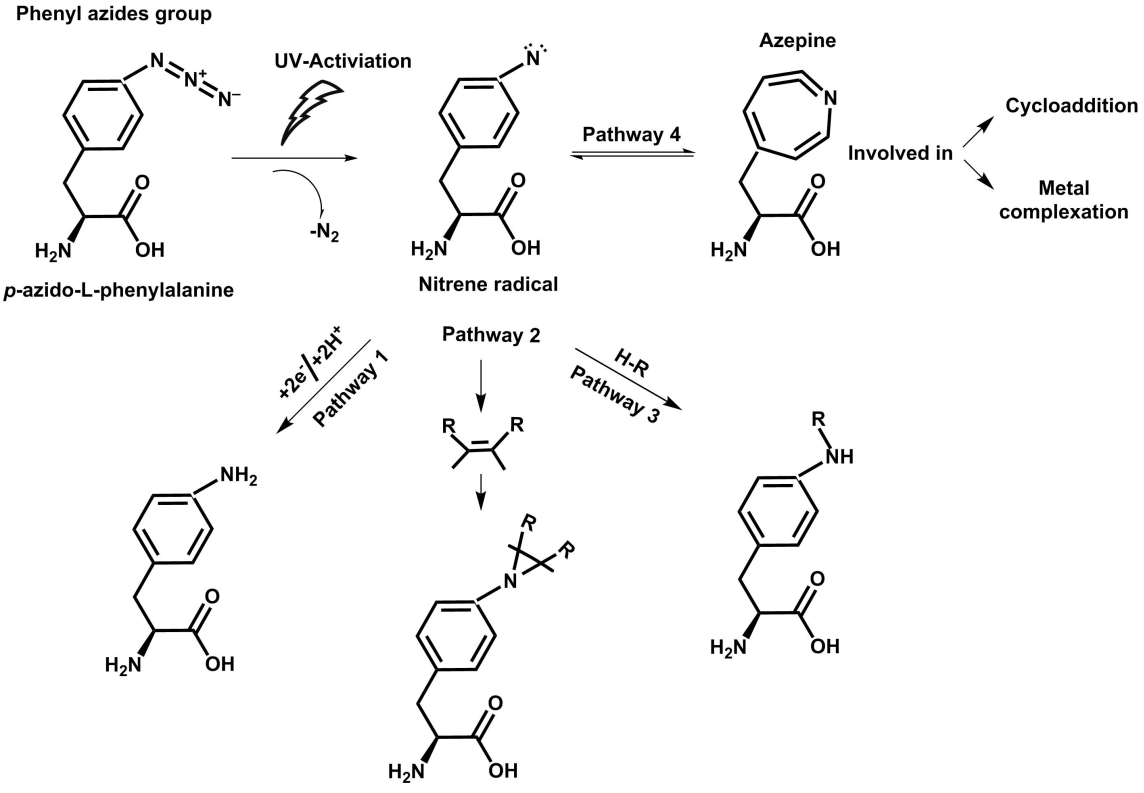

Figure 6.1. Photoreactivity of phenyl azides.

6.2 Results and Discussion

6.2.1 Selection of residues for ncAA incorporation.

In this study, five residues have been chosen to replacement with an ncAA: Y66, Y145, H148, Y203 and Q204. The position of theses residues with respect to the chromophore (CRO) is shown in Figure 6.2. Four of the five different residues (66, 145, 148 and 203) were chosen based on the previous results reported by Reddington (2013) which showed that these residues have a significant effect on sfGFP properties (Reddington 2013). Similar effects may be seen in Venus, a protein backbone with a further red shifted spectrum compared to sfGFP. Venus will have the advantage that there may be less interference from background cell excitation and cells can be exposed to lower energy during *in vivo* use. Furthermore, Y66 is a core residue of the chromophore that has been targeted by others to alter GFP's properties. The other residues (H148 and Y145) lie close in space to the CRO, with H148 involved in hydrogen bond network to Y66. Y203 is an important deviation away from GFP as it plays a key role in the yellow fluorescent properties of Venus and other YFPs. Y203 pi-stacks on top of the phenol group of the chromophore so reducing the energy required to excite the CRO, resulting in the red shift. By placing a new aromatic ncAA (e.g. azF) at this position, it is predicted that there should be changes the inherent fluorescent properties of Venus.

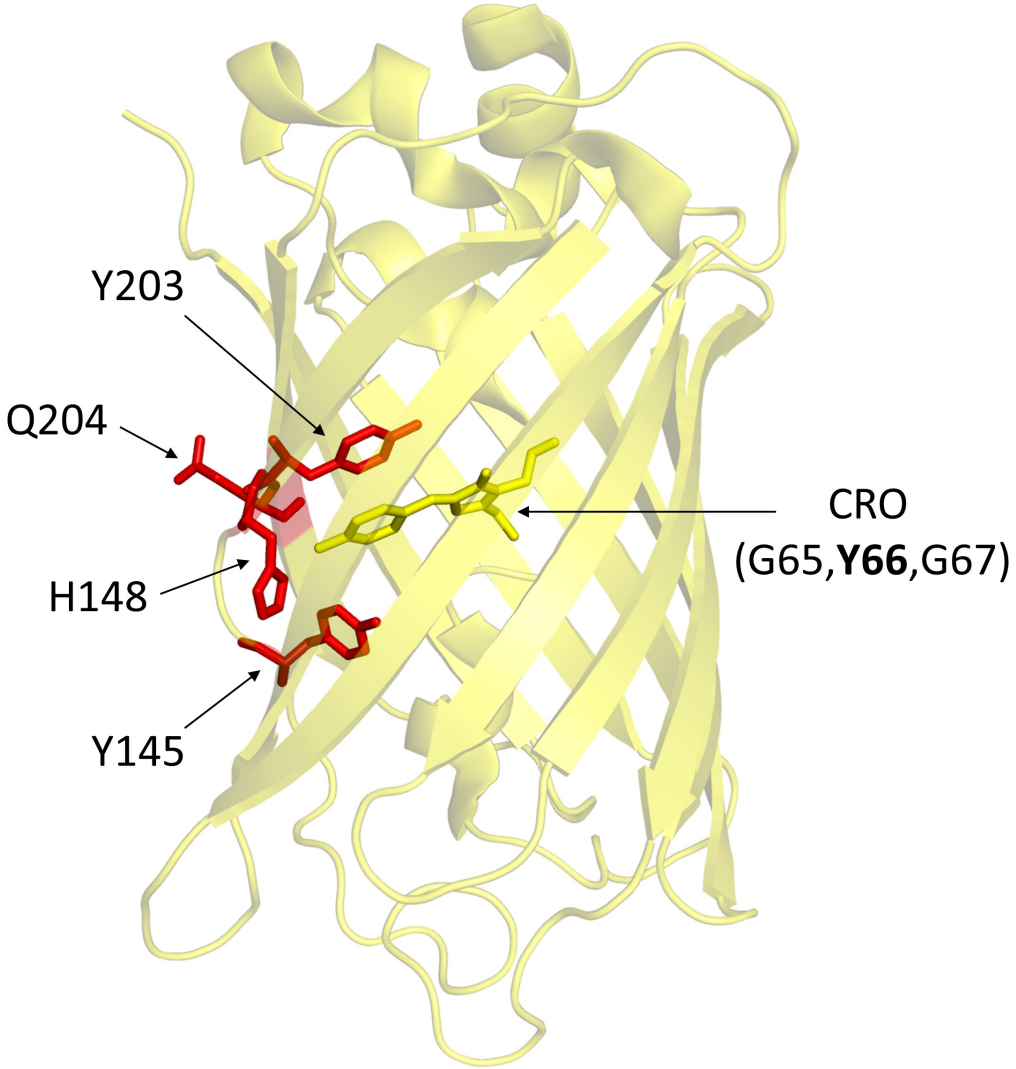

Figure 6.2. Cartoon representation of Venus structure showing targeted positions for ncAA incorporation.
A yellow central residue corresponds to the CRO, which incorporates Y66. Red coloured residues correspond to Y145, H148, Y203 and Q204.

6.2.2 Site directed mutagenesis to introduce required mutations

As mentioned in Section 1.7, incorporation of ncAA requires reprogramming the genetic code. As the amber stop codon (TAG) is the least used codon and has proved the most successful for genetic code expansion for ncAA incorporation, it was, thus, used in this study. TAG codons were generated in the *Venus* gene at the designed positions (66TAG, 145TAG, 148TAG, 203TAG and 204TAG, where the number represents the residue number in Venus) by whole vector site-directed mutagenesis (SDM) as described

in Section 2.2.9.3. The plasmid vector containing the *Venus* gene was amplified by PCR using a pair of oligonucleotide primers containing the TAG mutation (Table 2.3).

Analysis of PCR products by agarose gel electrophoresis showed a DNA band of ~5.8 kb that corresponded to the original size of pBAD-Venus. The linear DNA products were phosphorylated and ligated to recircularise the vector as described in Sections 2.2.7 and 2.2.8 respectively. Re-circularised vector was transformed into a Top10^{TM} electrocompetent *E. coli* as described in Section 2.2.10.2 and subsequently plated on LB agar supplemented with ampicillin for vector selection and 1% arabinose for TAG mutant selection. Yellow fluorescent colonies corresponded to the wild-type Venus so would not contain the premature stop codon while non-fluorescent colonies might contain the TAG mutant so were selected for gene sequencing. Introduction of TAG mutants in *Venus* gene are illustrated in Figure 6.3.

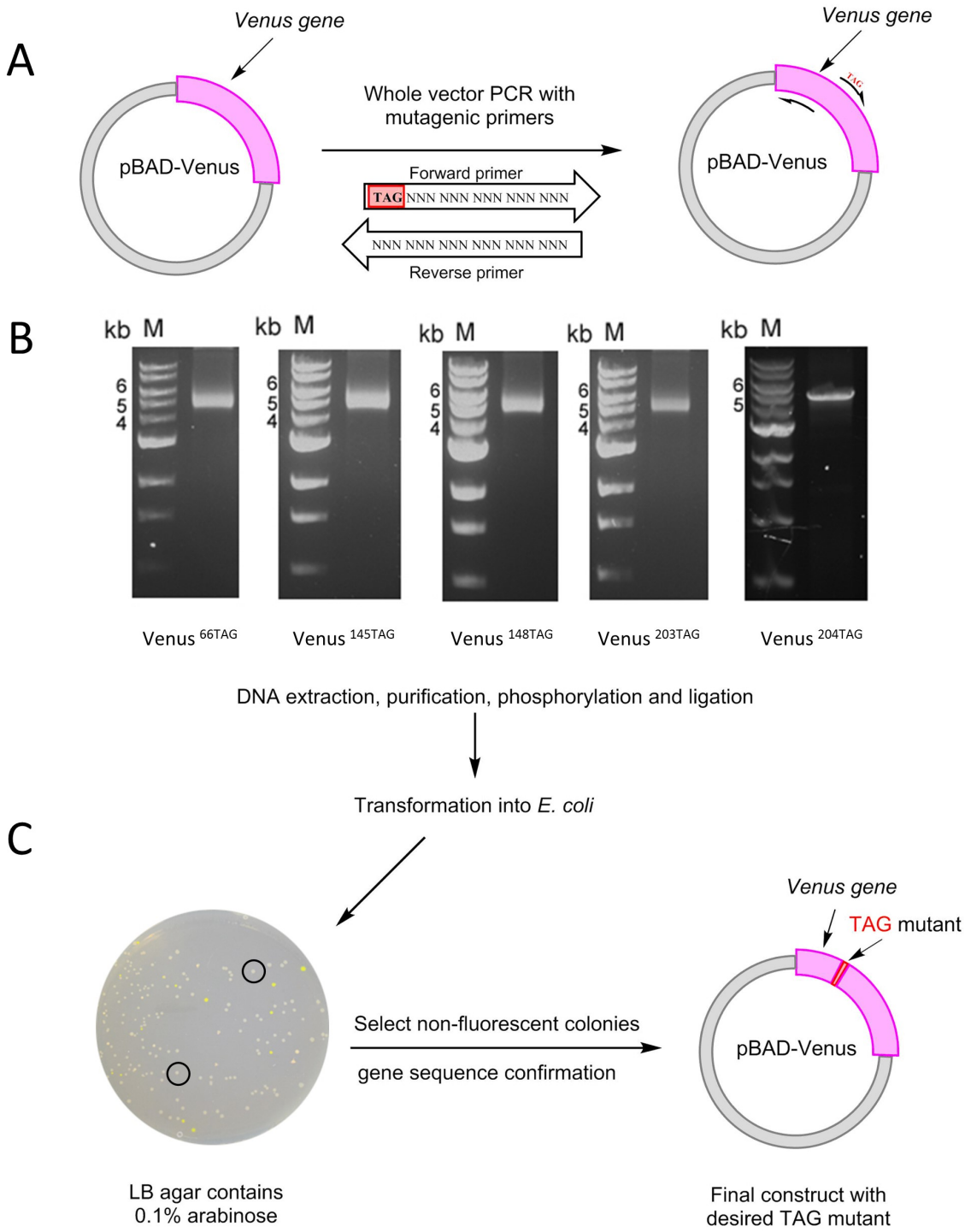

Figure 6.3. Introduction of TAG mutants in Venus by site-directed mutagenesis.
A) Schematic showing PCR with TAG mutagenic primers of interest. **B)** Agarose (1%) gel electrophoretic analysis of PCR products of pBAD-Venus (~5.8 kb) to introduce TAG codons encoding residues 66, 145, 148, 203 and 204, run against 1 kb DNA marker (M). **C)** Schematic showing transformation of purified PCR products into Top10^{TM} electrocompetent E. *coli*. Yellow fluorescent colonies were neglected (non-TAG mutant colonies) and non-fluorescent colonies were selected for gene sequence analysis.

6.2.3 Incorporation of ncAA into proteins

The requirements for incorporation of ncAAs were described previously for Vip1Ac1 (Section 3.2.8.2), the expression system needs two vectors: the first containing the mutated gene of interest resident in the pBAD backbone and a second vector, termed pDULE carrying the engineered tRNA/aaRS pair for the desired ncAA. This was achieved by transforming by electroporation electrocompetent *E. coli* Top10^{TM} with the two vectors (see Figure 2.1 for outline vector maps and Table 2.2).

Incorporation of a ncAA, specifically azF was assessed as described in Section 2.3.4 using a "plus-negative" approach that utilises the original role of the amber stop codon. In the absence of the ncAA from the growth medium, truncated protein will be produced; in the presence of the ncAA, full length protein will be produced. Results showed that full-length protein was produced only in the presence of ncAA, indicating successful read-through of the amber stop codon. In the absence of azF, full-length protein was not produced as the amber stop codon was read as a stop signal to terminate the translation process as shown in Figure 6.4. A general issue with some single residue mutations is the potential impact on protein structure that may result in partially/fully unfolded protein and lead the protein to accumulate as insoluble inclusion bodies. The same could be true when using ncAAs. For example, mutation at residue 222 in sfGFP with azF (E222azF) caused incorrect folding of the protein and expression as inclusion bodies (Reddington 2013). Here, 3 of the 5 chosen residues are either fully (e.g. Q204) or partially (e.g. F145 and H148) surface exposed; the remaining residues (Y66 and Y203) are in the core of the protein, so could affect folding. To test this, the soluble protein fraction was assessed for the presence of the protein compared to the whole cell lysates. In all cases, the azF-containing proteins were successfully produced in the soluble form, indicating that the incorporation position of ncAA has not had a negative effect of protein structure. All 5 Venus containing azF (Venus ^{azF}) variants were thus taken forward for purification and characterisation.

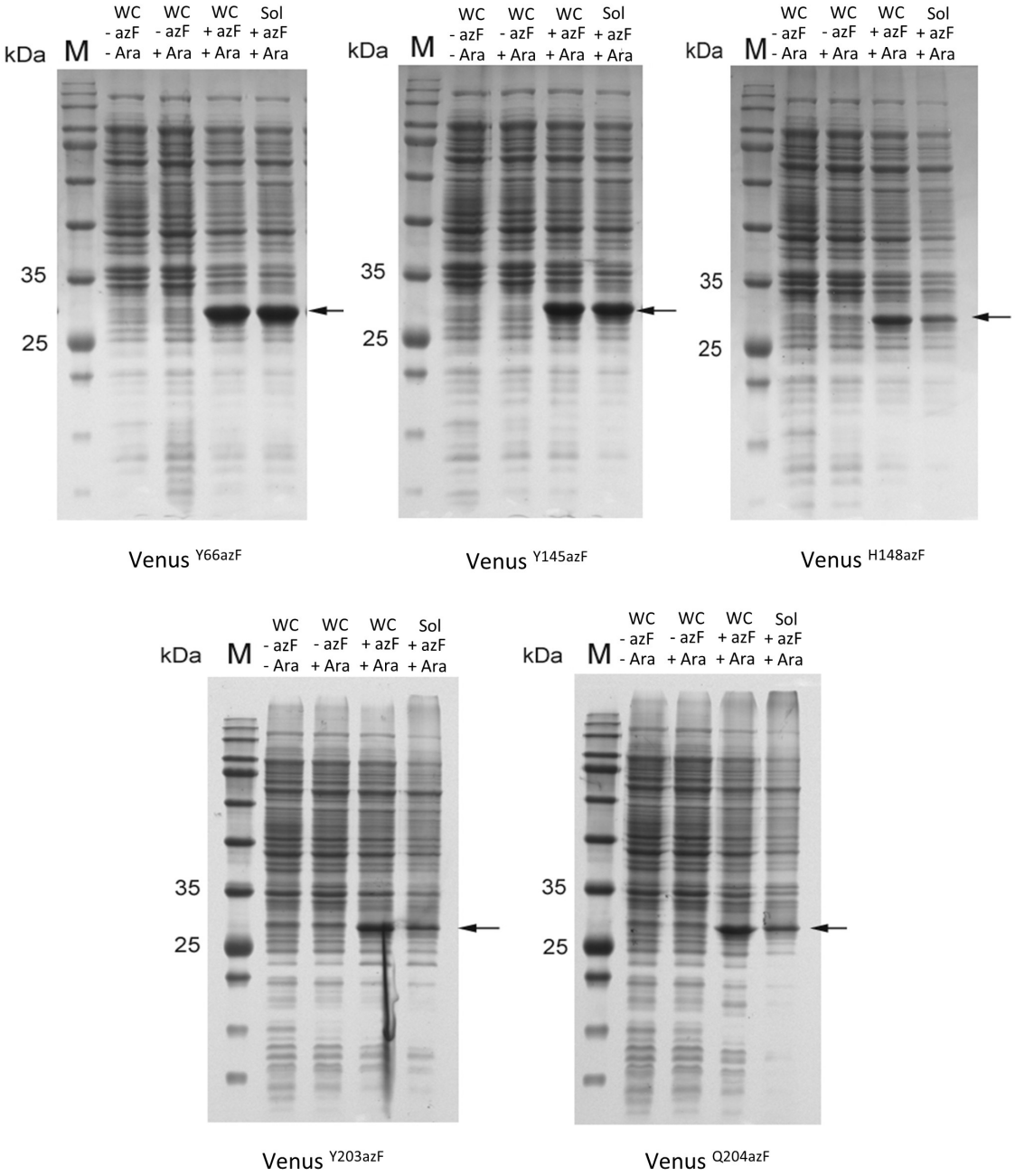

Figure 6.4. Incorporation of azF into Venus.
SDS-PAGE (12.5%) analysis of incorporating azF at position 66, 145, 148, 203 and 204. Lane M corresponds to the protein marker of known molecular weights. WC corresponds to whole cell lysate and Sol corresponds to soluble whole cell lysate. -azF corresponds to absence of azF in the culture. +azF corresponds to presence of azF in the culture. -Ara corresponds to absence of arabinose (1%) in the culture. +Ara corresponds to presence of arabinose (1%) in the culture. Black arrows indicate the Venus variants.

6.2.4 Expression and purification of proteins containing ncAA

To study the effect of incorporation of ncAA on Venus' properties, pure protein samples were prepared. All Venus variants, including wild-type protein, where prepared as described in Sections 2.3.4.2 and 2.3.2 respectively, and purified in two steps. The first step involved nickel affinity chromatography (Section 2.3.8.1). The presence of an N-terminal polyhistidine-tag in Venus allows the protein to bind onto the nickel column. After cell lysis by French Press (Section 2.3.6.3), the His tagged protein was purified from the cell lysate as described in Section 2.3.8.1. Three elution fractions were collected from the Protino^{R} Ni-TED 2000 Packed affinity Columns (Figure 6.5A) and analysis by SDS-PAGE showed that high purity was achieved. In some variants, some protein eluted with the wash fraction and flow-through fraction as shown in Figure 6.5A and Figure 6.6; this is likely due to the column already being saturated with bound Venus protein. Protein present in the flow-through fraction was re-purified by a second application to the Ni column. Pure protein fractions were pooled and subjected to a second step of purification by size exclusion chromatography (SEC).

Purified protein from nickel affinity purification was concentrated to 2 ml and loaded into a Hiload^{TM} 16/600 Superdex™ S75 pg as described in Section 2.3.8.3. The elution of protein was monitored by absorption at 280 nm. Two peaks were observed as shown in Figure 6.5B. A minor peak, which corresponds to the aggregation as it eluted in the void volume of the column, and a major peak which corresponds to Venus, were observed. Fractions corresponding to the major peak were collected and analysed by SDS-PAGE as shown in Figure 6.5C. Fractions containing pure protein were pooled for subsequent use. All Venus^{azF} variants were purified in the same manner as shown in Figure 6.6.

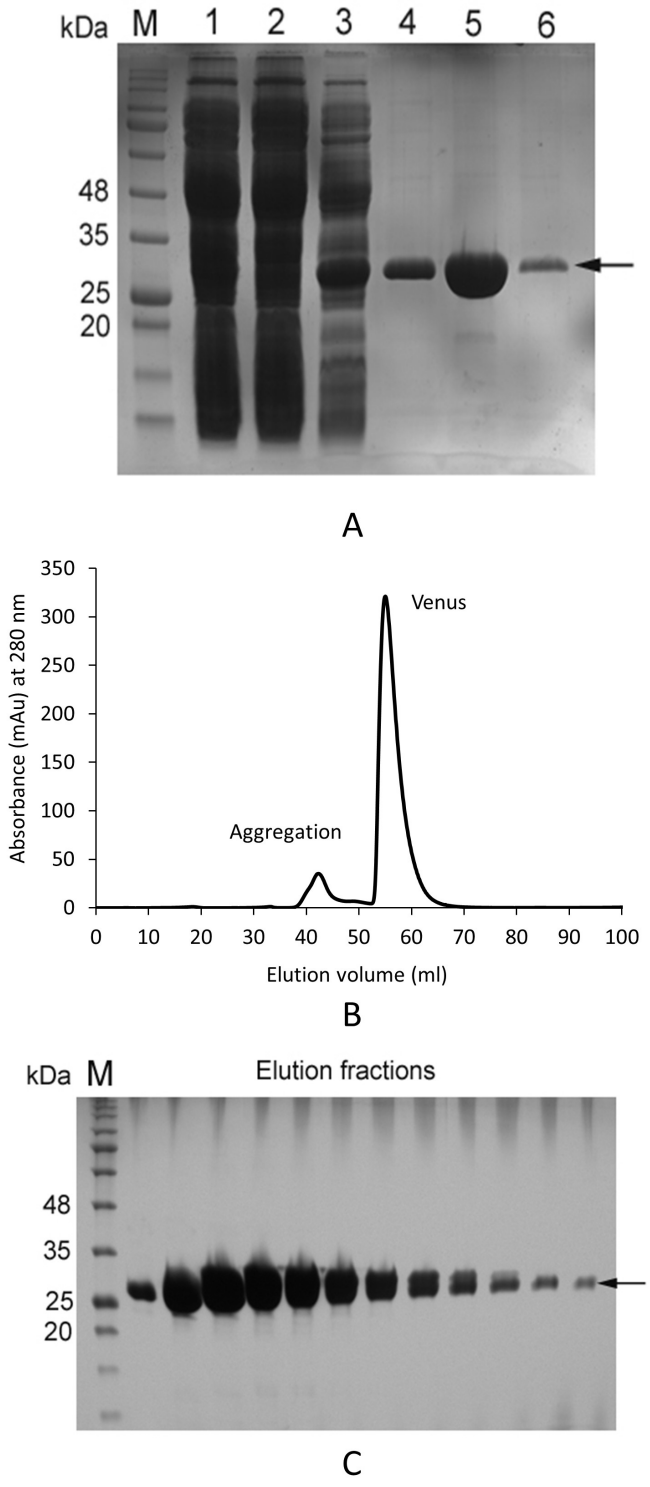

Figure 6.5. Purification profile of Venus wild-type protein.
A) SDS-PAGE (12.5%) analysis of nickel affinity purification. Lane M corresponds to the protein markers of known molecular weights. Lane 1 corresponds to whole cell lysate, lane 2 corresponds to flow-through fraction, lanes 3 and 4 correspond to wash fractions respectively and lanes 5 and 6 correspond to elution fractions respectively. **B)** Chromatogram of SEC showing absorbance (280 nm, black line) over total volume run through the Hiload™ 16/600 Superdex™ S75 pg column. First peak represented protein aggregation and eluted in the void volume of the column and the second peak represented Venus protein. **C)** SDS-PAGE (12.5%) analysis of elution fractions from SEC (second peak). Black arrows indicate the Venus protein.

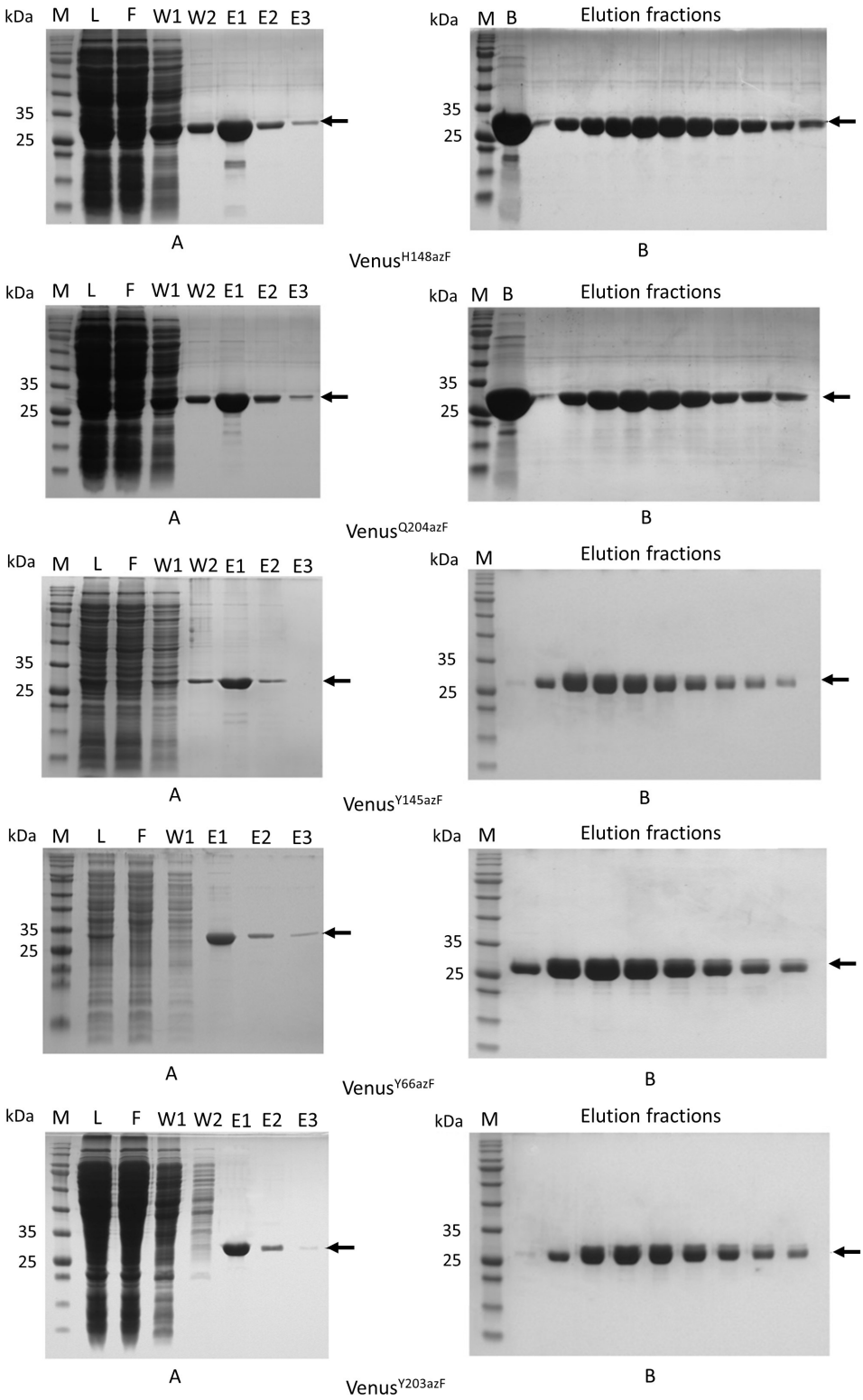

Figure 6.6. Purification profile of Venus^{azF} variants.
A) SDS-PAGE (12.5%) analysis of nickel affinity purification. Lane M corresponds to the protein markers of known molecular weights. Lane L corresponds to whole cell lysate, lane F corresponds to flow-through fraction, lanes W1 and W2 correspond to wash 1 and wash 2 fractions respectively, lanes E1, E2 and E3 correspond to elution 1, elution 2 and elution 3 fractions respectively. **B)** SDS-PAGE (12.5%) analysis of elution fractions from SEC. Lane B corresponds to sample before loading into the SEC column. The Black arrows indicate the Venus^{azF} variant.

6.2.5 Influence of azF incorporation on Venus

Traditional techniques of protein engineering such as deletion, insertion and substitution of amino acids have been successfully used to alter protein structure and functions, but they are inherently limited by the natural amino acid's chemistry. Therefore, introduction of new chemistry not existing in nature by utilising ncAAs incorporated at a specific position in a target protein can result in new properties not accessible through the standard 20 amino acids, most notably new reaction types (Reddington *et al.* 2015). One desirable new chemistry in protein engineering is photoreactivity (or light responsive activity). Since light is orthogonal with biological systems and can be easily controlled and dosed, together the available spatial and temporal control of the use of molecules with photo-controlled activity is rapidly expanding in biological studies (Hoorens and Szymanski 2018). When light responsive properties can be genetically encoded, it is known as optogenetics. The ncAAs such as azF have several useful properties (more information about azF can be found in Section 1.7.1). The most common properties of phenyl azides like azF are photoreactivity and selective reactivity via Click chemistry. In this study these properties have been utilised to alter Venus properties and generate a heterodimer (chapter 7). The effect of photoreactivity of azF at different positions on Venus properties was studied by fluorescence and absorption spectroscopy: Venus variants were analysed before and after photolysis using UV light as described in Section 2.4.8.

The wild type Venus ($Venus^{WT}$) was initially assessed for its response to light and used as control to confirm azF-dependent effects in the variants. $Venus^{WT}$ has a $\lambda_{max/ex}$ at 515 nm and λ_{em} at 528 nm with a molar absorbance (ε) 92200 $M^{-1}cm^{-1}$ at 515 nm and quantum yield (QY) 0.65. Results showed that UV light has little effect on the Venus fluorescence in terms of deactivation, activation and shifting the excitation/emission wavelengths (Figure 6.7). The ε was decreased from 92200 $M^{-1}cm^{-1}$ to 74458 $M^{-1}cm^{-1}$ and the emission peak was dropped by about 15% as result of UV-light exposure after one hour. The effect of azF on Venus at different positions is described below.

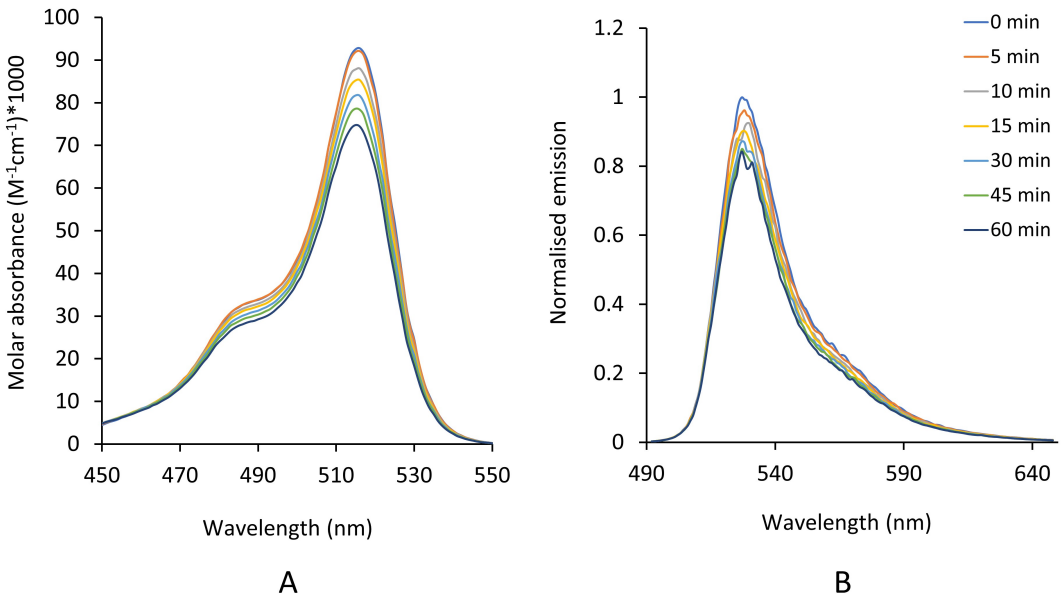

Figure 6.7. Effect of UV-light on wild-type Venus.
A) Absorbance spectra in the absence and presence of illumination with UV light (302 nm handheld 6 W UV lamp) for different lengths of time (up to 60 min) showing the effect of UV-light on molar absorbance**. B)** Normalised emission spectra showing the effect of UV-light on fluorescent properties (emission spectra were normalised to the maximum values of 0 min UV exposure sample.

6.2.5.1 Influence and UV-dependent properties of azF at position 148

Histidine 148 (H148) is located in β-strand 7, near to the tyrosyl moiety of the chromophore (CRO; see Figure 6.8). It is important to function as H148 is H-bonded directly to the CRO: the hydroxyl group of the phenol group in the CRO connects with the imidazole ring of histidine via a hydrogen bond as shown in Figure 6.8. In GFP, H148 plays a critical role in determining the ground state, ionisation state and photodynamic properties. The CRO of GFP can exist in two states based on the protonation state of the OH group in the Y66 phenol ring: neutral state (CRO-A), which absorbs at approximately 395 nm and anionic state (CRO-B) absorbs at approximately 485 nm. The neutral state is the predominant state with a ratio of 6:1 in the wild type GFP (Tsien 1998; Remington 2011). Structural studies have also shown that H148 may be dynamic and can be present in two conformations, with one conformation pointing to the CRO and the second pointing to the solvent (Seifert *et al.* 2002; Seifert *et al.* 2003). Also, it can impact on protein folding as its substitution to arginine or lysine can, in some

cases, cause poor expression and insolubility (Campanini *et al.* 2013). It is believed that incorporation of azF with photo-controlled activity could control the conformation of H148 that will have an impact on the CRO and subsequently the fluorescent properties.

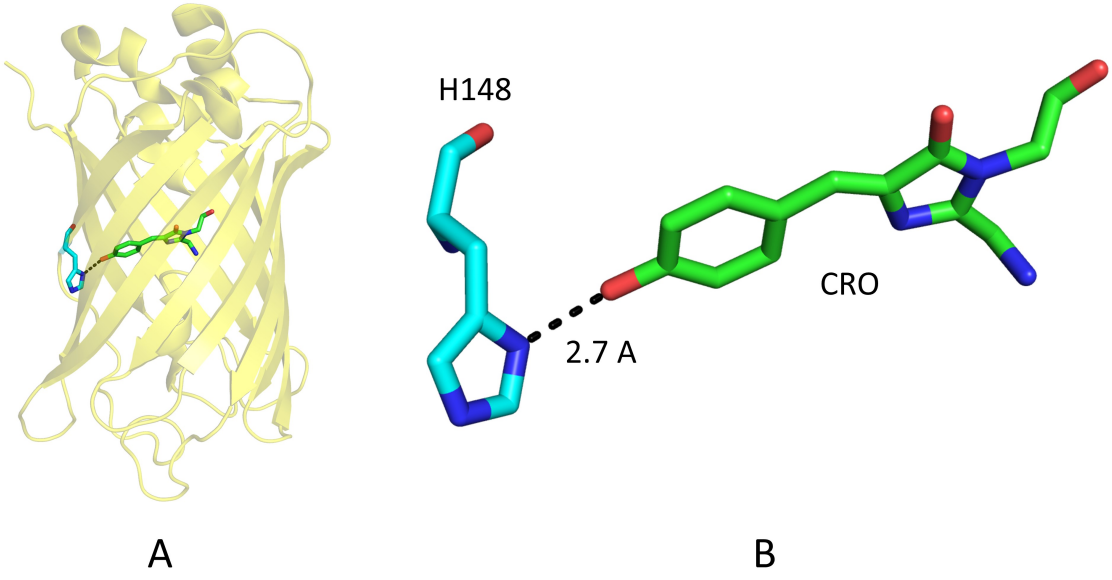

Figure 6.8. Structural position of H148 in Venus.
A) Overall structure of Venus highlighting residue 148 and CRO as a cyan sticks and green sticks respectively. **B)** Position of H148 to the CRO. Connection between hydroxyl group (coloured red) of CRO and nitrogen (coloured blue) of the imidazole group of H148 shown in dashed black.

