De novo designed proteins for applications in research and biotechnology

By George Andrew Sutherland

Department of Molecular Biology and Biotechnology

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

September, 2018

In loving memory of Megan and Derek Sutherland.

Acknowledgements

First and foremost I would like to thank Prof. Neil Hunter for his continued support and supervision over the last four years; Neil's enthusiasm for science and dedication to the success of his students are remarkable and will stay with me throughout my future career. A special mention must go to Dr. Andrew Hitchcock; without his support and tutelage, I would not have been able to complete this PhD. Another mention goes to Matthew 'Sharky' Proctor for suffering through the challenges of a PhD with me and for putting up with me for four years.

I would like to thank all members of the Hunter group, both past and present, for their friendship and guidance since joining the lab. Special mentions must go to Jack Chidgey for his help in pigment analysis and introduction to the website 'reddit'; Nathan Adams for his training in a plethora of biophysical techniques; Amanda Brindley for helping with protein production and purification; Phil Jackson for his help in mass spectrometry, Daniel Canniffe for his help in the final conclusions and David Swainsbury for explaining molecular dynamics simulations to a technophobe.

I would also like to thank Dr. Jenny Clark for what turned out to be a fruitful collaboration, as well as her feedback on the spectroscopy and physical chemistry in this thesis. I would also like to thank Daniel Polak and Jozra Garrido Velasco for their help in the acquisition and interpretation of the transient absorption spectroscopy, as well as Andrew Musser for his explanations of physical chemistry.

Most of all I would like to thank my family for their love and support over the years. To Jonny, for being a constant source of inspiration. To my Father, for your boundless positivity and encouragement to work hard – I do occasionally 'go the extra mile' now. To Uncle David for your continued interest in the work I'm doing and hugely helpful support. And to my Mother, for your constant, unwavering support.

Τελευταίο αλλά εξίσου σημαντικό, στην αγάπη μου. Σε ευχαριστώ για όλα όσα έχεις κάνει στη ζωή μου. Ανυπομονώ για το μέλλον μας.

Abstract

While the advances of the scientific community have enabled extraordinary improvements in the capabilities of synthetic biology, there is a continued desire in biotechnology for enhanced or entirely novel biological functions. As proteins are either directly or indirectly responsible for the vast majority of naturally occurring biological activities, the modification of peptide structures constitutes a promising approach to address the ambitions of biotechnology. Central to the work in this thesis is the recognition that naturally occurring protein structures are intangibly complex due to the relics of evolutionary processes, accumulated from years of blind natural selection. Chapter 1 introduces *de novo* protein design strategies that circumvent the use of naturally occurring peptide scaffolds, offering examples of tractable systems that have been generated to perform various biological functions, thus forming the justification for the experimental approach undertaken here.

The experimental work detailed in chapters 3 and 4 aimed to develop a system that would enable the incorporation of carotenoids and acenes into the internal cavity of *de novo*-designed 'maquette' proteins by hydrophobic partitioning alone. The results of these sections demonstrated that the protein chassis had little or no effect in the photophysical properties of the incorporated chromophores, whilst providing enhanced stability and solubility in entirely aqueous solutions. Conversely, the experimental strategy outlined in chapter 5 aimed to introduce nuclei of high atomic mass into the maquette proteins in order to directly affect the photophysical properties of the bound chromophores through the spin-orbit coupling interaction. The results of the final experimental chapter demonstrated that *de novo* designed proteins could effectively interface with native biological systems and provide a mechanism to enable cofactor incorporation *in vivo*. Where appropriate, the results of each experimental section are discussed in relation to their impact on specific areas of research and potential applications in biotechnology.

Contents

List of figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

List of tables

Abbreviations and short-hand nomenclature

Abbreviation	Description
A.U.	Absorbance units
ACN	Acetonitrile
AEC	Anion exchange chromatography
AXT	Astaxanthin
C8-BTBT	2,7-dioctyl[1]benzothieno[3,2-b][1]benzothiophene
Ccm	Cytochrome c maturation
CD	Circular dichroism
CpcA	Phycocyanin subunit a
D_2O	Deuterated water
DEAE	Diethylaminoethyl
DMSO	Dimethyl sulphoxide
ECH	Echinenone
EDTA	Ethylenediaminetetraacetic acid
E_m	Redox midpoint potential
EPR	Electron paramagnetic resonance
ESI	Electrospray ionisation
eTet	Elongated TIPS-tetracene
FPLC	Fast protein liquid chromatography
FRET	Förster
FT	Freeze-thaw
GFP	Green fluorescent protein
H-aggregate	Hypsochromic-aggregate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOMO	Highest occupied molecular orbital
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
J-aggregate	Jelley-aggregate
K_d	Dissociation constant
LB	Luria-Bertani
LC	Liquid chromatography
LH2	Light harvesting complex 2
LUMO	Lowest unoccupied molecular orbital
m/z	Mass to charge ratio
MD	Molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MS	Mass spectrometry
MW	Molecular weight
NMR	Nuclear magnetic resonance
NTMet	N-terminal methionine

OD	Optical density
OLED	Organic light emitting diode
PCR	Polymerase chain reaction
PDB ID	Protein databank identification
PelB	Pectate lyase B
PIA	Photoinduced absorption
PL	Photoluminescence
PMF	Proton motive forve
PV	Photovoltaic
PVDF	Polyvinylidene fluoride
QH_2O	Millipore filtered water
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
Sec	Secretory
SEC	Size exclusion chromatography
SeM	Selenomethionine
SF	Singlet fission
S_n	Singlet spin state
SOC	Spin orbit coupling
TA	Transient absorption
Tat	Twin-arginine translocase
TFA	Trifluroacetic acid
TIPS	Triisopropylsilyl
T_m	Melting transition temperature
T_n	Triplet spin state
TOF	Time-of-flight
TorA	trimethylamine-N-oxide reductase
TTA	Triplet-triplet annihilation
UV	Ultraviolet
UV-Vis	Ultraviolet-visible
βC	β-carotene

Chapter 1

Introduction

1.1 Overview

The field of synthetic biology encompasses a broad range of scientific research areas and can generally be described as the predictable alteration or adaptation of native biological systems to generate novel products or functions (Purnick and Weiss, 2009; Khalil and Collins, 2010). The challenges of such man-made design can often be described by Richard Feynman's axiom "what I cannot create, I do not understand" (Paz, 1989), where creation can only be brought about when a biological process is sufficiently understood. The principle of informed creation is evident in the adaptation of many biological processes, from genetic regulation and sequence identity (Khalil and Collins, 2010), to protein design (Fletcher et al., 2012) and the adaptation of whole biochemical networks (Chen et al., 2018). In this thesis, recent progress in artificial protein design will be discussed with a particular focus on the application of function for use in biotechnological processes. Central to the work described here is the recognition that natural protein structures are intangibly complex and the relationship between sequence, structure and function in such protein scaffolds remains insufficiently understood to facilitate de novo creation.

In this chapter some of the main approaches to artificial protein generation are introduced, with emphasis on protein design strategies that circumvent the use of natural peptide sequences. The scope of this thesis will not provide a comprehensive account of historical protein design, but rather introduce some of the recent advances that inform the experimental strategies detailed in chapters 3-6 and highlight some of the achievements in related areas to aid discussion in chapter 7. This chapter will also provide the necessary background information to understand the areas of biotechnology in which the artificial proteins may be applied.

1.2 Artificial protein rationale

Despite the diversity and impressive catalytic functionality of naturally occurring proteins, there is a desire in many areas of biotechnology for improved and entirely novel enzymatic reactions (Korendovych and DeGrado, 2014). Computational and non-computational design strategies (reviewed in Khoury et al., 2014 and Maeda et al., 2016, respectively) have made remarkable breakthroughs, generating native protein scaffolds with altered substrate specificity (Headstrom et al., 1992), modified allosteric regulation (Yang et al., 2012) and novel catalytic functions (Siegel et al., 2010; Arnold, 2015). The modification of natural protein scaffolds typically involves the process of directed evolution, involving successive rounds of generating and selecting for variation (Goldsmith and Tawfik, 2012; Arnold, 2015). Although shown to be successful in a number of cases, modifying natural proteins can be considered an extension of evolutionary processes and its use in *de novo* protein design has been questioned (Dutton and Moser, 2011; Lichtenstein et al., 2012, Huang et al., 2016a).

Muller (1964) states that over time genetic systems accumulate irreversible mutations and interdependency leading to a one-directional evolutionary force. This ‘ratchet’ mechanism of evolution, driven by rounds of blind natural selection, leads to intangible complexity and fragility in natural protein sequences (Kraut et al., 2003; Dutton and Moser, 2011). Thus, the incremental nature of evolution means that the structure and make up of naturally occurring proteins does not cover the entire sequence space available, instead grouping proteins into small families of similar structure and function. It can therefore be argued that the design space available in the generation of *de novo* proteins would be restricted if evolutionary methods were used in isolation (Huang et al., 2016a).

Redundancy and degeneracy can be observed at all levels of a biological system. Redundancy can be defined as a process or function that is performed by identical elements, for example, the duplication of genes resulting in functionally identical products. Degeneracy occurs when structurally different elements are able to carry out the same function or yield the same product, for example, the ability of different codons to encode identical amino acids (Edelman and Gally, 2001). Although such features of biological systems are necessary to ensure adaptation and robustness to environmental changes (Barkai and Leibler, 1997; Csete and Doyle, 2002), it has been argued that this complexity is not a prerequisite for achieving catalytic function in protein engineering (Collman et al., 2004; Koder et al., 2009). Rather, it has been proposed that simple protein structures, free from the degeneracy and evolutionary relics of natural structures, may offer an alternative in generating enzymes with improved and novel biological functions. This ‘bottom-up’ approach to *de novo* protein design involves the rational engineering of simple peptide structures to fulfil a desired function. Unburdened by the complexity of evolved proteins, such an approach permits the formulation of general rules in artificial protein design and allows the full peptide sequence space to be sampled (Ghadiri and Case 1993; Chroma et al., 1994; DeGrado et al., 1999; Gibney et al., 1999; Hill et al., 2000; Koder and Dutton, 2006).

One of the driving forces behind the field of *de novo* protein design was a desire to test current understanding of native protein folding in a controlled, iterative experimental manner, with the aim of generating some fundamental rules for folding that would facilitate accurate structural prediction from primary amino acid sequence alone (Regan and DeGrado, 1988; DeGrado et al., 1989). Since its inception in the late 1980s, the field of protein design has seen rapid progress; largely attributable to the increasing power of computation and the reducing cost of DNA synthesis (reviewed in Regan et al., 2015 and Huang et al., 2016a).

Beginning with the generation of relatively simple, self-assembling α-helix bundle proteins (Regan and DeGrado, 1988) and model β-sheet systems (reviewed in Gellman, 1998; Khakshoor and Nowick, 2008), fundamental rules surrounding protein design could be empirically determined. In conjunction to this experimental strategy, computational modelling has seen great advances in folding prediction with constraints largely derived from native protein topology. This has led to the generation of peptides with entirely non-native primary amino acid sequences that fold predictably into native-like (Dahiyat and Mayo, 1997) and entirely novel tertiary structures (Kuhlman et al., 2003). The culmination of this understanding has resulted in the predictable formation of sophisticated, multi-component structures, demonstrated in figure 1.1 (Bale et al., 2016; Huang et al., 2016b).

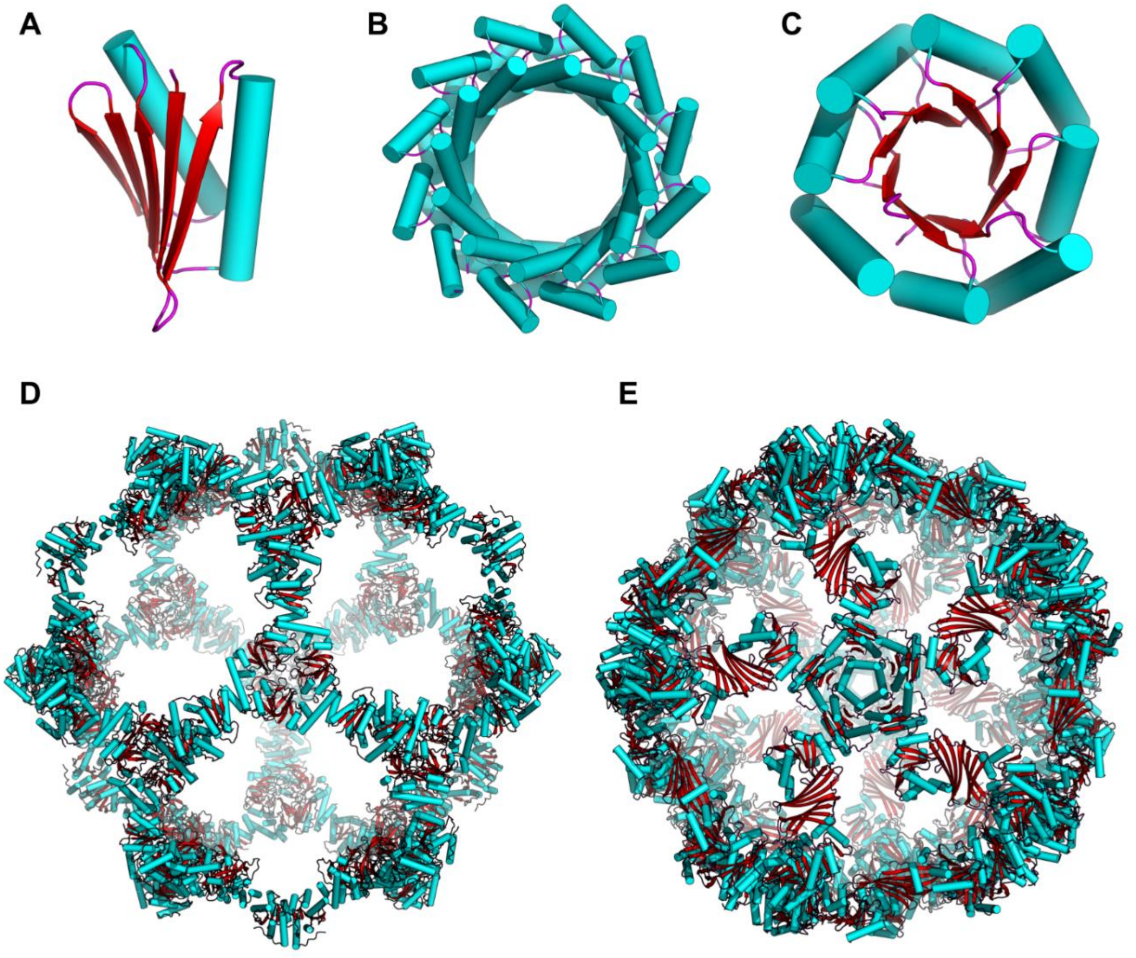


Figure 1.1 X-ray crystal structures of *de novo* designed proteins. (A) The artificial "Top7" peptide fold not found in natural protein structures (PDB ID: 1QYS; Kuhlman et al., 2003). **(B)** Structure of a *de novo* designed, left-handed α-helical closed-solenoid structure with no similarities found in the protein data bank (PDB ID: 5BYO; Doyle et al., 2015). **(C)** Structure of a *de novo* designed TIM-barrel protein (PDB ID: 5BVL; Huang et al., 2016b) **(D & E)** Large, co-assembling, two-component *de novo* designed protein nanostructures (Bale et al., 2016). All images were generated in PyMol (PyMol Molecular Graphics System Version 1.7, Schrödinger, LLC); secondary structure motifs are coloured in red (β-sheet), cyan (α-helix) and purple (loop).