Substitution of H148 with azF did not affect the folding of the sfGFP but did affect the spectral characteristics by blue shifting the excitation wavelength due to changing the chromophore charged state; removal of the proton accepting imidazole disrupts the proton transfer from the CRO and the associated hydrogen bond network (Hartley *et al.* 2016).

Results showed that replacement of H148 with azF in Venus did not affect the solubility or folding of the protein as the protein expressed as the soluble form as shown in Figure 6.4. However, this mutation has a significant effect on spectral properties as expected. Prior to UV exposure (dark state), the excitation spectra (measuring emission at 528 nm) of Venus^{H148azF} had the λ_{ex} at 517 nm, similar to Venus^{WT} (515 nm) but with intensity decreased by ~72 % (Figure 6.9A). In terms of fluorescence emission, the λ_{em}

remains at 528 nm but the fluorescence intensity decreased by ~72% (Figure 6.9C). The ε and QY also decreased from 95200 to 30100 $M^{-1}cm^{-1}$ (~65% of $Venus^{wt}$) and from 0.65 to 0.45 (~30% of $Venus^{wt}$) respectively (Table 6.1). This result indicates that incorporation of azF at position 148 has a significant effect on the CRO.

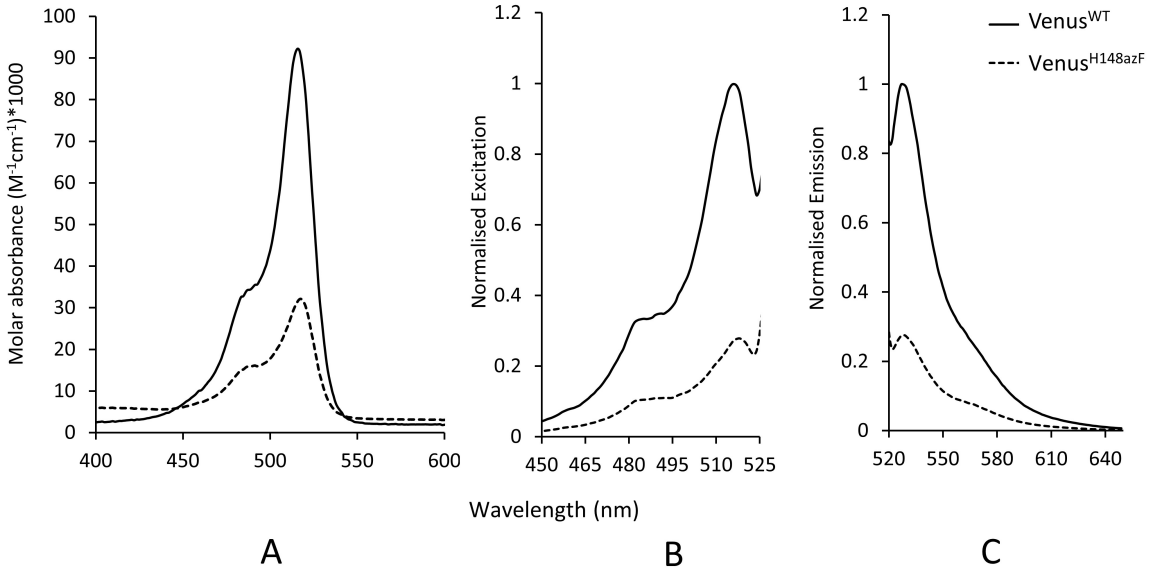

Figure 6.9. Absorbance and fluorescence spectra of $Venus^{H148azF}$ compared with $Venus^{WT}$.
A) Absorbance spectra shown as molar absorbance. **B**) Normalised excitation spectra (emission monitored at 530 nm). **C**) Normalised emission spectra (excitation at 515 nm). Excitation and emission spectra were normalised to the maximum values of $Venus^{WT}$.

The observed changes in fluorescence properties of Venus are unsurprising, as H148 plays a vital role in the protonation state of the chromophore and involved in hydrogen bond network. Incorporation of azF at residue 148 had a significant effect on the CRO, most likely through changing the hydrogen bond H148 normally made with the CRO. The hydrogen bond acceptors in the His side chain are placed correctly to hydrogen bond with the CRO compared to the azF that could have different conformation. The azide group in azF, which is a polar group, may force the azF to occupy a different conformation from the original H148 conformation in the protein towards the solvent and hence disrupt the hydrogen bond with the CRO in Venus (Figure 6.8B). This theory is supported by the ability of $Venus^{H148azF}$ to react with the orthogonal fluorescence DBCO-

TAMRA dye as outlined in chapter 7, which indicated the location of the azide group was in fact free and available to react. This suggests that it is pointed away from the core of the protein to the surface. Additionally, the crystal structure of sfGFP^{H148azF} revealed the local conformational change as a result of 148 azF incorporation, which results in the loss of two hydrogen bonds to the CRO; one from the original 148 imidazole and another from the side chain of T203, (Hartley *et al.* 2016). In the same study, the crystal structure of modified sfGFP^{H148azF} with orthogonal DBCO-amine (a Click chemistry tool containing a short aliphatic amine group) showed that the position of 148azF pointed away from the CRO.

Exposure of Venus^{H148azF} to UV-light for different periods of time showed that the photolysed Venus^{H148azF} (irradiated state) exhibited a different behaviour in terms of absorbance (excitation) and emission spectra. After 5 min λ_{ex} shifted by 10 nm (λ_{ex} 507 nm) and the λ_{em} shifted by 6 nm (λ_{em} 522 nm) with a slight decreased in molar absorbance by ~5% and fluorescence intensity by ~13%. After 60 min of UV-light exposure it displayed photo-switching (switched off) as shown in Figure 6.10. The peak intensity of emission was dropped by ~57% of the dark state. Also, molar absorbance decreased from 31000 to 9963 $M^{-1}cm^{-1}$ (~50% of the dark state). Research by Hartley *et al.* also showed that incorporation of azF at position 148 in sfGFP displayed photo-switching. However, photolysis of sfGFP^{H148azF} activated the protein and switch on not off (Hartley *et al.* 2016).

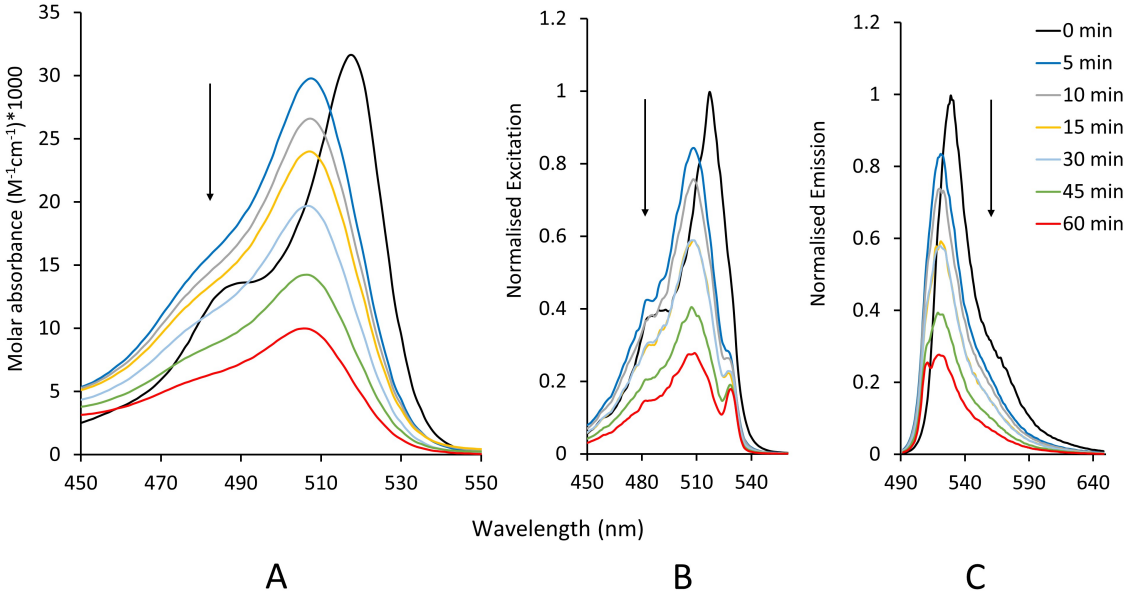

Figure 6.10. Absorbance and fluorescence spectra of Venus^{H148azF} photolysis for 60 min. **A**) Absorbance spectra shown as molar absorbance. **B**) Normalised excitation spectra (emission monitored at 530 nm). **C**) Normalised emission spectra (excitation at 515 nm). 0 min corresponds to the sample before exposure to the UV-light (dark state). 5 min, 10 min, 15 min, 30 min, 45 min, and 60 correspond to the UV-light exposure time respectively. Excitation and emission spectra were normalised to the maximum values of Venus ^{H148azF} (dark state). Black arrow indicates the direction of UV-light effect.

Photolysis of Venus^{H148azF} by UV-light generates a singlet reactive nitrene and releases a molecular nitrogen. One possibility is that reduced the polarity of the azide group after release of a molecular nitrogen might lead the azF to take the original conformation of H148 towards the core of the protein and this could allow the reactive nitrene radical to react with CRO or surrounding residues, hence disrupting the CRO in Venus and explaining the deactivation of fluorescence after UV-light exposure. Possibly the crosslink of azF could occur with either Y145 or Y203, as both residues have an OH group (a nucleophile) and are close to the nitrene radical of azF (an electrophile), which creates a perfect crosslink environment. The crosslink with Y203 would disrupt the π-π stacking interactions with CRO and this could explain the shifting in excitation and emission spectra (Figure 6.10). However, it is hard to know the actual change in azF conformation and the endpoint of the azide group after UV-light treatment without solving the molecular structure.

Other possible pathways for the UV activated azide could produce phenyl amine through reduction of the nitrene radical or the ring expansion via rearrangement of the azide group. Neither of these outcomes would be expected to deactivate fluorescence since neither is expected to affect the CRO. As mentioned earlier, photolysis of sfGFP^{H148azF} activates the protein (switches on), the opposite effect to that seen here with the Venus^{H148azF} variant (switches off). Structural analysis of sfGFP^{148azF} (irradiated state) confirmed the rearrangement of its azide group and ring expansion to dehydroazepine (Hartley *et al.* 2016). Ring expansion led to creation of a water channel and network directly to CRO as present in wildtype sfGFP by formation, together with N146, of hydrogen bonds with the R168 backbone amide and carbonyl groups respectively. As mentioned, the final fate of the nitrene radical depends on the local environment and this explains the different behaviour of H148azF in Venus and sfGFP. The differences in 145 and 203 residues between Venus (Y145 and Y203) and sfGFP (F145 and T203) may give the opportunity to the nitrene radical to crosslink in Venus rather than ring expansion as in sfGFP and this could explain the deactivation of Venus^{148} after photolysis.

6.2.5.2 Influence of azF at position 145

Residue Y145 is an important neighbouring residue to the chromophore that has been targeted in order to influence the fluorescent properties of wild type GFP, blue and yellow versions of GFP (Griesbeck *et al.* 2001; Kajihara *et al.* 2005; Reddington 2013). It is located close to the CRO, with its phenol ring pointing towards the phenol ring of CRO as shown in the Figure 6.11. Incorporation of azF at position 145 in Venus showed little effect on fluorescent properties. Venus^{Y145azF} (dark state) had similar spectral properties to Venus^{WT}; the λ_{max} (λ_{ex}) = 515 nm and λ_{em} = 528 nm but the intensity peak of both excitation and emission had dropped slightly by 8.5% (Figure 6.12). The molar absorbance and QY also dropped slightly from 95200 to 82073 $M^{-1}cm^{-1}$ and from 0.65 to 0.54 respectively compared with Venus ^{WT} (Table 6.1).

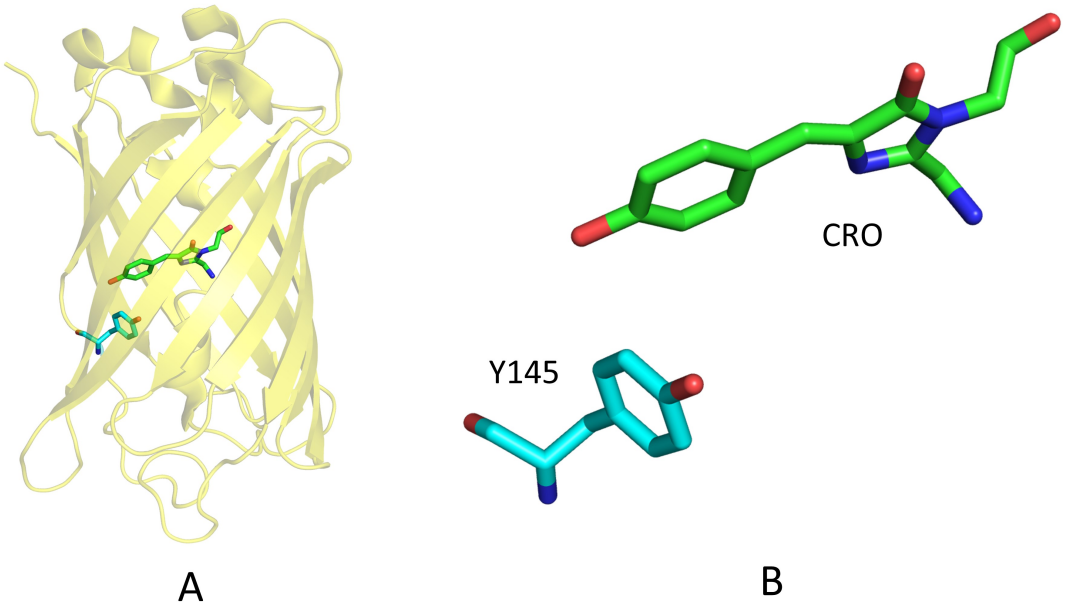

Figure 6.11. Structural position of Y145 in Venus.
A) Overall structure of Venus highlighting residue 145 and CRO as a cyan sticks and green sticks respectively. **B)** Position of Y145 to the CRO.

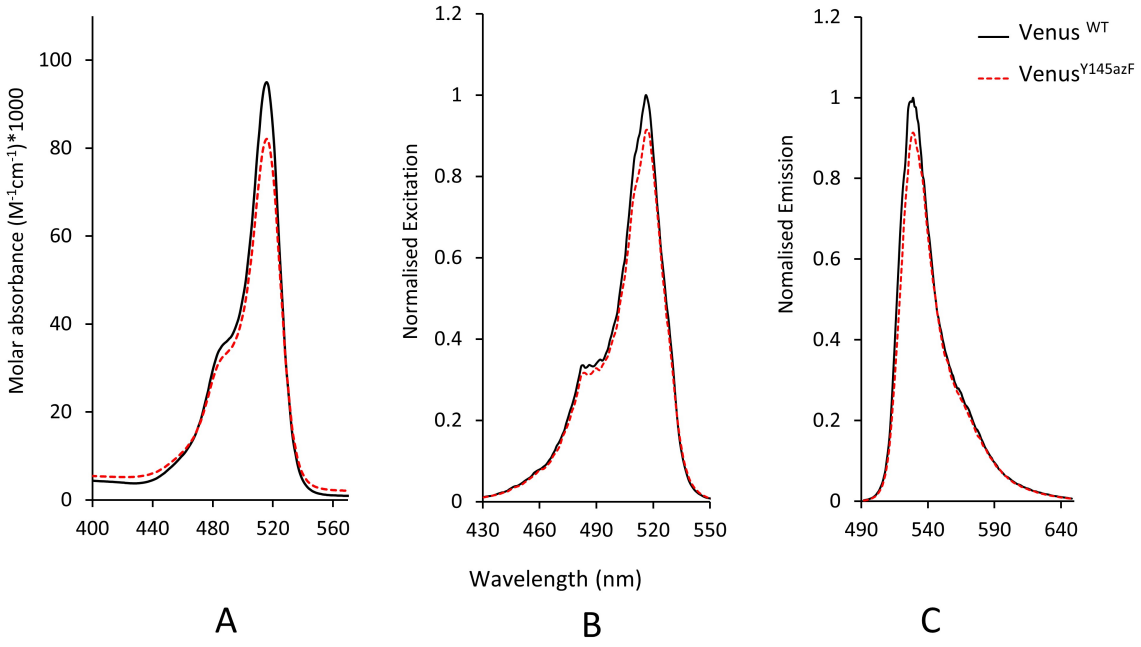

Figure 6.12. Absorbance and fluorescence spectra of Venus^{Y145azF} against Venus ^{WT}.
A) Absorbance spectra shown as molar absorbance. **B**) Normalised excitation spectra (emission monitored at 530 nm). **C**) Normalised emission spectra (excitation at 515 nm). Excitation and emission spectra were normalised to the maximum values of Venus ^{WT}.

Exposure of Venus^{Y145azF} to UV-light for different periods of time showed that the photolysed Venus^{Y145azF} (irradiated state) displayed photo-deactivation (switch off). Photo-deactivation occurred quickly, during the initial 5 min of UV-light exposure; the fluorescence intensity dropped by 86% as shown in Figure 6.13. The molar absorbance decreased to 32820 $M^{-1}cm^{-1}$ (59% of the dark state), and the fluorescence intensity and molar absorbance dropped slightly after 5 min of UV-light treatment.

The similarity of spectra properties of Venus^{Y145azF} (dark state) and Venus^{WT} indicates that introduction of an azido group at the para position of the phenyl side chain of Y145 had little effect on CRO environment. However, the photolysis of the azide group has a significant effect on CRO and deactivate the protein. It is likely that the deactivation occurred as a result of cross-linking between the nitrene group derived by photolysis of 145azF with the CRO. The crystal structure of sfGFP^{F145azF} (irradiated state) that displayed photodeactivation similar to that of Venus^{Y145azF} (irradiated state) showed that the nitrene radical formed an N-phenyl crosslink with the *meta* carbon of the chromophore tyrosyl moiety (Reddington *et al.* 2013). The general feature of the photolyzed Venus^{Y145azF} (irradiated state) in terms of fluorescent level and spectra properties suggests that it too follows the same pathway as sfGFP^{F145azF} (irradiated state, PDB code 4J8A) and forms a cross-link with CRO and deactivates the fluorescence.

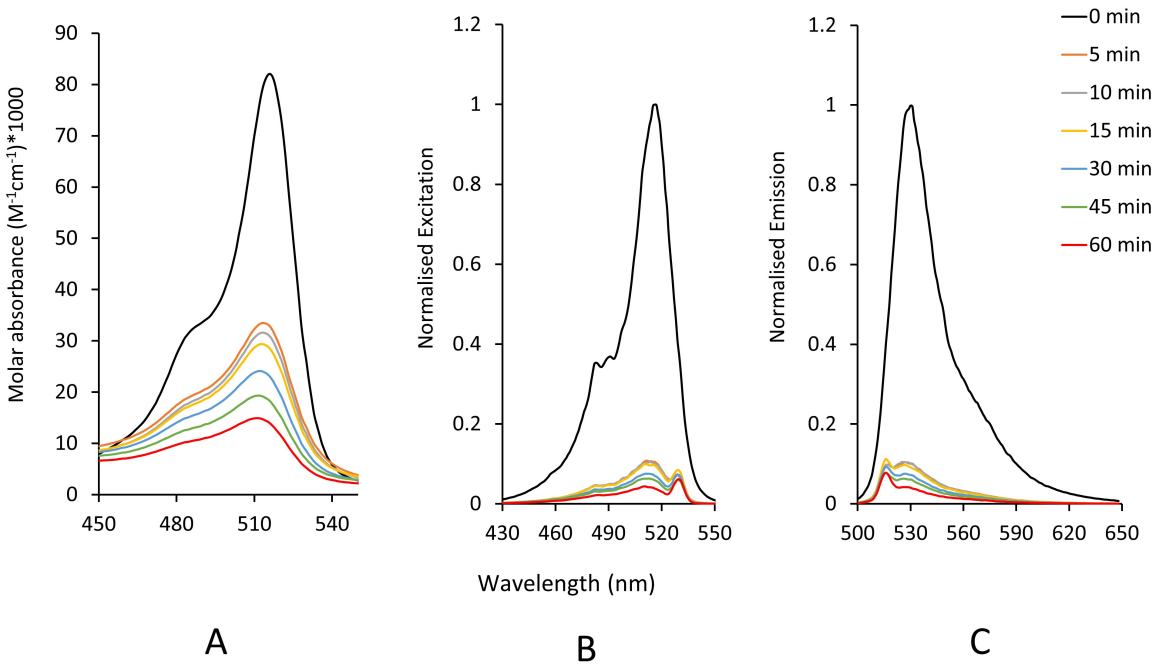

Figure 6.13. Absorbance and fluorescence spectra of Venus^{Y145azF} photolysis for 60 min.
A) Absorbance spectra shown as molar absorbance. **B**) Normalised excitation spectra (emission monitored at 530 nm). **C**) Normalised emission spectra (excitation at 515 nm). 0 min corresponds to the sample before exposure to the UV-light. 5 min, 10 min, 15 min, 30 min, 45 min, and 60 correspond to the UV-light exposure time respectively. Excitation and emission spectra were normalised to the maximum values of Venus ^{Y145azF} (dark state) before exposure to the UV-light. Black arrow indicates the direction of UV-light effect.

6.2.5.3 Influence of azF at position 203

Threonine 203 residue (Tyrosine in Venus) is one of three critical residues (together with S65 and E222) in the vicinity of the chromophore of wild-type GFP that play an important role in the photophysical behaviour of the protein. Mutation to these residues results in disruption of the hydrogen-bonding network and hence fluorescence properties (Jung *et al.* 2005b). Mutation of T203 to tyrosine is seen as the key mutation for shifting fluorescence into the yellow regions due to pi-stacking of the introduced Y203 phenol group with the CRO (Figure 6.14). Tyrosine 203 in Venus is located above the CRO and its phenol ring and connects with E222 via a water molecule as shown in Figure 6.14. Thus, replacement of Y203 with a phenyl azide moiety (azF) should elicit changes to Venus's fluorescent properties.

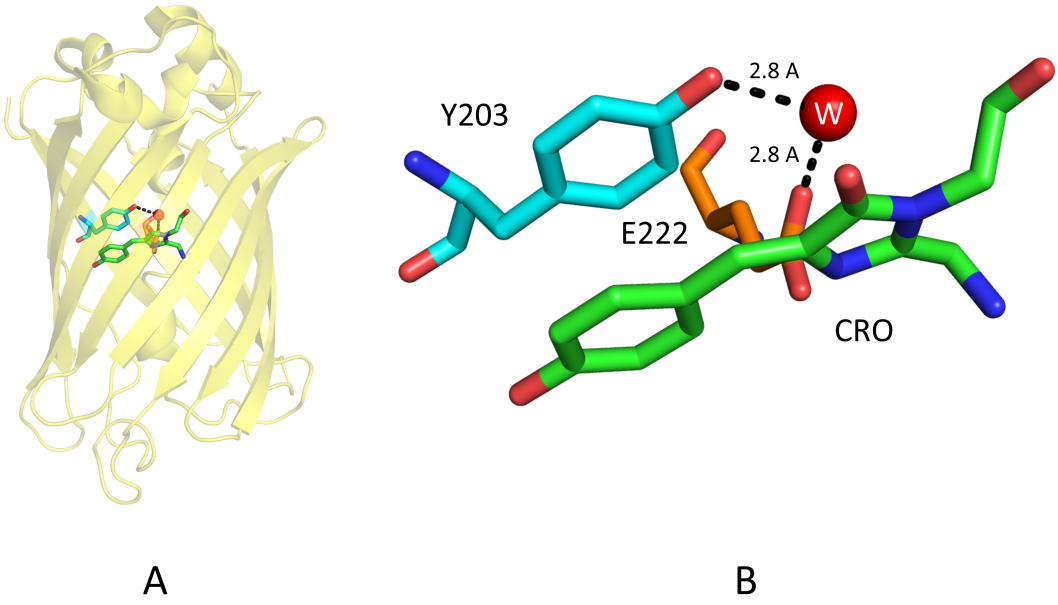

Figure 6.14. Structural position of Y203 in Venus.
A) Overall structure of Venus highlighting residue Y203, E222 and CRO as a cyan sticks, orange sticks and green sticks respectively. **B)** Position of Y203 to the CRO and hydrogen bond connection with E222. Connection between hydroxyl group (coloured red) of Y203 and hydroxyl group (coloured red) of the E222 with water molecule (W; red sphere) shown in dashed black.

Results showed that incorporation of azF at position 203 had a significant effect on absorbance and fluorescence properties, including displaying photo-switching. Prior to UV or ambient light exposure (dark state), Venus^{Y203azF} had λ_{max} (λ_{ex}) at 514 nm shifted from Venus ^{wt} (515 nm) by 1 nm and λ_{em} shifted by 8 nm (522 nm). The protein was essentially non-fluorescent with the observed intensity ~92 % of lower than Venus^{WT} as shown in Figure 6.15. The molar absorbance and QY were also substantially lower, dropping from 92200 to 6683 $M^{-1}cm^{-1}$ (~93 % of Venus^{wt}) and from 0.65 to 0.002 respectively as illustrated in Table 6.1.

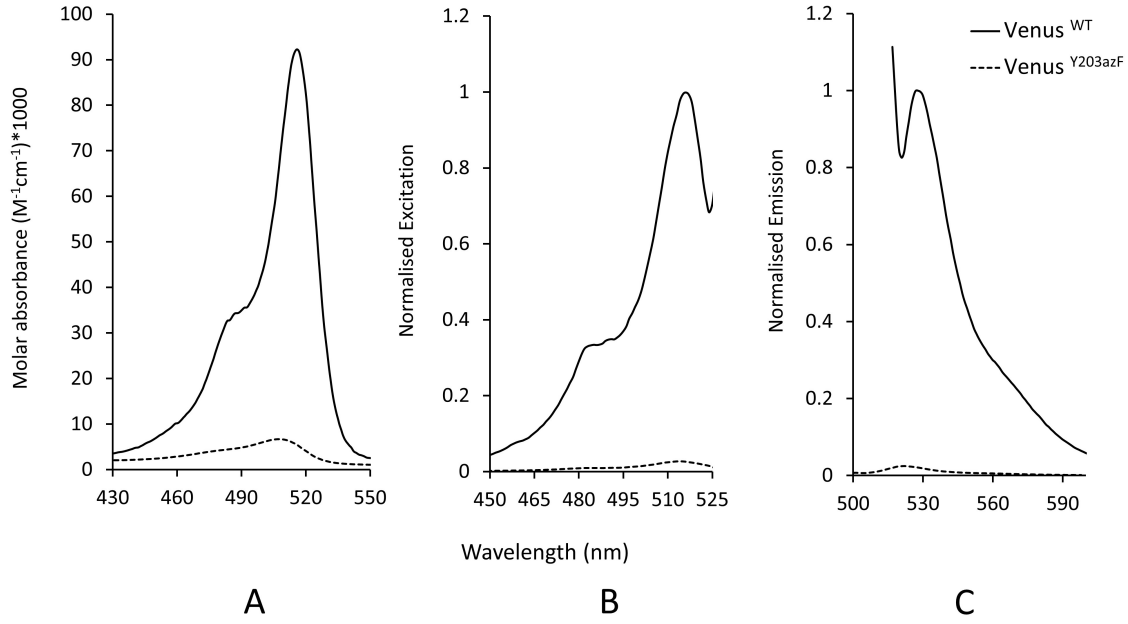

Figure 6.15. Absorbance and fluorescence spectra of Venus^{Y203azF} against Venus ^{WT}.
A) Absorbance spectra shown as molar absorbance. **B**) Normalised excitation spectra (emission monitored at 530 nm). **C**) Normalised emission spectra (excitation at 515 nm). Excitation and emission spectra were normalised to the maximum values of Venus ^{WT}.

Venus^{Y203azF} was responsive to ambient light as well as UV light but in distinct ways. Unlike other variants, Venus^{Y203azF} (dark state) was very sensitive to ambient light. Venus^{Y203azF} (dark state) already shows only low levels of fluorescence (Figure 6.15) and exposure to ambient light deactivates the Venus^{Y203azF} further, effectively converting it to a non-fluorescent form (Venus^{Y203azF} ambient light state) as can be clearly seen in Figure 6.16D. The fluorescence intensity decreased by ~38% of that of the dark state (switch off), λ_{ex} at 511 nm slightly shifted by 3 nm and λ_{em} slightly shifted by 2 nm (518 nm). Molar absorbance dropped from 6683 $M^{-1}cm^{-1}$ to 2894 $M^{-1}cm^{-1}$ (~57% of Venus^{Y203azF} dark state) as shown in Figure 6.16A and B.

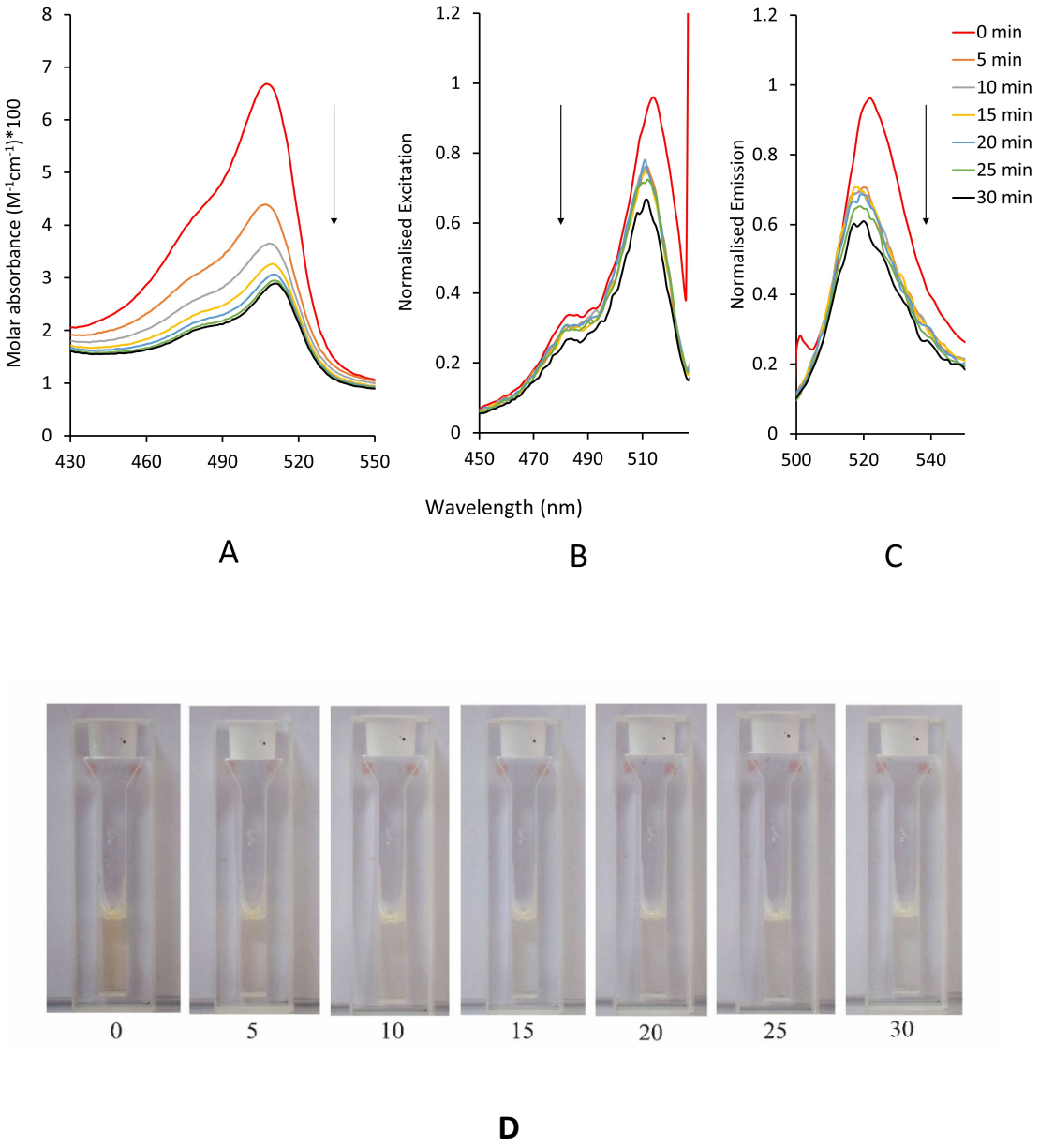

Figure 6.16. Absorbance and fluorescence spectra of ambient light photo-switching of Venus^{Y203azF}.

A) Absorbance spectra shown as molar absorbance. **B)** Normalised excitation spectra (emission monitored at 530 nm). **C)** Normalised emission spectra (excitation at 515 nm). 0 min corresponds to the Venus^{Y203azF} (dark state) before exposure to the ambient light. 5 min, 10 min, 15 min, 30 min, 45 min, and 60 correspond to the ambient light exposure time respectively. Excitation and emission spectra were normalised to the maximum values of Venus ^{Y203azF} (dark state) before exposure to ambient light. Black arrow indicates the direction of ambient light effect. **D)** Image showing photo-switching (switch off) of Venus^{Y203azF} (dark state, 20 µM) during 30 min of ambient light exposure.

On exposure to UV-light, the non-fluorescent form of Venus^{Y203azF} activated to a fluorescent form (Figure 6.17D). After 30 min of UV-light treatment, the fluorescence intensity increased by ~53% (switch on), λ_{ex} at 507 nm slightly shifted by 4 nm and λ_{em} shifted by 2 nm (516 nm) as shown in Figure 6.17. Molar absorbance increases significantly to 16614 $M^{-1}cm^{-1}$ (~82.5% of non-fluorescent form) as illustrated in Table 6.1.

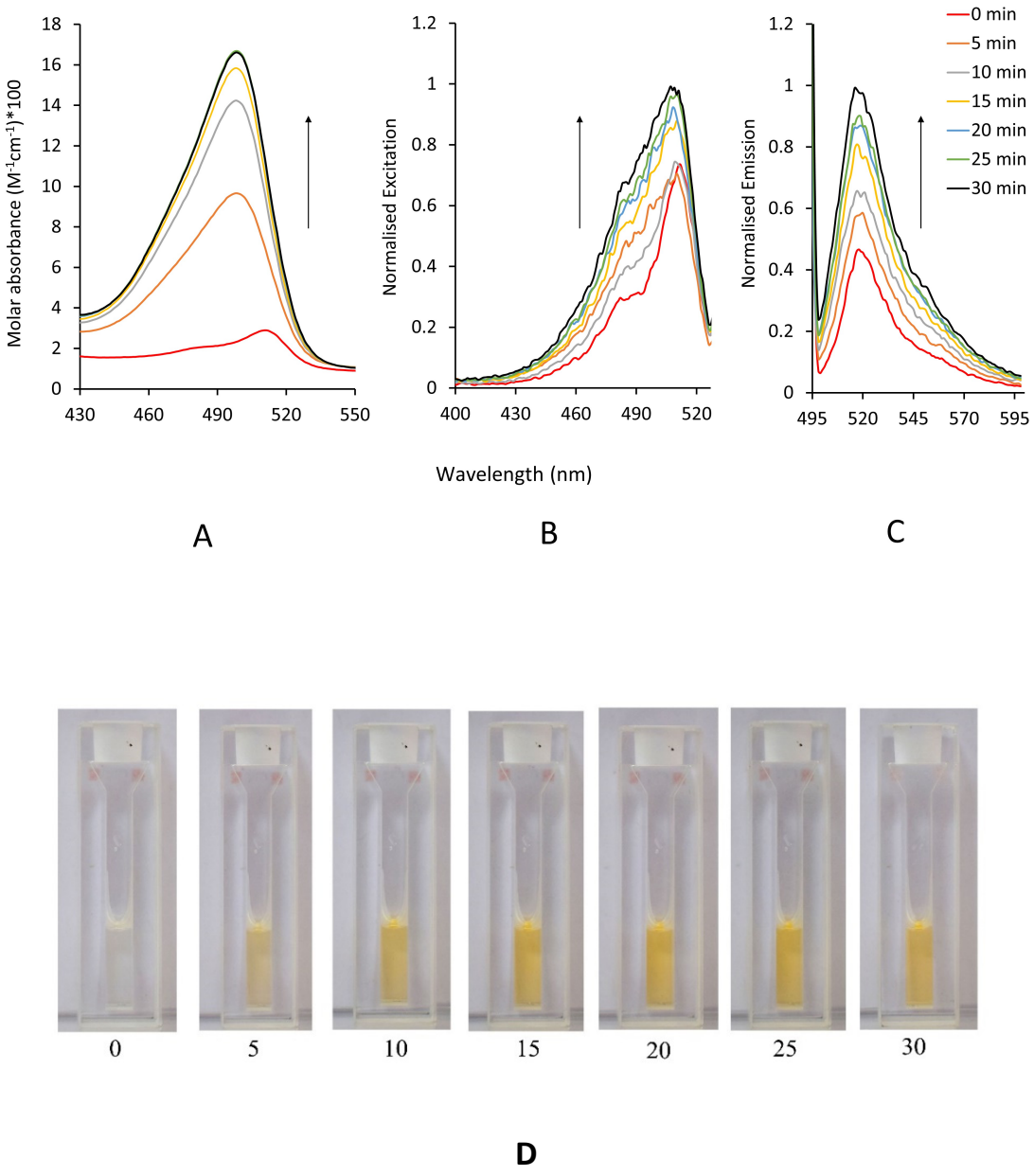

Figure 6.17. Absorbance and fluorescence spectra of UV-light photo-switching of Venus^{Y203azF}. **A)** Absorbance spectra shown as molar absorbance. **B)** Normalised excitation spectra (emission monitored at 530 nm). **C)** Normalised emission spectra (excitation at 515 nm). 0 min corresponds to the Venus ^{Y203azF} (ambient light state) before exposure to the UV-light. 5 min, 10 min, 15 min, 30 min, 45 min, and 60 correspond to the UV-light exposure time respectively. Excitation and emission spectra were normalised to the maximum values of Venus ^{Y203azF} (ambient light state) before exposure to UV-light. Black arrows indicate the direction of UV-light effect. **D)** Image showing photo-switching (switch on) of Venus ^{Y203azF} (ambient light state, 20 µM) during 30 min of UV-light exposure.