1.3 Imparting functionality onto artificial protein structures

The advances described above demonstrate that fundamental understanding of protein folding and engineering has reached such a point that the generation of protein scaffolds with arbitrarily chosen structures is becoming possible (Huang et al., 2016a). Despite the range of artificial scaffolds available, the assignment of function has largely occurred in simpler structures, more reminiscent of the early helical bundle proteins created by DeGrado and co-workers (Regan and DeGrado 1988; DeGrado et al., 1989; Moser et al., 2016). As a comprehensive history of *de novo* protein function and design is beyond the scope of this thesis, the reader is directed to Moser et al., (2016), Ennist et al., (2017a), Mocny and Pecoraro (2015), and Huang et al., (2016a).

1.3.1 Maquette design

The generation of catalytically proficient *de novo* protein scaffolds was largely founded by the work of the Dutton laboratory, utilising the 'maquette' approach to protein engineering (Robertson et al., 1994; reviewed in Moser et al., 2016). The maquette approach draws on the long-established methods used by architects and sculptors in which simple models, known as maquettes, are progressively altered to give more detailed characteristics (Ennist et al., 2017a). The core principle of maquette design is to have the simplest primary amino acid sequence in order to achieve the desired protein structure and function. Starting from a simple structure, details, such as cofactor binding and functionality, can be included intuitively through an iterative process of experimentation and redesign. The result of this approach is that it becomes possible to reconcile the purpose of every amino acid in the structure and subsequent alterations become more predictable than in natural scaffolds (Dutton and Moser, 2011).

Maquette designs have solely adopted α-helical secondary structures for several reasons. First, the stability of helices in solution is well established (Bierzynski et al., 1982), as is the propensity of each amino acid to form helical structures (Monera et al., 1995). The intramolecular hydrogen bonding between backbone amide oxygens and nitrogens results in two complete helical turns every seven residues. This not only facilitates patterning over seven residue segments (heptad repeats), but also limits the potential for intermolecular hydrogen bond formation and aggregation, which is commonly found in model β-sheet structures (Gellman, 1998; Moser et al., 2018). Finally, the binary patterning of polar and non-polar residues within helical sections affords a simple design approach that can facilitate predictable self-assembly and structure (DeGrado et al., 1989; Kamtekar et al., 1993).

As the experimental work in chapters 3-6 solely employs tetrahelical bundle proteins, the discussion of design features will be limited to this arrangement; for a comprehensive review on other helical assemblies see Woolfson (2005). First described by Regan and DeGrado (1988), a simple amphiphilic α-helix can self-assemble into a tetrameric complex, with hydrophobic residues packing into the protein interior and polar residues projecting into the aqueous surroundings. Binary patterning of hydrophobic (H) and polar (P) residues in the helical sections of the protein followed the general HPPHXPX repeat motif (X may be either H or P). As shown in figure 1.2, the amphiphilic characteristic of the α-helix is achieved by the arrangement of hydrophobic residues in the first and fourth positions (*a* and *d*, respectively) of the heptad and polar residues in the second, third and sixth positions (*b*, *c* and *f*, respectively). A minimal protein sequence, solely comprised of Leucine (Leu, L) in hydrophobic positions and either Glutamate (Glu, E) or Lysine (Lys, K) in polar positions, was sufficient to form a self-assembling homotetrameric complex (DeGrado et al., 1989). The identical helices could then be tethered together through the introduction of proline-arginine (Pro, P; Arg, R) loop regions. Di-helical peptides co-assembled to form dimeric tetrahelical complexes, and the combination of four helical sections on a single polypeptide resulted in a self-folding tetrahelical protein. Maquette structures have subsequently developed from the principles described here, favouring the use of dimeric or monomeric tetrahelical assemblies.

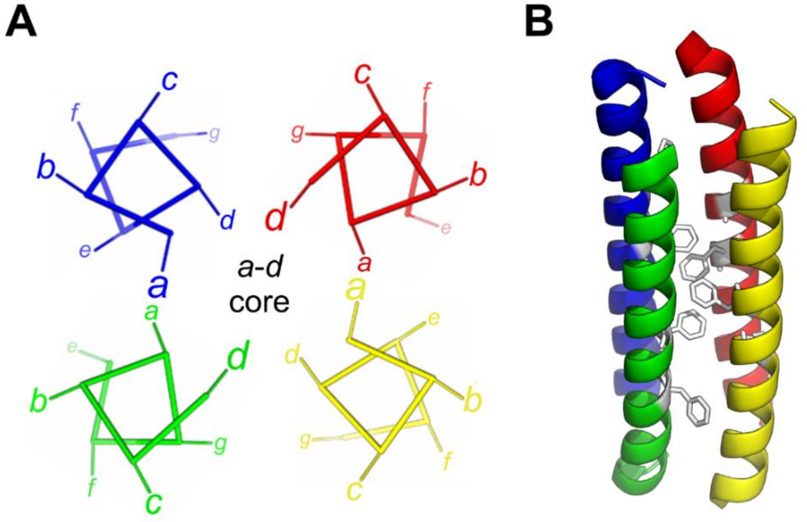


Figure 1.2 Tetrahelical protein design strategy (A) Intramolecular hydrogen bonding in α-helical peptides results in two helical turns every seven residues (*a*-*g*). Binary patterning of polar (P) and hydrophobic (H) residues according to the general HPPHXPX repeat motif results in a hydrophobic *a*-*d* core that drives spontaneous protein folding. **(B)** The crystal structure of the homotetrameric H10H24 maquette; hydrophobic phenylalanine residues are highlighted in white (PDB ID: 1M3W).

1.3.2 Maquette function

As cofactors are required for many protein-catalysed reactions, the integration of functionality into maquette scaffolds is largely dependent on the inclusion of cofactor binding sites (Koder and Dutton, 2006). Drawing inspiration from the non-covalent haem *b* coordination observed in the *b* subunit the cytochrome *bc*$_1$ complex (figure 1.3A, Xia et al., 1997; Iwata et al., 1998), Robertson et al., (1994) describe the synthesis of the first maquette protein, H10H24, a di-haem *b* binding tetrahelical bundle protein composed of two subunits. The inclusion of haem cofactors into the relatively simple structure was sufficient to impart redox functionality onto the protein, resulting in a tractable system to investigate energy-transfer processes.

However, further development of maquette proteins was largely hindered by the lack of structural information. Nuclear magnetic resonance (NMR) spectroscopy revealed typically molten globular characteristics in the *apo* and *holo* form of the H10H24 maquette, hindering efforts towards crystallisation and structural determination (Gibney et al., 1997; Huang et al., 2003). To promote structural uniqueness in the H10H24 maquette, an iterative redesign process was employed; it was found that minimal changes into the internal binding pocket (L6I and L13F) were sufficient to impose a uniquely structured core in the *apo* state (Gibney et al., 1999). The *apo* structure of a triple mutant (L6I, L13F, L31M) was subsequently solved by X-ray crystallography (figure 1.2B; Huang et al., 2003).

Despite imparting structural uniqueness in the *apo* state, the ligation of haem *b* into the maquette structure resulted in a protein with molten globule characteristics. Huang et al., (2004) describe the design of a structured haemoprotein (HP-1), based on the structurally determined L6I, L13F, L31M *apo* protein. Bis-histidine ligation of iron-porphyrins occurs through the coordination of the metal centre by the epsilon nitrogen of the two imidazole rings on either side of the molecule, with the ligating side chains oriented perpendicular to the plane of the macrocycle (figure 1.3B). Consequently, the coordination of haem into the maquette structure requires significant structural reorganisation, achieved through a 55° rotation of the helices and subsequently affecting the positioning of hydrophobic and polar residues as set out from the binary pattern design. The restructuring of the maquette following haem ligation resulted in previously buried hydrophobic residues becoming exposed to the aqueous surroundings and polar side chains packing into the protein core. The design of the HP-1 maquette accounted for the haem-induced helical rotations in the positioning of hydrophobic and hydrophilic residues, resulting in an unstructured *apo* protein that becomes structurally unique following haem insertion, as assessed by one-dimensional proton NMR (1D ^{1}H NMR).

Although shown to have numerous cofactor binding and oxidoreductase functions, the sequence duplication and symmetry of the homotetrameric and homodimeric maquette structures limits the ability of the protein to form sophisticated, asymmetric systems (Farid et al., 2013). Single polypeptide maquettes provide a means to create asymmetry between the four helices, permitting, for example, the positioning of single binding sites or redox-active amino acids (Moser et al., 2018). Farid et al., (2013) describe the design of single-chain tetrahelical maquettes (hereafter named BT6), similar in sequence to the HP-1 maquette, but with four helices joined by three identical loop regions (figure 1.3C). Relatively small changes in amino acid composition, affecting the hydrophobicity of core packing residues, were shown to greatly affect protein stability, as determined by temperature-dependent circular dichroism (CD) spectroscopy. The maquette scaffolds were shown to bind a range of cofactors, including Fe-porphyrins, Zn-porphyrins, Zn-chlorins and flavins (see section 1.3.3), resulting in single-chain multi-component artificial proteins. These findings demonstrated that it was possible for artificial proteins, with minimal alterations to sequence composition, to perform diverse oxidoreductase functions dependent on the identity of the cofactors incorporated.

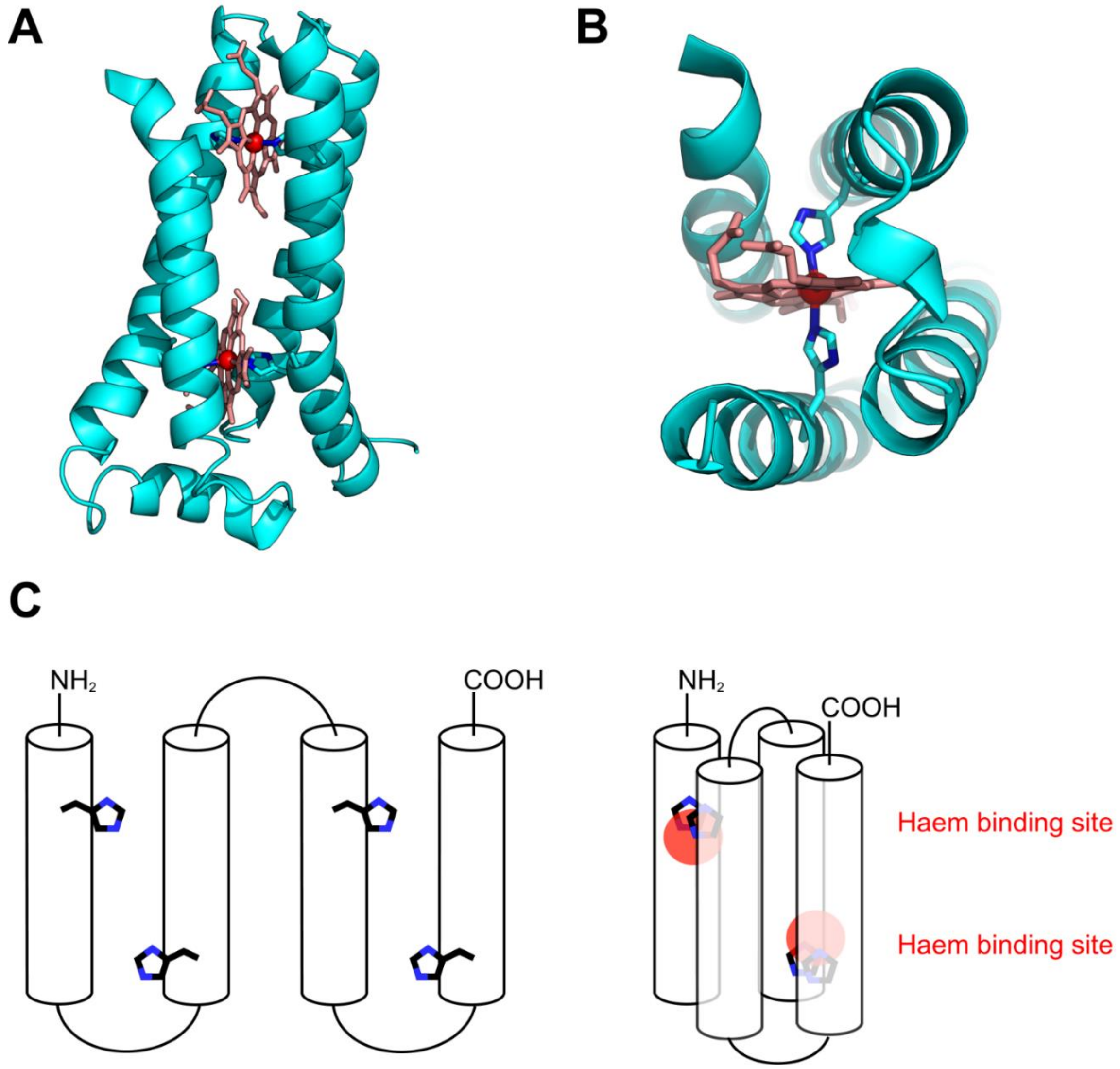


Figure 1.3 Haem maquette design strategy. The haem-binding domain of the cytochrome *b* subunit in the cytochrome *bc*$_1$ complex shown in parallel **(A)** and perpendicular to **(B)** the plane of the membrane. A section of the *b* subunit is shown in cyan, the haem macrocycle in pink, the ferric metal ion in red and imidazole nitrogens in the axial coordinating His residues are shown in blue. **(C)** Schematic diagram showing the proposed haem binding mechanism in the single-polypeptide BT6 maquette; the approximate locations of the coordinating His residues are shown in black, with nitrogen groups in blue and Fe-porphyrin binding sites in red. Structural images were generated in PyMol (PDB ID: 1BGY).

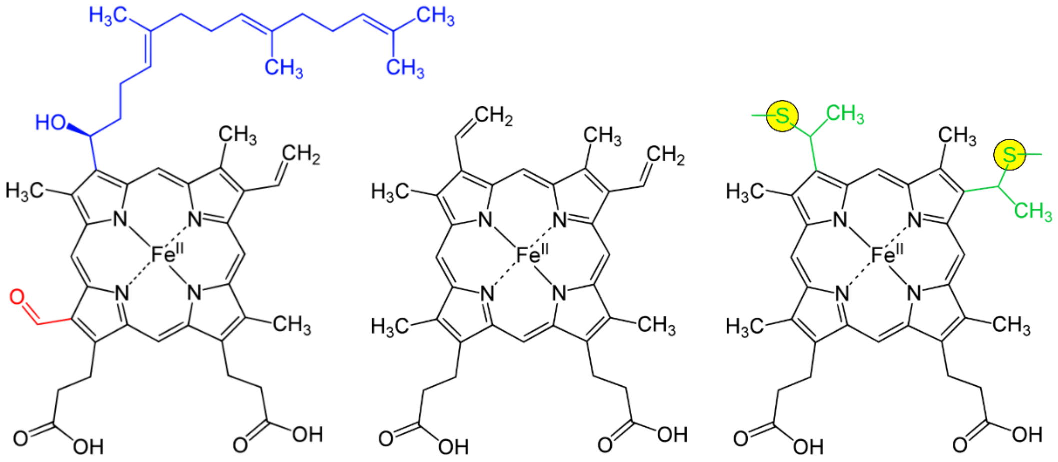
1.3.3 Maquette cofactors

Alongside the developments in structural design there was also a desire to extend the palette of available cofactors and binding motifs beyond the bis-histidine ligation of haem *b* discussed so far. Bis-histidine ligation of haem *a* was shown by Gibney et al., (2000) and Solomon et al., (2014) demonstrated the non-covalent incorporation of ten Fe-porphyrin variants into maquette constructs. The addition of the consensus CXXCH motif into a maquette structure, similar in sequence to BT6, was sufficient to form covalent thioester bonds between the two vinyl groups on the haem *c* macrocycle and the cysteine residues in the protein (Anderson et al., 2014).

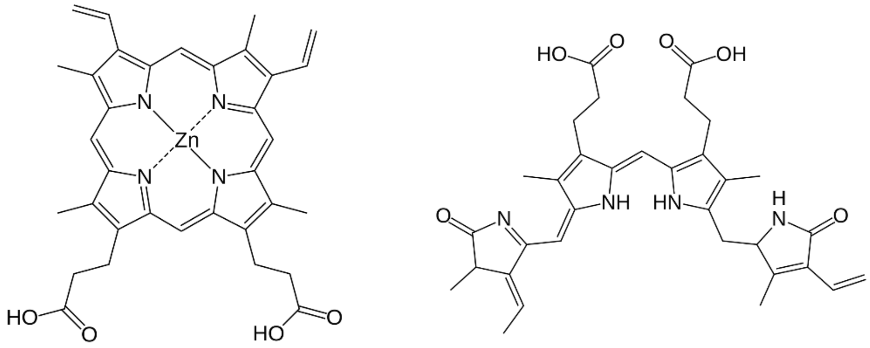
Zn-tetrapyrroles are found natively in the phototrophic bacterium *Acidiphilium rubrum* (Wichlacz et al., 1986; Jaschke et al., 2011) and provide a means to generate light-activated redox functionality in maquettes (see section 1.3.5) through mono-histidine ligation (Koder et al., 2006; Cohen-Ofri et al., 2011; Farid et al 2013; Kodali et al., 2017). Light-active, open-tetrapyrrole bilin pigments can also associate with maquette structures either covalently, through thioether bonds between vinyl groups and cysteine residues, or non-covalently through spontaneous partitioning into the hydrophobic protein core (Moser et al., 2018).

Flavins and quinones facilitate one- or two-electron transfer reactions in a variety of natural systems (Moser et al., 2016), and can also be found as light active redox cofactors in DNA photolyase (Sancar, 1994) and the plant blue-light photoreceptor (Lin et al., 1995). Although non-covalently bound in approximately 90% of native proteins (Moser et al., 2018), *de novo* maquette designs have employed the attachment of flavins by thioether linkage to cysteine residues (Sharp et al., 1998) and through protein semi-synthesis to incorporate the unnatural napthoquinone amino acid into maquette structures (Lichtenstein et al., 2015).

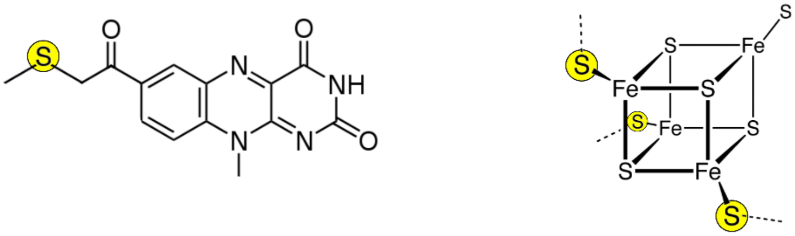
Tetranuclear iron-sulphur clusters (4Fe-4S) can be incorporated into maquette structures through the inclusion of a consensus peptide motif, approximately sixteen residues long, which forms an elongated loop region between otherwise unperturbed helical sections (Gibney et al., 1996). Iron-sulphur clusters can therefore be introduced into previously developed maquette assemblies without alteration to the existing helical sections.



Haem *a* Haem *b* Haem *c*



Zn-protoporphyrin Phycoerythrobilin



7-acetyl-10-methylisoalloxazine 4Fe-4S

Figure 1.4 Structural diagrams of maquette cofactors. Changes to the porphyrin macrocycle of haem *b* are highlighted. In haem *a* an hydroxyethylfarnesyl group is added at the C3 position (blue) and formyl group at C18 (red). In haem *c*, modification of the C3 and C8 vinyl groups leads to the formation of thioester bonds to cysteine residues (green). The structures of a Zn-porphyrin (Zn-protoporphyrin), bilin (Phycoerythrobilin), flavin (7-acetyl-10-methylisoalloxazine) and a tetranuclear iron-sulphur cluster (4Fe-4S) are also shown. Atoms provided by protein residues are highlighted in yellow

1.3.4 Modulating the oxidoreductase functions of maquettes

Oxidoreductase functions account for approximately 25% of the enzymatically catalysed reactions in the natural world (Dutton and Moser, 2011). Maquette proteins can begin to fulfil aspirations for functionality by using the inherent redox potentials of the cofactors discussed above to catalyse oxidoreductase reactions.

Observations made in mutants of naturally occurring cytochromes revealed that the redox midpoint potential (E_m) of bound haem cofactors could be modulated by altering electrostatic interactions with protein residues and changing the charge-state of nearby amino acids (Caffrey and Cusanovich, 1994; Shifman et al., 2000). In agreement with these findings, the E_m of haem molecules incorporated into maquette scaffolds has been modulated by controlling the protonation state of external residues through solution pH, the inclusion of redox-active amino acids in the internal hydrophobic cavity and mutations affecting the net charge of the protein exterior (Shifman et al., 2000; Farid et al., 2013). The selection of a particular cofactor, modulated by the charge state of the surrounding protein matrix, therefore permits a wide range of redox states (+170 mV to -290 mV for haem *a*), and as mutations to maquette scaffolds are predictable, the E_m of a particular cofactor can be tuned without consequences for protein structure and folding.

With the palette of available cofactors and means to modulate specific redox states, maquette proteins offer a robust system to study fundamental intramolecular electron transfers. Critical in determining electron transfer dynamics is the driving force (ΔG) and edge-to-edge distance of the donor (D) and acceptor (A) (Moser et al., 1992). Both of these criteria can be tightly controlled in maquette scaffolds by selecting cofactors with particular redox properties, modulated where necessary by the protein microenvironment, and defining donor and acceptor positions by adjusting ligand binding sites. Given the ΔG value and D/A distance, the electron transfer rate can be calculated through simulations (described in detail in Moser et al., 2018), allowing the redox properties of various cofactor configurations to be estimated *in silico*.

1.3.5 Light-driven oxidoreductase functions

Experimental determination of energy-transfer processes can be resolved through excitation of light-activated redox cofactors, such as Zn-tetrapyrroles, flavins and bilins (Moser et al., 2016), or through photolysis of small-molecule (e.g. carbon monoxide) haem ligands (Fry et al., 2016). Light-driven electron-transfer was shown in a cofactor 'dyad' maquette where a haem 'acceptor' was able to accept an electron from a light-activated Zn-tetrapyrrole 'donor' (Farid et al., 2013). The asymmetric characteristic of the single-polypeptide BT6 maquette was essential to permit both mono- and bis-His binding sites for selective incorporation of both Zn- and Fe-tetrapyrroles, respectively. Similarly, the asymmetry of the BT6 maquette was used to facilitate the covalent attachment of a single flavin cofactor, with light-driven electron transfer shown between the flavin and haem.

Reminiscent of native light-harvesting antenna systems (Nelson and Junge, 2015), energy transfer was shown to occur in the BT6 maquette between blue-absorbing (Zn-chlorin) and red-absorbing chromophores (Alexafluor 750) (Farid et al., 2013). Building on this process, Ennist et al., (2017a; 2017b) describe the design of a single-polypeptide tetrahelical bundle maquette (MZH3), based on the function of the photosystem II complex of plants, algae and cyanobacteria (Barber, 2006). The cofactor 'triad' MZH3 maquette was engineered for bis-His ligation of haem, mono-His ligation of a Zn-porphyrin and manganese binding centre based on the iron binding site of bacterioferritin (Conlan et al., 2009; Ennist et al., 2017a). Photo-activation of the Zn-porphyrin leads to electron transfer and the reduction of ferric haem to ferrous haem. The di-manganese centre then reduces the Zn-porphyrin via a redox-active tyrosine residue in the maquette core. Although in the nascent stages of its development, the photo-induced reduction of manganese in a *de novo*-designed protein scaffold shows promise in creating an artificial water-splitting enzyme (figure 1.5).

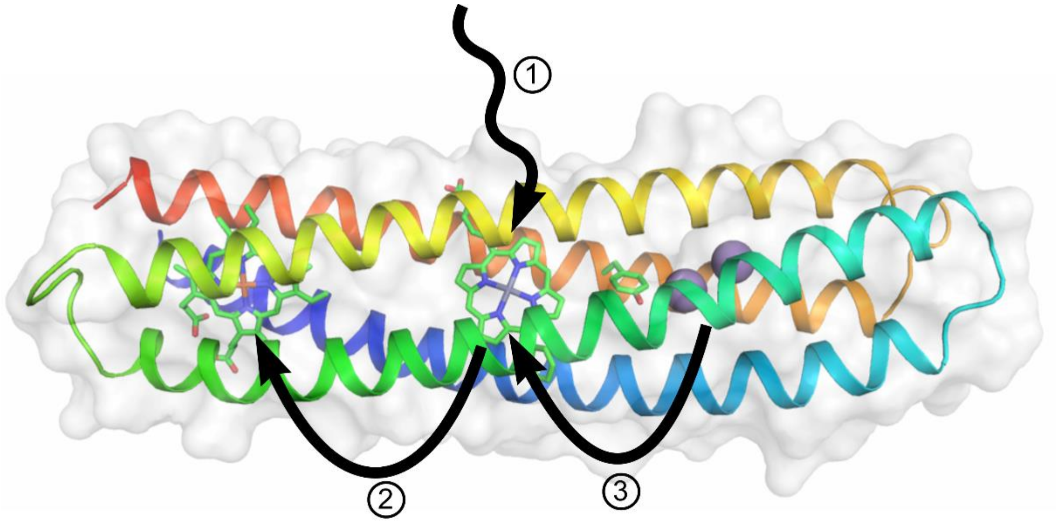


Figure 1.5 Structure and mechanism of the MZH3 maquette. (1) Photo-activation of a synthetic Zn-porphyrin results in the reduction of ferric haem (2). The oxidised porphyrin can then accept an electron from the reducing manganese metal cluster (3). Adapted from Ennist et al., (2017a), with permission.

1.3.6 Maquette-mediated enzymatic reactions

Extending functionality in *de novo* proteins further than simple oxidoreductase reactions has proved challenging and there are few examples of enzymatic reactions catalysed by maquettes (Watkins et al., 2017). Koder et al., (2009) describe the intuitive design of an oxygen transport haem-maquette with function akin to human neuroglobin. The inclusion of charged glutamate residues in the hydrophobic core of the protein was sufficient to cause strain on one of the two haem-ligating histidine residues, allowing competition from other ligands (such as O_2 and CO) for the Fe centre. Interestingly, the maquette showed comparable O_2 affinities to natural globins, but with reduced CO binding. This discrimination between O_2 and CO binding calls into question the design of naturally occurring globin folds, and gives evidence that *de novo* design can provide improved function over evolutionarily generated enzymes.

Further towards this point, it has been shown that thermostable cytochrome *c*-type maquettes can display kinetics that surpass those of naturally occurring peroxidases (Watkins et al., 2017). The inclusion of phenylalanine residues into the hydrophobic protein cavity was sufficient to impart stability onto the maquette scaffold, with a melting transition (T_m) of 84 °C allowing catalytic function to be observed across a range of temperatures (20-70 °C) and solvent concentrations.

1.3.7 Biological interfacing of maquette proteins

One of the future challenges for *de novo* protein design will be to develop complexes that are able to cooperate with and complement natural biological systems. Simple interfacing has been shown through diffusive intermolecular electron transfer between water-soluble haem-maquettes and cytochrome *c* (Fry et al., 2016) and from the insertion of amphiphilic tetrahelical maquettes into detergent membranes (Discher et al., 2005). Energy-transfer has also been shown to occur between the phycocyanin *α* subunit (CpcA) of the cyanobacterial phycobilisome and a cofactor dyad maquette (Mancini et al., 2017). Fusion of the CpcA protein to the C-terminus of a tetrahelical maquette was followed by production and purification of the artificial protein hybrid from *Escherichia coli* (*E. coli*). Phycoerythrobilin or phycocyanobilin was incorporated into the CpcA subunit *in vivo* and Zn-tetrapyrroles were added *in vitro*. Excitation of either bilin pigment resulted in downhill energy-transfer to a maquette-bound Zn-chlorin, followed by a second energy transfer step to a covalently bound bacteriochlorin, which was monitored by fluorescence emission (figure 1.6).