Venus^{Y203azF} is an interesting variant, which showed photo-switching (switch between two different states) in terms of fluorescent properties and colour change as can be seen in Figure 6.16D and Figure 6.17D. The same mutation in sfGFP (sfGFP^{T203azF}) does not result in photo-switching between two different states but it also exhibited high sensitivity to the ambient light with a change in colour, and photolysis, resulting in a substantial reduction in molar absorbance about 5 fold (Reddington *et al.* 2015). In Venus, two residues in the CRO region interact via a water molecule: E222 (which is involved in maturation of the CRO of GFP derivatives (Barondeau *et al.* 2003; Arpino *et al.* 2012a; Reddington, 2013)) and Y203 (which is involved in the pi stacking interaction that gives the yellow shift in Venus) (Figure 6.14). Introducing the azide group in Venus^{Y203azF} may lead to loss of the connection with E222 and could induce local conformational changes which affect the maturation of the CRO. Additionally, the presence of the azide group may force other residues around the CRO to change their conformation, particularly the residues Q69, L68 and Q67, and subsequently affect CRO maturation. These effects would explain the poor absorbance and shifting in emission spectra (Figure 6.15). *In silico* substitution of Y203 with azF using PyMol software shows the potential location of the azide group with respect to the CRO and surrounding environment (Figure 6.18). The reason for susceptibility of Venus^{Y203azF} to ambient light is not obvious.

Activation of the fluorescence of Venus^{Y203azF} after photolysis by UV-light and change in colour (Figure 6.17) are presumably as a result of local conformational modifications around the CRO environment through the formation of the nitrene radical, which then converts to one of the forms shown in Figure 6.1. This modification could lead the final end product derived from the azide group to generate a hydrogen bond network with E222 via a water molecule and induce local modification, thereby removing the negative effect on CRO maturation and this would explain the activation of protein and increased the absorbance and fluorescence spectra (Figure 6.17A and C). The endpoint of the azide group is unlikely to crosslink with adjacent residues because this would lead to disruption of the CRO and deactivate the fluorescence but results from UV exposure showed the activation of fluorescence. The endpoint of the azide group is possibly reduced to phenyl amine or rearrangement of azide group and ring expansion (Inui *et al.* 2013). The latter form could have a good a π-π stacking interaction but not as

strong a π-π stacking interaction as the original phenyl ring and this would explain the shift in spectral properties of Venus^{Y203azF} (irradiated state) compared with Venus^{WT}.

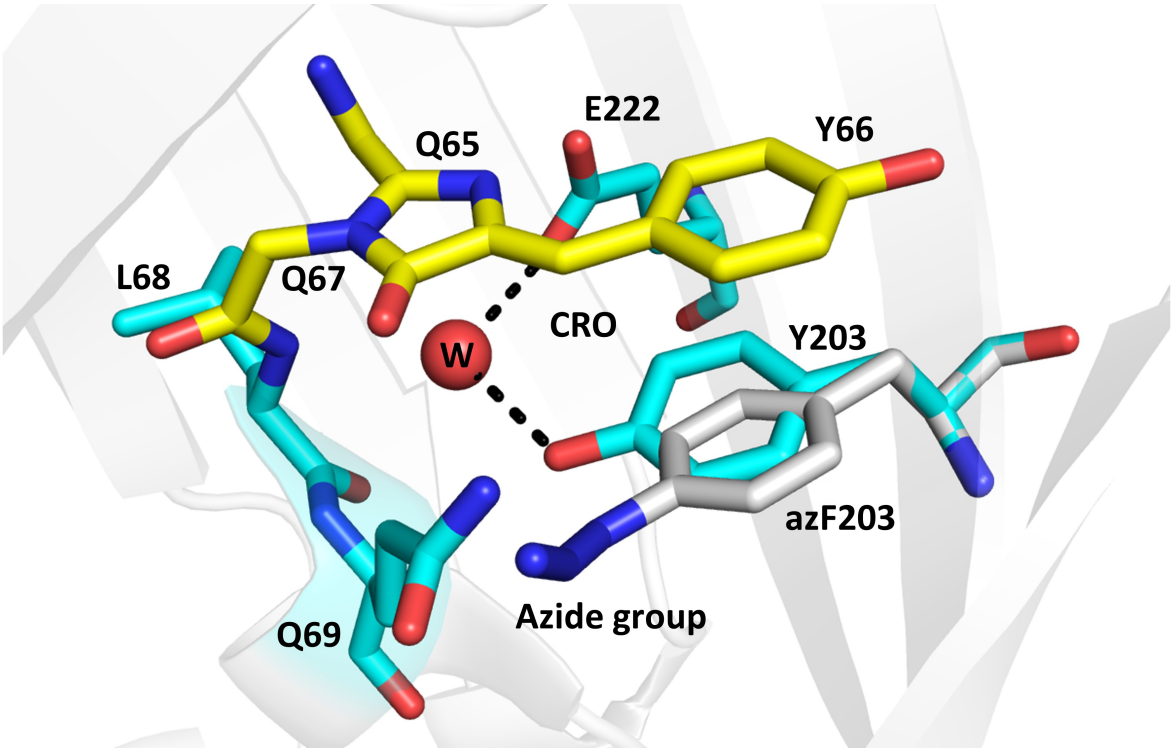

Figure 6.18. A model of azF incorporation at 203 residue in Venus.
Structure of Venus^{wt} showing the predicted position of 203azF and azide group in relation to the CRO and surrounding environment. CRO (carbon coloured yellow). Y203, E222, Q68 and Q69 carbon coloured cyan. 203azF carbon coloured grey. Water molecule (W) coloured red. PDB code IMYW.

6.2.5.4 Influence of azF at position 66

As mentioned in the introduction (Section 1.5.2.2), in wt-GFP the chromophore consists of the tripeptide (S65, Y66 and G67). Y66 is a key residue of the chromophore and, as such, protein engineering has targeted this residue to generate new colour variants that will expand the applications available to fluorescent proteins. Two different coloured versions were generated from GFP: blue fluorescent protein (BFP) with two substitutions, Y66H (in the chromophore)/Y145F; and cyan fluorescent protein (CFP) with the chromophore substitution Y66W (Heim *et al.*1994;Wachter *et al.* 1997). Based on these

findings, Kajihara *et al.* (2005) suggested the possibility to generate novel chromophores by incorporating ncAAs at 66 position in GFP. In 2005, Kajihara *et al.* incorporated 18 aromatic ncAAs at 66 position in GFP in an attempt to generate novel variants, of which 16 generated non-fluorescent proteins and two ncAA mutants (*p*-aminophenylalanine and *O*-methyltyrosine mutants), generated a weakly fluorescent, blue-shifted protein (Kajihara *et al.* 2005). In contrast, the study by Reddington *et al.* (2013) showed that incorporation of ncAA (azF) at position 66 in sfGFP generated useful new properties (photo-switching). This sfGFP^{Y66azF} generated a weakly fluorescent protein in the dark that could be quickly activated by UV light (switching on) (Reddington *et al.* 2013).

The chromophore of Venus differs from sfGFP through changes at residue 65 (T in sfGFP, G in Venus), as shown in Figure 6.19, as well as other important residues around the CRO such as Y203 and Y145. It is assumed that incorporation of azF at residue 66 could generate photo-switching properties.

Incorporation of azF at 66 position in Venus (Venus^{Y66azF}) has considerable effect on fluorescent properties, including the introduction of photo-activation properties (switch on) and a shift in the major peaks (λ_{ex} = 432 nm; λ_{em} = 505 nm). Venus^{Y66azF} was produced as a non-fluorescent protein (colourless protein) (Figure 6.20A). It appears to have an immature CRO, due to the characteristic absorbance peak at 352 nm, which has been observed previously (Abbyad *et al.* 2007); no emission intensity was detected as shown in Figure 6.20B and C. Exposure of Venus^{Y66azF} to ambient light for 10 hours did not exhibit any change, so Venus^{Y66azF} remains non-fluorescent.

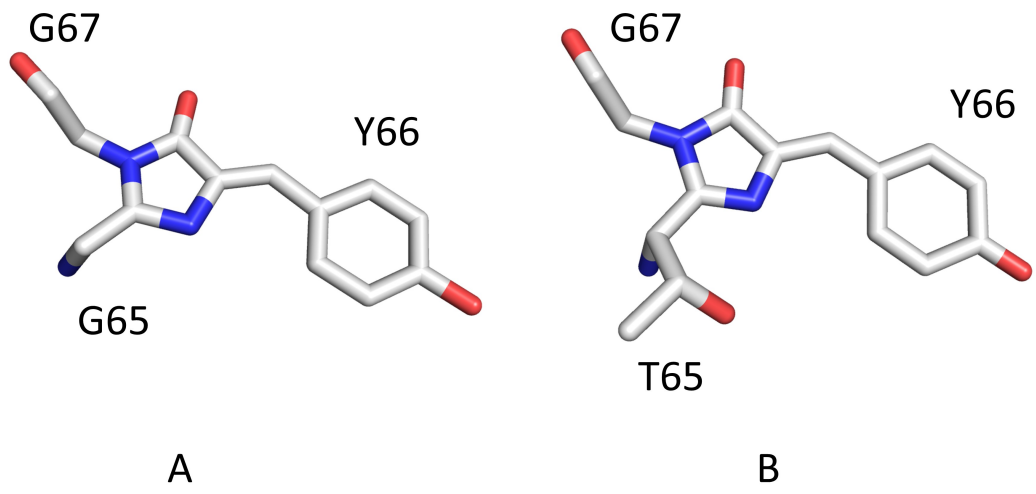

Figure 6.19. Chromophore residues of Venus and sfGFP wild type.
A) Chromophore residues of Venus wild type (PDB code 1MYW) **B**) Chromophore residues of sfGFP wild type (PDB code 2B3P). Carbon colored grey.

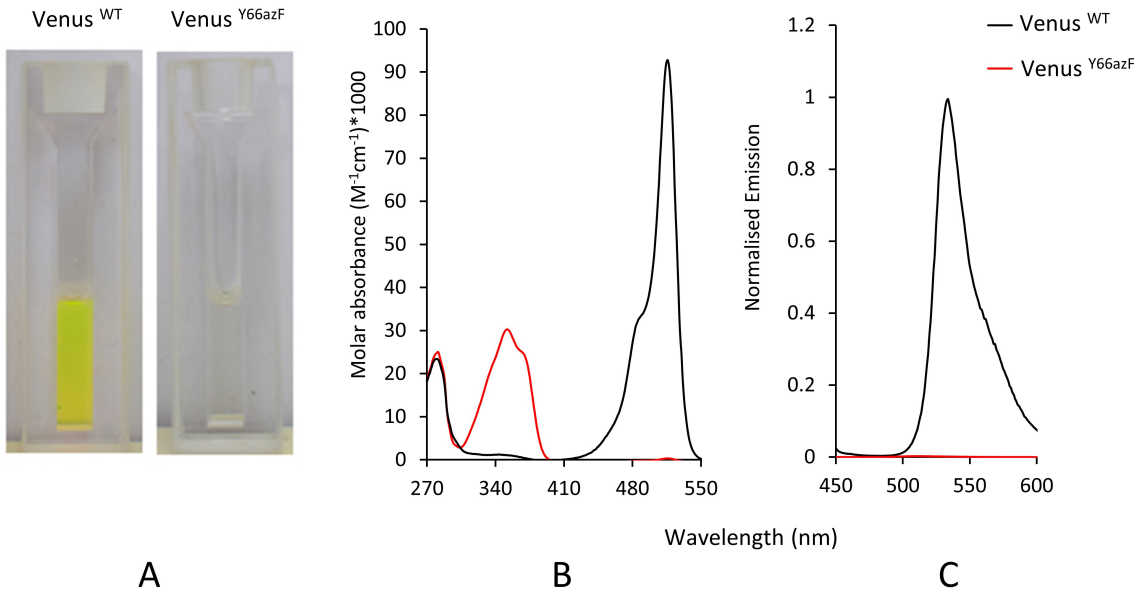

Figure 6.20. Absorbance and fluorescence spectra of Venus^{Y66azF} against Venus^{WT}.
A) Image showing the effect of azF incorporation at 66 position compared with Venus wild type. Venus^{Y66azF} is produced as a non-fluorescent protein (colourless protein; 20 µM). **B)** Absorbance spectra shown as molar absorbance. **C)** Normalised emission spectra (excitation at 515 nm). Emission spectra were normalised to the maximum values of Venus ^{WT}.

Interestingly, photolysis of Venus^{Y66azF} with UV-light activated the protein rapidly. The peak at 350 nm dropped during 30 min of UV-light exposure with a new peak increasing at 432 nm at the same time (Figure 6.21A). After 30 min of exposure, no further changes were observed. During this period, the protein's colour changed from colourless to yellow (switch on) after 1 min, as shown in Figure 6.21C. Emission spectra matched the increase in the absorbance peak at 432 nm and the emission intensity increased during 30 min of UV-light exposure (Figure 6.21B). λ_{em} was shifted by 23 nm (505 nm) compared with Venus^{WT}. These findings indicated that the Venus^{Y66azF} (non-fluorescent colourless protein) was switched on by UV-light to a yellow fluorescent protein with molar absorbance 18596 $M^{-1}cm^{-1}$ and QY 0.1.

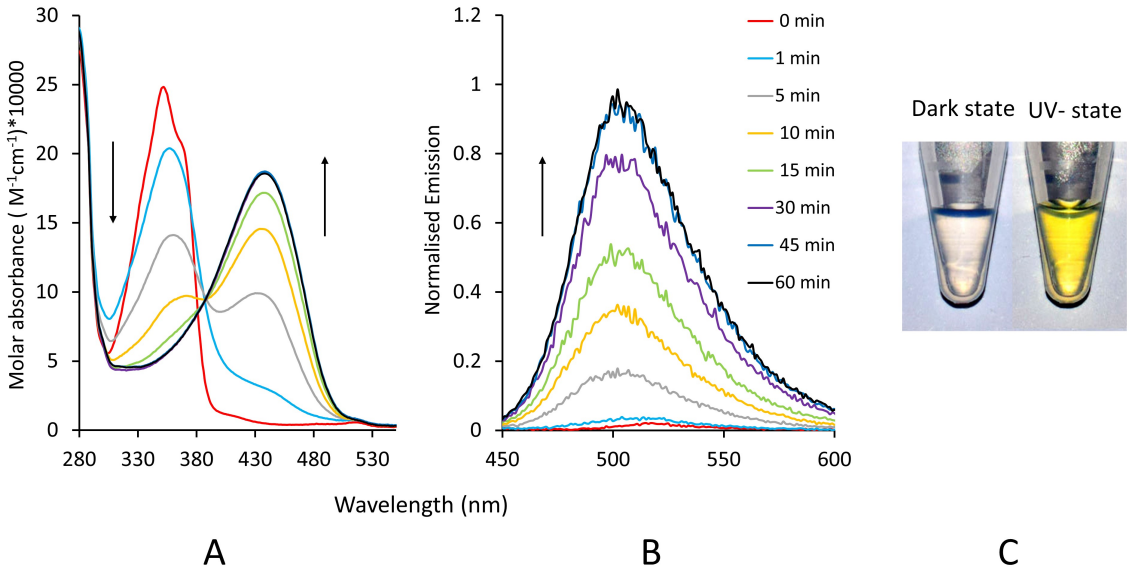

Figure 6.21. Absorbance and fluorescence spectra of UV-light photo-switching of Venus^{Y66azF}. **A)** Absorbance spectra shown as molar absorbance with increasing UV light exposure. **B)** Normalised emission spectra (excitation at 450 nm). 0 min corresponds to the Venus ^{Y203azF} (dark state) before exposure to the UV-light. 1 min, 5 min, 10 min, 15 min, 30 min, 45 min, and 60 correspond to the UV-light exposure time respectively. Emission spectra were normalised to the maximum value of Venus^{Y203azF} (irradiated state). Black arrows indicate the direction of UV-light effect. **C)** Image showing photo-switching (switch on) of Venus^{Y66azF} (dark state, 20 µM) after 30 min of UV-light exposure.

Table 6.1 Spectral Properties of Venus ^{azF} variants

Variant	λ_{ex} (nm)	λ_{em} (nm)	ε ($M^{-1}cm^{-1}$)	QY	Brightness (ε x QY)
Venus ^{WT}	515	528	92200[a]	0.65	59930
Venus^{H148azF} (dark state)	517	528	30100	0.45	13545
Venus^{H148azF} (irradiated state)	507	522	9963	0.082	816.9
Venus^{Y145azF} (dark state)	515	528	82073	0.54	44319
Venus^{Y145azF} (irradiated state)	511	528	32820	0.21	6892
Venus^{Y203azF} (dark state)	514	522	6683	0.0023	15.3
Venus^{Y203azF} (irradiated state)	507	516	16614	0.043	714
Venus^{Y66azF} (dark state)	-	-	-	-	-
Venus^{Y66azF} (irradiated state)	432	505	18596	0.1	1856

[a] Published previously by Nagai *et al* (Nagai *et al.* 2002) and measured in the current study as 95000 $M^{-1}cm^{-1}$. Given that the measured value is close to the reported value, the reported value has been used.

6.2.6 X-ray crystallography

To understand the molecular changes to Venus as a result of azF incorporation and to explain the molecular basis of photo-switching, the 3D structure of Venus^{Y66azF} was determined by X-ray crystallography. Due to time constraints, crystallography studies were focused only on Venus^{Y66azF} in the dark state. While crystallisation trials with Venus^{Y66azF} (irradiated state) were unsuccessful.

6.2.6.1 Protein crystals preparation

Venus^{Y66azF} was expressed and purified as outlined in Section 6.2.4. Purified proteins were concentrated to 20 mg/ml and used to setup a crystal trials as described in Section 2.12.1. A long, spiny crystal of Venus^{Y66azF} (dark state) was observed after a few days in the A12 condition of the PACT *premier*^{TM} HT-96 screen (0.01 M Zinc chloride, 0.1 M Sodium acetate, pH 5 and 20% w/v PEG 6000). The light activated, irradiated version of Venus^{Y66azF} (irradiated state) failed to produce any crystals. A similar outcome was observed by Reddington (2013) as the irradiated state of sfGFP^{Y66azF}, under a range of conditions, failed to produce crystals (Reddington 2013). This could be a result of disorder in some of the surface residues on exposure to UV-light, which may prevent packing of protein to form crystals.

Sample crystals of dark state Venus^{Y66azF} were harvested first, and the remaining crystals were irradiated for 5 min using UVM-57 mid-range UV lamp (UVP) as described in Section 2.4.8 before collection. All crystals were harvested by picking individual crystals in a mounted litholoop (Molecular Dimensions) and plunging them into liquid nitrogen. Crystals were kept in liquid nitrogen until the X-ray diffraction data were collected.

6.2.6.2 Data collection and structure determination

Diffraction data were collected at the Diamond Light Source, Harwell, UK, at beamline I04 as described in Section 2.12.1. Crystals of Venus^{Y66azF} (dark state) diffracted at 1.9 Å while irradiated crystals did not diffract at all as the crystals were disordered. Images of crystals mounted in a litholoop can be seen in Figure 6.22. Structures were determined by molecular replacement with sfGFP^{Y66azF} (PDB code 4J88) as a model using PHASER. The chromophore with azF was generated as a ligand by JLigand (in CCP4i suite) and adjusted manually with structure in COOT over several cycles. The structure was refined by TLS restrained refinement using RefMac (more information about structure determination can be found in Section 2.12.4). Full crystallographic data and refinement statistics are illustrated in Table 6.2.

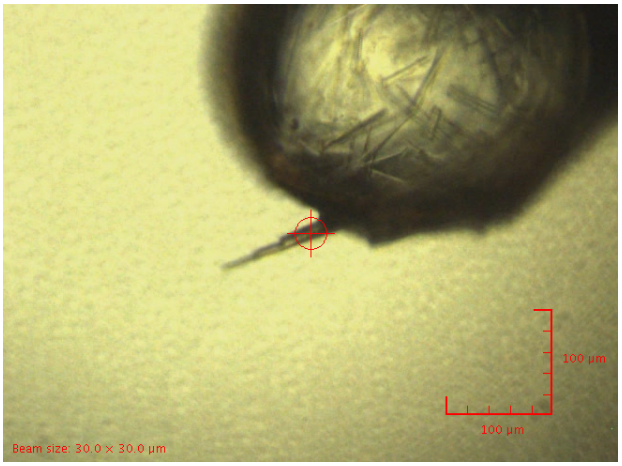

Figure 6.22. Venus ^{Y66azF} crystals.
Image of Venus ^{Y66azF} crystals in a mounted litholoop.

Table 6.2 Crystal diffraction and refinement statistics of Venus$^{Y66\ azF}$ (dark state)

PDB Code	6SM0
Data collection/reduction statistics	
Diamond Beamline	I04-1
Wavelength	0.91587
a, b, c (Å)	50.947, 63.251, 69.915
Space group	P 2_1 2_1 2_1
Resolution (Å)	1.909 – 63.26
Outer shell	1.909 – 1.96
R-merge (%)	16.4 (114.5)
CC1/2	0.989 (0.355)
I / σ	7.4 (1.8)
Completeness (%)	99.9 (99.9)
Total Measurements	103,864 (4,920)
Unique Reflections	18,146 (1,317)
Wilson B-factor(Å^2)	16.0
Refinement Statistics	
Non-H Atoms	2,118
R-work reflections	17,175
R-free reflections	1,238
R-work/R-free	19.4 / 24.2
rms deviations	
Bond lengths (Å)	0.013
Bond Angles (°)	1.854
[1]Coordinate error	0.135
Mean B value (Å2)	19.6
Ramachandran Statistics	
Favoured/allowed/Outliers	187 / 5 / 0
%	97.4 / 2.6 / 0.0

* Numbers in brackets refer to outer resolution shell, where applicable.
[1] Coordinate estimated standard uncertainty in (Å), calculated based on maximum likelihood statistics.

6.2.6.3 General features of Venus^{Y66azF} structure

Comparison between Venus^{WT} and Venus^{Y66azF} (dark state) in general, overall structure and in the molecular details around the CRO will help us to assess the molecular changes that give rise to the observed fluorescent properties and help understand the effect of azF incorporation on Venus structure and CRO environment. Based on the alignment of the two structures, Venus^{Y66azF} (dark state) was very similar to the Venus^{WT} in terms of general, overall fold, as can be seen in Figure 6.23, with an observed RMSD

of 0.61 Å across the backbone/Cα. However, the presence of the azide group in the centre of the protein had a significant effect on the rearrangement of side chains of some residues, which occupied different conformations. A local conformational change around the CRO was also observed. Water molecules have rearrangements and some of them occupied different positions, especially around the CRO.

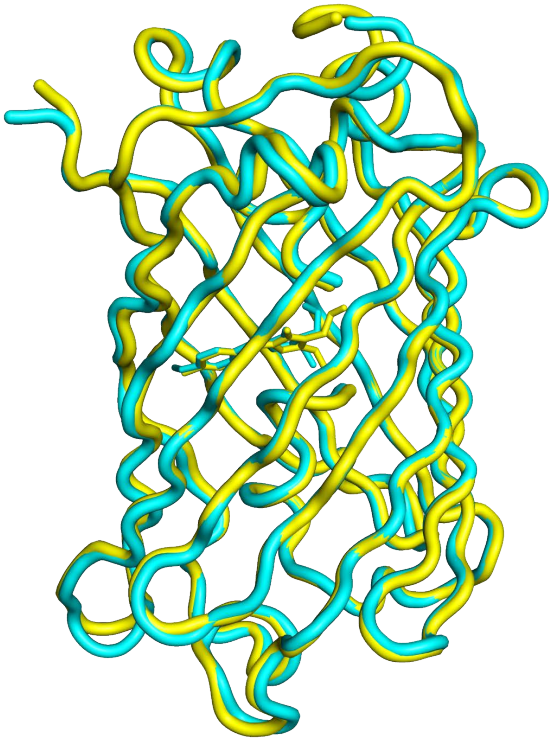

Figure 6.23. The effect of azF incorporation on Venus ^{Y66azF} structure.
Structural alignment of Venus^{WT} (yellow, PDB code 1MYW) and Venus^{Y66azF} dark state (cyan, PDB code 6SM0) shown as cartoon tubes.

A more detailed comparison of the CRO and surrounding residues showed that some of the side chains in Venus^{Y66azF} (dark state) were slightly shifted and others occupied different side chain conformations (Figure 6.24A-B). The electron density map confirmed the presence of the azido group at the *para* position of the tyrosyl moiety of the CRO (Figure 6.25). The presence of this extra bulk (N_3 compared to OH) forces other residues to rearrange and occupy different conformations as can be seen in Figure 6.24B. Structural analysis showed that the CRO itself had slightly changed compared to the CRO of Venus^{WT}. This change included rotation of the peptide bond of G67 by 180° and loss of the connection with N121 via a water molecule. The imidazole ring of H148, a key

residue in defining the fluorescence properties of the CRO (see Figure 6.8 above) also appears to flip out and away from the CRO, so losing its contact (Figure 6.26). While such a change in H148 conformation has been observed previously as part of a conformation population, the electron density associated with H148 in Venus^{Y66azF}, suggests that it occupies one conformation the flipped-out form (Figure 6.26).

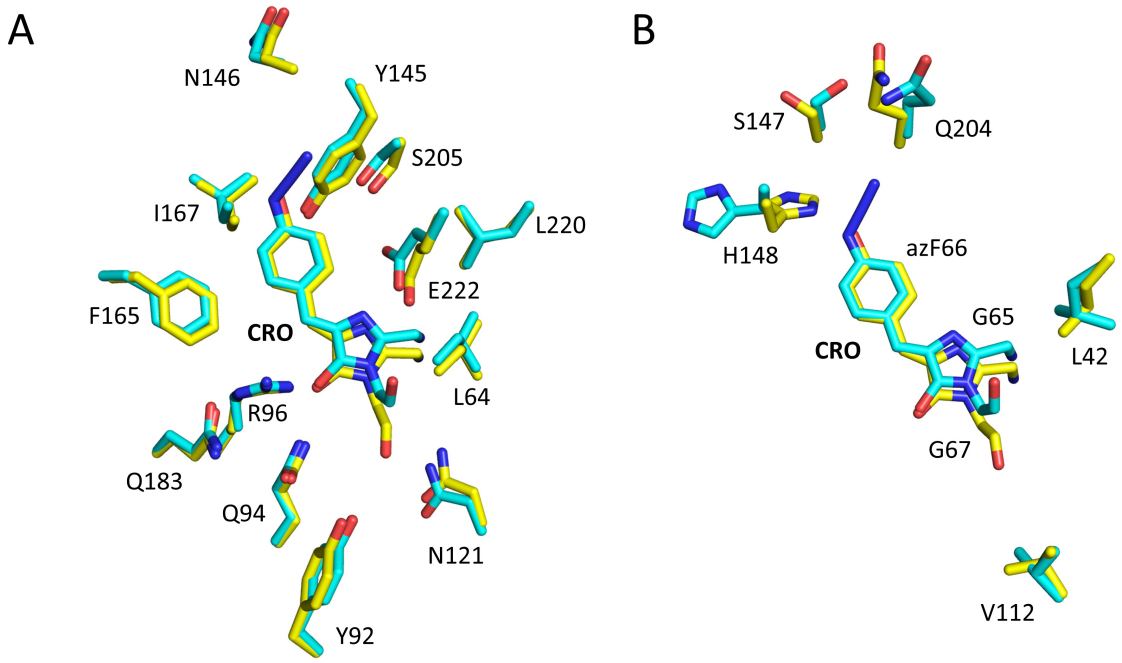

Figure 6.24. The effect of azF incorporation on CRO and surrounding residues of Venus^{Y66azF}. **A)** CRO and surrounding residues positions of which are slightly changed in Venus^{Y66azF} (dark state) structure compared with Venus^{WT} structure. **B)** CRO and surrounding residues that occupy different conformation in Venus^{Y66azF} (dark state) structure compared with Venus^{WT} structure. Venus^{Y66azF} (dark state) has carbons coloured cyan and Venus^{WT} has carbons coloured yellow, nitrogen is coloured blue and hydroxyl groups coloured red. Venus^{WT}, PDB code 1MYW; Venus^{Y66azF} (dark state), PDB code 6SM0.

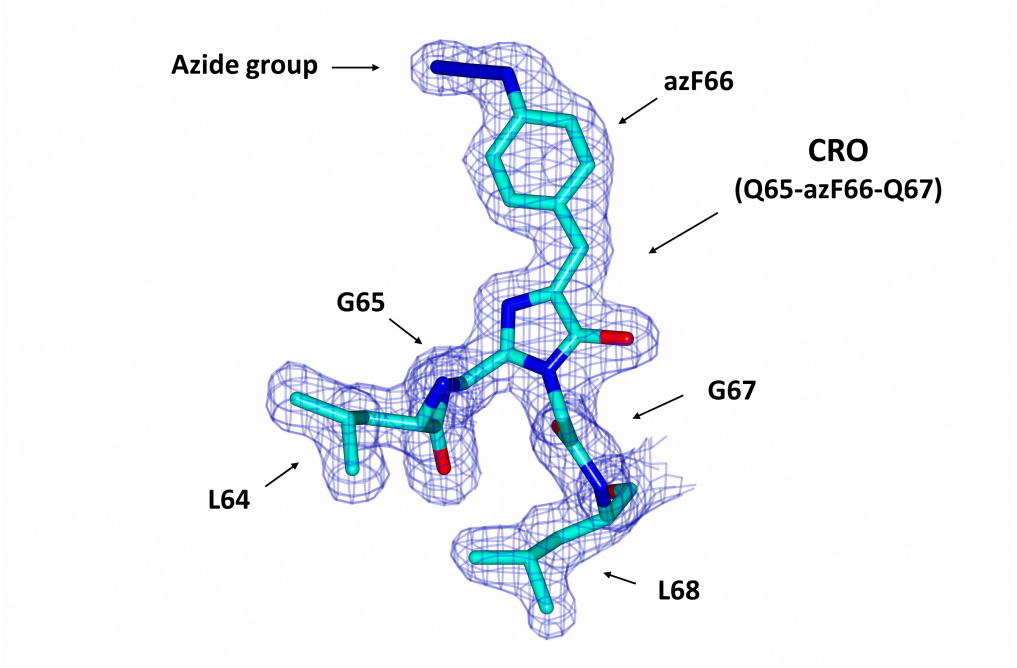

Figure 6.25. Electron density map of the CRO of Venus ^{Y66azF} dark state.
Electron density map of CRO of Venus^{Y66azF} dark state coloured blue showing the presence of azide group at the *para* position of the tyrosyl moiety of the CRO. Carbons coloured cyan, nitrogen coloured blue, hydroxyl group coloured red, (PDB code 6SM0).

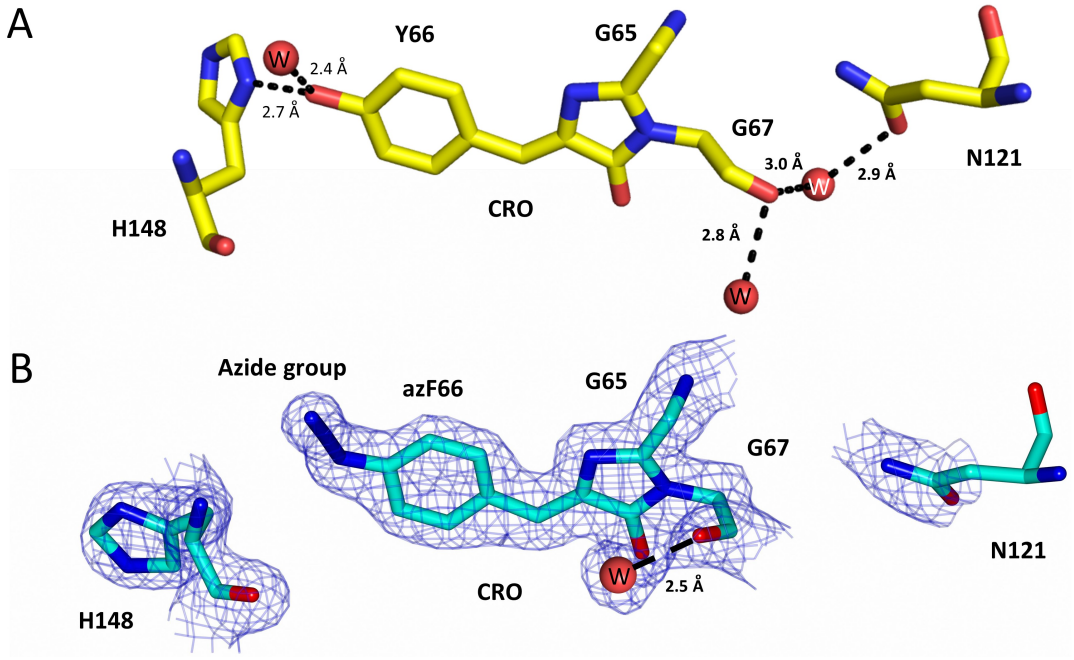

Figure 6.26 CRO of Venus and hydrogen bond network with H148 and N121.
A) CRO of Venus^{WT} showing the hydrogen bonds (black dashed) between the hydroxyl group (red) of the tyrosyl moiety of the CRO with a water molecule (W; red sphere) and nitrogen (blue) of the imidazole group of H148. The hydroxyl group of G67 connected with two water molecules (W; red sphere), one of them connected with the hydroxyl group of the N121. **B)** CRO of Venus^{Y66azF} (dark state) showing the hydrogen bond (black dashed) between hydroxyl group (red) of G67 and water molecule (W; red sphere) and disconnected from H148 and N121 residues. Venus^{WT}, PDB code 1MYW; Venus^{Y66azF} (dark state), PDB code 6SM0.

Other changes in surrounding residues as a result of the presence of the azide group were associated with Q204 and S147 (Figure 6.24B). In Venus^{Y66azF}, the side chains of Q204 and S147 were slightly twisted away from the azido group (Figure 6.24B). However, both residues are surface exposed and have no contact or involvement in the hydrogen bond network with the CRO. Q204 has been found play a critical role in pH stability and protein dimerisation in sfGFP (Roberts *et al.* 2016) while residue S147 has been found to be involved in the dimer interface of YFP (Rekas *et al.* 2002) with no effect on the fluorescent properties of the protein. All the changes in amino acid conformation in Venus^{Y66azF}, especially in H148 and G67, lead to changes in the hydrogen bond network around the CRO, hence loss of the fluorescent properties.

6.2.6.4 Molecular mechanism of photoactivation in Venus^{Y66azF}

As previously demonstrated, Venus^{Y66azF} was produced as a colourless non-fluorescent protein, indicating a significant effect of the azide group on the CRO. The electron rich azide group is known to act as a fluorescence quencher (Sivakumar *et al.* 2004). This is likely to be the main reason for the lack of fluorescence in Venus^{Y66azF}. Fluorescence quenching is a process that refers to the decrease of fluorescence intensity as a result of quenching of the light emission by interaction with closely contacted molecules, known as fluorescence quenchers. The structure of the dark state, as described above in Section 6.2.6.3, suggests a potential mechanism for the lack of fluorescence: the local rearrangement of residues, especially H148 and the backbone of G67 with disruption of the hydrogen-bond network around the CRO may play a role in blocking the fluorescence (Figure 6.26). The same mutation in sfGFP does not result in complete loss of fluorescence in the dark state but it is severely compromised (Reddington *et al.* 2013). The structure of sfGFP^{Y66azF} also does not show the same conformational changes around G67 that shifts the side-chain position of G67 (Figure 6.27) although in one molecule in the unit cell, H148 is shifted away from its normal position. Thus, while a similar mechanism of general quenching by the phenyl azide moiety is playing the major role, local structural changes further impact on function to make Venus^{66azF} a largely non-fluorescent protein before activation. This may be a very important property to possess, for example in applications such as single molecule cell imaging, whereby there is a need for a zero-background sign before light activation (Birk 2019).