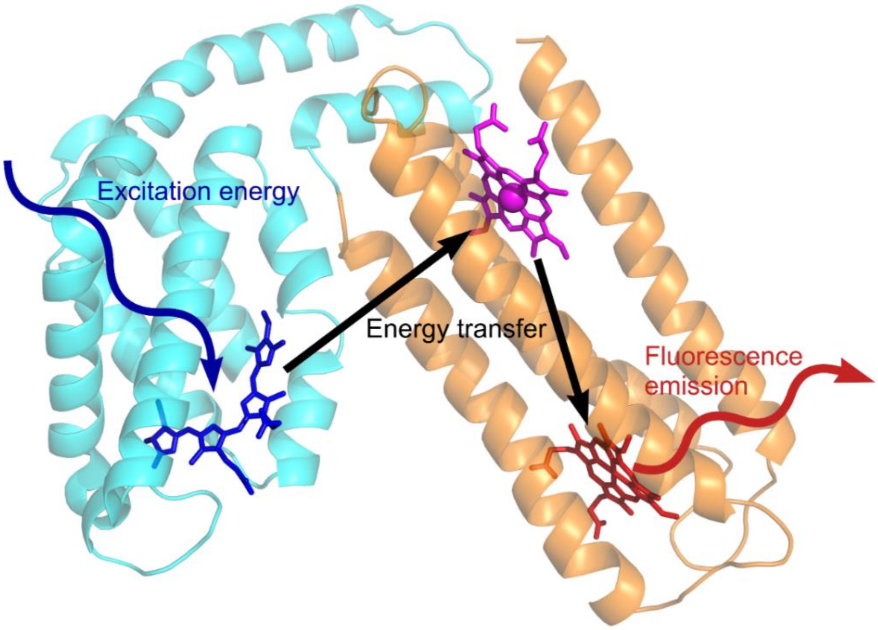


Figure 1.6 Maquettes interfacing with native biological systems. In the fusion of the native CpcA protein (cyan) to a tetrahelical maquette (orange), photo-excitation of a bilin pigment (blue) results in energy transfer to a His-ligated Zn-chlorin and subsequent transfer to a covalently bound emmisive bacteriochlorin (red). The structural model of the CpcA subunit (PDB ID: 4F0T) and tetrahelical maquette is for illustrative purposes and should not be regarded as a structurally determined complex. Based on Mancini et al., (2017).

The first interface between artificial proteins and *in vivo* cellular machinery was shown by Anderson et al., (2014) in the production of a *c*-type cytochrome maquette from *E*. *coli*. As covalent incorporation of haem *c* occurs in the periplasm (Stevens et al., 2004), the maquette scaffold was exported from the cytoplasm, through the general secretory system (see section 1.7), by the inclusion of the N-terminal signal peptide from the maltose binding protein. Efficient export and haem *c* incorporation were achieved through overproduction of Ccm proteins (CcmA-H), with CcmE responsible for the covalent attachment of the macrocycle to the proteins CXXCH motif (see section 1.3.3). The *in vivo* production of a c-type cytochrome maquette necessitates multiple recognition events between the native cellular machinery and the artificial protein structure, thus demonstrating the feasibility of the maquette approach to interface with biological systems.

1.4 The interaction of matter with light

Pivotal to much of the work throughout this thesis is the interaction of chromophores with light. This section will discuss how the structural features of chromophores relate to their photophysical properties and introduce some of the fundamental concepts governing interactions with light at a level that is accessible to readers without extensive knowledge of the area (for more detailed information see Barford, 2005; Musser, 2013; Köhler and Bässler, 2009; Bredas et al., 2004). The scope of the discussion provided here will be limited to the processes necessary to inform the subsequent experimental work in chapters 3-6.

1.4.1 Electronic properties of π-conjugated molecules

Critical to understanding the optical properties of chromophores is the consideration of the extensive π-conjugation in their molecular structure, the origins of which can simply be explained by considering the electronic configuration of ethylene. The configuration of atomic orbitals in the ground-state of elemental carbon is $1s^2\ 2s^2\ 2p^2$. In the formation of covalent bonds, as in methane (CH_4) for example, the *2s* and three *2p* orbitals hybridise to give four equivalent $2sp^3$ hybrid orbitals arranged in a tetrahedral manner around the carbon nuclei. The four valence electrons are then able to form covalent σ-bonds; strong interactions with narrow overlapping electron density directly between each nuclei. As the electrons in σ-bonds are so tightly interacting, they offer little to the electronic and light-absorbing properties of a molecule (Musser, 2013). In the formation of π-bonds, carbon hybridises the *2s* orbital with only two of the *2p* orbitals resulting in three $2sp^2$ hybrid orbitals and one unhybridised *2p* orbital. The three $2sp^2$ hybrid orbitals are arranged in a planar configuration around the central nuclei, with the *2p* orbital oriented perpendicular to the plane of the $2sp^2$ orbitals. The formation of a σ-bond between two $2sp^2$ hybridised carbons, as in ethylene (C_2H_4), results in the formation of a π-bond above and below the plane of the σ-bond due to the overlap of the two *2p* orbitals on each carbon (figure 1.7).

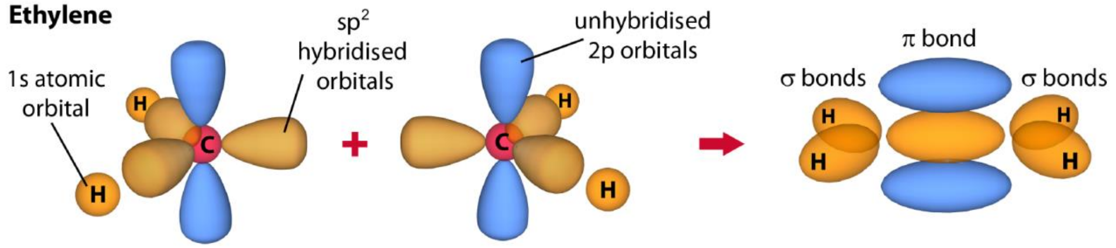


Figure 1.7 π-bond formation in ethylene. Hybridisation of three 2*sp*^{2} orbitals in carbon results in a planar configuration around the central nuclei, with the remaining electron in an unhybridised 2*p* orbital oriented perpendicularly to the plane of the hybridised orbitals. In ethylene, two of the *sp*^{2} hybridised orbitals from each form σ-bonds with hydrogen, with the remaining *sp*^{2} orbitals forming σ-bonds between the carbon nuclei. The remaining unhybridised 2*p* orbitals then form a π-bond above and below the plane of the C-C σ-bond. Reproduced with permission from Musser (2013).

The low density of electrons in π-bonds results in relatively weak association of the electrons to the originating nuclei, which is known as delocalisation. The presence of multiple π-bonded atoms in a molecular structure leads to conjugation of the delocalised electron density; the more extended a conjugated π-system is, the greater the delocalisation. Therefore, in extensively π-conjugated materials, the association of electrons with the originating nuclei becomes so weak that they are able to interact with external energy sources, such as incident light energy (discussed further in section 1.4.2).

In highly π-conjugated molecules the electrons become so weakly bound that they effectively dissociate from individual nuclei and can be described in terms of 'molecular orbitals' rather than atomic orbitals. Molecular orbital theories operate on the simplification that the electrons of high-energy σ-bonds are confined to the interatomic electron density. The position of the σ-bonds can be thought of as providing a track through which the π-delocalised electrons can move, with the originating atoms described as 'nodes'. The movement of π-electrons throughout the delocalised system can be described by a wavefunction, the energy of which is defined by the number of times the wavefunction changes phase. Therefore, the wavefunction of the lowest energy molecular orbital will not change phase throughout the molecule and the wavefunction of the highest energy orbital will change phase between every node.

Benzene provides a simple example to consider molecular orbital theory, having six *2sp*^{2}-hybridised carbon atoms (nodes) with π-electrons delocalised across the entire framework (figure 1.8). The presence of six atoms means that six distinct molecular orbitals must exist, however, as with atomic orbital theory, two electrons can occupy the same molecular orbital. This means that in the ground-state, with low energy, the six electrons occupy the three lowest molecular orbitals, which are known as ‘bonding’ orbitals (π). The three remaining orbitals, which are of higher energy, are known as the ‘anti-bonding’ orbitals (π*).

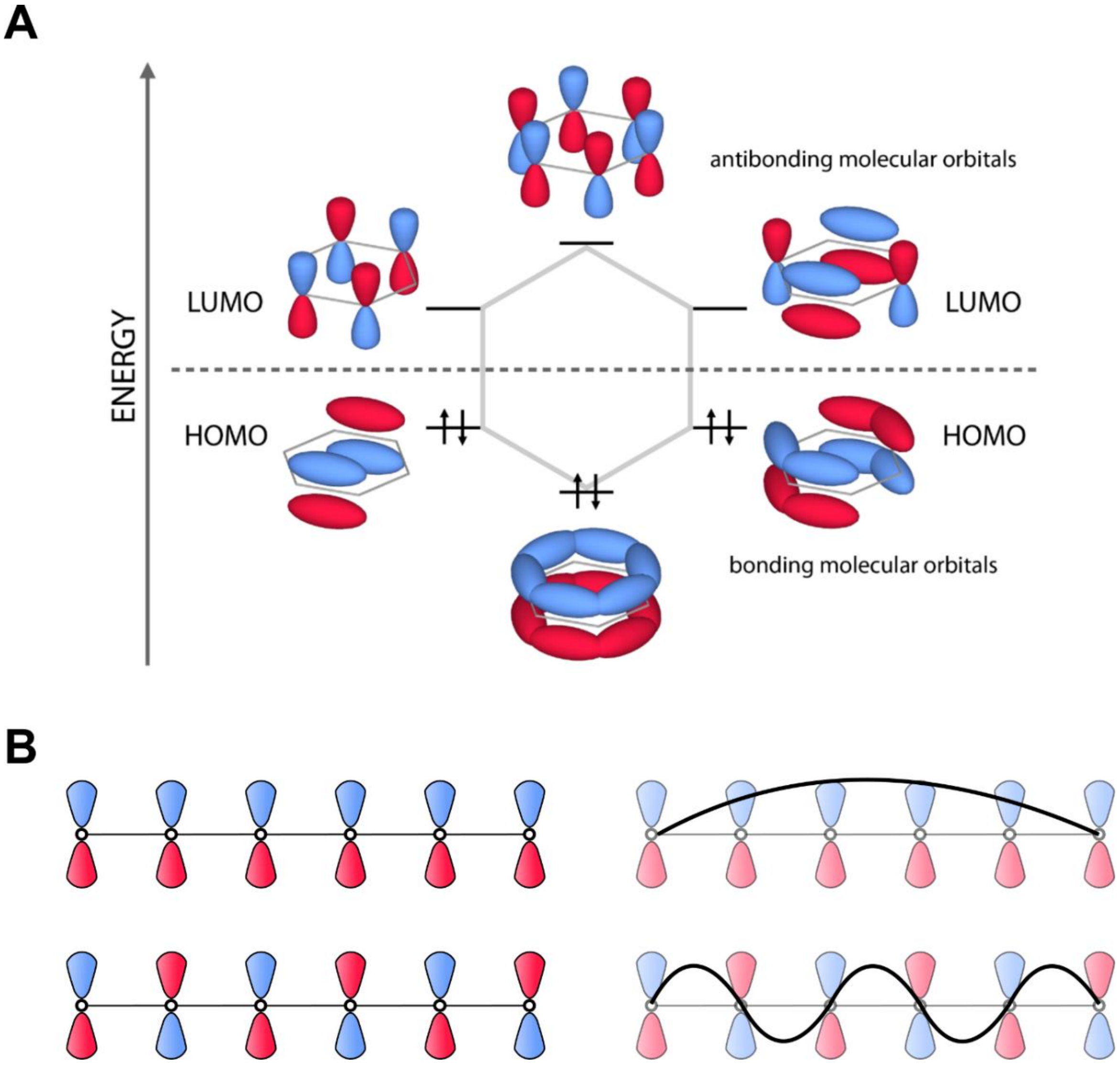


Figure 1.8 Molecular orbital schematic. (A) Benzene contains six independent molecular orbitals; the lower three orbitals correspond to the bonding orbitals and the upper three to the antibonding orbitals. The energy of a molecular orbital is determined by the number of times the phase changes between each node. The chemical structure of benzene is represented by grey solid lines and the positive or negative phase of the wavefunction indicated at each node in blue or red, respectively. **(B)** To more clearly illustrate the relationship between molecular orbital energy and wavefunction, the highest and lowest energy orbitals of benzene are shown in a linear diagram. In the lowest energy state (top) the wavefunction does not change phase across any node, which means that the wavefunction across all six nodes can be described by a half-sine wave. In the highest orbital state the phase of the wavefunction changes between each node and a half-sine wave describes the wavefunction between each node. Adapted from Clayden et al., (2012), with permission.

1.4.2 Absorption of light-energy

The process by which matter is able to absorb light is known as resonance and can be thought to behave in a similar manner to the resonance transfer that may be observed between tuning forks. Light energy travels in discrete 'wave-packets' called photons that have a defined electric and magnetic dipole that oscillates at a particular wavelength (λ) or frequency. The relationship between electrons and their nuclei can also be described by a discrete electromagnetic dipole oscillation, which is dictated by the energy state of the molecule. The oscillation of incident electromagnetic radiation may then resonate with the oscillation of the electrons and nuclei in a molecule, resulting in complete absorption of the photon energy (Price et al., 2001).

Several criteria must be met for energy transfer to take place. The first is that there must be at least one higher energy level available in the molecule for the energy of the photon to be accepted. Secondly, the energetic difference between the initial and final level must be matched by the energy of the incident photon. Finally, as dipolar irradiation will only induce dipolar changes, the transfer of energy must be met with a change in the dipole moment of the acceptor.

Of particular interest in optical spectroscopy of π-conjugated materials are the energy states of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), as transition between HOMO and LUMO typically corresponds to the lowest excitation energy required to generate an excited state in the molecule ($\pi \rightarrow \pi^*$). When a negatively charged electron is excited into the LUMO energy level, it leaves behind a positively charged space, or 'hole', in the originating HOMO. This electronically neutral electron-hole pair can be described as an 'exciton' which behaves as a quasi-particle (Musser, 2013). Important further considerations in $\pi \rightarrow \pi^*$ transitions are the implications of molecular symmetry, the vibronic state of the molecule and the implications of electron spin. As all of these factors contribute to the optical properties of a chromophore they are briefly outlined below.

1.4.3 The role of electron spin in optical transitions

The spin state of electrons can be described by their precession in a magnetic field. The two unpaired electrons in an exciton, arising from a single electron in each of the HOMO and LUMO, behave as two half-spin particles and result in four distinct quantum spin states (figure 1.9). The spins of the two electrons can be aligned in a parallel or anti-parallel manner. Where the alignment is anti-parallel the electrons are 180° out of phase with one another and result in a total angular spin momentum of zero ($S=0$); this is known as a 'singlet' excited state (S_n). There are three distinct parallel spin configurations each resulting in a total angular momentum of one; these three states are known as 'triplets' (T_n) (Köhler and Bässler, 2009). In symmetrical organic systems, the absorption of light may only result in transitions between states of the same spin (i.e. S_0 to S_1).

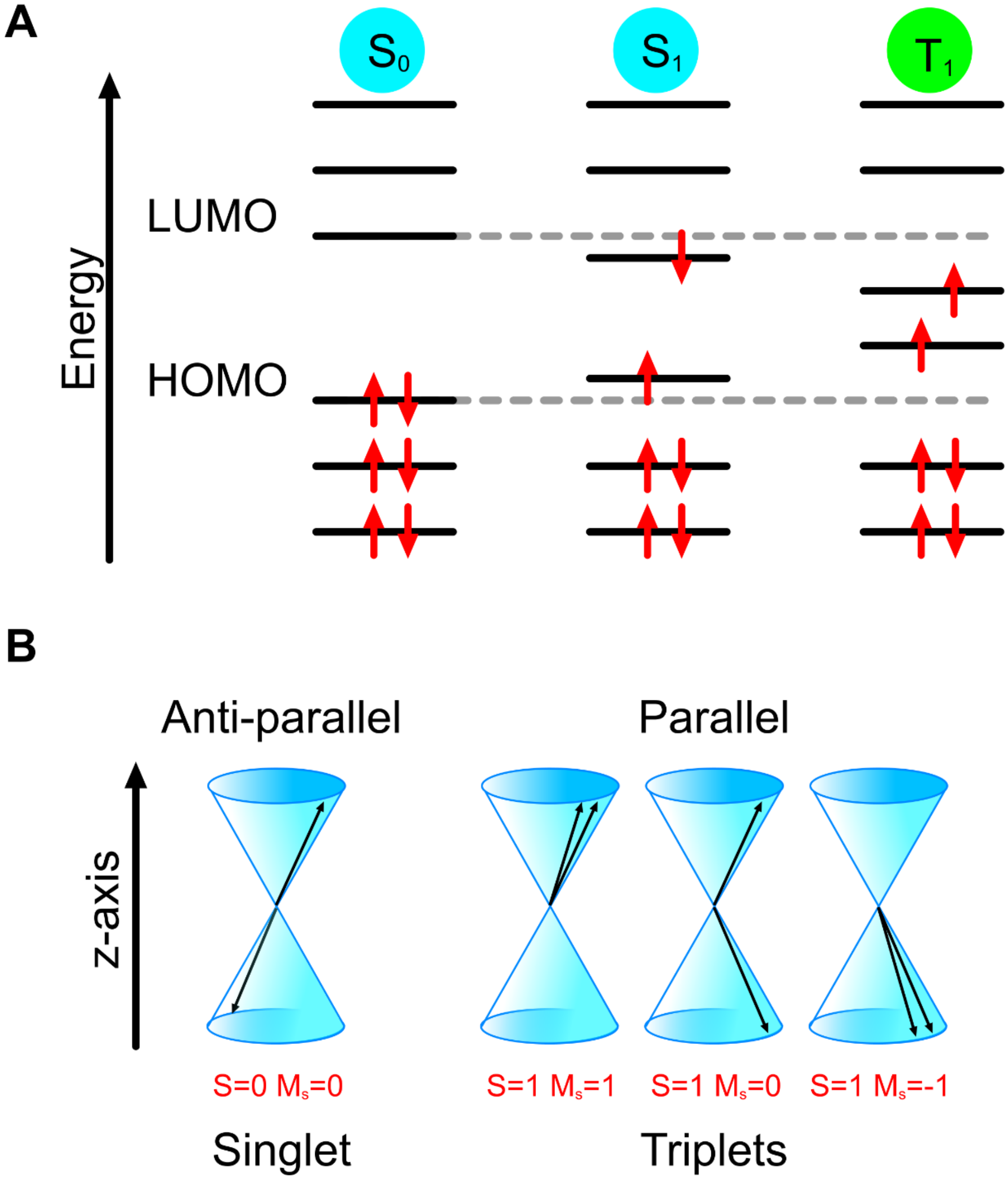


Figure 1.9 Orbital configurations of singlet and triplet energy states. (A) An example Jablonski diagram shows the relative energy and electron spin (red arrows) of singlet and triplet states. All energy levels are arbitrary. **(B)** Vector diagram of electron pair spins. Anti-parallel spin means the two electrons are 180° out of phase and the resulting angular momentum is zero (S=0). Three distinct parallel spins exist which result in a total angular momentum of one (S=1). An additional label is given (M_s) to describe the projection of spin onto the z-axis. Adapted from Köhler and Bässler, 2009.

1.4.4 Molecular orbital symmetry

So far the nomenclature of the ground, singlet and triplet states has been limited to a single letter and subscript integer (e.g. S_0, S_1 and T_1). However, for more detailed analysis this simple naming convention is insufficient to describe the differences in the symmetry characteristics of these excited states. The following discussion therefore requires a more detailed nomenclature that takes into account the rotational and inversional symmetry of the excited states, as well as describing the symmetry of the molecule with respect to the remaining 'hole'. First, an integer is used to describe the energy of symmetrically identical states; for example, the $2A_g^-$ state is identical to the $1A_g^-$ state in symmetry but of higher energy. The capital letter corresponds to symmetry (A) or anti-symmetry (B) with respect to rotation of the molecule by 180°. The subscript letters correspond to gerade (g) and ungerade (u) which describe symmetry or anti-symmetry, respectively, upon inversion of the molecule. Finally, the superscript notation (+ or -) may be used to describe the symmetry of the excited state with respect to the symmetry of the remaining hole (Barford, 2005). Although shown to be of importance in theoretical descriptions of energy transitions, the relevance of this final notation in practical systems remains unclear.

It is essential to describe excited states according to their symmetry because one-photon transitions (both absorption and emission) may only occur between states of opposite symmetry. By definition, molecules in the ground state with a centre of symmetry have a label of $1A_g^-$. This means that absorption from, or emission to, the $1A_g^-$ ground state must occur from a B_u^+ state. The requirement for opposing symmetry in photon transitions has important implications in the lifetimes of each energy level, which not only dictates the photophysical properties of the molecule but also affects its interaction with other materials (discussed in section 1.4.6).

1.4.5 Vibronic coupling and Stokes shift

The distance between two nuclei effects the strength (or energy) of the σ-bond. Under normal conditions the nuclei are not at a set distance but moving relative to one another, vibrating around an average bond length (bond equilibrium). In a classical model of bond vibration the length of the bond equilibrium is set by the Coulombic attraction between the bonding electrons and the positive nuclei. Shifting the bond length away from this equilibrium is associated with an energy cost; moving the nuclei further apart requires energy input against the Coulombic attraction and moving the bond length closer together causes repulsion of the core electrons (Price et al., 2001). The relationship between the coordinates of the nuclei and the energetic cost is depicted in figure 1.10A. In π-conjugated molecules, this classical view of vibrational bond energy is also affected by the energy state of the delocalised electrons. The combination of the vibrational and electronic state of the molecule thus gives the ‘vibronic’ character. It is therefore implicit that upon the transition of an electron to a higher excited state, the overall vibronic bond energy of a molecule changes. Particularly important to note is the change in the equilibrium geometries (ΔQ_k) between the two states.

The Born-Oppenheimer approximation is founded on the fact that nuclei are significantly larger and vibrationally slower than the electrons around them; differences sufficient enough to allow the wavefunctions of nuclei and electrons to be treated independently. As further detail on this approximation is beyond the scope of this work, the reader is directed to Barford (2005) for a detailed explanation. The essence of the Born-Oppenheimer approximation is that the vibrational energy of the nuclei is dictated by the energy of the electron cloud surrounding them and can be thought to behave as a spring. The wavefunction of the nuclear coordinates can thus be described by a quantum anharmonic oscillator and, as such, may only exist in discrete vibrational levels.

The absorption of light energy will typically occur from the lowest vibrational level in the ground state according to the semi-classical Franck-Condon principle (Lax, 1952) (figure 1.10B). This principle states that as the movement of nuclei occurs on much slower timescales than electronic transitions, the nuclear coordinates of the excited state will be the same as that of the ground state equilibrium. This means that the molecule must be excited into a wavefunction that is applicable to the current nuclear coordinates. As the nuclear coordinates of an excited state are often different than that of the ground state, the transition often occurs to a high vibrational level within the excited state. The high vibrational level of the excited state then quickly relaxes to the lowest vibrational state by non-radiative nuclear relaxation. Emission of photons also occurs from the lowest vibrational level in the excited state, resulting in transition back to the ground state in much the same way as with absorption. Similarly, this also results in transition to a higher vibrational level in the ground state following the transition, which then relaxes non-radiatively to the bond-equilibrium (figure 1.10C). On this basis, excitation from the ground state to the lowest vibrational level in the excited state should result in the emission of a photon that is of identical energy of the photon absorbed. However, in many isolated molecules a shift in absorption and emission energies (known as the Stokes shift) is observed for the lowest energy transition, which mainly derives from coupling to low energy vibrational modes through torsional motion and the effects of solvent reorganisation.

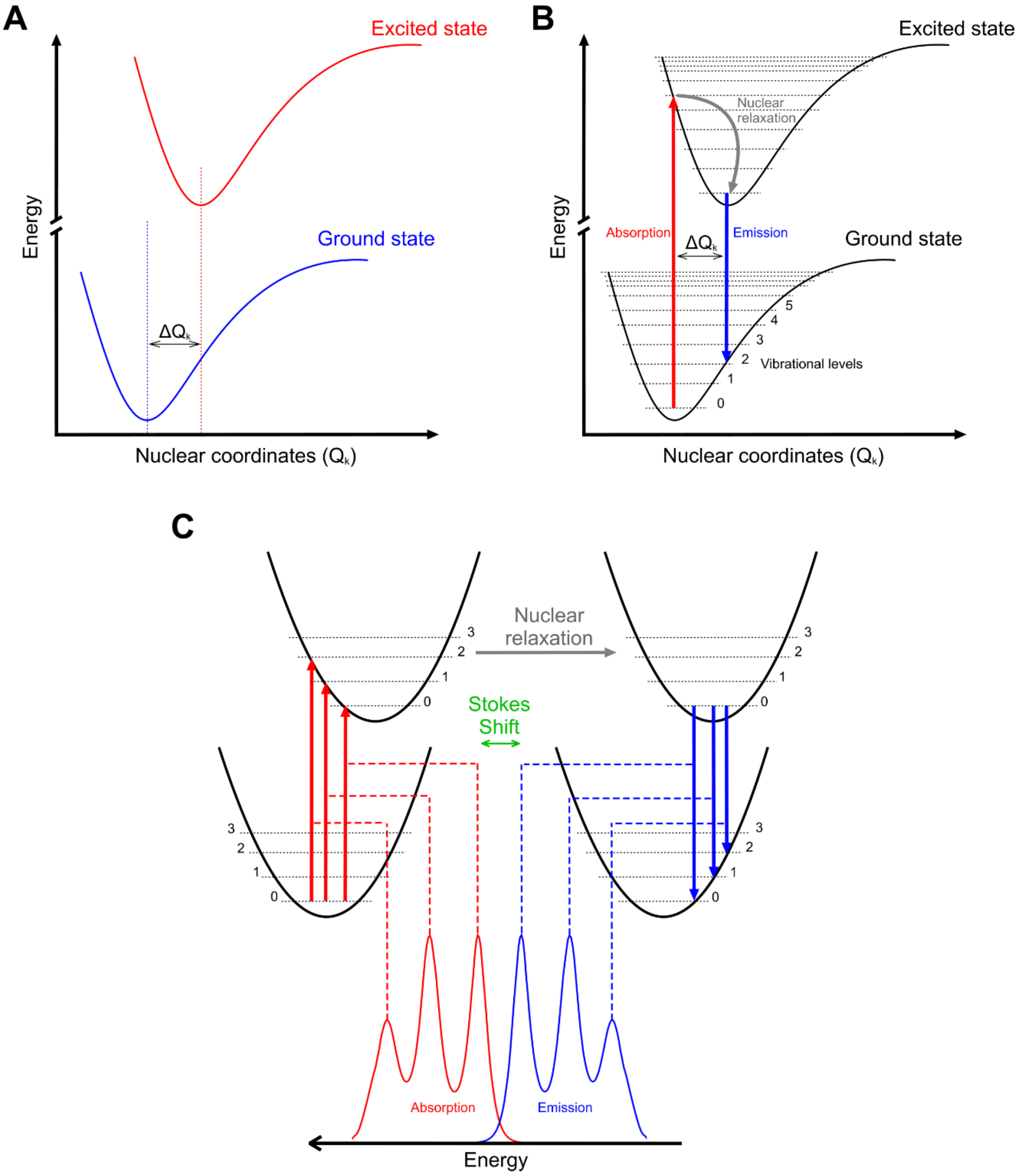


Figure 1.10 The Franck-Condon principle and Stokes shift. (A) The relationship between nuclear coordinates (Q_k) and energy is shown. Energy input is required to make the bond length of the two nuclei either shorter or longer than the equilibrium bond length (dashed vertical lines). The equilibrium bond length for excited states is often different than that of the ground state (ΔQ_k). **(B)** Energy states are made up of discrete vibrational levels (dashed black lines). Absorption occurs from the lowest vibrational level in the ground state to a high vibrational level in the excited state (red arrow). The high vibrational level relaxes to bond equilibrium by non-radiative decay (grey arrow). Emission then occurs from the lowest vibrational level in the excited state to a high vibrational level in the ground state (blue arrow). **(C)** Simplified vibronic structures are shown for clarity. The absorption and emission of photons (red and blue arrows, respectively) to different vibrational levels is shown relative to arbitrary absorption and emission spectra. Adapted from Musser (2013).

1.4.6 Energy transfer from excited states

So far, only the general rules governing the absorption and emission of light have been covered, but many other processes exist to facilitate the transfer of excited state energy. Of most relevance to the biological processes discussed in section 1.3 is the mechanism of intermolecular energy transfer. Intermolecular energy transfer may occur by two separate non-radiative processes; Förster resonance energy transfer (FRET) (Förster, 1965) and Dexter transfer (Dexter, 1953).

FRET occurs in a mechanism similar to that described for the absorption of light (section 1.4.2); namely, a Coulombic resonance transfer where the vibrational oscillation in the dipole of one excited molecule (donor) may be transferred to another nearby molecule (acceptor). The efficiency of this energy transfer is highly dependent on the distance between two molecules, the rate being inversely proportional to the sixth power of the distance (Blankenship, 2014). For energy conservation, the spectrum of donor emission must overlap with the spectrum of acceptor absorption for FRET to take place. It must be noted that although it is useful to use optical spectra characterise the energetics of the transfer reaction, there is no actual photon emission or absorption. As FRET involves a dipole interaction it is subject to the same rules discussed above for the transitions following absorption of light. Therefore, only optically-allowed transitions are permitted, meaning triplet excited states are not able to transfer energy via FRET (Laquai et al., 2009).

Dexter transfer involves the exchange of an electron from donor to acceptor. The efficiency of electron transfer is dependent on direct spatial overlap between donor and acceptor electron orbital wavefunctions. The rate of electron transfer is therefore heavily reliant on the distance between donor and acceptor. However, as this interaction does not involve changes in the dipole moment of donor and acceptor, Dexter transfer may occur between states that are not optically-permitted such as triplet to singlet transfer (Musser, 2013).