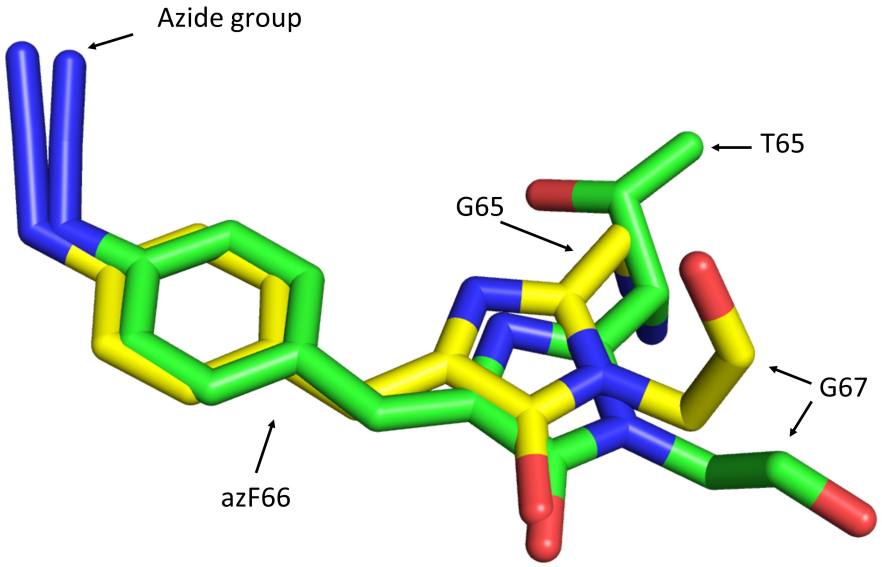

Figure 6.27. Comparison between CRO in Venus^{Y66azF} and sfGFP^{Y66azF} dark state.
Structural alignment of CRO of Venus^{Y66azF} (yellow, PDB code 6SM0) and sfGFP^{Y66azF} (green, PDB code 4j88) showing the difference in G67 conformation as a result of azF incorporation.

The photo-activation is more clear for Venus^{Y66azF} than other Venus variants. Photolysis of Venus^{Y66azF} led to activation of the protein, presumably through the formation of the nitrene radical, which then decays to one of the forms shown in Figure 6.1. Primarily, this is probably due to the removal of the quenching properties of the original azide group. A similar result was observed by Reddington *et al.*(2013) on incorporation of azF at position 66 in sfGFP (Reddington *et al.* 2013). The endpoint of the azide group is unlikely to be crosslinking with adjacent residues or rearrangement and ring expansion because the crosslinking led to disruption of the CRO and deactivates the fluorescence. Ring expansion is likely to disrupt the conjugated bond system, hence deactivate the fluorescence, while results from UV exposure showed the activation of fluorescence. As mentioned, the quencher properties of the electron rich azide group need to be removed for fluorescence activation. It is most likely that the nitrene radical reduced to phenyl amine. The general spectral characteristics of the photolyzed Venus^{Y66azF} are similar to those observed previously for sfGFP^{Y66azF} (Reddington *et al.* 2013) and other GFPs (Wang *et al.* 2003) with an amPhe at residue 66. This is typified

by the broad absorbance peak between 400-490 nm peaking at around 440 nm. The crystal structure of sfGFP^{Y66AzF} (irradiated state, PDB code 4J89) confirmed the reduction of the azide group to amine (Reddington *et al.* 2013), suggesting that Venus^{Y66azF} also follows the same pathway as sfGFP^{Y66azF} and is reduced to phenyl amine. Additional conformational changes playing a role cannot be ruled out. The fact that, in crystallography, illumination destroyed the crystal quality, suggests that some additional conformational change other than the simple release of N_2 is at play.

6.2.7 Live cell imaging of bacteria using Venus^{Y66azF}

Amongst Venus^{azF} variants, Venus^{Y66azF} exhibited a clear photo-switching in terms of fluorescent properties. To consider this variant as a useful tool for cell biology it must exhibit photo-switching properties inside living cells. *E. coli* Top10 competent cells were used to demonstrate the photo-switching behaviour of Venus^{Y66azF} *in vivo*. Twenty ml of induced bacterial cells (Section 2.3.4.2) were separated into two tubes (10 ml each), one tube kept in the dark and the other exposed to UV-light for 1 min at a distance of about 1 cm. Cells were centrifuged at 5000 x*g* for 15 min for both tubes. The result showed the ability of Venus^{Y66azF} to photoactivate inside the bacterial cells (Figure 6.28). Additionally, the induced bacterial cells were imaging with fluorescence widefield microscopy. Slide with induced live bacterial cells was prepared as described in Section 2.9.1. As a control, uninduced cells were treated in the same manner. Both bacterial cultures were exposed to UV-light for 1 min prior to imaging for photo-switching activation. Results from transmitted and fluorescence widefield imaging show that induced cells exhibit photo-switching and appeared as fluorescent bacteria while uninduced cells did not (Figure 6.29). The same mutation in sfGFP (sfGFP^{Y66azF}) shows similar results, the induced cells are rapidly activated by a 430 nm excitation filter in fluorescence widefield microscopy and become a fluorescent (Reddington *et al.* 2013). These results indicate the ability of Venus^{Y66azF} to display photo-switching in bacterial cells and not only in solution.

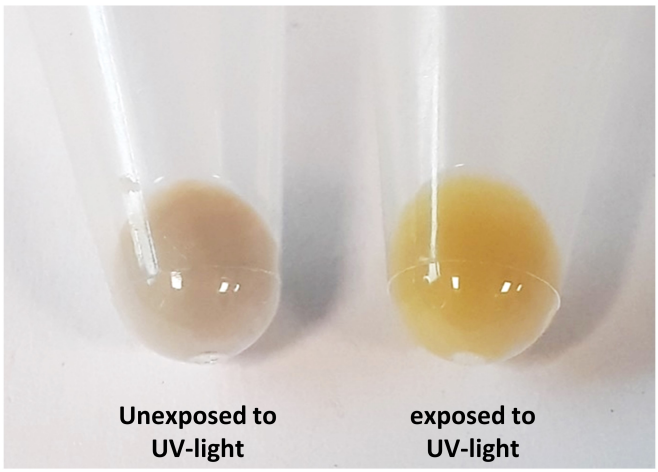

Figure 6.28. Photo-switching of Venus^{Y66azF} in induced bacterial culture.
Each Eppendorf tube contains a pellet derived from 10 ml induced Top10 *E. coli* cells. UV-light exposure for 1 min at a distance of about 1 cm.

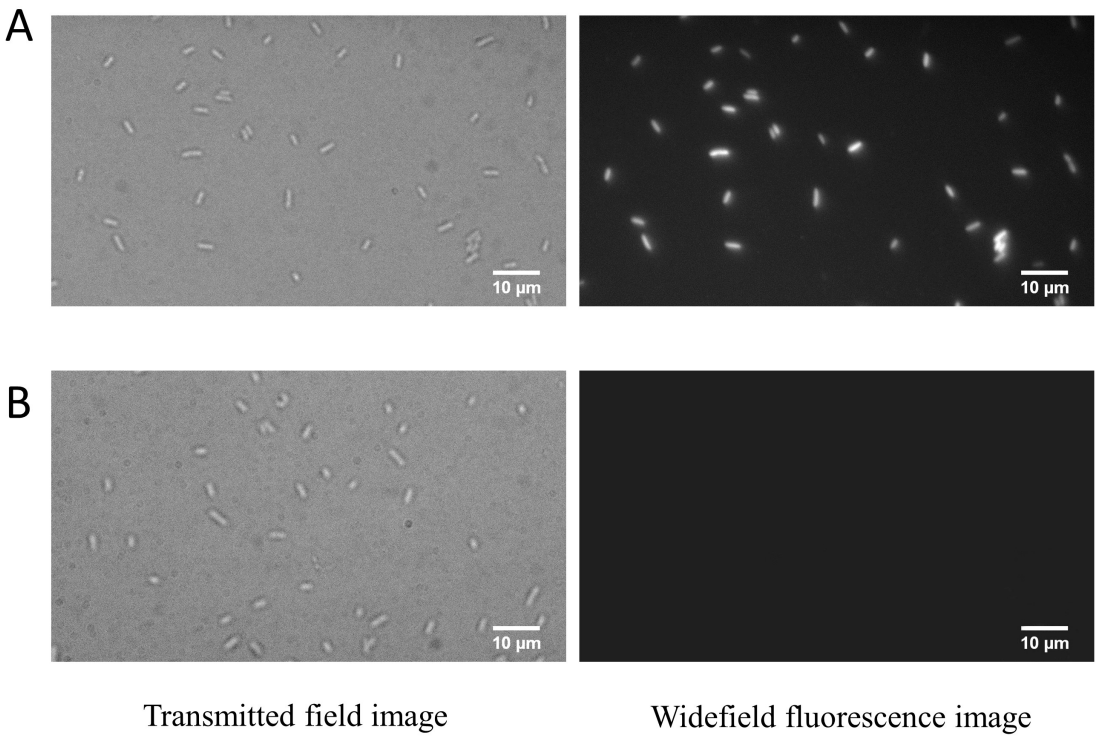

Figure 6.29. Live cell imaging of bacteria using Venus^{Y66azF}.
A. Induced Top10 *E. coli* cells **B.** Uninduced Top10 *E. coli* cells. A wavelength of 450 nm was used for the excitation.

6.2.8 Conclusion

Introduction of a new photoactivable group such as an azide group into the yellow fluorescent protein (Venus) cause a significant change in fluorescence properties before and after photo activation by UV-light. The azide group at residues 66, 148 and 203 disrupts the inherent function of the protein. This is borne out in the observed spectral properties compared with the Venus wildtype. In terms of incorporation at residues 66, non-fluorescent protein was produced (Figure 6.20), probably as a result of quenching properties of the azide group. Furthermore, the structure of Venus^{Y66azF} (dark state) reveals the structural changes in the CRO itself and surrounding residues: a relatively small change in terms of size (-OH to -N_3), forces the imidazole ring of H148 to flip and also causes relocation of other amino acids, particularly G67. These two changes cause loss of the connection of the CRO with H148 and loss of the connection between G67 and N121 that is normally mediated via a water molecule (Figure 6.26). Replacement of H148 with azF led to disrupt the hydrogen bonds that normally connect the H148 with CRO, hence affecting the spectral properties. Incorporation of azF at position 203 led to disruption of the hydrogen bond with E222 and is predicted to bring the azide group closer to the CRO environment and lead to local conformational changes that affect the maturation of the CRO and reduce the absorbance and emission spectra. No significant change in fluorescence properties were observed as a result of substitution of Y145 with azF, which indicates the location of azide group has no effect on CRO in this case.

In general, photo activation of the azide group of azF by UV-light leads to loss of N_2 molecular nitrogen to generate a singlet reactive nitrene radical. Based on the surrounding environment, the nitrene radical can undergo different chemical reactions as illustrated in Figure 6.1. Incorporation of azF at different positions in Venus surrounding the CRO, introduced the azide group into different environments that ultimately affect the pathway taken by the reactive nitrene radical, which, in turn, affected the fluorescent properties. Photolysis of the phenyl azide group has different effects on the CRO and thus overall fluorescent properties of Venus based on the location of this group and surrounding environment. Venus^{H148azF} and Venus^{Y145azF} exhibit photo-switching (switch off) and this could be a result of cross-linking with the CRO or surrounding residues. Venus^{Y66azF} displays photo-switching (switch on) presumably due to the removal of the azide group's quenching properties. Photolysis of Venus^{Y203azF} is predicted to induce

local conformational changes and remove the negative effect on CRO maturation, hence improving the absorbance and fluorescence spectral properties.

Based on the above findings the position of photoactivable ncAA to the CRO and surrounding environment define the final fluorescent properties of the target fluorescent protein. Mutation at same residue in two different versions of avGFP (sfGFP and Venus) with azF have different effects on fluorescent properties and protein as observed with mutations at residue 66,148 and 203. Such a result suggests that incorporation of azF at the same residue in other avGFP variants (such as blue, cyan and red R10-3 variants) could generate further novel fluorescent properties. Additionally, replacement of a residue around the CRO with no direct contact with CRO such as residue 145 azF affects the protein's properties and generates a variant with photo-switching properties. This result could open avenues to generate more new variants by mutating other residues around CRO with azF.

Chapter seven

Creation of functionally linked artificial heterodimers by Click chemistry

7. Creation of functionally linked artificial heterodimers by Click chemistry

7.1 Introduction

Oligomerisation of proteins (or quaternary structure) to form high-order complexes is a natural process in biology. Indeed, most proteins form temporary or permanent complexes as a part of their function, such as transcription factors and cell receptors (Ali and Imperiali 2005). Proteins that form permanent complexes have benefits compared to monomers in terms of stability, resistance to denaturation and degradation, complexity and cooperative functionality (eg multiple active sites) (Monod *et al.* 1964; Larsen *et al.* 1998). To take advantage of these benefits, creation of artificial protein complexes (homo or hetero complexes) has become of interest within the field of protein engineering. Constructing protein oligomers or dimers opens up avenues to produce novel structures with new properties that cannot be sampled by monomeric forms (Worthy *et al.* 2019). A variety of approaches that mimic natural mechanisms for oligomerisation have been implemented to create artificial protein complexes. These include disulphide bridging (Leibly *et al.* 2015), metal ion bridging (Song and Tezcan 2014), β-sheet-β-sheet interactions (Khakshoor *et al.* 2010), helix-helix interactions (Huang *et al.* 2014), and domain-swapped oligomers (Kobayashi and Ryoichi 2017). However, these approaches are limited to the nature of the proteins themselves and required some modification in monomeric protein before applying modifications, which may deactivate the proteins' native functionalities. Additionally, the yield of the protein complex usually low (Kim *et al.* 2015).

The most common of the above approaches for dimerisation is formation a disulphide bridge. This is a powerful method with proteins that lack cysteine in their natural sequences (otherwise misfolding could occur by forming disulphide bridges between introduced mutant cysteine and existing cysteine residues). An additional issue is lack of selective orthogonality, which is critical for heterodimerisation. As critical thiol chemistry is present in the starting monomers, the sulphydryl group linking the two proteins together can generate a mixed population of hetero and homodimers that may be difficult to separate. For example, monomers A-SH and B-SH intended to link together as A-S-S-B, will also produce a population of dimers A-S-S-A and B-S-S-B. Furthermore, this technique suitable only *in vitro* due to the reducing environment of the

cytosol of living cells (Van Amsterdam *et al.* 2002) and these dimers could be disassociated as a result of reduction and cleavage of the S-S bond in the presence of reducing agents.

While most of the above techniques can be used to successfully create a protein complex, one main feature of artificial protein oligomer structures is that they usually lack functional synergy between the individual monomers, so complexes manifest the properties of the starting components. The link is normally a passive linkage so two remain functionally independent. So, it is simply a process of assembly with no communication beyond the interface region, rather than true, full structural integration. For example, a variety of GFP oligomers (homooligomers) have been generated using both metal coordination and disulphide bridging techniques but these oligomers showed little communication between the functional centres (Leibly *et al.* 2015). Thus, creating a complex of unrelated protein monomers (hetero complex) with functional synergy between the monomers is a big challenge in terms of integrating them both structurally and functionally.

One approach that has been used to couple very different proteins is a process called domain insertion (Aroul-Selvam *et al*. 2004; Kanwar *et al.* 2013). This involves inserting one protein within another so that events in one protein are coupled to the other. For example, Edwards *et al.* (2008) have successfully generated a protein complex with functional synergy by linking cytochrome b_{562} (cyt b_{562}) and TEM β-lactamase via domain insertion. This complex generates haem dependent ampicillin resistance in bacteria (Edwards *et al.* 2008). While domain insertion is a powerful approach, inserting one protein into another can result in a destabilised complex and care needs to be taken with linker sequences; too long and flexible and there is no functional coupling; too short and rigid can cause stability issues (Arpino *et al.* 2012b).

As mentioned in chapter five, I have shown that a protein can be labelled at a desirable position through strain promoted azide alkyne cycloaddition (SPAAC) by incorporation of ncAA with azide chemistry that can undergo a bioorthogonal reaction with a strained cyclooctyne moiety present in a fluorescent dye. What if the cyclooctyne chemistry could also be genetically encoded and incorporated as an ncAA in a similar manner to the azide chemistry? Whole proteins could then be linked together via SPAAC so allowing the construction of artificial protein dimers (Figure 7.1) by forming stable triazole linkage at

desirable position near to the functional centre of protein. SPAAC is a one to one, excellent selectivity reaction without needing any toxic cofactors or catalysts such as copper catalysed azide-alkyne reaction (CuAAC). This chapter focuses on using SPAAC as an alternative method to overcome the issues of classical techniques to generate functionally linked heterodimers, in this case fluorescent proteins. This builds on previous work in the Jones lab concerning use SPAAC to generate homodimers of sfGFP (Worthy 2018).

Venus and super-folder green fluorescent protein (sfGFP) were used in this study as target proteins (sequence alignment shown in Figure 7.2) to create heterodimers and study the interactions between two functional centres, given their capacity to communicate through processes such as FRET (Bajar *et al.* 2016) (Figure 7.3). Azide group as a Click chemistry was introduced into Venus by the genetically encoded ncAA *p*-azido-L-phenylalanine (azF). A strained cyclic alkyne, as the orthogonal Click chemistry to azide group, was introduced into sfGFP by the genetically encoded ncAA s-cyclooctyne-L-lysine (SCO-K). The azide group in Venus reacts selectively with strained cyclic alkyne in sfGFP to form a stable triazole linkage between Venus and sfGFP and generate a dimer as shown in Figure 7.1. The result is that two heterodimers were generated termed GFVen^{148} (Venus^{H148azF} and sfGFP$^{H148SCO\text{-}K}$) and GFVen^{204} (Venus^{Q204azF} and sfGFP$^{Q204SCO\text{-}K}$) with fluorescent properties different from each of the monomeric units.

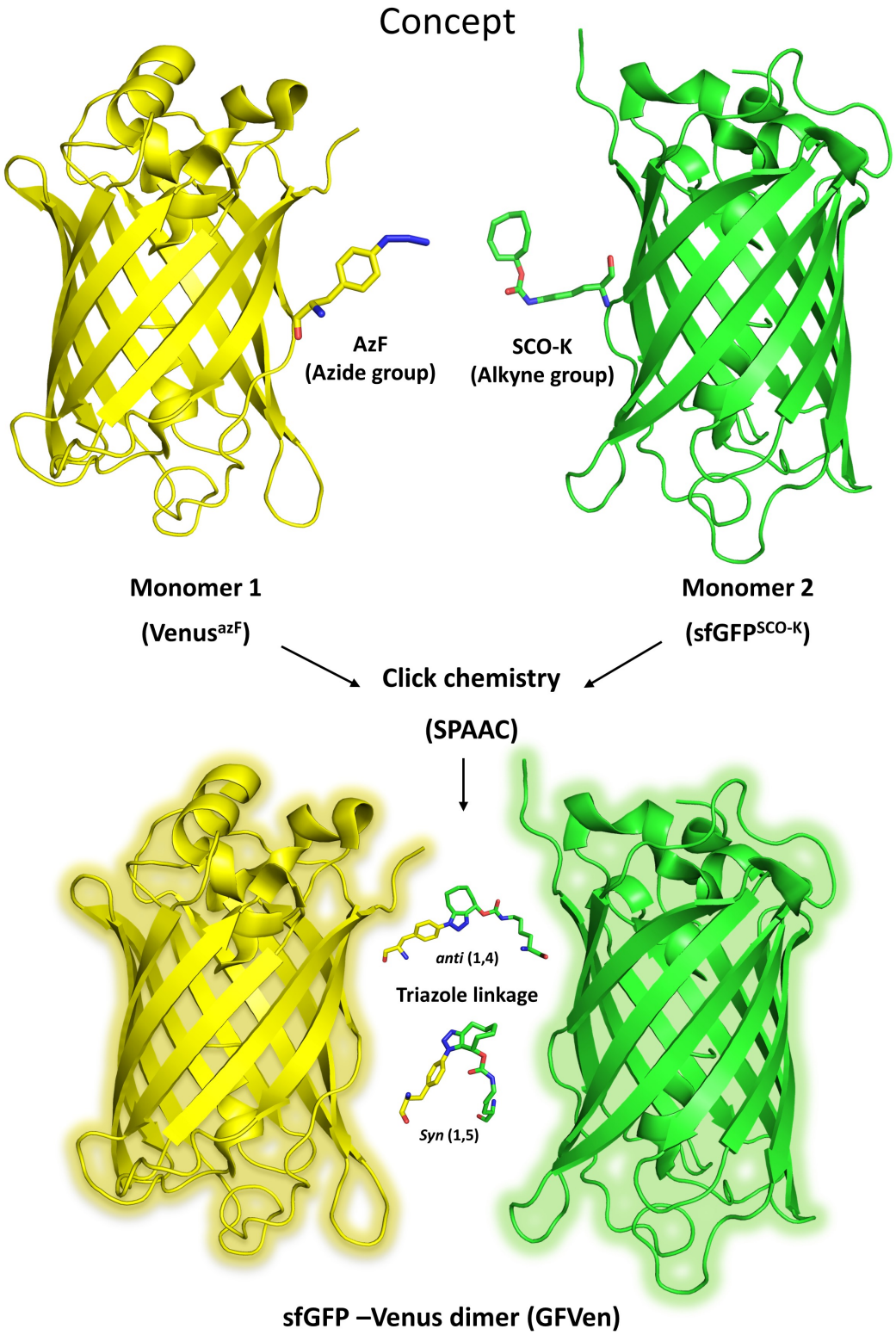

Figure 7.1. Concept of SPAAC reaction for heterodimer construction.
SPAAC reaction between genetically modified fluorescent protein with ncAAs (azF or SCO-K), Venus with azido group coloured yellow and sfGFP with alkyne group coloured green. Heterodimer generated via triazole link (syn 1,5 or anti 1,4 linkage based on the sfGFP$^{azF+SCO-K}$ homodimer; PDB code 5NI3 and 5NHN) between two monomeric proteins.

```
Venus      -35 MRGSHHHHHHGMASMTGGQQMGRDLYENLYFQGSSMVSKGEELFTGVVPI      14
                                                 X|||||||||||||
sfGFP                                          1-MSKGEELFTGVVPI      14

Venus       15 LVELDGDVNGHKFSVSGEGEGDATYGKLTLKLICTTGKLPVPWPTLVTTL      64
               |||||||||||||||X|||||||X||||||X||||||||||||||||||
sfGFP       15 LVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTL      64

Venus       65 GYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKF     114
               X||X|||X|||||||X||||||||||||||||||X|||||X|||||||||
sfGFP       65 TYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKF     114

Venus      115 EGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKAN     164
               ||||||||||||||||||||||||||||X||||||||||||||||||||
sfGFP      115 EGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKAN     164

Venus      165 FKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKDPNEK     214
               ||||||X|||X|||||||||||||||||||||||||||X||X||||||||
sfGFP      165 FKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEK     214

Venus      215 RDHMVLLEFVTAAGITLGMDELYK--------     274
               |||||||||||||||X|||||||
sfGFP      215 RDHMVLLEFVTAAGITHGMDELYKLEHHHHHH     246
```

Figure 7.2. Sequence alignment between Venus and sfGFP.
The mutated H148 and Q204 are shown in bold and highlighted in yellow. Blue, red and green highlighted residues correspond to His tags, TEV protease cleavage motif and CRO forming residues, respectively. Residues that differ between the two proteins are highlighted by an X.

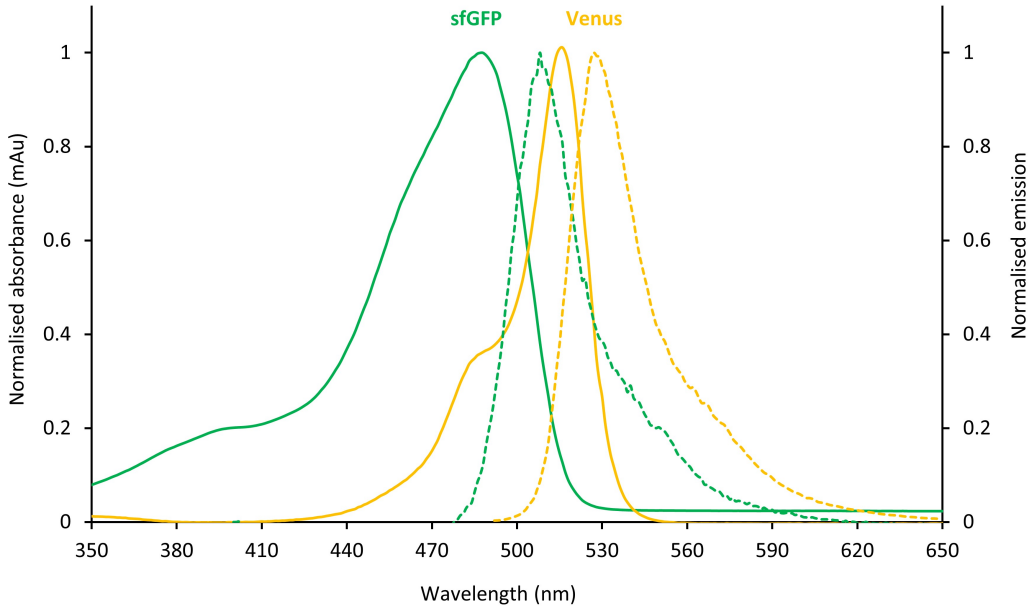

Figure 7.3. Overlapping spectral properties of sfGFP and Venus.
A. Normalised absorbance spectra (normalised to high λ_{max} value for each protein) shown as solid lines. Normalised emission spectra (normalised to high λ_{em} value for each protein) shown as dashed lines. Green colour corresponds to the sfGFP and yellow colour corresponds to Venus. The emission spectra of sfGFP overlap the absorbance spectra of Venus.

7.2 Results and Discussion

7.2.1 Design of the linking residues for dimerisation

Designing the sites for linking the two proteins is a critical step in dimerisation, especially when aiming to build functionally linked dimers. Residues need to be at least partly surface exposed for the chemistry to be accessible for forming the cross link. The link should ideally be close to the functional centre or allow the functional centres to come into close contact. Also, for an intimate structural interaction, the protein surfaces at the interface site should be compatible to some degree (e.g. local hydrophobic patches, H-bond acceptor/donors available). In previous studies, different surface residues were chosen for cysteine mutation to create a GFP disulphide homodimer including: K26C, D102C, D117C, Q157C, and D190C. These residues were successful in creating a dimer but there was no evidence of any communication between the functional centres of GFP (Leibly *et al.* 2015). Previously published studies on FRET donor–acceptor pairs in sfGFP showed that among four different residues mutated to azF (K26, E34, E132 and Q204) the residue Q204azF exhibited about 90% FRET after click with the dibenzylcyclooctyne (DBCO) compatible dye by SPAAC (Reddington *et al.* 2012). Another study by Hartley *et al.*(2016) showed that replacement of H148 with azF in sfGFP and click with dibenzylcyclooctyne-amine (DBCO-amine), improved quantum yield which indicates the communication between the functional centre of sfGFP and DBCO-amine (Hartley *et al.* 2016).

Initially the H148 and Q204 residues were chosen as a dimerisation point because they are accessible for small Click chemistry tool, which means they are free to react, and associated with the functional centre. However, in the above previous studies, 148 and 204 sites were used as click point with small Click chemistry tool which might be not compatible with protein due to protein-protein interface interaction. It has recently been shown that *in silico* protein-protein interface prediction approaches such as the ClustPro server (cluspro.org) (Vajda *et al.* 2017; Kozakov *et al.* 2017), can be used to identify mutually compatible dimer interface regions. Analysis of the models created by ClustPro server (Section 2.11.2), showed that residues H148 and Q204 were routinely found at the predicted dimer interface in both monomers in all top 10 models (Figure 7.4). As a result,

these residues were chosen to replace with the two Click compatible ncAAs, azF (azide group) and SCO-K (strained alkyne group) for dimerisation.

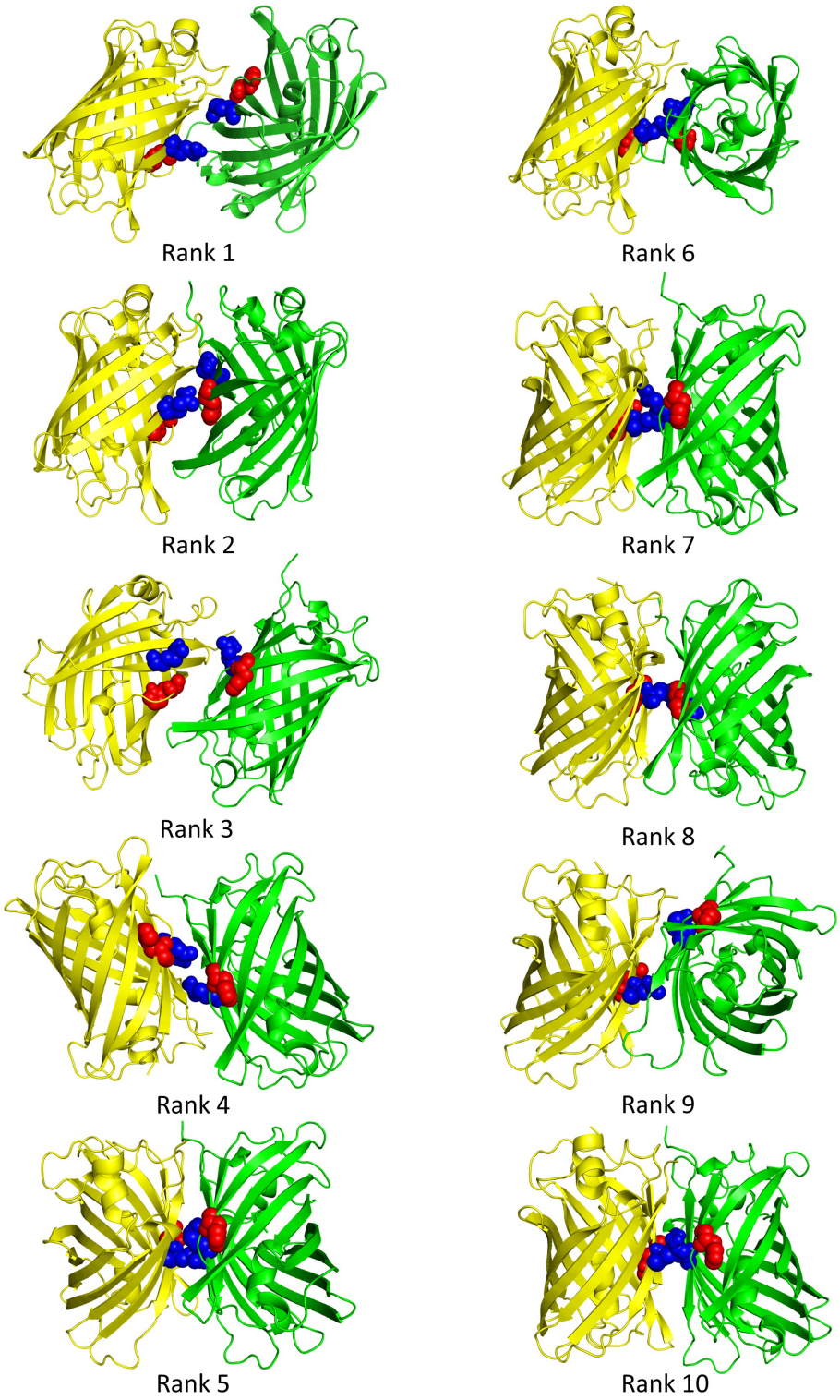

Figure 7.4. Top 10 models generated by ClustPro server.
Dimer models were generated and ranked by the ClustPro server. Venus coloured yellow (PDB code 1myw) and sfGFP coloured green (BDP code 2b3p). Residues H148 and Q204 shown as spheres red and blue respectively.

7.2.2 Expression of monomers with Click compatible ncAAs at desired positions

As stated previously (chapter 2; Section 2.3.4), genetically encoded ncAAs require a reprogrammed codon to encode the incorporation of an ncAA by using the amber codon (TAG) suppression technology. Mutation of the codons for H148 and Q204 in Venus to TAG were prepared as described in Section 6.2.2. The pBAD vector carrying either the *sfGFP*^{148TAG} or *sfGFP*^{204TAG} gene were donated by Harley Worthy and Jacob Pope (DDJ lab). Venus azF variants (Venus^{H148azF} and Venus^{Q204azF}) were previously expressed and purified as outlined in Section 6.2.4. All the above stages were carried out in the dark to prevent photolysis of azF by ambient light. To assess the compatibility of H148^{azF} and Q204^{azF} for applications involving the SPAAC reaction, a fluorescence strained alkyne-containing fluorescent dye, DBCO-TAMRA dye was used as described in Section 2.5.1. Results showed that both residues can react with the DBCO-TAMRA dye (Figure 7.5). These results indicate that the azide group is available to react, at least with small dye molecules, and is not buried and inaccessible. H148^{SCO-K} and Q204^{SCO-K} in sfGFP have previously been reported as Click chemistry compatible residues and successfully used to generate a sfGFP homodimer with a sfGFP$^{H148,204\ azF}$ variants (Worthy 2018).

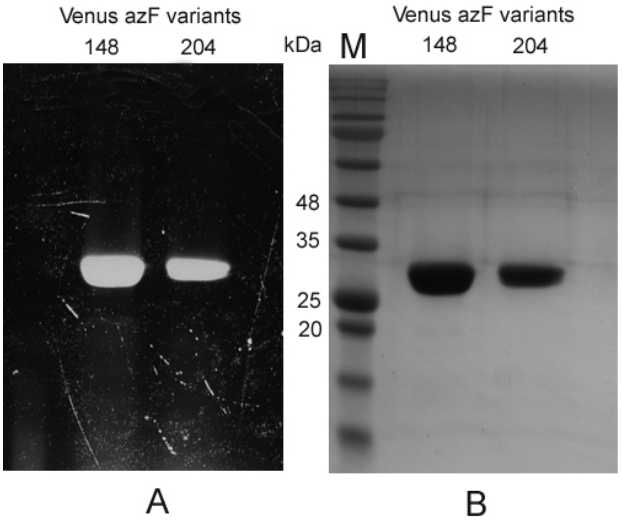

Figure 7.5. Labelling efficiency of azF-containing Venus variants.
A. SDS-PAGE (12.5%) gel imaged under UV light before staining showing fluorescent protein bands which indicate the click reaction between the protein and compatible fluorescence dye DBCO-TAMRA. **B.** Gel after staining with Coomassie blue R250 showing the protein bands. M, molecular weight markers, with weights of the relevant bands shown in the figure.

To incorporate SCO-K into sfGFP, Top10^{TM} electrocompetent *E. coli* were co-transformed with pBAD vector contains *sfGFP* gene with desired TAG mutations (encoding residue 148 or 204) and the pEVOL vector carrying the engineered tRNA/aaRS pair for SCO-K incorporation, as described in Section 2.2.10.2. The sfGFP^{SCO-K} variants (sfGFP$^{H148SCO-K}$ and sfGFP$^{Q204SCO-K}$) were prepared in a larger scale (500 ml), as described in Section 2.3.4.3, and extracted as described in Section 2.3.6.3.

7.2.2.1 Purification of sfGFP^{SCO-K} variants

In preparation for dimerisation, purified monomer protein preparations are required. The sfGFP^{SCO-K} variants were prepared as described in Section 2.3.4.3 and purified first by nickel affinity chromatography (Section 2.3.8.1). The presence of C-terminal hexahistidine-tag in the sfGFP variants allows the protein to bind to the nickel column. After cell lysis by French Press (Section 2.3.6.3), the His tagged protein was purified from the cell lysate as described in Section 2.3.8.1. Three elution fractions were collected from the ProtinoR Ni-TED 2000 Packed affinity Columns. Analysis by SDS-PAGE showed that high purity was achieved as shown in Figure 7.6A and D; most of the protein eluted in the first fraction. Elution fractions containing sfGFP^{SCO-K} variants were pooled and subjected to a second step of purification by size exclusion chromatography (SEC).

The pooled fractions after nickel affinity purification were concentrated to 2 ml and loaded onto a Hiload^{TM} 16/600 Superdex™ S75 pg (Section 2.3.8.3). The elution of protein was monitored by absorption at 280 nm as shown in Figure 7.6B and E. Two peaks were observed; a minor peak that eluted in the column void volume (~ 40 ml), which likely corresponds to the protein aggregates and, the second major peak that corresponds to the sfGFP^{SCO-K} variant, which was confirmed by SDS-PAGE analysis (Figure 7.6.C and F). Fractions containing pure protein were pooled and concentrated for dimer production experiments.