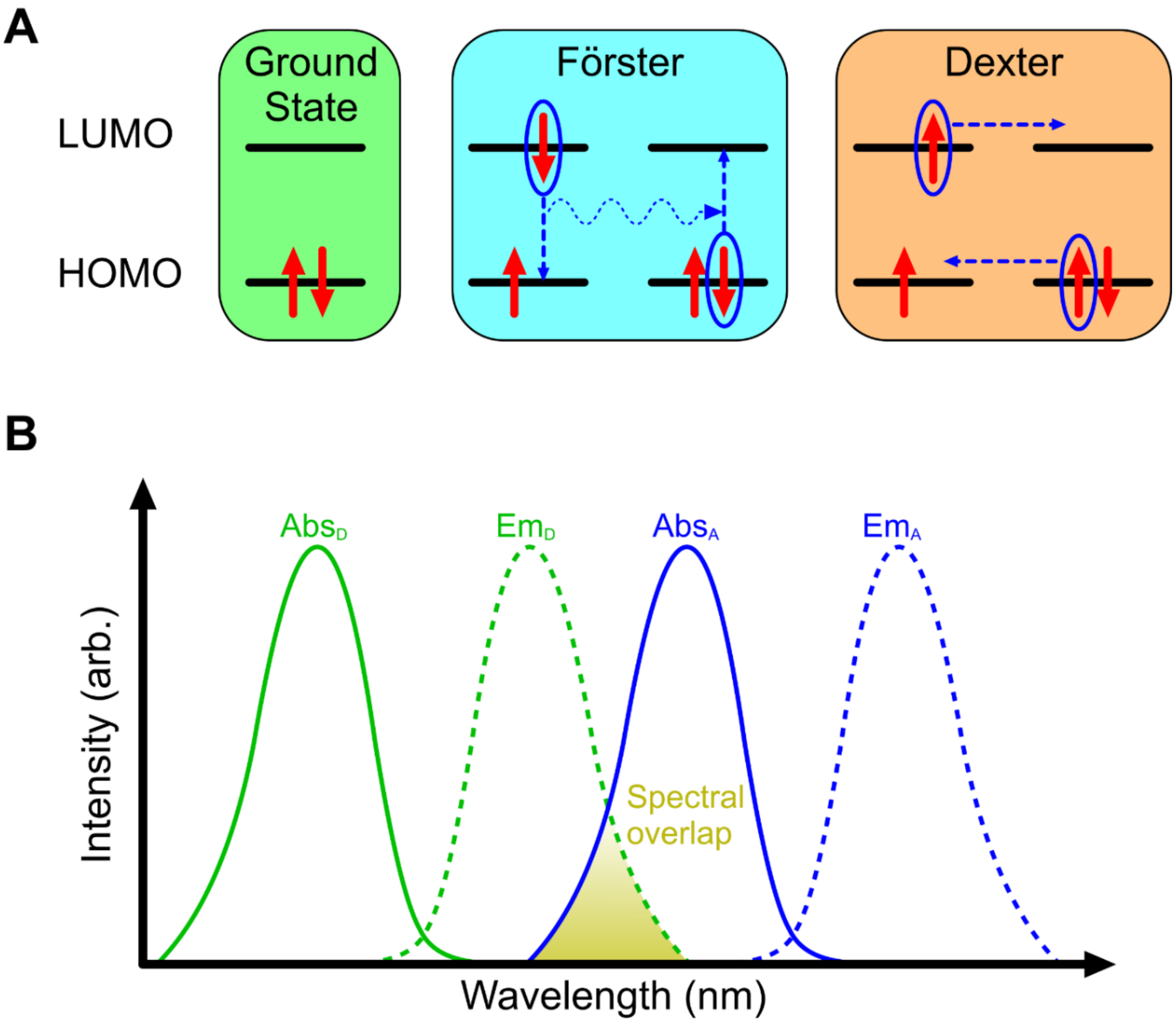


Figure 1.11 Energy transfer processes. (A) Schematic of the orbital transitions in Förster (FRET) and Dexter energy transfer. Following excitation of an electron into the LUMO, energy transfer can occur through resonance energy transfer (FRET) or electron transfer (Dexter). Either mechanism results in an electron returning to the 'hole' in the HOMO of the donor, and excitation of an electron in the acceptor. **(B)** Absorption and emission spectra (solid and dashed, respectively) of donor (green) and acceptor (blue) chromophores. The spectral overlap (shaded yellow area) of the donor emission and the acceptor absorption is a key determinant in the rate of FRET. The relative intensities and spectral shapes of the absorption and emission spectra are arbitrary.

1.4.7 Internal conversion of excited states

The final concept covered here will be the processes by which an excited state may transition without emission of a photon or the transfer of energy. There are three main ways in which this can be done; internal conversion, intersystem crossing and decay through a conical intersection. As the last category has little bearing on any of the work discussed here, the reader is directed to Yarkony (1998).

In the 1950s, the work of Michael Kasha showed that photon emission only appears to occur from the lowest singlet excited state (S_1) (Kasha, 1950). Vavilov's law also states that the energy of photon emission is independent of the excitation wavelength (Williams and Verhoeven, 1994). The conclusion that can be drawn from these results is that excitation to energy levels higher than that of the lowest excited singlet state, must result in rapid, non-radiative transition to the lowest energy state; this process is known as internal conversion. The decay process can initiate where there is an energetic overlap between the equilibrium geometry of an excited state and a high vibrational level of a lower lying state. This 'horizontal' transfer process is driven by factors omitted from the Born-Oppenheimer approximation, and essentially requires the simultaneous alteration of nuclear coordinates and electronic wavefunction (figure 1.12A)(Barford, 2005). Because of the relatively small energetic spacing between the high energy states, overlaps are common and internal conversion happens rapidly to the lowest electronic state. If the energetic spacing between the S_1 and ground state is large, the transition will occur more rapidly through photon emission than internal conversion, if it is dipole-permitted. If the energetic spacing is small or if the emissive decay is not optically-permitted, then internal relaxation becomes the primary decay process (Musser, 2013).

The final energy transfer process detailed here is the process of intersystem crossing, which singularly constitutes the rationale for the experimental work undertaken in chapter 5. As described previously (section 1.4.3), optical transitions can only occur between excited states of the same spin; i.e. S_0 to S_1. Therefore, the formation of triplet states requires non-radiative conversion from singlets in a process known as intersystem crossing. Spin-orbit coupling (SOC) is the most established mechanism for achieving such intersystem crossing and is governed by the interaction between the spin-state of a molecule and its orbital angular momentum (Beljonne et al., 2001). As the SOC interaction is proportional to the fourth power of the nucleus charge, it does not play a significant role in organic hydrocarbon systems, but is greatly affected by the presence of heavy atoms, such as platinum (Beljonne et al., 2001). The SOC interaction perturbs both the singlet and the triplet wavefunction, allowing a slight degree of mixing between the two spin states. Although minor, the reciprocal mixing of wavefunctions allows transition from a singlet state to a triplet state. As with internal conversion, the transition in SOC-mediated intersystem crossing occurs in a 'horizontal' manner, meaning that an energy overlap must occur between the singlet and triplet states (Figure 1.12B). Given the large difference in the S_1 and T_1 energy states the transition is suggested to occur to higher energy triplet states, which subsequently non-radiatively relax to form T_1. Due to the presence of singlet wavefunction in the triplet state, radiative transition from T_1 back to the ground state becomes dipole-permitted, and may occur through a process known as phosphorescence (Musser, 2013).

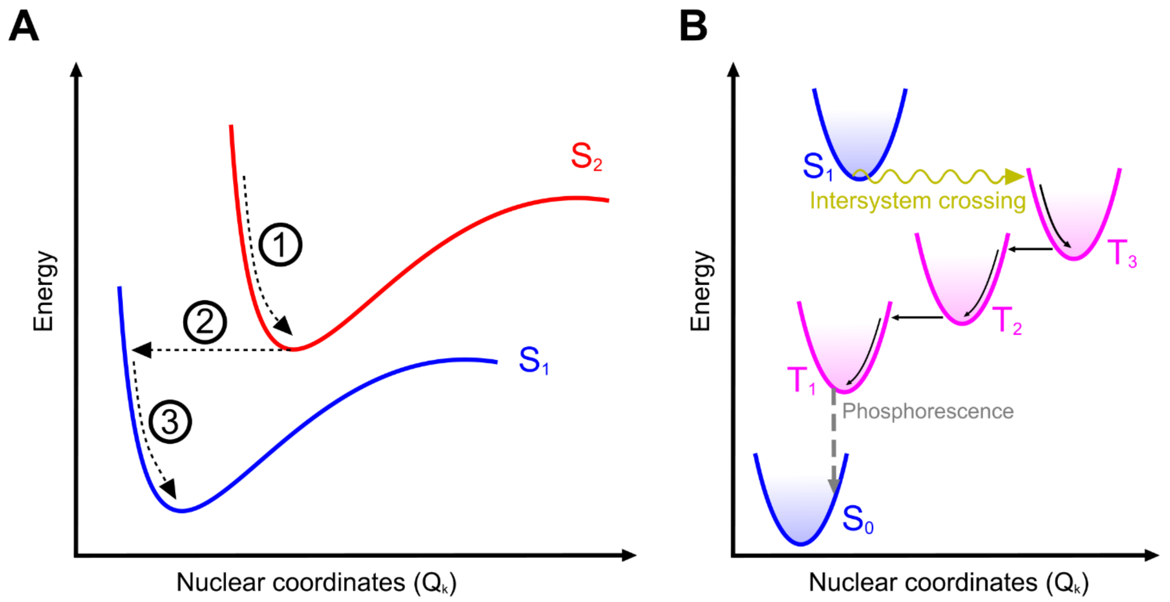


Figure 1.12 Internal conversion and intersystem crossing. (A) Excitation to a high excited state (S_2) is met with vibrational relaxation to the bond equilibrium (1). If there is overlap between the bond equilibrium of the S_2 state and a high-vibrational level in a lower energy state (S_1), then a horizontal transition can occur (2). The following high vibrational level then relaxes to bond equilibrium (3). **(B)** Intersystem crossing (yellow oscillating line) can occur through the spin-orbital coupling (SOC) interaction. The transition typically takes place from the lowest excited singlet state (S_1) to a high triplet state (T_3). Rapid internal conversion occurs in the triplet to reach the lowest energy triplet state (T_1) through a system analogous to that described in panel A. The SOC interaction also permits T_1 to transition radiatively back to the ground state (S_0), resulting in phosphorescence (grey dashed line). In both panels the nuclear coordinates are arbitrary and have been positioned for clarity.

1.4.8 Intermolecular interactions in aggregated chromophores

The optical properties and electronic transitions of chromophores can dramatically change upon aggregation. The effect of aggregation on the photophysical properties of a chromophore is more easily described by considering the electronic properties of an idealised dimeric molecule, in which the polarisation of the transition dipole runs along the long molecular axis. The alignment of dimers can occur in three ways: parallel, head-to-tail and oblique (Kasha, 1963). The scope of this discussion will only consider parallel and head-to-tail alignments, for details on the optical characteristics of oblique aggregates see Kasha (1963).

The coupling between the two aggregated chromophores affects the interaction of the molecules with light. In Kasha's example, the strong electronic coupling of the chromophores means that vibronic coupling may be ignored. In this limiting case a high and low energy state can be calculated for each alignment, with only one of the two states being optically accessible (figure 1.13). In the case of parallel alignment, the only dipole permitted transition is of greater energy than that of the same excited state in the monomer; this aggregation state is therefore associated with a characteristic blue-shift in the absorption spectra compared to the monomer. The reverse is true for head-to-tail alignments, where only a lower lying energy state is optically accessible, resulting in a characteristic red shift.

For the molecules discussed in the following experimental sections, the vibrionic coupling is of similar strength to the electronic coupling and a more complex description is required (see Spano and Silva, 2014 for review). While further discussion on the mechanistic basis for the optical transitions in such aggregates is beyond the scope of this thesis, a brief summary would conclude that the aggregate arrangement effects the proportion of transitions permitted to the lowest and highest vibrational levels in the excited state. Parallel aggregation states lead to weaker transition to a low vibrational level and head-to-tail arrangements show a stronger transition to a low vibrational level. Note that, in line with conventional spectroscopic nomenclature, parallel and head-to-tail alignments are referred to as hypsochromic (H-) or Jelley (J-) aggregates, respectively (Spano and Silva, 2014).

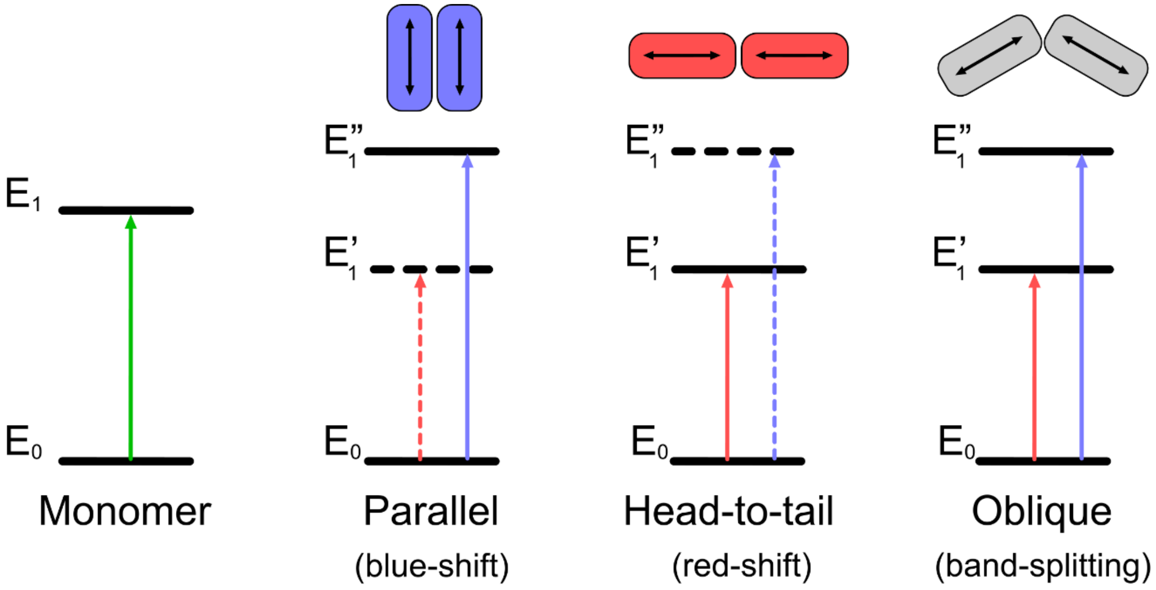


Figure 1.13 Molecular dimer arrangements. The relative energy levels are shown for dimers in parallel, head-to-tail and oblique arrangements. Dashed lines correspond to dipole-forbidden energy states and transitions. Adapted from Kasha (1963).

1.5 The photophysical properties of the molecules used in this study

Following on from the fundamental concepts previously introduced, this section will cover how such rules can explain the photophysical properties of the specific classes of molecules used in the experimental chapters. First discussed are the organic, naturally-occurring tetrapyrrole and carotenoid classes, which inform chapters 6 and 3, respectively. A brief introduction to artificial poly-acene and benzothiophene derivatives follows, to further understand the work conducted in chapters 4 and 5.

1.5.1 The tetrapyrroles

Tetrapyrrole molecules are found ubiquitously across all domains of life with roles in wide-ranging biological processes, such as photosynthesis, methanogenesis, respiration, metabolism, signalling, electron transport and transcription (extensively reviewed in Scheer, 1991). With the exception of bilins, tetrapyrroles are found as cyclic, often metallated, molecules. Cyclic tetrapyrroles are typically classified according to the extent of the π-conjugation in the macrocycle, forming subcategories known as porphyrins, chlorins or bacteriochlorins (figure 1.14A).

The electrochemical properties of tetrapyrroles are affected by the chemical structure and planarity of the macrocycle, the identity and oxidation state of the coordinated metal centre, and axial ligation of the central ion by ligands such as carbon monoxide and imidazole (Kadish and Caemelbecke, 2003). The spectroscopic properties of all metalloporphyrins are similar, with their optical absorption dominated by two types of bands; the Q-bands ($\lambda > 500$ nm) and the Soret bands ($\lambda \sim 400$ nm). The characteristic spectra of the cyclic tetrapyrroles can largely be explained by the so-called "four orbital" model, first proposed by Martin Gouterman (Gouterman, 1961). The Soret bands (also known as the B-bands) correspond to the highest energy absorption in the visible range; as orbital symmetry labels are not appropriate for all the molecules discussed here, this transition will simply be referred to as $S_0 \rightarrow S_2$. The absorption of the Q-bands corresponds to a lower energy transition ($S_0 \rightarrow S_1$).

Critical to the optical properties of cyclic tetrapyrroles is the role of symmetry. As extensive discussion on this subject is beyond the scope of this work, the reader is directed to Gouterman (1978) for review. To simply summarise this work, the breakdown of symmetry results in the splitting of the Q-band, attributable to the polarization of light along the molecular x- or y-axis (figure 1.14A). Symmetry can be broken in porphyrins by the removal of the metal centre and more dramatically in the case of chlorins and bacteriochlorins by the reduction of π-bonds in the macrocycle. For the highly symmetrical metallated porphyrin molecule, the Soret energy transition is strongly allowed and the weaker Q-band transition only quasi-allowed, resulting in the relatively intense absorption of the Soret band compared to the Q-band (Gouterman, 1978). For the less symmetrical chlorin and bacteriochlorin molecules, the weaker $S_0 \rightarrow S_1$ transition becomes more permitted and the absorption bands more intense (Figure 1.14B).

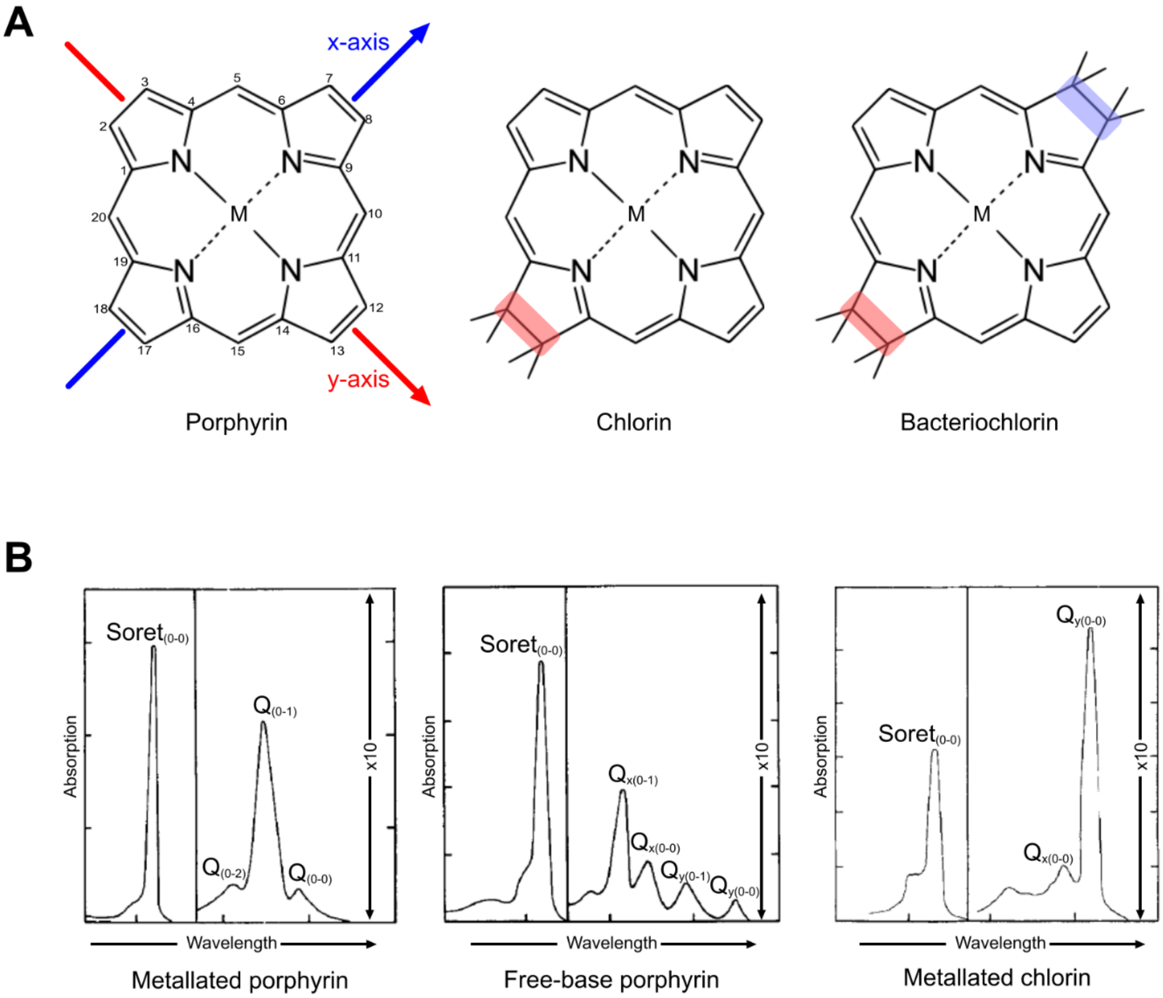


Figure 1.14 Chemical structures of cyclic tetrapyrroles and absorption spectra. (A) The IUPAC numbering scheme for tetrapyrrole macrocycle carbons is shown. A desaturation at C17 characterises the asymmetric chlorin from the symmetric porphyrin, with a further desaturation at C7 designating bacteriochlorins. **(B)** The absorption spectra of metallated porphyrin, metal-free (free-base) porphyrin and metallated chlorin are shown with the corresponding absorption bands indicated. The symmetrical metallated porphyrin shows strong Soret band absorption and weaker Q-band absorption. Removal of the metal centre breaks the symmetry and the Q-band absorption splits according to the polarization along the x- and y-axes (Q_x and Q_y, respectively). In the less symmetrical chlorin, Q-band absorption is more permitted, resulting in greater absorption intensity. Additional subscript labels are given (e.g. $Q_{(0-0)}$, $Q_{(0-1)}$) to denote transitions to different vibrational levels in the same excited state. Absorption spectra are arbitrary and adapted from Gouterman (1961). Note that the absorption spectra in (B) do not correspond with the chemical structures shown above in (A).

1.5.2 The carotenoids

The structurally diverse carotenoids can be found in many organisms and fulfil roles in important biological processes, such as light-harvesting, nutrition, photoprotection and colouration (Britton et al., 1995). The carotenoids are comprised of over 600 different naturally occurring molecules which can be traced back to the same C_{40} isoprenoid backbone (Weedon and Moss, 1995). The impressive structural diversity of the carotenoids is demonstated by the examples in figure 1. 15, and can be attributed to substitution, elimination, addition and cyclisation modifications to the precursor C_{40} backbone skeleton. The conventional numbering system of carbon atoms in carotenoids is annotated in figure 1.15; for further information on the nomenclature of carotenoid structures see Weedon and Moss (1995).

The optical properties of carotenoids derive from the extent of the conjugation in the polyene backbone and additional chemical groups that may be associated. Therefore, light absorption alone can be used as a powerful tool to give information on the identity of a carotenoid in solution and also provide an indication of modification or decomposition (Schiedt and Liaaen-Jensen, 1995). The highly conjugated structure of carotenoids, however, renders them susceptible to oxidation, electrophilic attack and light- or heat-induced isomerisation.

Over the past half century there has been heated debate surrounding the exact nature of the electronic transitions in carotenoids (summarised in Polívka and Sundström, 2009). What is accepted is that due to the symmetry of the π-conjugated polyene backbone in carotenoid molecules, the strongly allowed one-photon transition must occur from the $1A_g^-$ ground state to the $1B_u^+$ excited state. It was also shown, originally in non-carotenoid polyenes, that an additional excited state ($2A_g^-$) existed that was of lower energy than the $1B_u^+$ state (Schulten and Karplus, 1972). As the $2A_g^-$ state has the same symmetry as the ground state, direct transition ($1A_g^- \rightarrow 2A_g^-$) is not dipole-permitted. It is worth noting at this point that inconsistencies arise in the nomenclature of these excited states in the literature and in the following experimental chapters; the ground state may be referred to as $1A_g^-$ or S_0, the lowest energy singlet energy state called $2A_g^-$ or S_1 and the higher-energy absorbing state known as $1B_u^+$ or S_2.

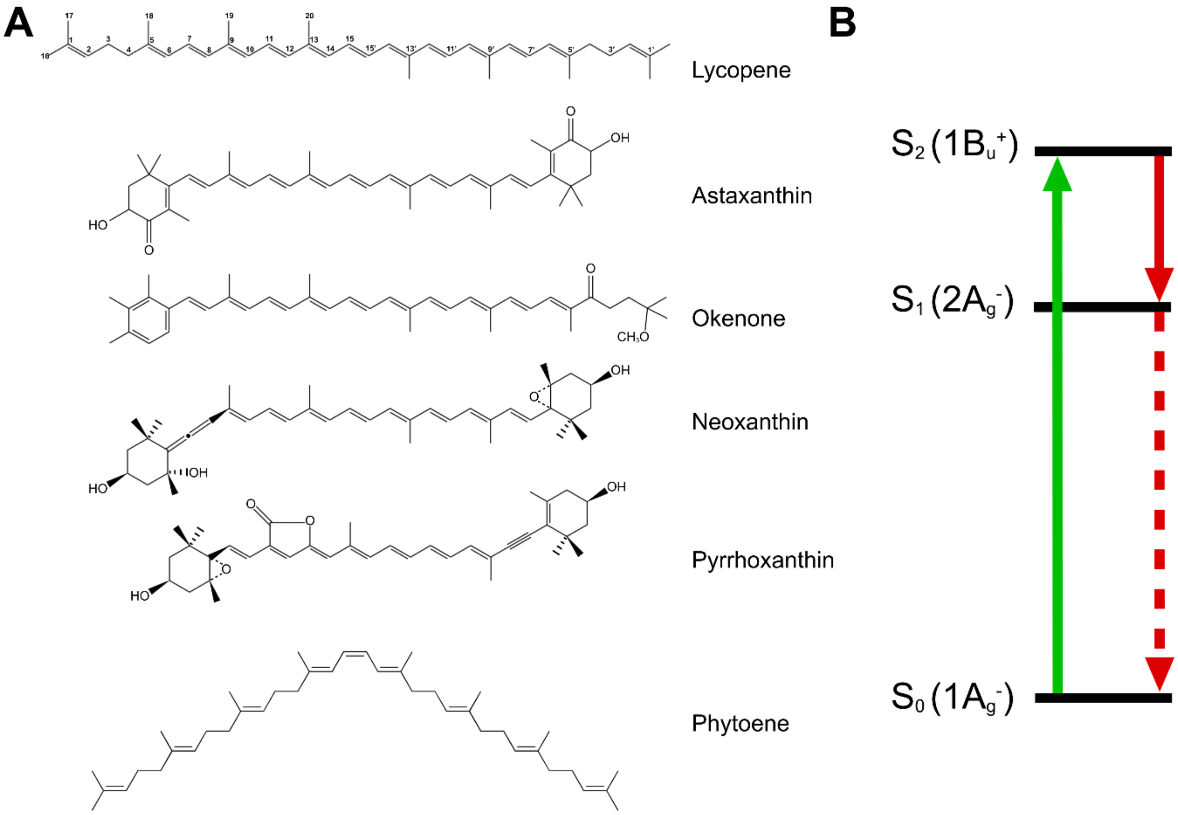


Figure 1.15 Diverse chemical structures of carotenoids and simplified energy diagram. (A) The C_{40} carotenoid lycopene can be thought of as the forerunner to which all other carotenoids are generated. Modifications to the polyene backbone results in the generation of chemically and functionally distinct molecules; such as cyclisation in astaxanthin, the addition of keto and methoxy groups in okenone. Neoxanthin is an example of a C_{40} epoxyxanthophyll and pyrrhoxanthin demonstrates the incorporation of acetoxy, butenolide and acetylenic groups into a carotenoid structure. Phytoene (otherwise known as 15-*cis*-phytoene) is an example of a C_{40} carotenoid in a *cis* conformation. **(B)** In monomeric carotenoids a one-photon transition must occur from the $1A_g^-$ ground state to the $1B_u^+$ excited state as the $2A_g^-$ state has the same symmetry as the ground state. Formation of the $1B_u^+$ state results in rapid (~100 fs) internal conversion into the $2A_g^-$ state. As emission from the $2A_g^-$ state is forbidden, the transition back to the ground state occurs by slow, non-radiative decay (~10 ps).

In monomeric carotenoids excitation from the ground state ($1A_g^-$) to the $1B_u^+$ state results in rapid (~100 fs) internal conversion into the $2A_g^-$ state. As emission from the $2A_g^-$ state is forbidden, the transition back to the ground state occurs by slower, non-radiative decay (~10 ps). Although emission from the $1B_u^+$ state is possible, the rapid internal conversion to the $2A_g^-$ state occurs in hundreds of femtoseconds (Macpherson and Gillbro, 1998), outcompeting other decay mechanisms and effectively rendering carotenoids non-emissive (Gillbro and Cogdell, 1989). The mechanism for the conversion between $1B_u^+$ and $2A_g^-$ is still contentious, with the transitions suggested to occur via so called ‘dark-states’; the most widely discussed being the $1B_u^-$ or S* states (Cerullo et al., 2002; Balevičius et al., 2016). Details of the nature and plausibility of these states is beyond the scope of this work and the reader is directed to the following publications: Polívka and Sundström (2009), Niedzwiedzki et al., (2016) and Niedzwiedzki et al.,(2017).

The extensive conjugation present in carotenoids results in poor solubility in polar solvents and practical insolubility in water. The addition of water to monomeric carotenoid in solvent results in the formation of aggregates, largely driven by hydrogen bonding interactions (Billsten et al., 2005; Musser et al., 2015). The aggregation of carotenoids can occur in a head-to-head or parallel configuration, as discussed in section 1.4.8, leading to H- or J-aggregates with concomitant absorption shifts often visible by eye. A dramatic example of this can be observed in nature by looking at the carotenoids in lobster (*Homarus americanus*). The native pigmentation is caused by the presence of astaxanthin molecules in two types of carotenoprotein in the lobster shell; the blue crustacyanins and the so called 'yellow protein'. The orientation of the two astaxanthin molecules, in either H- or J-aggregate conformation, is responsible for the difference in colour between the crustacyanin and the yellow protein. Denaturation of these proteins releases the carotenoid resulting in the characteristic red colour of non-aggregated astaxanthin (Salares et al., 1977; Christensson et al., 2013).