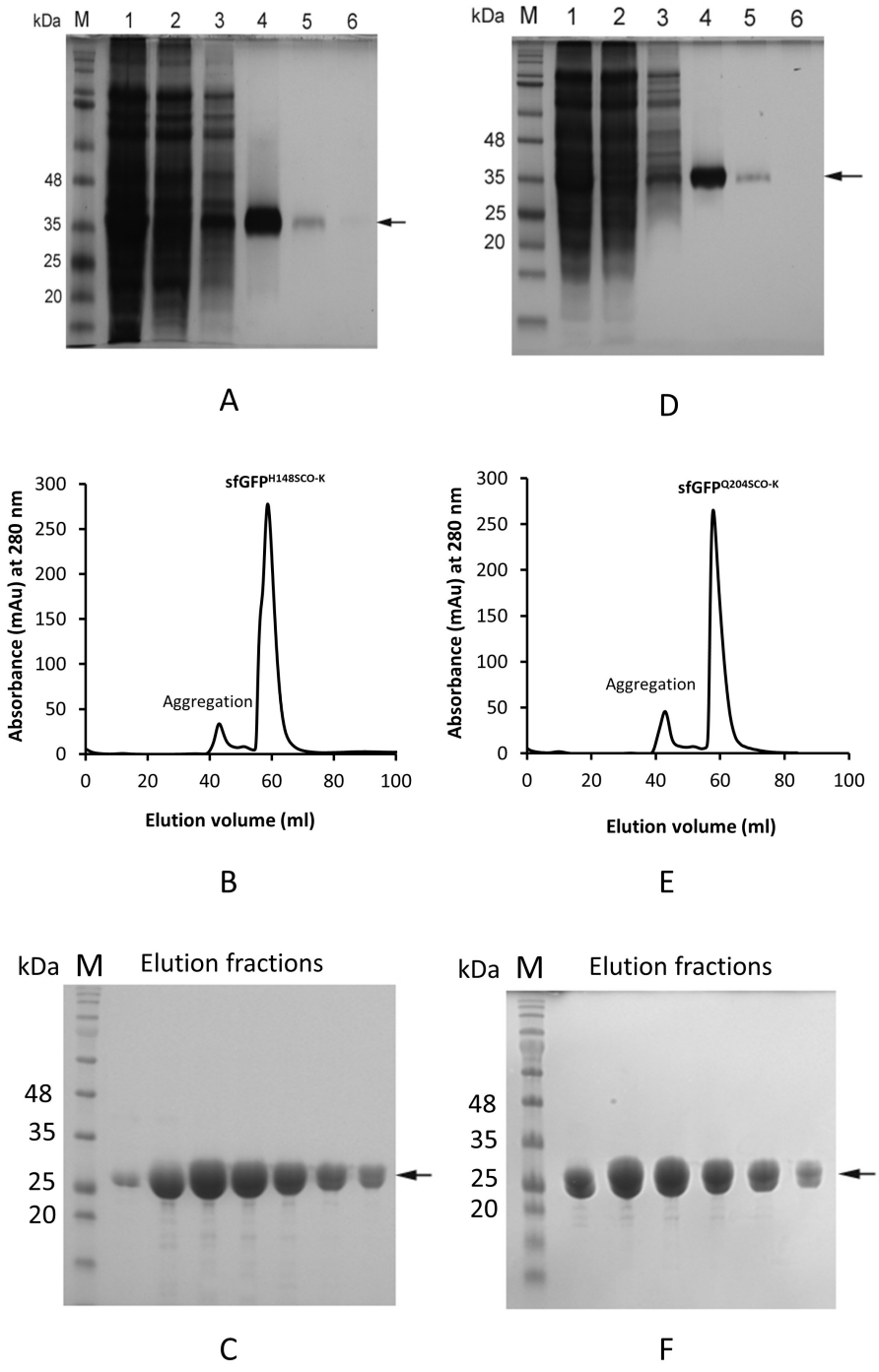

Figure 7.6. Purification of the sfGFP^{SCO-K} variants.
A, B and **C** represent the purification of sfGFP$^{148SCO-K}$. **D, E** and **F** represent the purification of sfGFP$^{204SCO-K}$. **A** and **D** are SDS-PAGE (12.5%) analyses of the fractions eluted on nickel affinity chromatography. Lane M corresponds to the protein marker. Lane 1 corresponds to whole cell lysate, lane 2 corresponds to follow-through fraction, lane 3 corresponds to wash fraction and lanes 4,5 and 6 correspond to elution fractions respectively. **B** and **E** show the chromatogram from SEC chromatography (Hiload^{TM} 16/600 Superdex™ S75 pg column), with absorbance (280 nm, black line). First peak represents protein aggregates that eluted in the void volume of the column and the second peak represented sfGFP^{SCO-K} protein. **C** and **F** are SDS-PAGE (12.5%) analyses of elution fractions from SEC (second peak). Black arrows indicate the sfGFP^{SCO-K} protein.

7.2.3 Creation of artificial heterodimers sfGFP-Venus (GFVen)

To implement dimer design (Section 7.2.1), it was necessary to demonstrate that heterodimers of sfGFP and Venus could be formed by Click chemistry. To do this, a 1:1 molar ratio of purified azF-Venus (azide group) and the equivalent purified SCO-K-sfGFP (alkyne group), were mixed together at a concentration of 150 µM and incubated in the dark with gentle mixing at room temperature for 24 h. Two dimer-types were tested: (1) sfGFP$^{148SCO-K}$ with Venus^{148azF}, generating a dimer termed GFVen^{148}; (2) sfGFP$^{204SCO-K}$ and Venus^{204azF}, generating the dimer termed GFVen^{204}. To confirm the conjugation of the two proteins and dimer formation, samples of monomers were compared with the click reaction mixture by SDS-PAGE. Results showed that a high level of dimer (a band with an estimated molecular weight of ~57 kDa) was successfully formed after 24 h incubation as can been seen in Figure 7.7A and B. Successfully formed dimer was subjected to purification to separate it from the monomers.

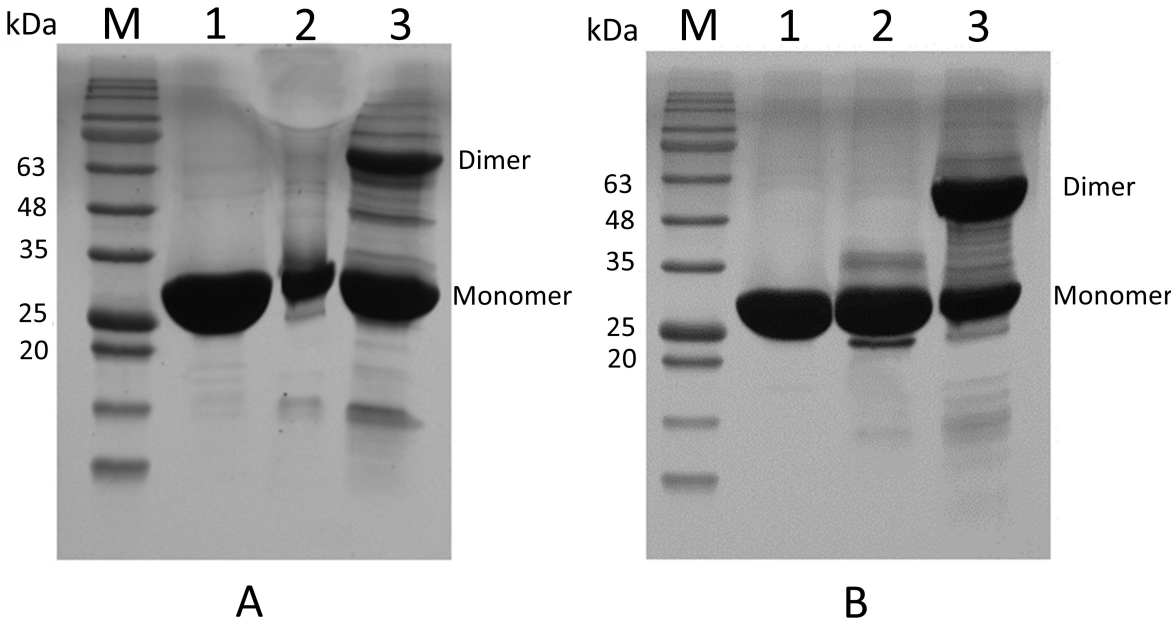

Figure 7.7. Constructing sfGFP-Venus dimers.
SDS-PAGE (12.5%) analysis of heterodimers (**A**) GFVen^{148} production **(B)** GFVen^{204}. Lane M corresponds to the protein molecular weight marker. Lane 1 corresponds to the Venus ^{azF} variant shown as monomer, lane 2 corresponds to the sfGFP^{SCO-K} variant shown as monomer and lane 3 corresponds to click reaction mixture and showed the monomers and the dimer as labelled.

7.2.4 Purification of heterodimer from the click reaction mixture

To characterise the dimers and study the communication between the functional centres (the chromophores), highly pure dimers are required to avoid the monomers contributing to any observed signal. The concentrated mixture of the click reaction mixture was diluted and purified by SEC (Section 2.3.8.3). Results from the SEC purification of GFVen[148] showed two major peaks corresponding to dimer and monomers respectively, and several minor peaks corresponding to contaminations (Figure 7.8). The same elution profile was observed with GFVen[204]. SDS-PAGE analysis of elution fractions for the major peaks showed some fractions have a high purity of dimer and others have both dimer and mixture of monomers (Figure 7.9). Fractions with high dimer purity were collected and pooled. Fractions with both dimer and monomers were pooled and re-purified by a second application to the SEC, after 2 rounds of SEC purification, high purification of both dimers was achieved, as can be observed in Figure 7.10.

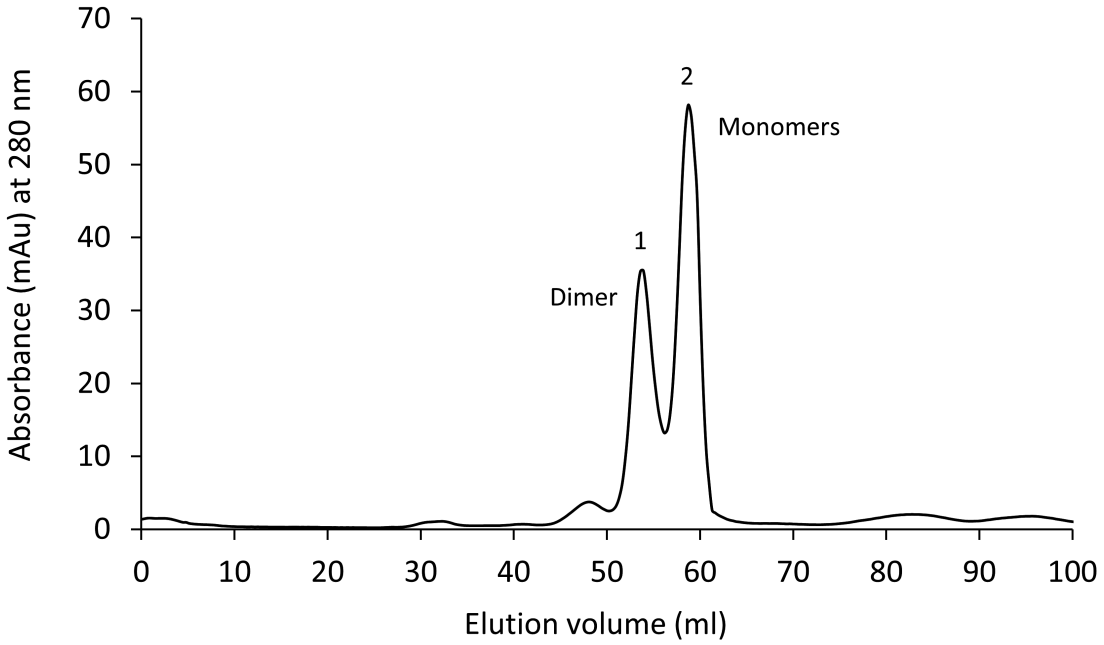

Figure 7.8. Separation of GFVen[148] from monomer by SEC chromatography.
SEC chromatogram of GFVen[148] purification with absorbance (280 nm, black line) shown. SEC was performed with a Hiload™ 16/600 Superdex™ S75 pg column. The first major peak represented the heterodimer and the second major peak represented the monomers of Venus ^{azF} and sfGFP$^{SCO\text{-}K}$ proteins (see figure 7.9 for SDS PAGE analysis).

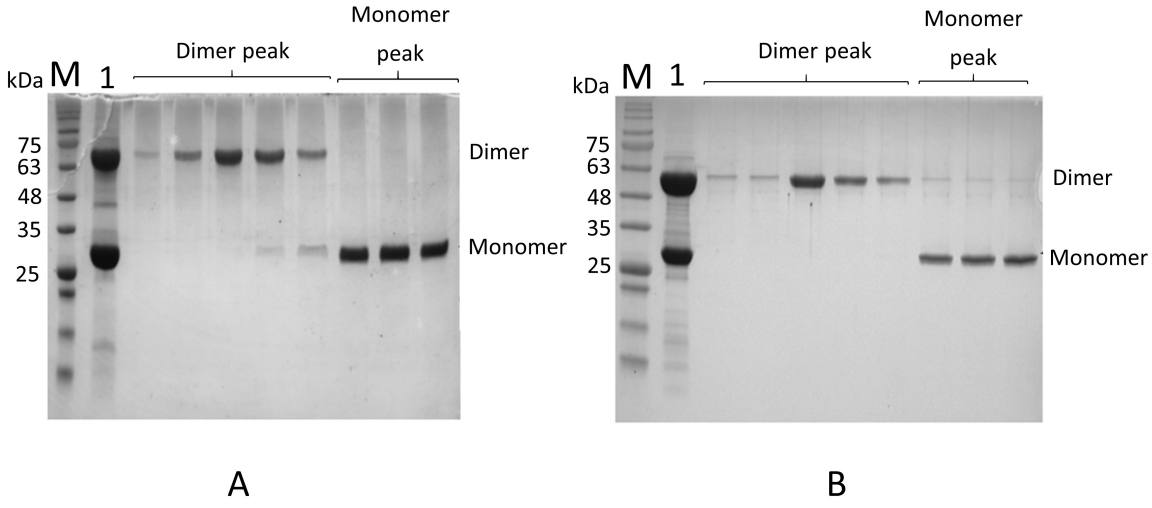

Figure 7.9. Analysis of dimer separation by SEC.
A. SDS-PAGE (12.5%) analysis of elution fractions from SEC purification of GFVen^{148} dimer. **B.** SDS-PAGE (12.5%) analysis of elution fractions from SEC purification of GFVen^{204} dimer. Lane M corresponds to the protein marker of known molecular weights. Lane 1 corresponds to the mixture of the click dimerisation reaction before loading onto the SEC column. The dimer corresponds to the first major elution peak and monomer corresponds to the second major elution peak on the chromatogram SEC profile.

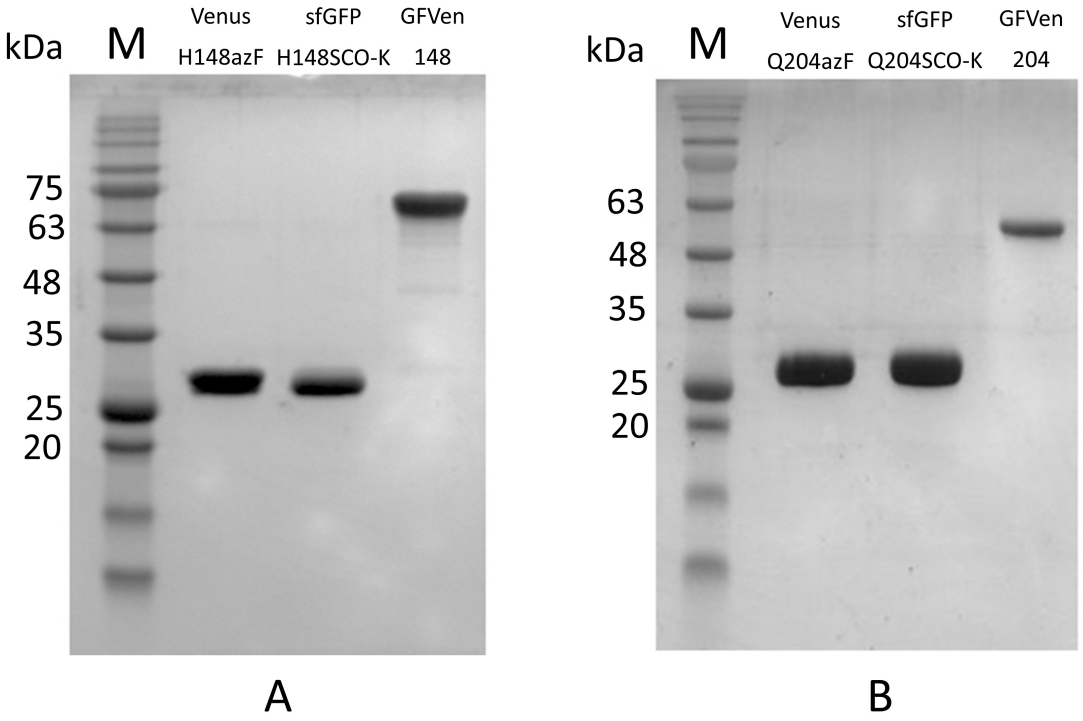

Figure 7.10. Purity of isolated Venus, sfGFP monomers and GFVen dimers.
SDS-PAGE (12.5) analysis of pure: **A**. Venus^{148azF} and sfGFP$^{148SCO-K}$ monomers and GFVen^{148} dimer; **B.** Pure Venus^{204azF} and sfGFP$^{204SCO-K}$ monomers and GFVen^{204} dimer.

7.2.4 Mass spectrometric analysis of dimers

Pure dimers were analysed by mass spectrometry to confirm they were covalently linked and that they were composed of the correct monomer starting points. Samples of 10 μM for each dimer were analysis as described in Section 2.4.11. Results from mass spectrometry showed major peaks at 58827 Da and 54924 Da, corresponding to GFVen^{204} and GFVen^{148} respectively (Figure 7.11A-B). For the GFVen^{204} dimer, the calculated mass was 58820 Da (Venus^{204azF}, 30843 Da; sfGFP$^{204SCO-K}$, 27977 Da). The measured mass was 58827 Da, a difference of 7 Da higher than the calculated mass (0.012% difference). Based on this result it can be concluded that the GFVen^{204} dimer was successfully linked covalently by Click chemistry.

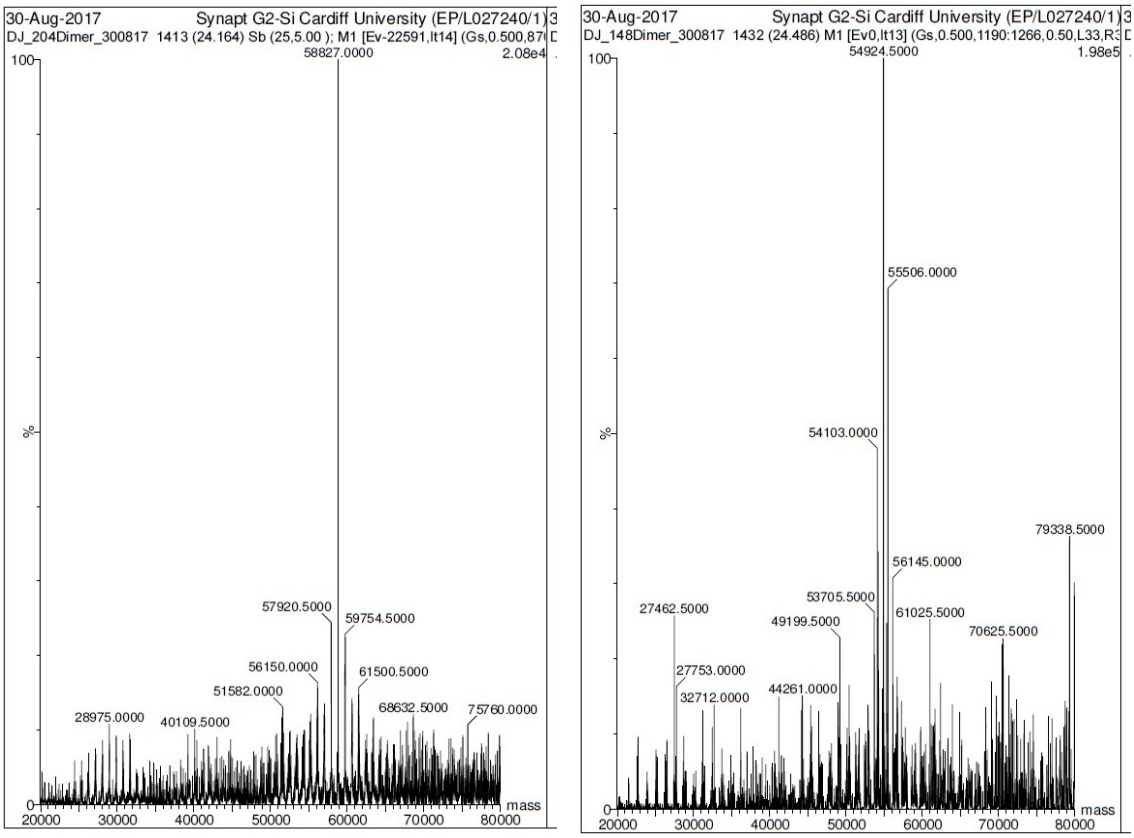

Figure 7.11. Mass spectrometric analysis of heterodimers.
A. Mass spectrometric analysis of GFVen^{204} dimer. **B.** Mass spectrometric analysis of GFVen^{148} dimer.

For GFVen^{148}, the mass spectrum showed a major peak at 54924 Da and other two peaks at 55506 Da and 54103 Da (Figure 7.11B). All these peaks are significantly different from the theoretical mass of 58802 Da (Venus^{148azF}, 30834 Da; sfGFP$^{148SCO-K}$,27968 Da) by 3891 Da, 3309 Da and 4712 Da respectively. All the measured peaks are too large to be either Venus^{148azF} (30834 Da) or sfGFP$^{148SCO-K}$ (27968 Da) and smaller than the predicted full-length GFVen^{148} heterodimer. This could be explained by processing from either N- or C-terminus of either protein. The major peak at 54924 Da is consistent with to sfGFP$^{148SCO-K}$ (27698 Da) and Venus^{148azF} minus the added N-terminal extension (MRGSHHHHHHGMASMTGGQQMGRDLYENLYFQGS, Figure 7.2; 26956 Da), with a calculated molecular mass of 54924 Da. The second peak at 55506 Da, is consistent with sfGFP$^{148SCO-K}$ and Venus^{148azF} minus the N-terminal extension (MRGSHHHHHHGMASMTGGQQMGRDLYENLYFQG, Figure 7.2; 27539 Da), with a calculated mass of 55507 Da, a difference of 1 Da higher than calculated mass. The third peak at 54103 Da is consistent with sfGFP$^{148SCO-K}$ with the loss of its C-terminal His-tag (27145 Da) and Venus^{148azF} minus the N-terminal extension (MRGSHHHHHHGMASMTGGQQMGRDLYENLYFQGS, Figure 7.2; 26956 Da), which has a calculated molecular mass of 54101 Da. Similar results have been reported by Worthy (2018) with regards to His-tag processing from the C-terminus, as he found that both monomeric units of the homodimer sfGFP^{148x2} lost the C-terminal His-tag after dimerisation as shown by mass spectrometry analysis (Worthy 2018). However, it appears that dimerisation via position 148 may have a significant effect on peptide bonds close to the His-tag, which leads to their removal. It may well be that incorporating azF at residue 148 is somehow promoting cleavage close to the termini. Given that the termini are likely to be flexible (and are absent from most crystal structures of GFP variants (Hartley *et al*. 2016)), they may come close in space at some point to residue 148. Nevertheless, based on the results above, it can be concluded that the heterodimer GFVen^{148} was successfully linked covalently via SPAAC at position 148 as predicted. Furthermore, the generation of the GFVen^{148} dimer and the GFVen^{204} dimer were confirmed independently by SDS-PAGE analysis as described in Section 7.2.3 and shown in Figure 7.7.

7.2.5 Characterisation of monomers with ncAA

7.2.5.1 Characterisation of Venus^{H148azF} and Venus^{Q204azF}

The absorbance and fluorescence spectral properties for Venus monomers containing azF need to be characterised to assess the effect that the ncAA *per se* has on the protein's fluorescent properties and how these properties change after dimerisation. In the previous chapter (chapter 6), the effect of azF incorporation at position 148 in Venus was described in detail. In brief, incorporation of azF at position 148 has a significant effect on fluorescent properties of Venus: the fluorescence intensity decreased by ~72% and the excitation wavelength shifted by 2 nm (517 nm), the molar absorbance (ε) and QY were decreased from ~95000 to ~30100 $M^{-1}.cm^{-1}$ and from 0.65 to 0.45 respectively (chapter 6; Figure 6.9).

Q204 in Venus is located in β-strand 10, is partially surface exposed and its side chain points away from the phenol moiety of the CRO (Figure 7.12). In sfGFP, Q204 has been found to play a critical role in pH stability and protein dimerisation. For example, the Q204H mutation resulted in dimerisation of sfGFP and increased pH stability (Roberts *et al.* 2016). Incorporation of azF at residue 204 (generating the variant from now on termed Venus^{Q204azF}), produced a highly fluorescent protein with only slight differences in spectral properties to Venus^{WT} (Figure 7.13). The λ_{max} (λ_{ex}) = 515 nm and λ_{em} = 528 nm. The intensity peaks of both excitation and emission were essentially the same as for Venus^{WT}, with only slight decreases of 2.4% and 4.5% respectively, as shown in Figure 7.13B and C. The molar absorbance and QY also dropped slightly more to 87600 $M^{-1}cm^{-1}$ and 0.42 respectively compared with Venus^{WT}. Photolysis of Venus^{Q204azF} with UV-light showed no effect on fluorescent properties. This finding indicates that the side chain of azF had little effect on the CRO environment and, overall, incorporation of azF at position 204 had a little effect on the fluorescent properties of Venus.

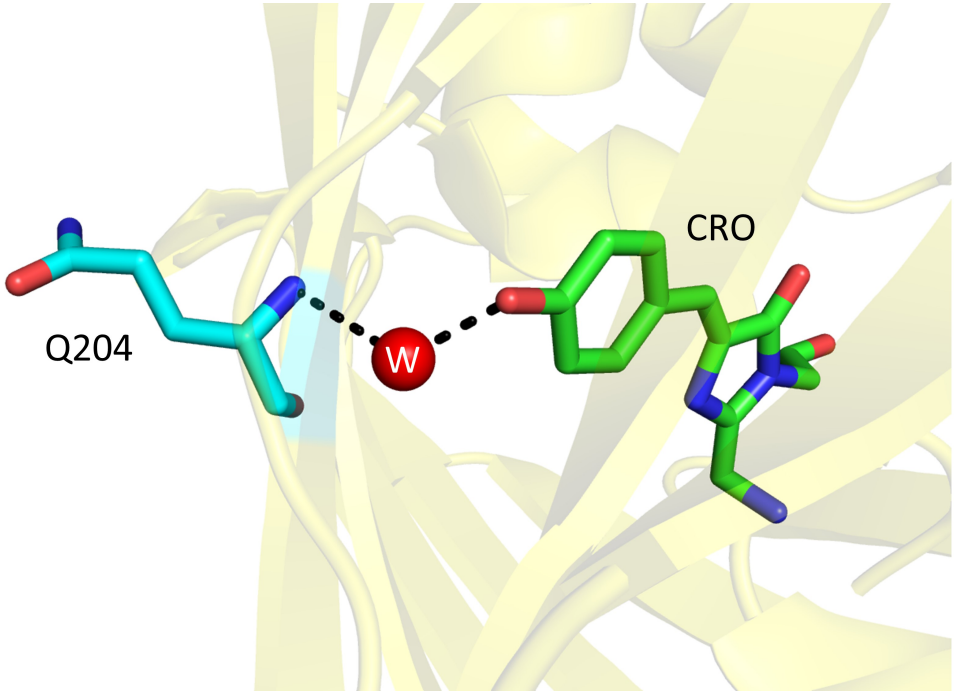

Figure 7.12. Structural position of Q204 in Venus.
Gross structure of Venus highlighting residue Q204 and CRO as cyan sticks and green sticks respectively. Water molecule (W) coloured red.

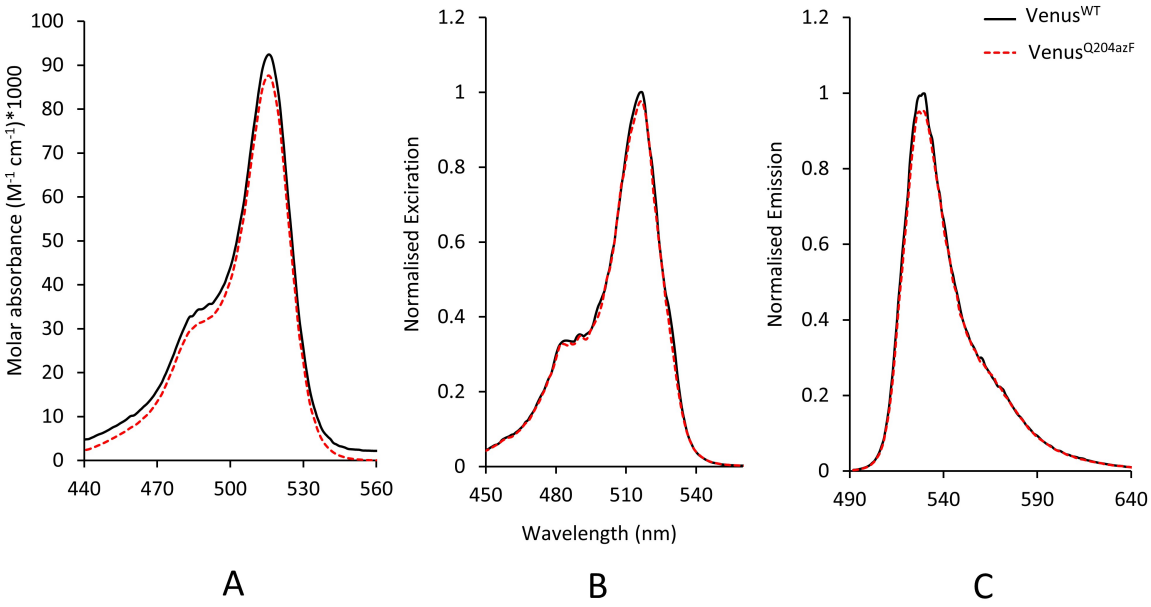

Figure 7.13. Absorbance and fluorescence spectra of Venus^{Q204azF} compared to Venus ^{WT}.
A) Absorbance spectra shown as molar absorbance. **B**) Normalised excitation spectra (emission monitored at 530 nm). **C**) Normalised emission spectra (excitation at 515 nm). Excitation and emission spectra were normalised to the maximum values of Venus^{WT}.

The ability of Venus^{Q204azF} to click with DBCO-TAMRA fluorescence dye (Figure 7.5) indicates the presence of a free azide group pointing away from the core of the protein and accessible for reaction with a strained alkyne. Previous work on sfGFP^{Q204azF} shows that Q204 is accessible for forming the cross link with DBCO compatible dye by SPAAC and exhibited about 90% FRET (Reddington *et al.* 2012). In 2017, Marth *et al.* demonstrated the ability of sfGFP to link into a modified signal strand DNA (containing a terminal strained bicyclononyne) via Q204azF (Marth *et al.* 2017). Thus, in keeping with previous results, Q204 is amenable to modification with a strained alkyne.

7.2.5.2 Characterisation of sfGFP$^{H148SCO-K}$ and sfGFP$^{Q204SCO-K}$

In chapter 6, results showed that incorporation of ncAA in Venus at different positions has a different effect on the protein in terms of spectral and fluorescent properties. So, the effect of SCO-K on sfGFP needed to be established before characterising the heterodimers. Based on the results of azF incorporation at 148 and 204 positions in Venus, it is predicted that incorporation will have a similar effect on the chromophore and fluorescence. As previously mentioned, the H148 residue is involved in a hydrogen bond network with the CRO that promotes formation of the CRO B state. Replacement of H148 with SCO-K is predicted to disrupt this network and the population of the CRO B state, similar to azF incorporation. The effects of incorporation of SCO-K at either of the two residues (148 and 204) in sfGFP have been previously studied in our lab by Harley Worthy. His findings showed that incorporation of SCO-K at position 204 in sfGFP had no effect on spectral properties of sfGFP apart from the reducing molar absorbance by about 20% compared with sfGFP^{WT} (49,000 $M^{-1}cm^{-1}$ to 40,000 $M^{-1}cm^{-1}$;Worthy, 2018;Worthy *et al.* 2019).

Incorporating SCO-K at position 148 in sfGFP had a complex effect on sfGFP. It altered the fluorescence and absorbance significantly, with spectral features of the neutral protonated state CRO A (λ_{max} 395 nm) dominating (Worthy 2018), as shown in Table 7.1. As mentioned, the Y66 residue in GFP can be present in two protonation states, based on the ionisable phenol OH group. A charged state (CRO B), which excites at 475 nm and a neutral, protonated state (CRO A), which excites at 395 nm (Tsien 1998). The result of the H148SCO-K mutation in sfGFP is to switch the protonation state of the CRO to the A form.

The spectral properties of both sfGFP$^{H148SCO-K}$ and sfGFP$^{Q204SCO-K}$ were determined for comparison with heterodimer spectra. Results of spectral properties of both sfGFP^{SCO-K} variants, as shown in Figure 7.14, were similar to previous spectral properties that were determined by Worthy (Worthy 2018).

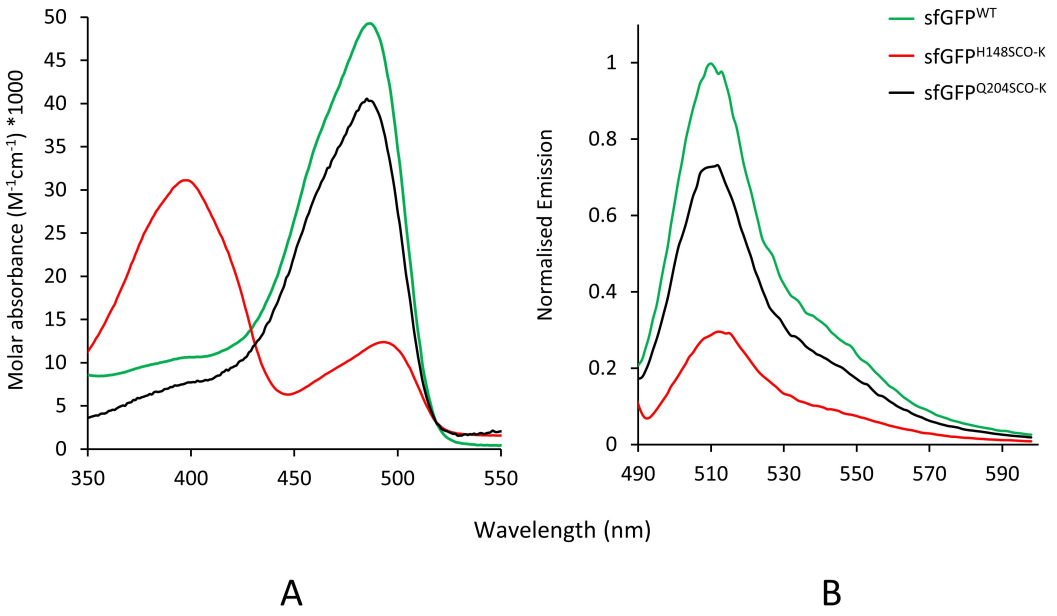

Figure 7.14. Absorbance and fluorescence spectra of sfGFP^{SCO-K} against sfGFP ^{WT}.
A) Absorbance spectra shown as molar absorbance. **B**) Normalised emission spectra (excitation at 485 nm). Emission spectra were normalised to the maximum value of sfGFP ^{WT}.

7.2.6 Characterisation of heterodimer GFVen

Based on the spectral properties of sfGFP and Venus (see Figure 7.3 above), different wavelengths were used to study the spectral properties and monitor communication between two monomers in GFVen dimers: 400 nm (excited sfGFP only); 450 nm (excited sfGFP and minor excitation of Venus); 490 nm (sfGFP λ_{max}, Venus shoulder); 510 (excited Venus and minor excitation of sfGFP), as depicted in Figure 7.15A-B.

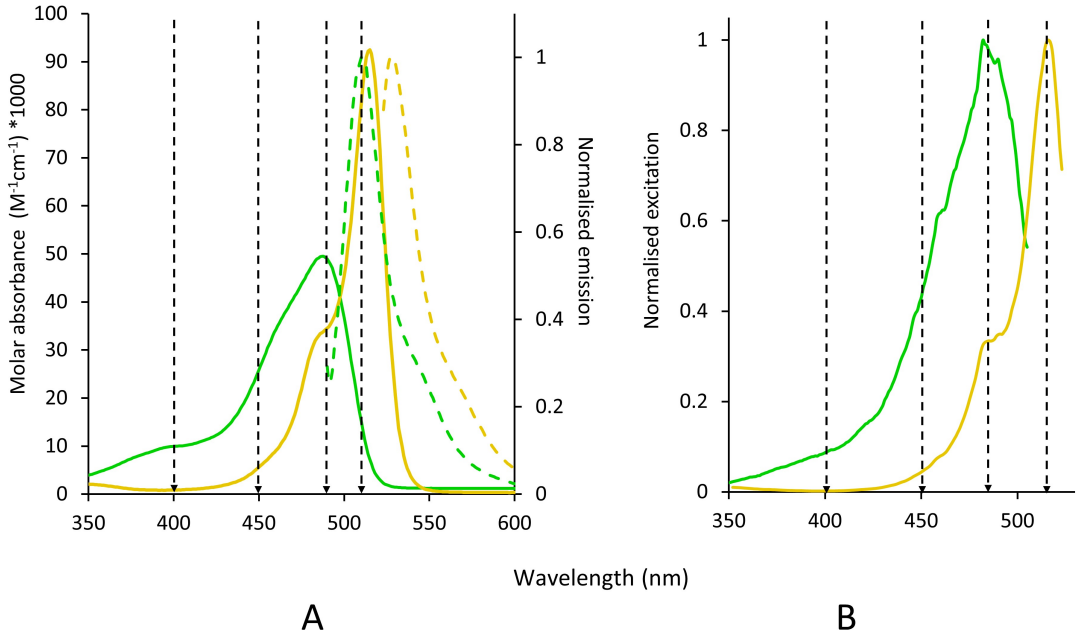

Figure 7.15. Comparison of the sfGFP (green) and Venus (gold) absorbance and fluorescence spectra.
A. Absorbance spectra shown as molar absorbance (solid lines) and normalised emission spectra (dashed lines) excitation at 485 nm for sfGFP and 515 nm for Venus (emission spectra were normalised to the maximum values of each protein). **B.** Normalised excitation spectra (on monitoring emission at 511 nm for sfGFP and at 528 nm for Venus). The black dashed arrows indicate wavelengths used to monitor communication between two monomers in GFVen dimers.

Basically, assembling of two different fluorescent proteins (having different excitation /emission wavelengths) may display classical Förster resonance energy transfer (FRET). If the functional centre is contiguous during assembly between the two fluorescent monomers and if the emission wavelength of the first fluorescent protein overlaps with the excitation wavelength of the second fluorescent protein, conditions exist such that the energy is transferred in a radiationless manner. In other words, the excitation of the first monomer leads to emission by the second monomer due to FRET (Nguyen and Daugherty 2005).

7.2.6.1 Characterisation of heterodimer GFVen^{148}

When Venus^{azF} combines with sfGFP$^{SCO\text{-}K}$ through the 148 SPAAC linkage, new properties emerged that appear to integrate the spectral features (in terms of λ_{max} and λ_{em}) of the original monomeric proteins. GFVen^{148} dimer was functionally superior to either the GFP$^{H148SCO\text{-}K}$ or Venus^{H148azF} in terms of their spectral properties (switch on) as can

be seen in Figure 7.16 and Table 7.1. Interestingly, results showed that the dimer has spectral properties intermediate between the sfGFP$^{H148SCO-K}$ and Venus^{H148azF} monomers. The major excitation wavelength (λ_{max}) of GFVen^{148} dimer is 505 nm, intermediate between individual monomers; sfGFP$^{H148SCO-K}$ (λ_{max} 492 nm) and Venus^{H148azF} (λ_{max} 517 nm). The molar absorbance increases significantly (96000 $M^{-1}cm^{-1}$ at λ_{max} 505 nm), higher than both monomeric units; Venus^{H148azF} (λ_{max} 517 nm; 30100 $M^{-1}cm^{-1}$) and sfGFP$^{H148SCO-K}$ (λ_{max} 492 nm; 17300 $M^{-1}cm^{-1}$) as can be seen in Figure 7.16A and Table 7.1. A single emission peak (λ_{em} 517 nm) that is also intermediate between the two individual monomers; sfGFP$^{H148SCO-K}$ (λ_{em} 492 nm) and Venus^{H148azF} (λ_{em} 525 nm) was observed, irrespective of the excitation wavelength (λ_{em} at 517 nm; Figure 7.16B-C); a single peak rather than a broadened or double peak was observed on excitation at the wavelength (λ_{max} 490 nm; Figure 7.15) that is capable of exciting both CROs (Figure 7.16 C). Further analysis of the emission spectra revealed that there is no double peak nor significant peak broadening (Figure 7.16A-B and Figure 7.17). Furthermore, based on the addition emission spectra of sfGFP and Venus monomers, the major peak is in fact shifted (λ_{ex} 515 nm).

Based on these results, it is believed that the two functional centres (CRO) of sfGFP$^{H148SCO-K}$ and Venus^{H148azF} have become functionally integrated in terms of fluorescence emission. These results suggest that the GFVen^{148} dimer is emitting as a single species, although it cannot be ruled out at this stage if the λ_{em} of sfGFP$^{H148SCO-K}$ together with λ_{max} for Venus^{H148azF} has been shifted so the major peak represent FRET from sfGFP to Venus. The molar absorbance equivalent to the B CRO form (490-510 nm region) increases considerably (~4-5-fold), higher than the simple additive spectra of the monomers (~1.7-fold), while the A CRO population decreases but is still observed (Figure 7.16A). This is matched by a ~4-fold increase in emission intensity on excitation at 505 nm (Figure 7.16B).

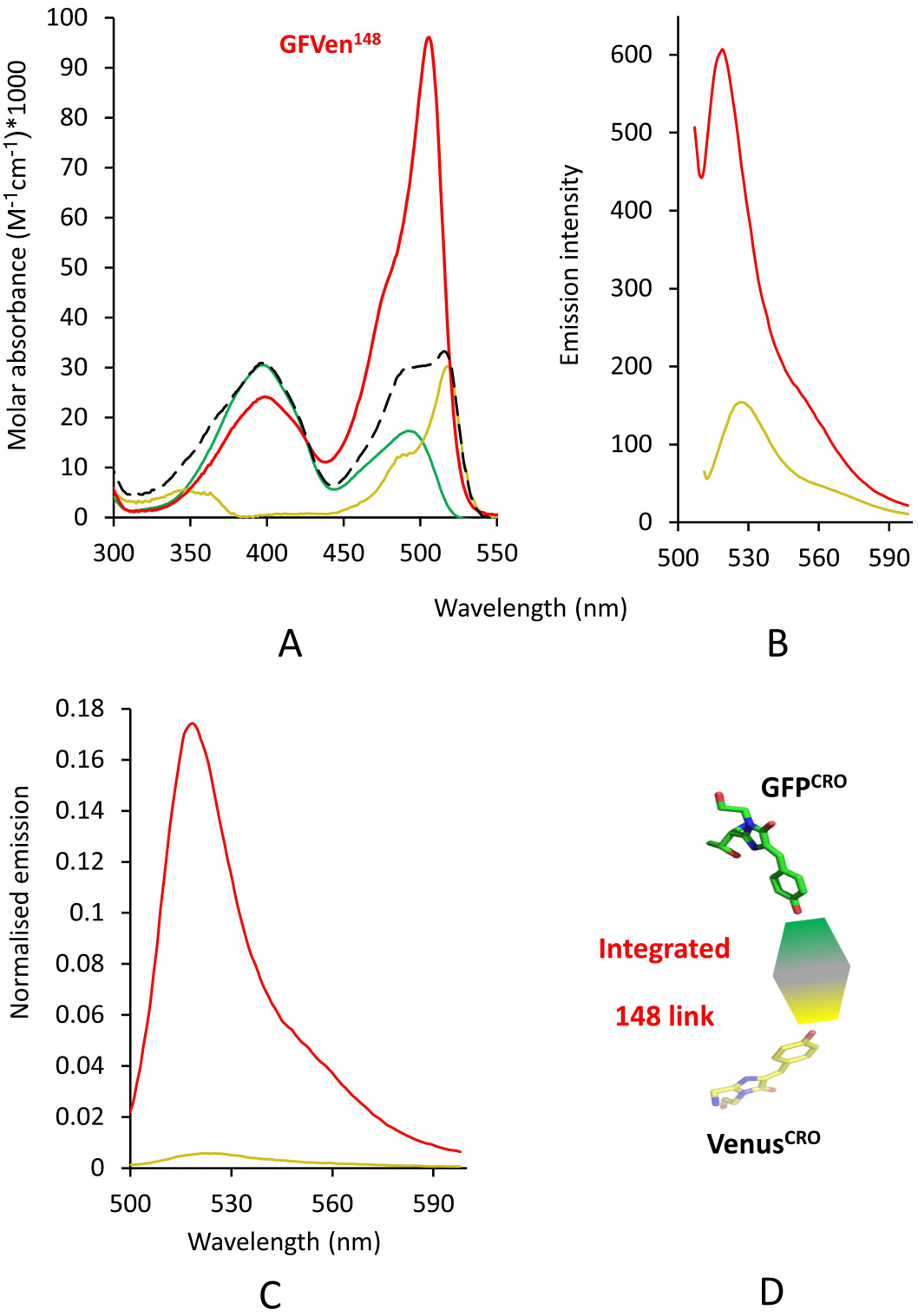

Figure 7.16. Communication between GFVen^{148} heterodimers.
A. Absorbance spectra shown as molar absorbance of GFVen^{148} (red), Venus^{H148azF} (gold), sfGFP$^{H148SCO-K}$ (green) and monomer addition spectrum (Venus^{H148azF}+ sfGFP$^{H148SCO-K}$) (black dashed lines). **B**. Emission intensity of 0.5 µM GFVen^{148} (red) and Venus^{H148azF} (gold) on excitation at 505 nm. **C**. Normalised emission of GFVen^{148} (red) and Venus^{H148azF} (gold) on excitation at 490 nm. **D**. Spatial arrangement of sfGFP$^{H148SCO-K}$ and Venus^{H148azF} CRO based on the homodimer sfGFP^{148x2} structure (PDB code: 5NHN).

Emission spectra on excitation at 400 nm (capable of exciting sfGFP CRO A form only) was also measured for both Venus^{H148azF} and GFVen^{148}. Results showed that Venus^{H148azF} has little absorbance when excited at 400 nm compared to GFVen^{148}. Interestingly, emission intensity was 30-fold higher for GFVen^{148} compared to Venus^{H148azF} with emission still peaking at intermediate 517 nm wavelength (Figure 7.17C).

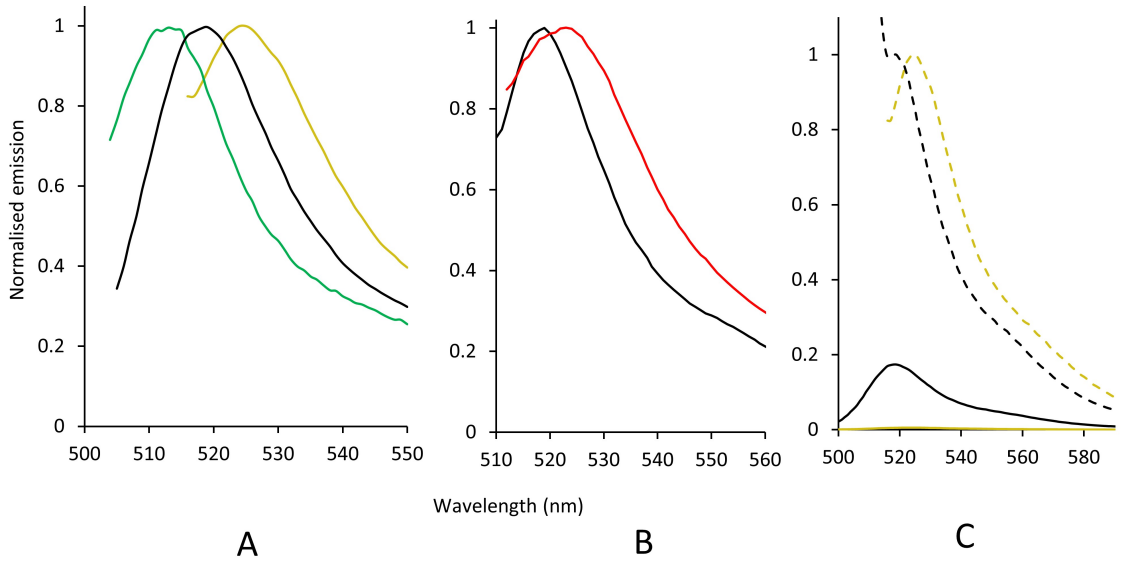

Figure 7.17. Comparison of emission spectra between GFVen^{148} and monomeric units.
A. Normalised emission of sfGFP$^{H148SCO-K}$ (green; excitation at 490 nm), Venus^{H148azF} (gold; excitation at 510 nm) and GFVen^{148} (black; excitation at 490 nm) normalised to high value for each variant. **B**. Comparison of measured emission spectra (on excitation at 505 nm) of GFVen^{148} (black line) and additive emission spectrum of sfGFP$^{H148SCO-K}$ (excitation at 490 nm) and Venus^{H148azF} (excitation at 505 nm). **C**. Normalised emission of Venus^{H148azF} (gold) and GFVen^{148} (black) on excitation at 400 nm (solid line) or 510 nm (dashed line).

Based on the absorbance and fluorescence spectra results (Figure 7.16) of GFVen^{148} compared with individual monomers, GFVen^{148} did not display a classical FRET, where both fluorescent proteins retain their individual spectral profile on energy transfer. While this could suggest that the dimer appears to be acting as a single entity in terms of fluorescence output (λ_{em} 517 nm), as stated above dimerisation may have shifted the spectral properties of the Venus unit. It is believed that there are local conformational changes around Venus and sfGFP CRO as a result of forming the triazole linkage. This

theory is supported by the similarity of spectral properties of modified Venus^{H148azF} with DBCO-TAMRA dye with spectral properties of GFVen^{148} in terms of shifting the λ_{max} (Venus^{H148azF}- DBCO-dye; λ_{max} 507 nm and GFVen$^{148;}$ λ_{max} 505 nm) as shown in Figure 7.18. Moreover, previous work (chapter 6) showed that the photolysis of Venus^{H148azF} by UV-light, which removes the effect of the azide group, causes shifting in λ_{max} (507 nm) as a result of local rearrangement around the CRO as explained in Section 6.2.5.1 (Figure 7.18). Nevertheless, this result suggests that both functional centres (CRO) of sfGFP$^{H148SCO-K}$ and Venus^{H148azF} in GFVen^{148} are acting predominantly as one species. However, the presence of significant neutral protonated state (CRO-A), which excited at (λ_{max} 395 nm) in GFVen^{148} (Figure 7.16A) does suggest that the two monomeric units are not fully synchronised. The new intermediate spectral properties may be the result of structural changes in Venus^{H148azF} on dimerisation. However, it is clear that dimerisation had a significant positive effect with regards to GFVen^{148}.

In an attempt to understand the molecular basis of functional switching in GFVen^{148}, I attempted to solve by X-ray crystallography. Unfortunately, no crystals for GFVen^{148} formed using two types of screens: PACT *premier*^{TM} HT-96 and JCSG-plus^{TM} HT-96.

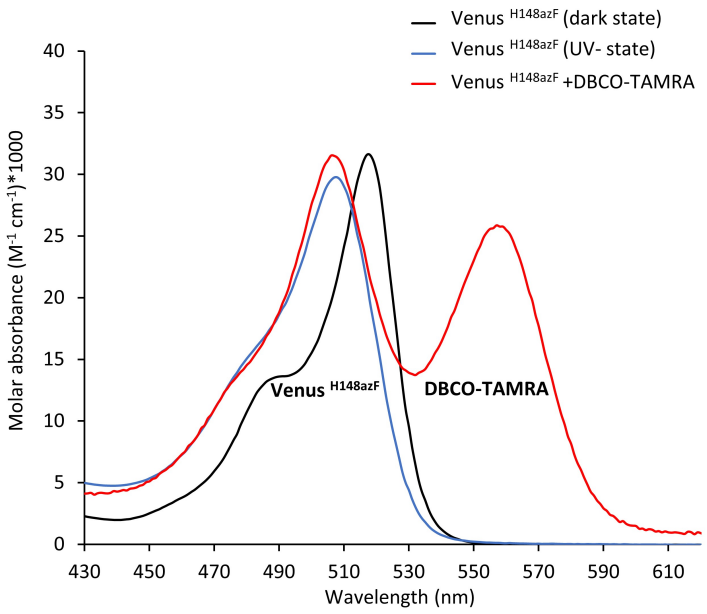

Figure 7.18. Absorbance spectral properties of Venus^{H148azF} modified with DBCO-TAMRA dye. Absorbance spectra shown as molar absorbance for Venus^{H148azF} (dark state), Venus^{H148azF} (UV- state) and Venus^{H148azF} modified with DBCO-TAMRA dye.

In terms of activation (switching on) and enhancing the spectral properties of GFVen^{148} this can, in part, be explained using the homodimer sfGFP^{148x2} structure (PDB 5NHN) and activation mechanism reported by Worthy *et al*. (Figure 7.19; Worthy *et al.* 2019). The structure of the sfGFP^{H148azF} monomer shows that incorporation of azF at position 148 in sfGFP leads to loss of the hydrogen bond with the CRO phenol group that promotes formation of the CRO-B state (Hartley *et al*. 2016). Loss of this key hydrogen bond largely deactivates the protein in terms of its ~490 nm major excitation peak. The structure of the dimeric sfGFP^{148x2} shows that dimerisation (sfGFP^{H148azF} + sfGFP$^{H148SCO-K}$) induces a number of conformational changes around the CRO (Worthy *et al.* 2019). Insertion of a water molecule to take the place of His148 since azF148 is involved in dimer cross-linking. A long-range interaction network is formed in the dimer, composed predominantly of water molecules that will also contribute to activation and enhancement of the fluorescent properties of the dimer. These water molecules have the potential to replace the hydrogen bond interactions that are lost as a result of ncAA incorporation, hence activating the fluorescence in the sfGFP^{148x2} dimer. Formation of a triazole link between 148azF and 148SCO-K causes changes in the backbone and side-chain position of both residues so generating a hole for the water molecule to occupy. This water molecule can now aid with promoting deprotonation of the CRO. It is believed that a similar conformational modification occurs during the dimerisation of the heterodimer GFVen^{148}. The water molecule could occupy the hole that would be generated after triazole link formation. However, this water occupation may not be complete or may be open to limited dynamic exchange with the solvent, at least in the SCO-K containing sfGFP unit, as the dimer still has a prominent excitation at ~400 nm synonymous with the CRO-A form.

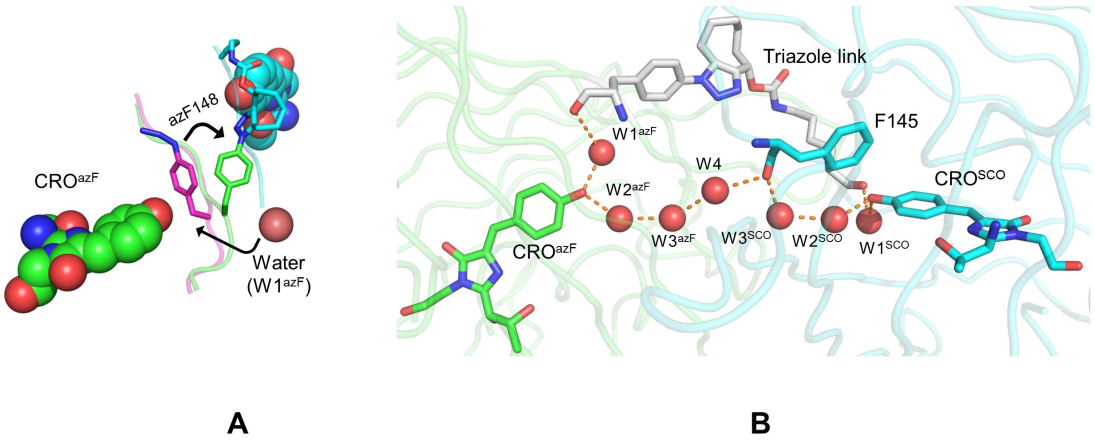

Figure 7.19. Conformational changes and activation of sfGFP^{H148x2} dimer.
A. Conformational change to azF 148 as a result of triazole link formation and dimerisation. **B.** Connection of two CRO of azF and SCO-K modified sfGFP by water dominated long-range hydrogen bond network. (figure was taken from Worthy *et al.* 2019).

Table 7.1 Spectral properties of sfGFP variants, Venus variants and heterodimers

Variant	λ_{max}	λ_{em}	ε ($M^{-1}cm^{-1}$)	QY	Brightness
sfGFP ^{WT*}	485	511	49000	0.75	36750
Venus^{WT}	515	528	92200[a]	0.65	59930
sfGFP$^{148SCO-K*}$	395	511	31000	0.52	16120
	492	511	17300	0.84	14532
Venus^{148azF}	517	525	30100	0.45	13545
GFVen^{148}	400	517	24000	0.42	10080
	505	517	96000	0.46	44160
Venus^{204azF}	515	528	87600	0.42	36792
sfGFP$^{204SCO-K*}$	485	511	39800	0.66	26268
GFVen^{204}	492	530	102000	0.70	71400
	514	530	118000	0.60	70800

[a] Published previously by Nagai *et al*. (T Nagai *et al.* 2002) and measured in the current study as 95000 $M^{-1}cm^{-1}$. Given that the measured value is close to the reported value, the reported value has been used. * Value reported by Worthy (Worthy 2018).

7.2.6.2 Characterisation of heterodimer GFVen^{204}

Covalent linkage via 204 SPAAC successfully generated a dimer (GFVen^{204}) (Figure 7.5). The new artificial dimer displayed enhanced spectral properties compared to the monomers that combined the features of both (Figure 7.20A-B). GFVen^{204} generated a species with λ_{max} at both 490 and 514 nm (ratio of 1:1.2). The monomeric Venus^{Q204azF} absorbs at the same wavelengths but in a different ratio (1:2.7), as shown in Figure 7.13A and Table 7.1. Based on the absorbance spectra of both GFVen^{204} and individual monomers (sfGFP$^{Q204SCO-K}$ and Venus^{Q204azF}), GFVen^{204} has an enhanced molar absorbance compared with monomeric units, increasing by ~26,000 $M^{-1}cm^{-1}$ (~27%) compared to Venus^{Q204azF} associated λ_{max} (514 nm) where there is little contribution for sfGFP. To confirm that the enhancement in the molar absorbance of GFVen^{204} dimer is due to dimerisation, an additive spectrum of individual monomer spectra that simulates two independently acting proteins showed lower molar absorbance compared to the measured GFVen^{148} dimer molar absorbance with λ_{max} at both 490 and 514 nm (ratio of 1:1.27), which is similar to the ratio of the GFVen^{204} dimer (Figure 7.20A). The ratio of the two entwined major dimer peaks (490 and 514 nm) is similar to that of the additive spectra suggesting that while the CRO environments are themselves similar to the monomer, there has been an overall enhancement in the ability of dimer to interact with light. Thus, unlike in GFVen^{148}, dimerisation caused local modifications around the CRO and improved its ability to interact with light.

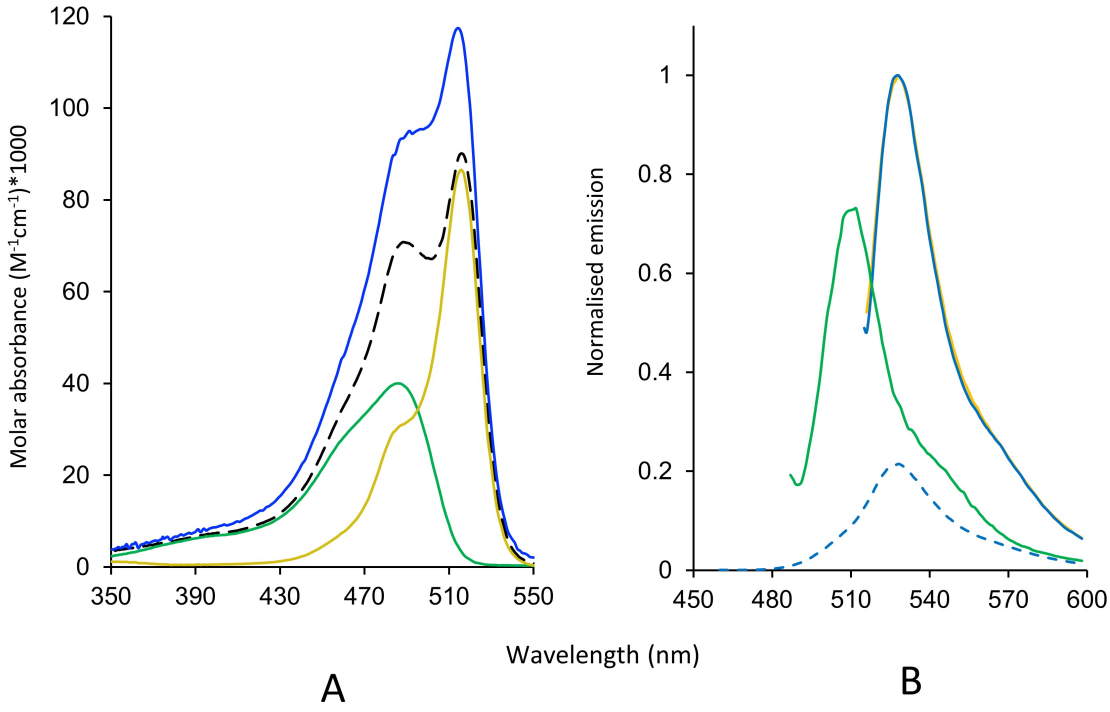

Figure 7.20. Absorbance and fluorescence spectra of GFVen^{204} dimer and monomeric units. **A.** Absorbance spectra shown as molar absorbance of GFVen^{204} (blue), Venus^{Q204azF} (gold), sfGFP$^{Q204SCO-K}$ (green) and monomer addition spectrum (black dashed lines). **B.** Normalised emission spectra for GFVen^{204} (blue) and Venus^{Q204azF} (gold) excited at 510 nm. sfGFP$^{Q204SCO-K}$ (green) excited at 490 nm, blue dashed line represents GFVen^{204} excitation at 450 nm.

To investigate any potential communication between the functional centre of monomers (sfGFP$^{Q204SCO-K}$ and Venus^{Q204azF}), emission spectra on excitation at four separate wavelengths were recorded: 400 nm (excited sfGFP only); 450 nm (excited sfGFP and minor excitation of Venus); 490 nm (sfGFP λ_{max}, Venus shoulder); 510 nm (excited Venus and minor excitation of sfGFP) (Figure 7.15). Interestingly, only a clear emission peak corresponding to Venus at 528 nm was observed at all excitation wavelengths, as shown in Figure 7.21A-B. This result suggests the communication of the functional centres by FRET. As the GFVen^{204} dimer is excited at wavelengths specific for sfGFP but emission is only at 528 nm (Venus^{Q204azF} λ_{em}). No emission peak corresponding to sfGFP$^{Q204SCO-K}$ (at ~510 nm) was observed at any of the above excitation wavelengths, even at the lower wavelengths specific for sfGFP$^{Q204SCO-K}$ (Figure 7.21B). The calculated relative FRET efficiency after spectral decomposition was *circa* 90%. These results suggest that the two functional centres of sfGFP$^{Q204SCO-K}$ and Venus^{Q204azF} in the GFVen^{204} dimer are communicating through energy transfer in a highly efficient manner (Figure 7.22B).

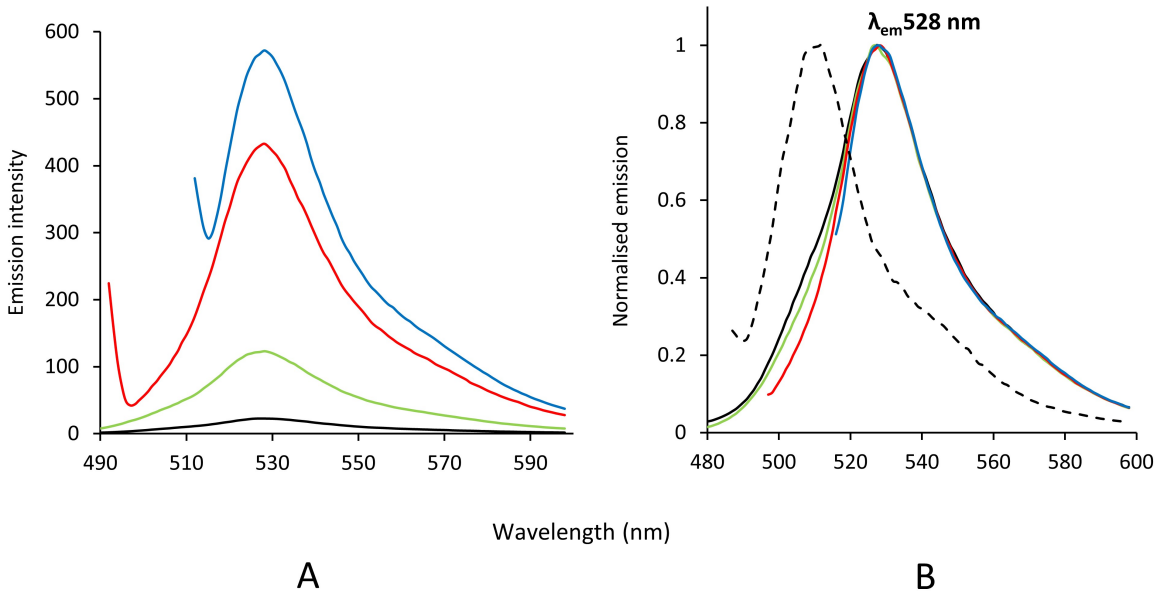

Figure 7.21. Emission spectra of GFVen^{204} on excitation at 4 separate wavelengths.
A. Emission spectra of GFVen^{204}. Black, green, red and blue lines represent emission spectra on excitation at 400 nm, 450 nm, 490 nm and 510 nm, respectively. **B**. Normalised emission spectra at a highest value for each measured to 1 to compare the emission peaks, at all excitation wavelengths the GFVen^{204} has λ_{em} at 528 nm (corresponding to Venus). Black dashed line represents sfGFP$^{Q204SCO\text{-}K}$.

Interestingly, as a result of high FRET efficiency between the monomeric units of GFVen^{204}, it displays improved fluorescence emission at lower wavelengths compared to Venus^{Q204azF} (Figure 7.22A). The emission intensity at 528 nm for the GFVen^{204} compared to Venus^{Q204azF} increased by 27.4%, 68.4% and 91.2% at the excitation wavelengths 490 nm, 450 nm and 400 nm respectively.

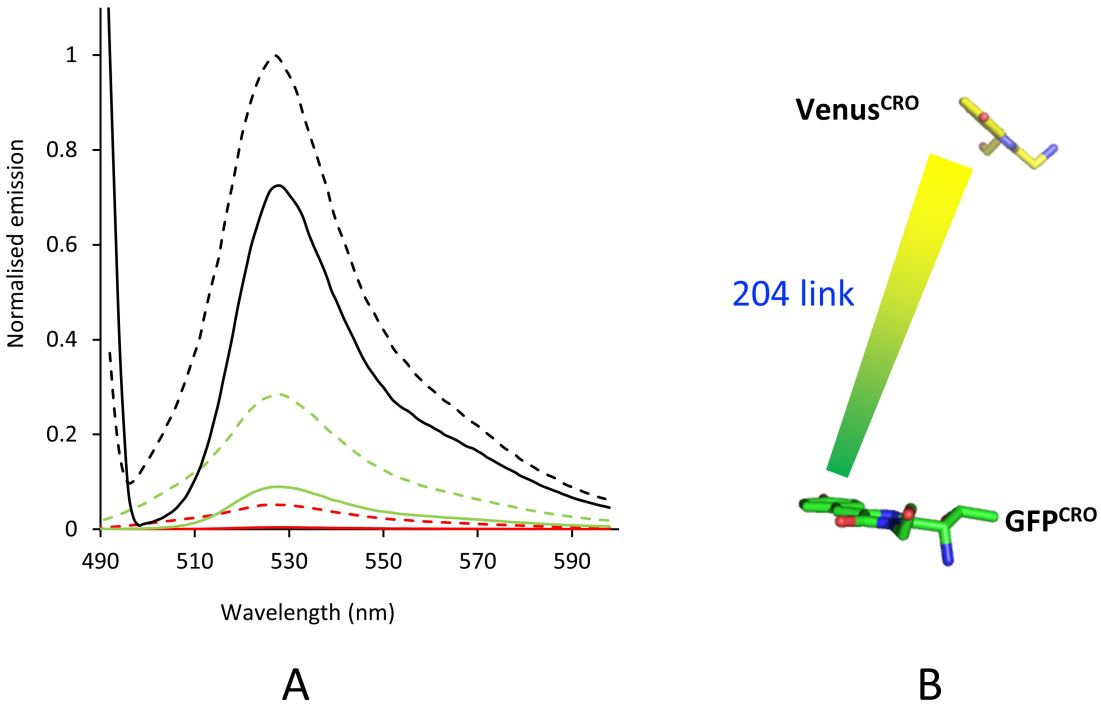

Figure 7.22. Comparison of emission spectra of GFVen^{204} and Venus ^{204azF}.
A. Normalised emission spectra of GFVen^{204} (dashed lines) and Venus^{Q204azF} (solid lines) excitation at different wavelengths. Red, green and black represent 400 nm, 450 nm and 490 nm respectively. **B.** Spatial arrangement of sfGFP$^{Q204SCO-K}$ and Venus^{Q204azF} CRO based on homodimer sfGFP^{204x2} structure (PDB code: 5NI3).

7.2.7 Conclusion

Homo-oligomerisation is a common process in nature and most proteins are present in their final structural state as a homodimer or homo-oligomer. In other cases, a single unit can form larger oligomeric complexes, as seen for Vip1 in chapter 5. In systems that need to recruit separate functions from distinct proteins, there is a need to form heterocomplexes. A good example is the classical AB binary toxins, in which the B subunit forms homo-oligomers and interact with A subunits to form heterocomplexes, which have the full toxicity effect. Generation of a protein heterocomplex could lead to new species with novel, complex functional properties.

Historically the interest in protein engineering of fluorescent proteins has focused on monomerisation of oligomers (Campbell *et al.* 2002). Nowadays, oligomerisation (homo or hetero) is a hot topic in protein engineering (Fallas *et al.* 2017), as it allows us to create new architectures and, thus, alter functional properties. Consequently, enhanced/integrated or new functional properties may emerge that are not present in

nature. In this chapter, protein engineering has been successfully used to incorporate ncAAs with orthogonal Click chemistry into fluorescent protein monomers at two different surface exposed positions using codon reprogramming. The engineered fluorescent monomers with ncAA were used successfully to create functionally linked heterodimers via SPAAC.

Heterodimers linked through either residues 148 or 204 were successfully constructed that displayed new properties not present in the monomeric proteins. Linking via residue 148 generated the GFVen^{148} dimer that switched on the fluorescence of both monomers by changing the protonation state of the CRO, thereby generating a dimer with novel fluorescent properties that are intermediate between the sfGFP$^{H148SCO\text{-}K}$ and Venus^{H148azF} monomers. The GFVen^{204} dimer, which was created via the residue 204 link, has an enhanced molar absorbance for the Venus^{Q204azF} associated λ_{max} (514 nm) and enhanced fluorescence intensity at a wide range of excitation wavelengths starting from 400 nm to 510 nm, which differ from either monomeric unit. Both heterodimers with these fluorescent properties could be used as new fluorescent tools in synthetic biology (Stark *et al.* 2018). In conclusion, the linking position defines the new functional properties in heterodimers, which is determined by the structural arrangement of the monomeric units to each other and the distance between the functional centres. It would be interesting to investigate more surface residues around the CRO for heterodimer creation, which could generate further species with new and interesting fluorescent properties.

Chapter eight

General discussion

8. General discussion

Protein engineering includes modification of proteins through gene mutation, posttranslational modification, or a combination of both, in order to study or improve proteins, thereby advancing our knowledge of the proteins and their functions. Bioinformatics and computer software are integrally linked with protein engineering through the use of database searches, modelling, structure prediction, data collection and analysis, three-dimensional structure determination, protein-protein docking and so on. In this work, a combination of different protein engineering technologies has been successfully applied to study the insecticidal proteins Vip1Ac1 and Vip2Ae3 and to generate useful variants of fluorescent protein (Venus).

8.1 Solubility issue of recombinant protein expression.

Nowadays most of proteins are produced recombinantly for research, pharmaceutical and industrial purposes, often because important proteins are produced naturally in small amounts and are hard to purify under native conditions. Genetic engineering technologies overcome these issues and allow the production of proteins in large amounts that may be easily purified. However, not all proteins can be successfully produced in recombinant form because these technologies involve transferring the gene encoding the protein from the original source into heterologous expression systems (prokaryotic or eukaryotic), which are different from the original sources in terms of production environment and level of production. These differences can lead, in some cases, to misfolding and insoluble aggregation of the recombinant protein as inclusion bodies.

Vip1 and Vip2 proteins are extracellular proteins secreted by *Bacillus sp.* to the surrounding environment, mediated by putative N-terminal signal peptides. Expression of Vip1Ac1 protein in *E. coli* using the pET28b vector as full-length protein was unsuccessful: Vip1Ac1 protein accumulated as inclusion bodies in the cytoplasm (chapter 3). Folding and posttranslational modification of secreted protein occurs after signal peptide cleavage (Braakman and Hebert 2013) and the presence of a non-compatible signal peptide could affect the solubility of the protein.

There is no universal strategy to solve solubility issues but different strategies have been implemented to enhance the solubility of recombinant protein expressed in the cytoplasm of *E. coli.* These include changing; expression vector, expression host,

expression parameter and fusion of the target protein with a soluble partner (Gopal and Kumar 2013). For each distinct protein, these strategies need to be evaluated. Several strategies were used to address the solubility problem for the Vip1Ac1 protein, including use of the soluble fusion partner glutathione S-transferase (GST). This soluble fusion partner was chosen because it has been successfully used to increase the solubility of anthrax protective antigen (PA), a homolog of Vip1Ac1, which expressed as inclusion bodies in *E. coli* (Wu *et al.* 2010). The above strategy did not enhance the solubility of Vip1Ac1 protein, but the yield of protein was increased (chapter 3). The nature of this expression system resulted in production of an extracellular protein in the reducing environment of cytoplasm, which may not be compatible with folding and solubility.

As a result, the protein was expressed in the periplasmic space. This required a signal peptide to be linked genetically to the N-terminus of the protein to direct it into periplasmic space. A number of expression vectors encoding signal peptides are commercially available and have been used to secrete recombinant proteins into periplasmic space (Sockolosky and Szoka 2013). In this study, the soluble protein, Cytochrome b_{562} (Cyt b_{562}), that is known to accumulate in the periplasm was used to generate the fusion protein (Cyt b_{562}-Vip1Ac1). Cytochrome b_{562} is a highly soluble, small (~12 kDa), coloured protein that is naturally expressed in *E. coli* as periplasmic protein (Mathews *et al.*1979) and which can be used as an expression marker. Interestingly, this fusion tag successfully improved the solubility of Vip1Ac1 and directed it into periplasmic space as a soluble protein after optimisation of the expression temperature. However, in order to study the activity of the recombinant protein in isolation, all tags need to be removed, although the difficulty of removing the tag is one of the most drawback of this strategy (Esposito and Chatterjee 2006). For the Cyt b_{562}-Vip1Ac1 fusion, it proved impossible to remove the fusion tag (Cyt b_{562}) after treatment with TEV protease using different incubation times and different ratios, presumably due to folding of the protein obscuring the TEV cleavage site between Vip1Ac1 and Cyt b_{562}.

To overcome this issue and to clarify whether the improvement of Vip1Ac1 solubility was attributable to the fusion protein or to the expression of protein in periplasmic space, the Cyt b_{562} was genetically removed to leave only its signal peptide linked to the Vip1Ac1 protein. Successful expression of Vip1Ac1 as a soluble protein in the periplasmic space indicated that the fusion tag Cyt b_{562} has no effect on the solubility of

the Vip1 protein while the signal peptide has a pivotal role in the solubility by directing Vip1Ac1 into the periplasm. The same strategy was effectively applied to express Vip2Ae3 in the periplasmic space (chapter 3). Thus, the periplasmic signal sequence of Cyt b_{562} appears to be a beneficial tool, which can be added to the set of recombinant protein technology tools that are available for expressing periplasmic recombinant proteins, and which might solve the solubility issue for some proteins.

8.2 Importance of molecular structure in protein studies and engineering.

Knowing the three-dimensional structure of a protein is the key component to understand a protein's properties and function, and is of exceptional value for rational protein engineering. The three-dimensional structure can be experimentally determined or predicted by molecular modelling using bioinformatic tools based on an established three-dimensional structure of a homologous protein (the template). Predicted structures can be very beneficial resources to explore protein structure-function and properties and models have been successfully used in rational protein engineering (Kreudenstein *et al.* 2014) although they sometimes miss significant information present in the actual structure (chapter 4). The range of modelling algorithms generate models slightly different from each other because they use different parameters for searching and choosing the template as well as building their models (Combet *et al.* 2002; Zhang 2008; Kiefer *et al.*2009). However, when applied to Vip1Ac1, these models have the same overall folding as observed in chapter 4. Based on the best model of Vip1Ac1 according to the validation parameters, two residues were selected as surface exposed residues for protein modification with fluorescent dye (chapters 3 and 5).

In silico prediction and creation of models of dimer interactions identifying a potential dimer interface area, provide a guideline for rational selection of residues or regions to be used in dimer modification (Yueh *et al.* 2017). In previous studies, modification of residues H148 and Q204 present in the predicted homodimer interface region, successfully generated a covalent homodimer of sfGFP while residue 132, which is not predicted to be at the dimer interface, did not show any potential for dimer formation (Worthy 2018). In the current study, prediction of the interface region in a heterodimer (sfGFP and Venus), confirmed the key residues (H148 and Q204) to be in a favourable interface region for post translational modification and these were successfully used to create a heterodimer (chapter 7). The *in silico* prediction of a

potential dimer interface area appears to be an effective approach for residue selection for artificial dimerisation.

The three-dimensional structures of both Vip1Ac1 and Venus^{Y66azF} dark state have been experimentally determined by X-ray crystallography. Analysis of the Vip1Ac1 structure reveals the significant similarity with PA in terms of folding and important motifs in domains 1-3, as well as the differences in domain 4, which explain the specificity for different targets. The similarity helps in prediction and interpretation of the functions of each domain and, thereby, guides experiments to investigate these functions (eg activation of toxin, oligomerisation of pre-pore oligomer, pore formation in artificial lipid bilayers, interaction of Vip2Ae3 with the oligomeric state of Vip1Ac1, transition of pre-pore oligomer to pore oligomer, rational mutation for oligomerisation inhibition and electron microscopy studies; chapter 5). As the structure of PA has been used successfully to interpret the function of Vip1Ac1, both structures (PA and Vip1Ac1) could be a good models to elucidate the structure/function of B subunits of iota-like binary toxins (produced by *Clostridium* species with toxic activity toward mammalian cells).

The molecular structure of the dark state of a genetically modified fluorescent protein with ncAA (Venus ^{Y66azF}) reveal the molecular changes in the CRO and surrounding residues as a result of the incorporation of azF at position 66 and this explains the changes in fluorescent properties and the switching off of the fluorescence of the protein (chapter 6). While in other modified variants of Venus^{azF} (residues 145, 148, and 203) for which structures are not available, changes in the observed spectral properties as a result of incorporation, and after photolysis of the azide group by UV-light were interpreted by comparison to the molecular structure of equivalent 145 and 148 azF variants of sfGFP. Similarly, enhanced fluorescence for the GFVen^{148} heterodimer produced in this work was interpreted on the basis of the sfGFP^{148} homodimer.

8.3 Genetically encoding ncAA as a protein engineering technique.

Non-canonical amino acids (ncAAs) are amino acid analogues with a functional group. Use of ncAA by genetically encoded reprogramming of stop codons into a protein of interest (Section 1.7) allows introduction of a new functional group at desirable positions in the protein. In addition, if natural amino acids are substituted with related

ncAAs (eg tyrosine with ncAA tyrosine derivatives), it is possible to introduce useful new chemistry into the protein without significant change to the amino acid composition or protein structure. A tyrosine derivative; 4-Iodo-L-phenylalanine, which is a ncAA containing a heavy atom was used to substitute 2 surface exposed tyrosine residues in Vip1Ac1 (Y358 and Y536) to prepare heavy atom derivative of Vip1Ac1 crystals in an attempt to allow phasing (chapter 3 and 4). Incorporation of 4-Iodo-L-phenylalanine to prepare heavy atom derivative was previously used to solve the structure of bacteriophage T4 lysozyme (Xie *et al.* 2004) although it proved unsuccessful for Vip1Ac1.

Site specific incorporation of ncAA into proteins has been widely used for post translational modification (Lim and Kwon 2016; Kwon and Yang 2017; Zaki *et al.* 2018). In the work described here, a tyrosine derivative, *p*-azido-L-phenylalanine (azF) containing a selective Click chemistry azide group, was used to substitute 2 surface exposed tyrosine residues in Vip1Ac1 (Y358 and Y536) then modified with DBCO-TAMRA fluorescent dye (containing an orthogonal strained alkyne Click chemistry group) via SPAAC reaction for labelling purposes (chapter 3 and 5). The same approach was applied to generate a heterodimer at a desirable dimerisation point to link the functional centre of two fluorescent protein monomers. Two copper-free click compatible ncAAs, azF (azide) and SCO-K (strained alkyne) were incorporated in Venus and sfGFP respectively (at positions 148 and 204) and the heterodimer was generated by SPAAC reaction (chapter 7).

Incorporation of ncAA has previously been used to modify and generate a useful variant of sfGFP (Reddington *et al.* 2015). The ncAAs have different effects on the protein based on the functional group of the ncAA and the site of incorporation. Incorporation of azF with its photoactivatable group in Venus showed different effects on its fluorescent properties, before and after photolysis by UV-light, based on the incorporation site to the chromophore (CRO; functional centre of fluorescence) (chapter 6).

The functional group of the ncAA used and the site of incorporation play vital roles in protein modification that can be achieved and the need for subsequent post translational modification. For example, genetically encoded 4-Iodo-L-phenylalanine will modify the protein directly with a heavy atom without further modification, while incorporation of azF or SCO-K will introduce photoactive and Click chemistry compatible groups that can

be used for further post translational protein modification such as labelling (chapter 5), photolysis (chapter 6) and dimer construction (chapter 7).

8.4 Usefulness of azF in protein engineering.

The tyrosine derivative ncAA, *p*-azido-L-phenylalanine (azF), has been widely used in protein engineering owing to its chemical properties. It differs from the tyrosine by replacing the (-OH) with the azide functional group ($-N_3$), which is characterised by photochemical reactivity (chapter 6) and bio-orthogonal selective Click chemistry (chapter 5 and 7). Photochemical reactivity of azF is dependent on the ability of azide to form a singlet reactive nitrene radical after photolysis by UV-visible light and loss of molecular nitrogen. Based on the surrounding environment, the nitrene radical can undergo different chemical reactions as illustrated in Figure 1.16, which lead to distinct modifications. The various reaction routes that can be taken by the nitrene radical after photolysis make the azF a flexible tool in protein engineering. In this study, incorporation of azF at different positions in Venus around the CRO introduced the azide group into different environments that ultimately affected the pathway taken after forming the nitrene radical, which, in turn, affected the fluorescent properties and generated derivatives with different spectral properties.

The azide group can selectively react with strained cyclic molecules containing alkynes by a cycloaddition reaction (SPAAC) resulting in a stable triazole linkage between the conjugated molecules (Baskin *et al.* 2007). This reaction is a bio-orthogonal metal-free Click chemistry reaction, which allows selective labelling of the protein *in vivo* with a small fluorescent dye (without needing a fluorescent protein as a fusion partner, which in some cases can cause folding problems) (Nikic *et al.* 2015). In this study, the azF modified Vip1Ac1 was labelled with the strained cyclic alkynes-containing fluorescent dye (DBCO-TAMRA), as outlined in chapter 5. One limitation to the use of azF is its light sensitivity, which means that all the work (incorporation, protein extraction, purification and post translational modification) must be carried out in the dark to avoid the photolysis of the azide group. An alternative is the incorporation of the strained alkyne SCO-K into the protein for post translational modification with commercially available probes containing azide, tetrazine or nitrone functional groups (Plass *et al.* 2011; Lang and Chin, 2014b; Dommerholt *et al.* 2016) which expands the

range of fluorescent probes that can be used without the requirement to be kept in the dark.

8.5 Future work.

8.5.1 Vip1Ac1 related future work.

The molecular structure of Vip1Ac1 (monomer state, ~82 kDa) has been elucidated in this study and helps us to understand the potential function of each domain through the significant similarity with anthrax PA and helps to interpret the protein's function as pore forming toxin. As a next step, it will be important to determinate the molecular structure of pre-pore and pore oligomers of $Vip1Ac1_{60}$ and the complex of $Vip1Ac1_{60}$ with Vip2Ae3. This will help us to understand the conformational changes in protein domains during the transition from pre-pore to pore oligomer, as well as the interaction with Vip2Ae3. This, in turn, would help develop our knowledge of the mode of action and the means by which Vip2Ae3 is translocated into cytosol of the target cell. As the oligomeric state of $Vip1Ac1_{60}$ has not, to date, successfully generated any crystals for structural studies, CryoEM seems to be best way to determinate the oligomer structure, especially since the TEM studies showed a good quality sample for CryoEM analysis.

Rational mutation of calcium binding motif residues to study the effect of calcium ions on the protein structure, before and after activation with trypsin will indicate their role in the oligomerisation and interaction with Vip2Ae3. In parallel studies, it has been demonstrated that the presence of Ca^{2+} ions play a role in the interaction of the enzymatic subunit of *Clostridium perfringens* iota-toxin (Ia) with the oligomeric state of the B subunit (Ib) (Kobayashi *et al.* 2008). Also, it would be interesting to identify the receptor in the insect targets of Vip1Ac1 and Vip2Ae3 as well as the specific target of the C-terminus (enzymatic domain) of Vip2Ae3 to help to understand the toxic effect on the target cells.

8.5.2 Venus related future work.

As mentioned in chapter 6, genetically encoded azF has a different effect on protein properties based on the incorporation site and the impact of this site on the functional centre of the protein (CRO). It would be interesting to investigate the effects of the replacement of other residues around the CRO with azF, especially the residues that have

been reported previously to have an effect on the spectral properties of the related sfGFP since these might have different effects on Venus as observed with 66, 203 and 148 in the current study. It will be important to determine the molecule structure of Venus ^{azF} variants before and after photolysis and understand the molecular basis for the effect of azF incorporation on Venus fluorescent properties. Having molecular structures of Venus ^{azF} variants will help for further rational engineering to generate further improvements. Since incorporation of azF at residues 66, 203 and 148 in Venus and sfGFP give different effects, it would be interesting to make equivalent changes in other GFP derivatives such as blue, cyan and red fluorescent proteins.

With regard to the artificial heterodimer (chapter 7), it would be interesting to solve the molecular structure of the heterodimer GFVen^{148} as this would help us to understand the effect of dimerisation and the conformational changes that led to the emergence of the new spectral properties that differ from each of the monomeric units. It would also be interesting to generate a Venus homodimer using position 148 as a dimerisation point and investigate the changes in fluorescent properties of the dimer in terms of switching on and fluorescence enhancement, as previously observed with a sfGFP^{148x2} homodimer (Worthy 2018). As the dimerisation between two GFP derivatives creates a heterodimer with new spectral properties, it would be useful to use a distinct fluorescent protein such as mCherry as a partner to create a further heterodimer. In fact, this work has already progressed in the Jones lab by linking sfGFP with mCherry.

8.6 Summary.

Vip1Ac1 expressed as a native soluble protein after the solubility issue was solved by genetically linking a periplasmic signal sequence from Cytochrome b_{562} at the N-terminus of Vip1Ac1 (lacking its own signal peptide). This lead to expression of the protein in the periplasmic space in soluble form, allowing affinity purification (His-tag) and subsequent TEV cleavage to remove the tag to purify Vip1Ac1. The same procedure was used to express Vip2Ae3 (chapter 3). Vip1Ac1 was modified via Click chemistry with a fluorescent dye (DBCO-TAMRA) for visualisation during lipid binding experiments. First, a Click chemistry azide group introduced into Vip1Ac1 by genetic incorporation of ncAA (*p*-azido-L-phenylalanine) at two surface exposed residues (Y358 and Y536, chosen based on predicted molecular structure; chapter4). The protein was then post translationally modified via Click chemistry with compatible DBCO-TAMRA

dye (chapter 3 and 5). Bioinformatic software were used to predict and generate a protein model, which was analysed and interpreted based on the protein homology. Later, experimental X-ray data allowed the three-dimensional structure of Vip1Ac1 to be solved at 1.47A resolution (chapter 4). Vip1Ac1 was activated by trypsin and the interaction with Vip2Ae3 was studied by SEC. The ability to oligomerise, form pores and transition from pre-pore to pore oligomer was studied by SEC, YO-PRO-1 uptake assay and electron microscopy (chapter 5).

Rational protein engineering using codon reprogrammed approaches was used to introduce a photoactivatable chemistry into an interesting site in the fluorescent protein (Venus) by incorporation ncAA and generate useful variants with new properties such as the ability to switch off / on (chapter 6). A combination of both genetic modification (incorporation of ncAA with compatible Click chemistry group, azide and alkyne groups) and posttranslational modification (Click chemistry) were successfully used to construct an artificial fluorescent heterodimer (sfGFP containing alkyne group +Venus containing azide group) with fluorescent spectral properties do not exist in either of the monomeric units (chapter 7).

8.7 Publications, conferences attendance, and posters associated with this thesis.

Chapter 7 (paper)

Worthy, H.L., **Auhim, H.S** [1], Jamieson, W.D., Pope, J.R., Wall, A., Johnson, R.L., Watkins, D.W., Rizkallah, P.J., Castell, O.K., & Jones, D.D. (2019).Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry, *Communications Chemistry*. Springer US, 2(83), pp. 1–12. doi: 10.1038/s42004-019-0185-5.

Chapter 4 (conference poster)

Auhim, H.S [1], Rizkallah, P., Paterson, N., Jones, D., and Berry, C. Molecular structure and engineering of insecticidal protein from *Bacillus thuringiensis.* Protein Engineering II: from New Molecules to New Processes. A biochemical society scientific meeting. University of York, UK, 15-17 July 2019.

[1] Joint 1st author

Other work (conference poster)

Amer, N., Worthy, H., **Auhim, H.S**, Embaby A., Saeed H., Hussein A. and Jones, D. Structure-guided engineering of a psychrophilic esterase. Protein Engineering II: from New Molecules to New Processes. A biochemical society scientific meeting. University of York, UK, 15-17 July 2019.

Conference attendance

South West Structural Biology Consortium 2017 (SWSBC 2017) conference, Cardiff university, UK, 3-4 July 2017.

Two minute presentation

Molecular Biosciences Division Postgraduate Away Day. Life Science Hub in Cardiff Bay, UK, 19 June 2017.

Internal poster (Introduction and research plan)

1st Year Research Students- Poster Event. Cardiff university, UK, 8 February 2016.

(Won the prize for the best poster for molecular biosciences division)

External workshop

- Chemical and Biological Security Training for Iraqi Nationals, funded by Sandia National Laboratories/ Liverpool Marriott City Centre, Liverpool, UK, 30/03/2018 - 01/04/2018.
- Effective tools and strategies for time management, tasks management and note management / University of Cardiff / Cardiff School of Engineering, UK. 18/11/2016.
- Effective tools and strategies for enterprises and projects management/ University of Cardiff / Cardiff School of Engineering, UK. 11/11/2016
- CCPBioSim Training - DAY 1: Enlighten: Tools for Enzyme-Ligand Modelling / University of Bristol. 6/06/2016

Chapter nine

References

References

1. Abbyad, P., Childs, W., Shi, X. and Boxer, S.G. (2007). Dynamic stokes shift in green fluorescent protein variants. *Proceedings of the National Academy of Sciences of the United States of America* **104**(51):20189–20194.
2. Adams, M., Blundell, T., Dodson , E., Dodson, G.,Vijayan, M., Baker, E.,… Sheat, S. (1969). Structure of Rhombohedral 2 Zinc Insulin Crystals. *Nature* **224**(5218):491.
3. Ahnert, S.E., Marsh, J.A., Hernández, H., Robinson, C. V and Teichmann, S.A. (2015). Principles of assembly reveal a periodic table of protein complexes. *Science* **350**(6266).
4. Aktories, K., Papatheodorou, P. and Schwan, C. (2018). Binary *Clostridium difficile* toxin (CDT) - A virulence factor disturbing the cytoskeleton. *Anaerobe* **53**:21–29.
5. Al-Mohanna, F. A.; Ohishi, I. and Hallett, M.B. (1987). Botulinum C2 toxin potentiates activation of the neutrophil oxidase Further evidence o f a role for actin polymerization.*FEBS Lett* **219**(1):40–44.
6. Ali, M.H. and Imperiali, B. (2005). Protein oligomerization : How and why. *Bioorganic & Medicinal Chemistry* **13**:5013–5020.
7. Alisaraie, L. and Rouiller, I. (2016). Molecular assembly of lethal factor enzyme and pre-pore heptameric protective antigen in early stage of translocation. *J Mol Mode* **7**(22).1-12.
8. Agard, N., Prescher, J. and Bertozzi, C. (2004). A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *Journal of the American Chemical Society* **126**(46):15046–15047.
9. Antikainen, N.M. and Martin, S.F. (2005). Altering protein specificity: techniques and applications. *Bioorganic and Medicinal Chemistry* **13**(8):2701–2716.
10. Arora, R. and Sandhu, S. (2017). *Breeding Insect Resistant Crops for Sustainable Agriculture*. Springer Nature. Singapore.
11. Aroul-Selvam, R., Hubbard, T. and Sasidharan, R. (2004). Domain insertions in protein structures. *Journal of Molecular Biology* **338**(4):633–641.
12. Arpino, J.A.J (2011). From single amino acid deletions to whole domain insertions ; Engineering GFP through polypeptide backbone mutations. PhD

Thesis.Cardiff University.

13. Arpino, J.A.J, Rizkallah, P.J. and Jones, D.D. (2012a). Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 A˚ Resolution Reveals Alternative Conformations for Glu222. *PloS One* **7**(10).
14. Arpino, J. A.J., Czapinska, H., Piasecka, A., Edwards, W.R., Barker, P., Gajda, M.J., … Jones, D.D. (2012b). Structural basis for efficient chromophore communication and energy transfer in a constructed didomain protein scaffold. *Journal of the American Chemical Society* **134**(33):13632–13640.
15. Bajar, B.T., Wang, E.S., Zhang, S., Lin, M.Z. and Chu, J. (2016). A guide to fluorescent protein FRET pairs. *Sensors (Switzerland)* **16**(9):1–24.
16. Bailey S. The CCP4 suite - programs for protein crystallography. *Acta Crystallogr D*. 1994. **50**: 760-763.
17. Barondeau, D.P., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. (2005). Understanding GFP chromophore biosynthesis: Controlling backbone cyclization and modifying post-translational chemistry. *Biochemistry* **44**(6):1960–1970.
18. Barondeau, D.P., Putnam, C.D., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. (2003). Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences* **100**(21):12111–12116.
19. Barth, H. and Aktories, K. (2011).New insights into the mode of action of the actin ADP-ribosylating virulence factors *Salmonella enterica* SpvB and *Clostridium botulinum* C2 toxin. *European Journal of Cell Biology* **90**(11):944–950.
20. Barth, H., Aktories, K., Popoff, M.R. and Stiles, B.G. (2004). Binary Bacterial Toxins : Biochemistry , Biology , and Applications of Common *Clostridium* and *Bacillus* Proteins. *Microbiology and Molecular Biology Reviews* **68**(3):373–402.
21. Barth, H., Blöcker, D., Behlke, J., Bergsma-Schutter, W., Brisson, A., Benz, R. and Aktories, K. (2000). Cellular uptake of *Clostridium botulinum* C2 toxin requires oligomerization and acidification. *Journal of Biological Chemistry* **275**(25):18704–18711.
22. Baskin, J.M., Prescher, J.A., Laughlin, S.T., Agard, N.J., Chang, P. V, Miller, I.A., … Bertozzi, C.R. (2007). Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences of the United States*

of America **104**(43):16793–16797.

23. Bateman, K.S., Brownie, E.R., Wolodko, W.T. and Fraser, M.E. (2002). Structure of the Mammalian CoA Transferase from Pig Heart. *Biochemistry* **41**(49):14455–14462.
24. Baumann, T., Schmitt, F., Pelzer, A., Spiering, V.J., Johannes, G., Sass, F. Von and Friedrich, T. (2018). Engineering ' Golden ' Fluorescence by Selective Pressure Incorporation of Non-canonical Amino Acids and Protein Analysis by Mass Spectrometry and Fluorescence. *Journal of Visualized Experiments*(134):1–14.
25. Belmonte, G., Menestrina, G., Pederzolli, C., Križaj, I., Gubenšek, F., Turk, T. and Maček, P. (1994). Primary and secondary structure of a pore-forming toxin from the sea anemone, *Actinia equina* L., and its association with lipid vesicles. *BBA - Biomembranes* **1192**(2):197–204.
26. Belmonte, G., Pederzolli, C., Maček, P. and Menestrina, G. (1993). Pore formation by the sea anemone cytolysin equinatoxin II in red blood cells and model lipid membranes. *The Journal of Membrane Biology* **131**(1):11–22.
27. Benelli, G., Jeffries, C.L. and Walker, T. (2016). Biological control of mosquito vectors: Past, present, and future. *Insects* **7**(4):1–18.
28. Benson, E.L., Huynh, P.D., Finkelstein, A. and Collier, R.J. (1998). Identification of residues lining the anthrax protective antigen channel. *Biochemistry* **37**(11):3941–3948.
29. Bi, Y., Zhang, Y., Shu, C., Crickmore, N., Wang, Q., Du, L., … Zhang, J. (2014). Genomic sequencing identifies novel *Bacillus thuringiensis* Vip1/Vip2 binary and Cry8 toxins that have high toxicity to Scarabaeoidea larvae. *Applied Microbiology and Biotechnology* **99**(2):753–760.
30. Birk, U.J. (2019). Super-Resolution Microscopy of Chromatin. *Gene* **10**(493):1–15.
31. Braakman, I. and Hebert, D.N. (2013). Protein Folding in the Endoplasmic Reticulum. *Cold Spring Harb Perspect Biol* **5**(5):1–17.
32. Brejc, K., Sixma, T., Kitts, P., Kain, ST., Tsien, R. and Ormo, M. (1997). Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl. Acad. Sci* **94**:2306–2311.
33. Bräse, S., Gil, C., Knepper, K. and Zimmermann, V. (2005). Organic Azides : An

Exploding Diversity of a Unique Class of Compounds. *Angew. Chem. Int. Ed* **44**:5188–5240.

34. Bradley, K ., Mogridge , J., Mourez, M., Collier,R. and Young, J. (2001). Identification of the cellular receptor for anthrax toxin. *Nature* **414**(6860):225–9.
35. Brossier, F., Lévy, M., Landier, A., Lafaye, P. and Mock, M. (2004). Functional analysis of *Bacillus anthracis* protective antigen by using neutralizing monoclonal antibodies. *Infection and Immunity* **72**(11):6313–6317.
36. Bowen, D. J., Chay, C. A., Flasinski, S., and Yin, Y. (2017).Novel insect inhibitory proteins. US Patent 2017/0044568.
37. Cai, S.X., Glenn, D.J., Kyle, R., Weber, E. and Keana, J.F.W. (1993). Chlorinated Phenyl Azides as Photolabeling Reagents . Synthesis of ortho , ortho ' Dichlorinated Arylazido PCP Receptor Ligand. *Bioconjugate Chem* **4**:545–548.
38. Campanini, B., Pioselli, B., Raboni, S., Felici, P., Giordano, I., D'Alfonso, L., … Bettati, S. (2013). Role of histidine 148 in stability and dynamics of a highly fluorescent GFP variant. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1834**(4):770–779.
39. Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A. and Tsien, R.Y. (2002). A monomeric red fluorescent protein. *Biochemistry* **99**(12):7877–7882.
40. Chakroun, M., Banyuls, N., Bel, Y., Escriche, B. and Ferré, J. (2016). Bacterial Vegetative Insecticidal Proteins (Vip) from *Entomopathogenic Bacteria*. *Microbiology and Molecular Biology Reviews* **80**(2):329–350.
41. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994). Green Fluorescent Protein as a Marker for Gene Expression. *Science* **263**(5148):802–805.
42. Chattopadhyay, A., Bhatnagar, N.B. and Bhatnagar, R. (2004). Bacterial insecticidal toxins. *Critical Reviews in Microbiology* **30**(1):33–54.
43. Chen, Y., Ebright, Y.W. and Ebright, R.H. (1994). Identification of the Target of a Transcription Activator Protein by Protein-Protein Photocrosslinking. *Science* **265**:90–92.
44. Chen, Z. and Zeng, A.P. (2016). Protein engineering approaches to chemical biotechnology. *Current Opinion in Biotechnology* **42**:198–205.
45. Chopra, A.P., Boone, S.A., Liang, X. and Duesbery, N.S. (2003). Anthrax Lethal

Factor Proteolysis and Inactivation of MAPK Kinase *. *The Journal of Biological Chemistry* **278**(11):9402–9406.

46. Chothia, C. and Lesk, A.M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO journal* **5**(4):823–826.
47. Collier, R.J. and Young, J.A.T. (2003). Anthrax Toxin. *Annu. Rev. Cell Dev. Biol* **19**:45–70.
48. Colovos, C. and Yeates, T.O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science* **2**(9):1511–1519.
49. Combet, C., Jambon, M., Deléage, G. and Geourjon, C. (2002). Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics* **18**(1):213–214.
50. Cohen, M. and Chang, P. (2018). Insights into the biogenesis, function, and regulation of ADP- ribosylation. *Nat Chem Biol* **14**(3):236–243.
51. Cormack, B.P., Valdivia, R.H. and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**(1):33–38.
52. Costa, S., Almeida, A., Castro, A. and Domingues, L. (2014). Fusion tags for protein solubility , purification , and immunogenicity in *Escherichia coli* : the novel Fh8 system. *Frontiers In Microbiology* **5**:1–20.
53. Crameri, A., Whitehorn, E.A., Tate, E. & Stemmer, W.P.. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature biotechnology* **14**(3):315–319.
54. Crickmore, N., Zeigler, D.R., Feitelson, J., Schnepf, H.E., Van Rie, J., Lereclus, D., … Dean, D.H. (1998). Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. *Microbiology and Molecular Biology Reviews : MMBR* **62**(3):807–813.
55. Crickmore, N., Zeigler, D.R., Schnepf, E., van Rie, J., Lereclus, D., Baum, J., Bravo, A. and Dean, D.H. *Bacillus thuringiensis* Toxin Nomenclature. Available online: http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/vip.html (accessed on 10 June 2019).
56. Cunningham, K., Lacy, D.B., Mogridge, J. and Collier, R.J. (2002). Mapping the lethal factor and edema factor binding sites on oligomeric anthrax protective antigen. *Proceedings of the National Academy of Sciences* **99**(10):7049–7053.

57. Deng, Q. and Barbieri, J.T. (2008). Molecular Mechanisms of the Cytotoxicity of ADP-Ribosylating Toxins. *Annual Review of Microbiology* **62**(1):271–288.
58. De la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M. and Sorzano, C.O.S. (2013). Xmipp 3.0: An improved software suite for image processing in electron microscopy. *Journal of Structural Biology* **184**(2):321–328.
59. Dimaio, F., Terwilliger, T.C., Read, R.J., Wlodawer, A., Oberdorfer, G., Wagner, U., … Baker, D. (2011). Improved molecular replacement by density- and energy-guided protein structure optimization. *Nature* **473**(7348):540–543.
60. Dirks, A.J.T., Berkel, S.S. Van, Hatzakis, N.S., Opsteen, J.A., Delft, F.L. Van, Cornelissen, J.J.L.M., … Nolte, R.J.M. (2005). Preparation of biohybrid amphiphiles via the copper catalysed Huisgen [3 + 2] dipolar cycloaddition reaction {. *Chem Commun*(33):4172–4174.
61. Dommerholt, J., Rutjes, F.P.J.T. and Delft, F.L. van (2016). Strain-Promoted 1,3-Dipolar Cycloaddition of Cycloalkynes and Organic Azides. *Top Curr Chem* **16**(374):229–241.
62. Donovan, W.P., Engleman, J.T., Donovan, J.C., Baum, J.A., Bunkers, G.J., Chi, D.J., … Walters, M.R. (2006). Discovery and characterization of Sip1A: A novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae. *Applied Microbiology and Biotechnology* **72**(4):713–719.
63. Doublié, S. (2007). Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. *Methods in Molecular Biology* **363**:91–108.
64. Delano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **40**: 82-92.
65. Edwards, W.R., Busse, K., Allemann, R.K. and Jones, D.D. (2008). Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Research* **36**(13).
66. Edwards, Y. and Cottage, A. (2003). Bioinformatics Methods to Predict Protein Structure and Function: A Practical Approach. *Molecular Biotechnology* **23**:139–166.
67. Esposito, D. and Chatterjee, D.K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology* **17**(4):353–358.

68. Estruch, J.J., Warren, G.W., Mullins, M.A., Nye, G.J., Craig, J.A. and Koziel, M.G. (1996). Vip3A, a novel *Bacillus thuringiensis* vegetative insecticidal protein with a wide spectrum of activities against lepidopteran insects. *Proc. Natl. Acad. Sci. USA* **93**(May):5389–5394.
69. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* **62**(1):72–82.
70. Evans, P. and McCoy, A. (2007). An introduction to molecular replacement. *Acta Crystallographica* **64**(1):1–10.
71. Fallas, J.A., Ueda, G., Shef, W., Nguyen, V., Mcnamara, D.E., Sankaran, B., … Baker, D. (2017). Computational design of self-assembling cyclic protein homo-oligomers. *Nature chemistry* **9**:353–360.
72. Farah, S., Korichi, H., Zendaoui, S., Saillard, J. and Zouchoune, B. (2009). The coordination of azepine to transition-metal complexes: A DFT analysis. *Inorganica Chimica Acta* **362**:3541–3546.
73. Fereja, T.H., Hymete, A. and Gunasekaran, T. (2013). A Recent Review on Chemiluminescence Reaction, Principle and Application on Pharmaceutical Analysis. *ISRN Spectroscopy* **2013**:1–12.
74. Ferré, J. and Escriche, B. (2017). Editorial for special issue: The insecticidal bacterial toxins in modern agriculture. *Toxins* **9**(12):3–6.
75. ffrench-Constant, R.H., Dowling, A. and Waterfield, N.R. (2007). Insecticidal toxins from *Photorhabdus* bacteria and their potential use in agriculture. *Toxicon* **49**(4):436–451.
76. Foot, J.S., Lui, F.E. and Kluger, R. (2009). Hemoglobin bis-tetramers via cooperative azide – alkyne coupling. *Chemical Communications*:7315–7317.
77. Gao-Sheridan, S., Zhang, S. and Collier, R.J. (2003). Exchange characteristics of calcium ions bound to anthrax protective antigen. *Biochemical and Biophysical Research Communications* **300**(1):61–64.
78. Garman, E. and Murray, J.W. (2003). Heavy-atom derivatization. *Acta Crystallographica - Section D Biological Crystallography* **59**(11):1903–1913.
79. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server;(In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press.571-607.

80. Geng, J., Jiang, J., Shu, C., Wang, Z., Song, F., Geng, L., … Zhang, J. (2019). *Bacillus thuringiensis* Vip1 Functions as a Receptor of Vip2 Toxin for Binary Insecticidal Activity against *Holotrichia parallela*. *Toxicon* **11**(440):1–12.
81. Geny, B. and Popoff, M.R. (2006). Bacterial protein toxins and lipids: pore formation or toxin entry into cells. *Biology of the Cell* **98**(11):667–678.
82. Gerding, D.N., Johnson, S., Rupnik, M. and Aktories, K.,(2013). *Clostridium difficile* binary toxin CDT Mechanism , epidemiology , and potential clinical importance *Gut Microbes*. **5**(1):15-27
83. Gomis-Cebolla, J., Ricietto, A.P.S. and Ferré, J. (2018). A genomic and proteomic approach to identify and quantify the expressed *Bacillus thuringiensis* proteins in the supernatant and parasporal crystal. *Toxins* **10**(5):1–18.
84. Goodsell, D.S. & Olson, A.J. (2000). Structural Symmetry and Protein Function. *Annual Review of Biophysics and Biomolecular Structure* **29**:105–153.
85. Gopal, G.J. and Kumar, A. (2013). Strategies for the production of recombinant protein in *Escherichia coli*. *Protein Journal* **32**(6):419–425.
86. Gordon, V.M., Leppla, S.H. and Hewlett, E.L. (1988). Inhibitors of receptor-mediated endocytosis block the entry of *Bacillus anthracis* adenylate cyclase toxin but not that of Bordetella pertussis adenylate cyclase toxin. *Infection and Immunity* **56**(5):1066–1069.
87. Griesbeck, O., Baird, G.S., Campbell, R.E., Zacharias, D.A. and Tsien, R.Y. (2001). Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *Journal of Biological Chemistry* **276**(31):29188–29194.
88. Grigorenko, B.L., Krylov, A.I. and Nemukhin, A. V. (2017). Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *Journal of the American Chemical Society* **139**(30):10239–10249.
89. Gubbens, J., Ruijter, E., Fays, L.E.V. De, Damen, J.M.A., Kruijff, B. De, Slijper, M., … Kroon, A.I.P.M. De (2009). Photocrosslinking and Click Chemistry Enable the Specific Detection of Proteins Interacting with Phospholipids at the Membrane Interface. *Chemistry & Biology* **16**(1):3–14.
90. Hammill, J.T., Miyake-Stoner, S., Hazen, J.L., Jackson, J.C. and Mehl, RA.(2007). Preparation of site- specifically labeled fluorinated proteins for ^{19}F-NMR structural characterization. *Nat Protoc*. **2** (10): 2601-2607.

91. Han, S., Arvai, A.S., Clancy, S.B. and Tainer, J.A. (2001). Crystal structure and novel recognition motif of Rho ADP-ribosylating C3 exoenzyme from *Clostridium botulinum*: Structural insights for recognition specificity and catalysis. *Journal of Molecular Biology* **305**(1):95–107.
92. Han, S., Craig, J.A., Putnam, C.D., Carozzi, N.B. and Tainer, J.A. (1999). Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. *Nature Structural Biology* **6**(10):932–936.
93. Hanuš, M. (2016). PhD Thesis. Charles University in Prague.
94. Hartley, A.M., Worthy, H.L., Reddington, S.C. and Rizkallah, J. and Jones .D (2016). Molecular basis for functional switching of GFP by modi fi cations of a phenyl azide reaction handle †. *Chemical Science* **7**:6484–6491.
95. Heim, R., Prasher, D. C. and Tsien, R.Y. (1994). Wavelenght mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U. S. A.* **91**:12501–12504.
96. Heim , R., Cubitt , A. and Tsien,T. (1995). Improved green fluorescence. *Nature* **373**(6516):663.
97. Heim, R. and Tsien, R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Current Biology* **6**(2):178–182.
98. Hendrickson, T. L. , Crécy-Lagard, V. and Schimmel, P. (2004). Incorporation of nonnatural amino acids into proteins. *Annu. Rev. Biochem* **73**(1):147–176.
99. Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal* **9**(5):1665–1672.
100. Hendrickson, W.A. and Ogata, C.M. (1997). Phase determination from multiwavelength anomalous diffraction measurements. *Methods in Enzymology* **276**(1993):494–523.
101. Hirota, S., Hattori, Y., Nagao, S., Taketa, M., Komori, H., Kamikubo, H., … Higuchi, Y. (2010). Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proceedings of the National Academy of Sciences* **107**(29):12854–12859.
102. Hoorens, M.W.H. and Szymanski, W. (2018). Reversible, Spatial and Temporal

Control over Protein Activity Using Light. *Trends in Biochemical Sciences* **43**(8):567–575.

103. Huang, A.P., Oberdorfer, G., Xu, C., Pei, X.Y. and Brent, L. (2014). High thermodynamic stability of parametrically designed helical bundles. *Science* **346**(6208):481–486.

104. Hutchison, C.A., Phillips, S., Edgell, M.H., Gillam, S., Jahnke, P. and Smith, M. (1978). Mutagenesis at a specific position in a DNA sequence. *Journal of Biological Chemistry* **253**(18):6551–6560.

105. Inui, H., Sawada, K., Oishi, S., Ushida, K. and McMahon, R.J. (2013). Aryl nitrene rearrangements: Spectroscopic observation of a benzazirine and its ring expansion to a ketenimine by heavy-atom tunneling. *Journal of the American Chemical Society* **135**(28):10246–10249.

106. Jackson, S.E., Craggs, T.D. and Huang, J.R. (2006). Understanding the folding of GFP using biophysical techniques. *Expert Review of Proteomics* **3**(5):545–559.

107. Jankielsohn, A. (2018). The Importance of Insects in Agricultural Ecosystems. *Advances in Entomology* **6**:62–73.

108. Jiang, J., Pentelute, B.L., Collier, R.J. and Hong Zhou, Z. (2015). Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521**(7553):545–549.

109. Jiang, R., Zhang, X. and Zhang, M.Q. (2013). *Basics of Bioinformatics*. Beijing: Tsinghua University Press, Springer.

110. Jisha, V.N., Smitha, R.B. and Benjamin, S. (2013). An Overview on the Crystal Toxins from *Bacillus thuringiensis. Advances in Microbiology* **03**(05):462–472.

111. Joseph, B.C., Pichaimuthu, S. and Srimeenakshi, S. (2015). An Overview of the Parameters for Recombinant Protein Expression in *Escherichia coli*. *Journal of Cell Science & Therapy* **06**(05).

112. Jung, K., Park, J., Maeng, P. and Kim, H. (2005a). Fluorescence Quenching of Green Fluorescent Protein during Denaturation by Guanidine. *Bull. Korean Chem. Soc* **26**(3):413–417.

113. Jung, G., Wiehler, J. and Zumbusch, A. (2005b). The photophysics of green fluorescent protein: Influence of the key amino acids at positions 65, 203, and 222. *Biophysical Journal* **88**(3):1932–1947.

114. Kaczanowski, S. and Zielenkiewicz, P. (2010). Why similar protein sequences

encode similar three-dimensional structures? *Theoretical Chemistry Accounts* **125**:643–650.

115. Kajihara, D., Hohsaka, T. and Sisido, M. (2005). Synthesis and sequence optimization of GFP mutants containing aromatic non-natural amino acids at the Tyr66 position. *Protein Engineering, Design and Selection* **18**(6):273–278.

116. Kanwar, M., Wright, R.C., Date, A., Tullman, J. and Ostermeier, M. (2013). Chapter Seventeen- Protein Switch Engineering by Domain Insertion. In *Methods in Enzymology*, Volume 523. Elsevier Inc.370-388.

117. Kapoor, S., Rafiq, A. and Sharma, S. (2017). Protein engineering and its applications in food industry. *Critical Reviews in Food Science and Nutrition* **57**(11):2321–2329.

118. Karasawa, A., Michalski, K., Mikhelzon, P. and Kawate, T. (2017). The P2X7 receptor forms a dye-permeable pore independent of its intracellular domain but dependent on membrane lipid composition. *eLife* **6**(1):1–22.

119. Khakshoor, O., Lin, A.J., Korman, T.P., Sawaya, M.R., Tsai, S.C., Eisenberg, D. and Nowick, J.S. (2010). X-ray crystallographic structure of an artificial β-sheet dimer. *Journal of the American Chemical Society* **132**(33):11622–11628.

120. Kiefer, F., Arnold, K. and Ku, M. (2009). The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research* **37**:387–392.

121. Kim, H., Siu, K.H., Raeeszadeh-Sarmazdeh, M., Sun, Q., Chen, Q. and Chen, W. (2015). Bioengineering strategies to generate artificial protein complexes. *Biotechnology and Bioengineering* **112**(8):1495–1505.

122. Kintzer, A.F., Thoren, K.L., Sterling, H.J., Dong, K.C., Feld, G.K., Tang, I.I., … Krantz, B.A. (2009). The Protective Antigen Component of Anthrax Toxin Forms Functional Octameric Complexes. *Journal of Molecular Biology* **392**(3):614–629.

123. Kozakov, D., Hall, D., Xia, B., Porter, K., Padhorny, D., Yueh, C., Beglov,D. and Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols* **12**(2):255–278.

124. Kobayashi, K., Nagahama, M., Ohkubo, N., Kojima, T., Shirai, H., Iwamoto, S., … Sakurai, J. (2008). Role of Ca2+-binding motif in cytotoxicity induced by *Clostridium perfringens* iota-toxin. *Microbial Pathogenesis* **44**(4):265–270.

125. Kobayashi, N and Arai, R. (2017). Design and construction of self-assembling

supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. *Current Opinion in Biotechnology* **46**:57–65.

126.Kobayashi, T., Sakamoto, K., Takimura, T., Sekine, R., Kelly, V.P., Kamata, K., … Yokoyama, S. (2005). Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion. *Proceedings of the National Academy of Sciences* **102**(5):1366–1371.

127.Kolb, H.C., Finn, M.G. and Sharpless, K.B. (2001). Click Chemistry : Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Edit* **40**(11):2004–21.

128.Kostallas, G., Löfdahl, P.Å. and Samuelson, P. (2011). Substrate profiling of tobacco Etch virus protease using a novel Fluorescence-Assisted whole-cell assay. *PLoS ONE* **6**(1):1–10.

129.Krishnan, Y., Rees, H.A., Rossitto, C.P., Kim, S.-E., Hung, H.-H.K., Frank, E.H., … Grodzinsky, A.J. (2018). Green fluorescent proteins engineered for cartilage-targeted drug delivery: Insights for transport into highly charged avascular tissues. *Biomaterials* **183**(June):218–233.

130.Kushibiki, T., Okawa, S., Hirasawa, T. and Ishihara, M. (2014). Optogenetics: Novel tools for controlling mammalian cell functions with light. *International Journal of Photoenergy*:1–10.

131.Kwon, I. and Yang, B. (2017). Bioconjugation and Active Site Design of Enzymes Using Non-natural Amino Acids. *Industrial and Engineering Chemistry Research* **56**(23):6535–6547.

132.Laemmli, U.K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**(5259):680–685.

133.Laing, S., Unger, M., Koch-Nolte, F. and Haag, F. (2011). ADP-ribosylation of arginine. *Amino Acids* **41**(2):257–269.

134.Lang, K. and Chin, Jason W (2014a). Bioorthogonal Reactions for Labeling Proteins. *ACS Chemical Biology* **9**:16–20.

135.Lang, K. and Chin, Jason W. (2014b). Cellular incorporation of unnatural amino acids and bioorthogonal Labeling of Proteins. *Chemical Reviews* **114**(9):4764–4806.

136.Larsen, T.A., Olson, A.J. and Goodsell, D.S. (1998). Morphology of protein – protein interfaces. *Structure*(6):421–427.

137.Leibly, D.J., Arbing, M.A., Pashkov, I., Devore, N., Waldo, G.S., Terwilliger, T.C. and Yeates, T.O. (2015a). A Suite of Engineered GFP Molecules for Oligomeric Scaffolding. *Structure* **23**(9):1754–1768.

138.Lee, M., Walters, F., Hart, H., Palekar, N. and Chen,J. (2003). The Mode of Action of the *Bacillus thuringiensis* Vegetative Insecticidal Protein Vip3A Differs from That of Cry1Ab delta-Endotoxin. *Applied and Environmental Microbiology* **69**(8):4648–4657.

139.Leuber, M., Orlik, F., Schiffler, B., Sickmann, A. and Benz, R. (2006). Vegetative insecticidal protein (Vip1Ac) of *Bacillus thuringiensis* HD201: Evidence for oligomer and channel formation. *Biochemistry* **45**(1):283–288.

140.Leppla, S. (1982). Anthrax toxin edema factor: a bacterial adenylatecyclase that increases cyclic AMP concentrations in eukaryotic cells. *Proc.Natl. Acad. Sci. USA* **79**(10):3162–3166.

141.Lim, S. and Kwon, I. (2016). Bioorthogonal Modification of Proteins Using Genetically Encoded Non-Natural Amino Acids. *Current Organic Chemistry* **20**(11):1232–1242.

142.Link, A.J. and Tirrell, D.A. (2003). Cell Surface Labeling of *Escherichia coli* via Copper (I) -Catalyzed [3 + 2] Cycloaddition. *J. AM. CHEM. SOC* **125**(37):11164–11165.

143.Ljubetič, A., Gradišar, H. and Jerala, R. (2017). Advances in design of protein folds and assemblies. *Current Opinion in Chemical Biology* **40**:65–71.

144.Los, F.C.O., Randis, T.M., Aroian, R. V. and Ratner, A.J. (2013). Role of Pore-Forming Toxins in Bacterial Infectious Diseases. *Microbiology and Molecular Biology Reviews* **77**(2):173–207.

145.Ma, Y., Sun, Q., Zhang, H., Peng, L., Yu, J.-G. and Smith, S.C. (2010). The mechanism of cyclization in chromophore maturation of green fluorescent protein: A theoretical study. *Journal of Physical Chemistry B* **114**(29):42–46.

146.Marianayagam, N.J., Sunde, M. and Matthews, J.M. (2004). The power of two: Protein dimerization in biology. *Trends in Biochemical Sciences* **29**(11):618–625.

147.Marques, S.M. and Esteves Da Silva, J.C.G. (2009). Firefly bioluminescence: A mechanistic approach of luciferase catalyzed reactions. *IUBMB Life* **61**(1):6–17.

148.Marth, G., Hartley, A.M., Reddington, S.C., Sargisson, L.L., Parcollet, M., Dunn, K.E., … Stulz, E. (2017). Precision Templated Bottom-Up Multiprotein

Nanoassembly through Defined Click Chemistry Linkage to DNA. *ACS Nano* **11**(5):5003–5010.

149. Mathews, F.S., Bethge, P.H. and Czerwinskis, E.W. (1979). The Structure of Cytochrome *b* $_{562}$ from *Escherichia coli* at 2,5 °A Resolution . *J. Biol. Chem* **254**(5):1699–1706.

150. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography* **40**(4):658–674.

151. McNicholas, S., Potterton, E., Wilson, K.S. and Noble, M.EM.(2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D* **67**: 386–394

152. Mei, G., Di Venere, A., Rosato, N. and Finazzi-Agrò, A. (2005). The importance of being dimeric. *FEBS Journal* **272**(1):16–27.

153. Miyawaki,A.,Grisbeck,O., Heim,R. and Tisen, R. (1999). Dynamic and quantitative Ca2 measurements using improved cameleons. *Cell Biology* **96**(5):2135–2140.

154. Mogridge, ., Cunningham, K., Lacy,D., Mourez, M. and Collier, R. (2002). The lethal and edema factors of anthrax toxin bind only to oligomeric forms of the protective antigen. *Proceedings of the National Academy of Sciences* **99**(10):7045–7048.

155. Milnes, J.C., Furlong, D., Hannas, P.C., Walls, J.S. and Collierh, R.J. (1994). Anthrax Protective Antigen Forms Oligomers during Intoxication of Mammalian Cells *. *The Journal of Biological Chemistry* **269**(32):20607–20612.

156. Mishin, A.S., Subach, F. V., Yampolsky, I. V., King, W., Lukyanov, K.A. and Verkhusha, V. V. (2008). The first mutant of the *Aequorea victoria* green fluorescent protein that forms a red chromophore. *Biochemistry* **47**(16):4666–4673.

157. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., … Finn, R.D. (2018). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research* **47**:1–10.

158. Mogridge, J., Mourez, M. and Collier, R.J. (2001). Involvement of domain 3 in oligomerization by the protective antigen moiety of anthrax toxin. *Journal of Bacteriology* **183**(6):2111–2116.

159.Molloy, S.S., Bresnahan, P.A., Leppla, S.H., Klimpel, K.R. and Thomas, G. (1992). Human furin is a calcium-dependent serine endoprotease that recognizes the sequence Arg-X-X-Arg and efficiently cleaves anthrax toxin protective antigen. *Journal of Biological Chemistry* **267**(23):16396–16402.

160.Monod, J., Wyman, J. and Changeux, J. (1964). On the Nature of Allosteric Transitions : A Plausible Model. *Journal of Molecular Biology* **12**:88–118.

161.Morise, H., Shimomura, O., Johnson, F. and Winant J. (1974). Intermolecular energy transfer in the bioluminescent system of Aequorea. *Biochemistry* **13**(12):2656–2662.

162.Morris, J.L., Reddington, S.C., Murphy, D.M., Jones, D.D., Platts, J.A. and Tippmann, E.M. (2013). Aryl azide photochemistry in defined protein environments. *Organic Letters* **15**(4):728–731.

163.Nagai, T, Ibata, K., Park, E., Kubota, M. and Mikoshiba, K. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* **20**:87–90.

164.Nagamune, T.(2017). Biomolecular engineering for nanobio/bionanotechnology. *Nano Convergence* **4**(1):9.

165.Nagaoka, M., Jiang, H.L., Hoshiba, T., Akaike, T. and Cho, C.S. (2010). Application of recombinant fusion proteins for tissue engineering. *Annals of Biomedical Engineering* **38**(3):683–693.

166.Nahar, P., Wali, N.M. and Gandhi, R.P. (2001). Light-Induced Activation of an Inert Surface for Covalent Immobilization of a Protein Ligand. *Analytical Biochemistry* **294**:148–153.

167.Nguyen, A.W. and Daugherty, P.S. (2005). Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nature Biotechnology* **23**(3):355–360.

168.Nikic, I., Kang, J.H., Girona, G.E., Aramburu, I.V. and Lemke, E.A. (2015). Labeling proteins on live mammalian cells using click chemistry.*Natural protocols*.10(5):780-791.

169.Nikkila, H., Gennis, R.B. and Sligar, S.G. (1991). Cloning and expression of the gene encoding the soluble cytochrome b 562 of *Escherichia coli*. *European Journal of Biochemistry* **202**(2):309–313.

170.Nilsson, M., Wang, X., Rodziewicz-Motowidlo, S., Janowski, R., Lindström, V.,

Önnerfjord, P., … Grubb, A. (2004). Prevention of domain swapping inhibits dimerization and amyloid fibril formation of cystatin C. Use of engineered disulfide bridges, antibodies, and carboxymethylpapain to stabilize the monomeric form of cystatin C. *Journal of Biological Chemistry* **279**(23):24236–24245.

171. Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y. and Remingtont, S.J. (1996). Crystal Structure of the *Aequorea victoria* Green Fluorescent Protein. *Science* **273**:1392–1395.

172. Ornelas, C., Broichhagen, J. and Weck, M. (2010). Strain-promoted alkyne azide cycloaddition for the functionalization of poly(amide)-based dendrons and dendrimers. *Journal of the American Chemical Society* **132**(11):3923–3931.

173. Ortega, J., Singh, S.K., Ishikawa, T., Maurizi, M.R. and Steven, A.C. (2000). Visualization of substrate binding and translocation by the ATP-dependent protease, ClpXP. *Molecular Cell* **6**(6):1515–1521.

174. Palma, L., de Escudero, I.R., Maeztu, M., Caballero, P. and Muñoz, D. (2013). Screening of vip genes from a Spanish *Bacillus thuringiensis* collection and characterization of two Vip3 proteins highly toxic to five lepidopteran crop pests n. *Biological Control* **66**(3):141–149.

175. Palma, L., Muñoz, D., Berry, C., Murillo, J., Caballero, P. and Caballero, P. (2014). *Bacillus thuringiensis* toxins: An overview of their biocidal activity. *Toxins* **6**(12):3296–3325.

176. Palma, L., Scott, D.J., Harris, G., Din, S.U., Williams, T.L., Roberts, O.J., … Berry, C. (2017). The Vip3Ag4 insecticidal protoxin from *Bacillus thuringiensis* adopts a tetrameric configuration that is maintained on proteolysis. *Toxins* **9**(5).

177. Papatheodorou, P., Wilczek, C., Nölke, T., Guttenberg, G., Hornuss, D., Schwan, C. and Aktories, K. (2012). Identification of the Cellular Receptor of *Clostridium spiroforme* Toxin. :1418–1423.

178. Paquette, L.A., Kuhla, D.E. and Barrett, J.H. (1969). Unsaturated heterocyclic systems. LIII. Thermochemical reactions of 1H-azepine derivatives. 2. Aromatization and sigmatropic migrations involving nitrogen. *The Journal of Organic Chemistry* **34**(10):2879–2884.

179. Pardo-López, L., Soberón, M. and Bravo, A. (2013). *Bacillus thuringiensis* insecticidal three-domain Cry toxins: Mode of action, insect resistance and

consequences for crop protection. *FEMS Microbiology Reviews* **37**(1):3–22.

180.Payne, C.C. (1988). Pathogens for the control of insects: where next? *Philosophical Transactions of the Royal Society B* **318**:225–248.

181.Pannifer,A., Wong,T., Schwarzenbacher, R., Renatus, M., Petosa, C., Bienkowska , J., … Liddington, R. (2001). Crystal structure of the anthrax lethal factor. *Nature* **414**(6860):225–229.

182.Pédelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **24**(1):79–88.

183.Perozzol, M.A., Ward, K.B., Thompsonq, R.B. and Wardll, W. (1988). X-ray Diffraction and Time-resolved Fluorescence Analyses of Aequorea Green Fluorescent Protein Crystals *. *Journal of Biological Chemistry* **263**(16):7713–7716.

184.Petosa , C., Collier , R., Klimpel, K., Leppla, S. and Liddington, R. (1997). Crystal structure of the anthrax toxin protective antigen. *Nature* **385**:833–838.

185.Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003) "A graphical user interface to the CCP4 program suite. *Acta. Cryst.D* **59**: 1131-1137

186.Pike, A.C.W., Garman, E.F., Krojer, T., Von Delft, F. and Carpenter, E.P. (2016). An overview of heavy-atom derivatization of protein crystals. *Acta Crystallographica Section D: Structural Biology* **72**(3):303–318.

187.Plass, T., Milles, S., Koehler, C., Schultz, C. and Lemke, E.A. (2011). Genetically encoded copper-free click chemistry. *Angewandte Chemie - International Edition* **50**(17):3878–3881.

188.Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G. and Cormier, M.J. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* **111**(2):229–233.

189.Radke, J., Pederson, K.J. and Barbieri, J.T. (1999). *Pseudomonas aeruginosa* exoenzyme S is a biglutamic acid ADP- ribosyltransferase. *Infection and Immunity* **67**(3):1508–1510.

190. Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. (1963). "Stereochemistry of polypeptide chain configurations". *Journal of Molecular Biology* **7**: 95–99.

191.Reddington, S.C. (2013). Introducing novel protein functionality using unnatural

amino acids. PhD Thesis.Cardiff Universtiy.

192.Reddington, S.C., Baldwin, A.J., Thompson, R., Brancale, A., Tippmann, M. and Jones, D.D. (2015a). Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence †. *The Royal Society of Chemistry* **6**:1159–1166.

193.Reddington, S., Driezis, S., Hartley, A., Watson, P., Rizkallahb,P. and Jones, D. (2015b). Genetically encoded phenyl azide photochemistry drives positive and negative functional modulation of a red fluorescent protein. *The Royal Society of Chemistry* **5**:77734–77738.

194.Reddington, S.C., Rizkallah, P.J., Watson, P.D., Pearson, R., Tippmann, E.M. and Jones, D.D. (2013). Different Photochemical Events of a Genetically Encoded Phenyl Azide Define and Modulate GFP Fluorescence **. *Angewandte*:5974–5977.

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

196.Rekas, A., Alattia, J.R., Nagai, T., Miyawaki, A. and Ikura, M. (2002). Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. *Journal of Biological Chemistry* **277**(52):50573–50578.

197.Remington, S.J. (2011). Green fluorescent protein: A perspective. *Protein Science* **20**(9):1509–1519.

198.Rigden, D.J., Mello, L. V, Galperin, M.Y., Kluma, B. and Epa, C. (2004). The PA14 domain , a conserved all- b domain in bacterial toxins , enzymes , adhesins and signaling molecules. *TRENDS in Biochemical Sciences* **29**(7):335–339.

199.Roberts, T.M., Rudolf, F., Meyer, A., Pellaux, R., Whitehead, E., Panke, S. and Held, M. (2016). Identification and Characterisation of a pH-stable GFP. *Scientific Reports* **6**:1–8.

200.Rocha, C.L., Coburn, J., Rucks, E.A. and Olson, J.C. (2003). Characterization of *Pseudomonas aeruginosa* Exoenzyme S as a Bifunctional Enzyme in J774A .1 Macrophages. *Infection and Immunity* **71**(9):5296–5305.

201.Roh, J.Y., Choi, J.Y., Li, M.S., Jin, B.R. and Je, Y.H. (2007). *Bacillus*

thuringiensis as a specific, safe, and effective tool for insect pest control. *Journal of Microbiology and Biotechnology* **17**(4):547–559.

202. Ros, U., Edwards, M.A., Epand, R.F., Lanio, M.E., Schreier, S., Yip, C.M., … Epand, R.M. (2013). The sticholysin family of pore-forming toxins induces the mixing of lipids in membrane domains. *Biochimica et Biophysica Acta - Biomembranes* **1828**(11):2757–2762.
203. Rosano, G.L. and Ceccarelli, E.A. (2014). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology* **5**(APR):1–17.
204. Rosenow, M.A., Huffman, H.A., Phail, M.E. and Wachter, R.M. (2004). The Crystal Structure of the Y66L Variant of Green Fluorescent Protein Supports a Cyclization - Oxidation - Dehydration Mechanism for Chromophore Maturation †,‡. *Biochemistry* **43**(15):4464–4472.
205. Rostovtsev, V. V, Green, L.G., Fokin, V. V and Sharpless, K.B. (2002). A stepwise Huisgen Cycloaddition process: Copper (i)-Catalyzed Regioselective Ligation of Azides and Terminal Alkynes. *Angew Chem Int Edit* **114**(14):2708–11.
206. Ruiu, L. (2013). *Brevibacillus laterosporus*, a pathogen of invertebrates and a broad-spectrum antimicrobial species. *Insects* **4**(3):476–492.
207. Sanahuja, G., Banakar, R., Twyman, R.M., Capell, T. and Christou, P. (2011). *Bacillus thuringiensis*: A century of research, development and commercial applications. *Plant Biotechnology Journal* **9**(3):283–300.
208. Sattar, S. and Maiti, M.K. (2011). Molecular characterization of a novel vegetative insecticidal protein from *Bacillus thuringiensis* effective against sap-sucking insect pest. *Journal of Microbiology and Biotechnology* **21**(9):937–946.
209. Sakurai, J., Nagahama, M., Oda, M., Tsuge, H. and Kobayashi, K. (2009). *Clostridium perfringens* Iota-Toxin: Structure and Function. *Toxins* **1**(2):208–228.
210. Schleberger, C., Hochmann, H., Barth, H., Aktories, K. and Schulz, G.E. (2006). Structure and Action of the Binary C2 Toxin from *Clostridium botulinum*. *Journal of Molecular Biology* **364**(4):705–715.
211. Schmid, A., Benzs, R., Ingo, J., Aktories and Klaus (1994). Interaction of *Clostridium botulinum* C2 Toxin with Lipid Bilayer Membranes. *The Journal of*

Biological Chemistry **269**(24):16706–16711.

212. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**(7):671–675.

213. Schrock, A.K. and Schuster, G.B. (1984). Photochemistry of Phenyl Azide : Chemical Properties of the Transient Intermediates. *J Am Chem Soc* **106**(18):5228–5234.

214. Schnepf, H., Narva, K., Stockhoff, B., Lee, S. and Mikki Walz, B.S. (2005). Pesticidal toxins and genes from *Bacillus laterosporus* strains. *United States Patent 6,956,116.*

215. Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R. and Dean D.H.(1998). *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiology and molecular biology reviews* **62**(3):775–806.

216. Schünemann, R., Knaak, N. and Fiuza, L.M. (2014). Mode of Action and Specificity of *Bacillus thuringiensis* Toxins in the Control of Caterpillars and Stink Bugs in Soybean Culture. *ISRN Microbiology* **2014**:1–12.

217. Scobie, H.M., Wigelsworth, D.J., Marlett, J.M., Thomas, D., Rainey, G.J.A., Lacy, D.B., … Young, J.A.T. (2006). Anthrax toxin receptor 2-dependent lethal toxin killing in vivo. *PLoS Pathogens* **2**(10):0949–0955.

218. Scott T. Laughlin, Jeremy M. Baskin,1 Sharon L. Amacher, C.R.B. (2008). In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* **320**(5876):664–668.

219. Seifert, M.H.J., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B. and Holak, T.A. (2003). Backbone dynamics of green fluorescent protein and the effect of Histidine 148 substitution. *Biochemistry* **42**(9):2500–2512.

220. Seifert, M.H.J., Ksiazek, D., Azim, M.K., Smialowski, P., Budisa, N. and Holak, T.A. (2002). Slow exchange in the chromophore of a green fluorescent protein variant. *Journal of the American Chemical Society* **124**(27):7932–7942.

221. Sellami, S., Jemli, S., Abdelmalek, N., Dabbéche, E. and Jamoussi, K. (2016). Localization and in silico study of the vegetative insecticidal proteins Vip2S-Vip1S of *Bacillus thuringiensis*. *International Journal of Biological Macromolecules* **91**:510–517.

222. Shi, Y., Xu, W., Yuan, M., Tang, M., Chen, J. and Pang, Y. (2004). Expression of vip1/vip2 genes in *Escherichia coli* and *Bacillus thuringiensis* and the analysis

of their signal peptides. *Journal of Applied Microbiology* **97**(4):757–765.

223. Shimomura, O., F.H. Johnson, and Y.S. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* **59**(3):223–240.

224. Shingote, P.R., Moharil, M.P., Dhumale, D.R., Jadhav, P. V., Satpute, N.S. and Dudhare, M.S. (2013). Screening of vip1/vip2 binary toxin gene and its isolation and cloning from local *Bacillus thuringiensis* isolates. *ScienceAsia* **39**(6):620–624.

225. Shukla, R., Dubey, A., Pandey, V., Golhani, D. and Jain, A.P. (2012). Chromophore- An Utility in UV Spectrophotometer. *Inventi Rapid: Pharm Ana & Qual Assur*(3):1-4

226. Siemering, K.R., Golbik, R., Sever, R. and Haseloff, J. (1996). Mutations that suppress the thermosensitivity of green fluorescent protein. *Current Biology* **6**(12):1653–1663.

227. Sigal, I.S., Harwood, B.G. and Arentzen, R. (1982). Thiol-beta-lactamase: replacement of the active-site serine of RTEM beta-lactamase by a cysteine residue. *Proceedings of the National Academy of Sciences* **79**(23):7157–60.

228. Singh, M.S., Chowdhury, S. and Koley, S. (2016). Advances of azide-alkyne cycloaddition-click chemistry over the recent decade. *Tetrahedron* **72**(35):5257–5283.

229. Singh, P., Sharma, L., Kulothungan, S.R., Adkar, B. V., Prajapati, R.S., Ali, P.S.S., … Varadarajan, R. (2013). Effect of Signal Peptide on Stability and Folding of *Escherichia coli* Thioredoxin. *PLoS ONE* **8**(5).

230. Sivakumar, K., Xie, F., Cash, B.M., Long, S., Barnhill, H.N. and Wang, Q. (2004). A fluorogenic 1,3-dipolar cycloaddition reaction of 3-azidocoumarins and acetylenes. *Organic Letters* **6**(24):4603–4606.

231. Smith, M. (1982). Site-directed mutagenesis. *Trends in Biochemical Science Science* **7**(12):440–442.

232. Sniegowski, J.A., Phail, M.E. and Wachter, R.M. (2005). Maturation efficiency, trypsin sensitivity, and optical properties of Arg96, Glu222, and Gly67 variants of green fluorescent protein. *Biochemical and Biophysical Research Communications* **332**:657–663.

233. Song, W.J. and Tezcan, F.A. (2014). A designed supramolecular protein

assembly with in vivo enzymatic activity. *Science* **346**(6216):1525–1528.

234. Sockolosky, J. and Szoka, F. (2013). Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. *Protein Expr Purif* **87**(2):129–135.

235. Spreter Von Kreudenstein, T., Lario, P.I. and Dixit, S.B. (2014). Protein engineering and the use of molecular modeling and simulation: The case of heterodimeric Fc engineering. *Methods* **65**(1):77–94.

236. Stark, J.C., Huang, A., Nguyen, P.Q., Dubner, R.S., Hsu, K.J., Ferrante, T.C., … Ottman, S. (2018). BioBits ™ Bright : A fluorescent synthetic biology education kit. *Science Advances* **4**(8):1–11.

237. Strub, M.P., Hoh, F., Sanchez, J.F., Strub, J.M., Böck, A., Aumelas, A. and Dumas, C. (2003). Selenomethionine and selenocysteine double labeling strategy for crystallographic phasing. *Structure* **11**(11):1359–1367.

238. Sun, J. and Collier, R.J. (2010). Disulfide Bonds in the Ectodomain of Anthrax Toxin Receptor 2 Are Required for the Receptor-Bound Protective-Antigen Pore to Function. *PLoS ONE* **5**(5):1–9.

239. Sun, J., Maresso, A.W., Kim, J.J.P. and Barbieri, J.T. (2004). How bacterial ADP-ribosylating toxins recognize substrates. *Nature Structural and Molecular Biology* **11**(9):868–876.

240. Taylor, G. (2003). The phase problem. *Acta Crystallographica - Section D Biological Crystallography* **59**(11):1881–1890.

241. Thamri, A., Létourneau, M., Djoboulian, A., Chatenet, D., Déziel, E., Castonguay, A. and Perreault, J. (2017). Peptide modification results in the formation of a dimer with a 60-fold enhanced antimicrobial activity. *PLoS ONE* **12**(3):1–12.

242. Torres-Quintero, M.C., Gómez, I., Pacheco, S., Sánchez, J., Flores, H., Osuna, J., … Bravo, A. (2018). Engineering *Bacillus thuringiensis* Cyt1Aa toxin specificity from dipteran to lepidopteran toxicity. *Scientific Reports* **8**(1):1–12.

243. Tounsi, S. and Jaoua, S. (2002). Identification of a promoter for the crystal protein-encoding gene cry1Ia from *Bacillus thuringiensis* subsp. kurstaki. *FEMS Microbiology Letters* **208**(2):215–218.

244. Tropea, J.E., Cherry, S. and Waugh, D.S. (2009). Chapter 19. Expression and Purification of Soluble His6 -Tagged TEV Protease. In: *Methods in Molecular*

Biology: High Throughput Protein Expression and Purification. Humana Press. Totowa, NJ **498**(1):297–307.

245.Tsien, R.Y. (1998). the Green Fluorescent Protein. *Annual Review of Biochemistry* **67**(1):509–544.

246.Vad, B.S., Bertelsen, K., Johansen, C.H., Pedersen, J.M., Skrydstrup, T., Nielsen, N.C. and Otzen, D.E. (2010). Pardaxin permeabilizes vesicles more efficiently by pore formation than by disruption. *Biophysical Journal* **98**(4):576–585.

247.Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S.E., Xia, B., … Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function and Bioinformatics* **85**(3):435–444.

248.Van Amsterdam, I.M.C., Ubbink, M., Einsle, O., Messerschmidt, A., Merli, A., Cavazzini, D., … Canters, G.W. (2002). Dramatic modulation of electron transfer in protein complexes by crosslinking. *Nature Structural Biology* **9**(1):48–52.

249.Van Wely, K.H.M., Swaving, J., Freudl, R. and Driessen, A.J.M. (2001). Translocation of proteins across the cell envelope of Gram-positive bacteria. *FEMS Microbiology Reviews* **25**(4):437–454.

250.Villa, J.K., Tran, H., Vipani, M., Gianturco, S., Bhasin, K., Russell, B.L., Harbron, E. and Young D.D. (2017). Fluorescence Modulation of Green Fluorescent Protein Using Fluorinated Unnatural Amino Acids. *molecules* **22**(1194):1–7.

251.Wachter, R.M., King, B.A., Heim, R., Kallio, K., Tsien, R.Y., Boxer, S.G. and Remington, S.J. (1997). Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. *Biochemistry* **36**(32):9759–9765.

252.Walde, P., Cosentino, K., Engel, H. and Stano, P. (2010). Giant Vesicles: Preparations and Applications. *ChemBioChem* **11**(7):848–865.

253.Walden, H. (2010). Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D: Biological Crystallography* **66**(4):352–357.

254.Waldo, G.S., Standish, B.M., Berendzen, J. and Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluoresecnt protein. *Nature Biotechnology* **17**(7):691–695.

255.Walker, J.M. (2010). *Computational Biological*. 673rd ed. Humana Press. Hatfield.

256.Wandrey, G., Wurzel, J., Hoffmann, K., Ladner, T., Büchs, J., Meinel, L. and Lühmann, T. (2016). Probing unnatural amino acid integration into enhanced green fluorescent protein by genetic code expansion with a high- throughput screening platform. *Journal of Biological Engineering* **10**(11):1–13.

257.Wang, F., Niu, W., Guo, J. and Schultz, P.G. (2012). Unnatural amino acid mutagenesis of fluorescent proteins. *Angewandte Chemie - International Edition* **51**(40):10132–10135.

258.Wang, L., Brock, A., Brad, H. and Schultz, P.G. (2001). Expanding the genetic code of *Escherichia coli*. *Science* **292**(5516):498–500.

259.Wang, L., Xie, J., Deniz, A.A. and Schultz, P.G. (2003). Unnatural amino acid mutagenesis of green fluorescent protein. *Journal of Organic Chemistry* **68**(1):174–176.

260.Warren, G.W. (1997). Vegetative Insecticidal Proteins: Novel Proteins for Control of Corn Pests. *In: Advances in Insect Control: The Role of Transgenic Plants (Carozzi, N. B., Koziel, M. G., Eds.)*. Taylor & Francis Ltd, London, United Kingdom.

261.West, T.S. (1971). Luminescent Molecules. *Nature* **230**:66.

262.Winter, G. (2010). Xia2: An expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography* **43**(1):186–190.

263.Wood, T.I., Barondeau, D.P., Hitomi, C., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. (2005). Defining the Role of Arginine 96 in Green Fluorescent Protein Fluorophore Biosynthesis. *Biochemistry*(44):16211–16220.

264.Worthy, H.L. (2018). Novel routes to defined post translational modifications using non-canonical amino acids . PhD Thesis. Cardiff University.

265.Worthy, H.L., Auhim, H.S., Jamieson, W.D., Pope, J.R., Wall, A., Batchelor, R., … Jones, D.D. (2019). Positive functional synergy of structurally integrated artificial protein dimers assembled by Click chemistry. *Communications Chemistry* **2**(83):1–12.

266.Wu, G., Feng, C., Hong, Y., Guo, A., Cao, S., Dong, J., … Liu, Z. (2010). Soluble expression and purification of the anthrax protective antigen in *E. coli* and identification of a novel dominant-negative mutant N435C. *Applied Microbiology and Biotechnology* **87**:609–616.

267.Xie, J., Wang, L., Wu, N., Brock, A., Spraggon, G. and Schultz, Peter G (2004).

The site-specific incorporation of *p*-iodo- L -phenylalanine into proteins for structure determination. *Nature Biotechnology* **22**(10):1297–1301.

268.Xie,J. and Schultz,P. (2006). A chemical toolkit for proteins an expanded genetic code. *Nature Reviews Molecular Cell Biology* **7**:775–782.

269.Yasuda, K., Sugimoto, H., Hayashi, K., Takita, T., Yasukawa, K., Ohta, M., … Sakaki, T. (2018). Protein engineering of CYP105s for their industrial uses. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1866**(1):23–31.

270.Yang, F., Moss, L. and Phillips, G. (1996). The molecular structure of green fluorescent protein. *Nature Biotechnology* **14**(10):1246–1251.

271.Yilmaz, N. and Kobayashi, T. (2016). Assemblies of pore-forming toxins visualized by atomic force microscopy. *Biochimica et Biophysica Acta - Biomembranes* **1858**(3):500–511.

272.Yu, X., Liu, T., Liang, X., Tang, C., Zhu, J., Wang, S., … Li, P. (2011). Rapid detection of vip1-type genes from *Bacillus cereus* and characterization of a novel vip binary toxin gene. *FEMS Microbiology Letters* **325**(1):30–36.

273.Yueh, C., Hall, D.R., Xia, B., Padhorny, D., Kozakov, D. and Vajda, S. (2017). ClusPro-DC: Dimer Classification by the Cluspro Server for Protein–Protein Docking. *Journal of Molecular Biology* **429**(3):372–381.

274.Zacharia , J.T. (2011). Identity, Physical and Chemical Properties of Pesticides. *Pesticides in the Modern World - Trends in Pesticides Analysis* **1873**.

275.Zaki, A.J., Hartley, A.M., Reddington, S.C., Thomas, S.K., Watson, P., Hayes, A., … Elliott, M. (2018). Defined covalent assembly of protein molecules on graphene using a genetically encoded photochemical reaction handle. *The Royal Society of Chemistry* **8**:5768–5775.

276.Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* **9**:40.

277.Zhang, Y., Baranov, P. V., Atkins, J.F. and Gladyshev, V.N. (2005). The relationship between protein structure and function: A comprehensive survey with application to the yeast genome. *Journal of Molecular Biology* **280**(21):20740–20751.

278.Zhao, K.Q., Hartnett, J. and Slater, M.R. (2007). Selenomethionine Protein Labeling Using the *Escherichia coli* Strain KRX. *Promega Notes* **96**(96):24–26.