Similar spectral shifts in astaxanthin have been replicated *in vitro* by generating aggregates in a protein-free environment (Musser et al., 2015) (figure 1.16). Although carotenoids display pronounced solvatochroism (Reichardt and Welton, 2010), the absorption characteristics of the astaxanthin molecules in figure 1.16 are firmly assigned to pigment aggregation by Musser et al., (2015) on the basis of their excited state dynamics. The absorption maximum of monomeric astaxanthin in DMSO is ~15 nm red-shift compared to the absorption of astaxanthin in acetone; however, both solvent environments produce identical excited state dynamics. As the electronic transitions of the five astaxanthin species shown in figure 1.16 are dramatically different than those of the monomer (section 3.2), they are firmly assigned to aggregation effects rather than solvatochroism.

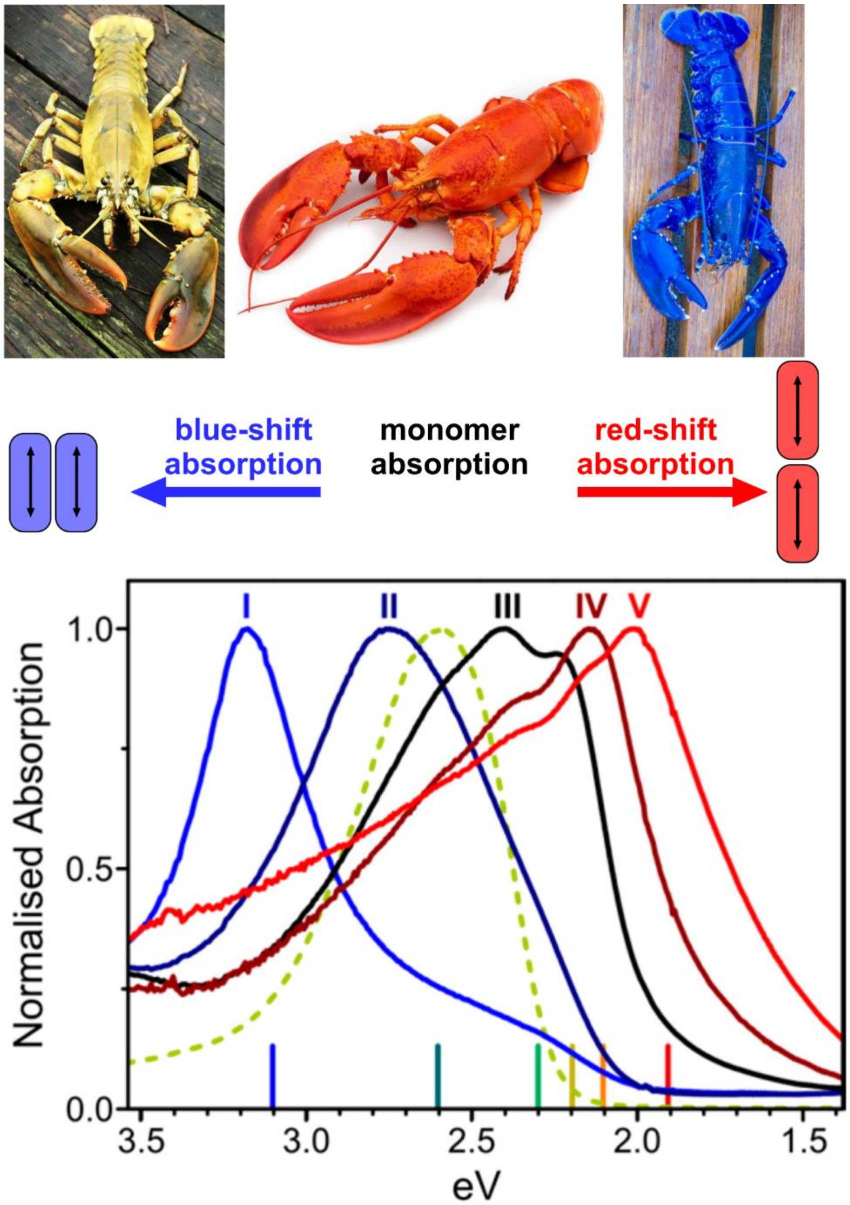


Figure 1.16 Astaxanthin aggregates. Images of rare yellow and blue mutant lobsters are displayed in the top panels with red, denatured lobster shown for comparison. The astaxanthin aggregation states responsible for the pigmentation can be replicated without protein *in vitro* as shown in the absorption spectra of five different aggregates (solid lines). Parallel H-aggregation results in blue-shifted absorption compared to monomeric astaxanthin (yellow, dashed) and head-to-head J-aggregation results in red-shifted absorption. Absorption spectra reproduced from Musser et al., (2015) with permission.

1.5.3 Artificial chromophores

The artificial chromophores used in this study include the polyacene tetracene and the benzothiophene derivative 2,7-dioctyl[1]benzothieno[3,2-b][1]benzothiophene (C8-BTBT); the structures of both these molecules are shown in figure 1.17 (for information on the optical characteristics of these molecules, the reader is directed to chapter 5). When crystalline or in the form of films both materials can be described as organic semi-conductors due to their high charge carrier mobility. There has been great interest in the development of organic semi-conductors for a variety of purposes; the polyacenes have attracted attention as a charge multiplication system (Rao et al., 2010; Smith and Michl, 2010; Zimmerman et al., 2011) through their interaction with light (section 1.6), while C8-BTBT has drawn interest for its applications to flat panel displays and biomedical sensors (Ebata et al., 2007; Yuan et al., 2014).

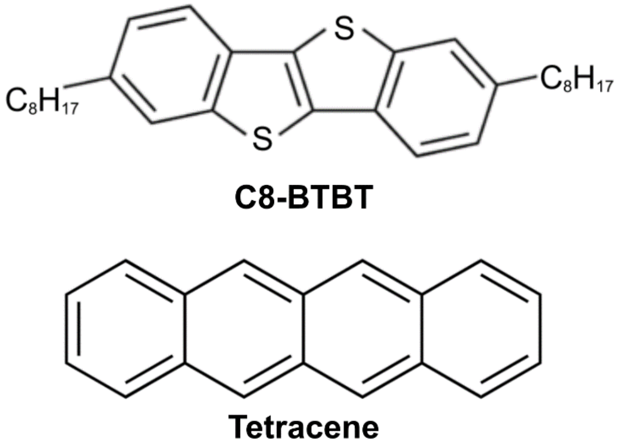


Figure 1.17 The chemical structures of C8-BTBT and tetracene.

The practical use of polyacenes is often stymied by their susceptibility to photooxidative damage and poor solubility; both of which become increasingly problematic with larger compound size. To address the issues of stability and solubility polyacenes are often chemically modified to incorporate triisopropylsilyl (TIPS) side groups (Anthony, 2008). As the photophysical properties of tetracene and its derivatives provide the best established mechanism for singlet fission, the excited state dynamics of these molecules will be discussed in the following section.

Despite the similarities in their structure, C8-BTBT has been reported to have high solubility in organic solvents (80 mg mL^{-1} in chloroform) and high thermal stability (Ebata et al., 2007; Yuan et al., 2014). However, as interest in this molecule is only nascent, little detail is available in the literature for its optical properties.

1.6 Singlet fission

The semi-conducting properties of the materials described in chapters 3 and 4 gives promise for their application in many novel biotechnological roles. One process that could conceivably be realised is the use of such materials in the study of singlet fission (SF) and the potential application of these SF sensitizers in the field of photovoltaics. Here, the general principles of singlet fission and its applications are briefly introduced.

1.6.1 Basic principles of singlet fission

The process of SF was first suggested in the 1960s to explain observations in crystals of anthracene (Singh et al., 1965) and has seen revived interest for its potential applications in the field of photovoltaics (Hanna and Nozik, 2006). SF is as a spin-permitted quantum mechanical process in which two independent triplet excited states are generated from one singlet excited state (Smith and Michl, 2010). As the two triplets are of opposing spin, the overall spin state is zero, and thus the transition is permitted from the singlet state (see section 1.4.3). SF may rapidly occur when the energy of the singlet excited state is more than double that of the lowest-lying triplet state (i.e. $S_n > 2xT_1$). Only rapid triplet formation, known as exothermic fission, will be considered here; for information on endothermic fission see Rao and Friend (2017).

The exact mechanism in which two triplet states are generated remains unclear (see Rao and Friend, 2017 and Smith and Michl, 2013 for review). What is generally accepted is that following excitation to the singlet state, the transition first proceeds to an intermediate state, which is believed to be an entangled triplet pair, and from this dark state the two independent triplet states are formed (Figure 1.18). In the classical view of SF, the formation of the two triplet excitons is thought to occur separately in neighbouring chromophores; however, recent evidence suggests the generation of two triplet states may occur in the same chromophore, though such observations have only been made in protein-pigment complexes (Rademaker et al., 1980; Gradinaru et al., 2001; Trinh et al., 2015; Yu et al., 2017).

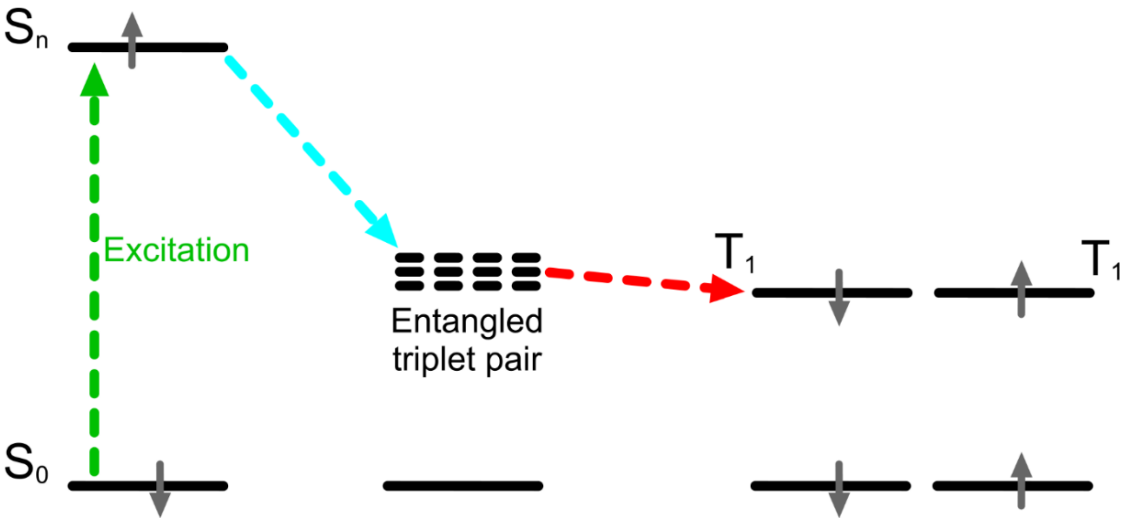


Figure 1.18 Jabolinski diagram of singlet fission. Excitation to a high singlet state (S_n) is coupled to the formation of an intermediate state thought to be an entangled triplet pair (TT). From this dark state, two triplet states are formed, each with opposing spin (grey arrows). The energy levels shown here are arbitrary.

1.6.2 Applications of SF in photovoltaics

Photovoltaic (PV) devices capable of converting solar energy into electrical energy are showing great promise as systems for renewable energy generation (Lewis and Nocera, 2007; Lewis, 2016). However, universal implementation of solar cells is only likely to occur when PV devices become an economically viable option for energy generation. Reducing the cost of production and installation of PV systems, and/or increasing the power conversion efficiency of PV modules will enable greater energy capture per unit cost (Rao and Friend, 2017). While the power conversion efficiency of PV modules is approaching that of the best laboratory devices, the fundamental energy loss mechanisms in conventional single-junction cells limits the potential for improvement (Rao and Friend, 2017). Regardless of whether an incident photon has energy that is equivalent or in excess of the bandgap of the absorber (E_g) the net result is a single electron-hole pair, with excess energy lost through thermalisation. The process of thermalisation is the largest contributor to the overall Shockley-Quiesser power-conversion efficiency limit, which for an idealised silicon cell is approximately 30% (Shockley and Quiesser, 1961).

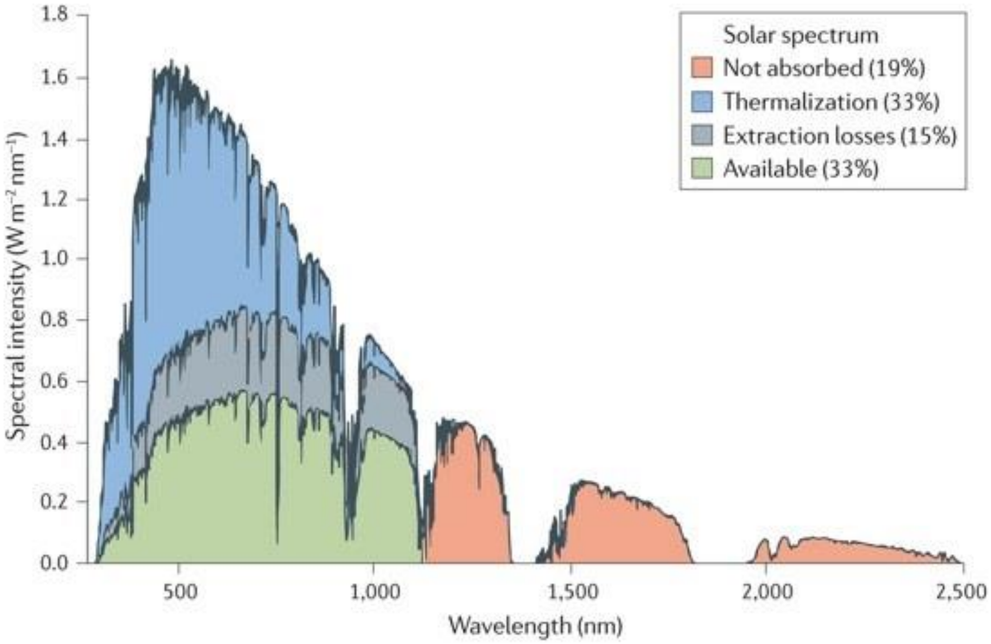


Figure 1.19 Energy loss mechanisms in a conventional silicon solar cell. The energetic output from the sun that reaches the Earth's surface is displayed. The bandgap of a conventional silicon solar cell is 1.1 eV (wavelength ~1100 nm), meaning incident photons of lower energy cannot be absorbed (orange region). Whether an incident photon has energy equivalent to or greater than 1.1 eV the net result is a single electron-hole pair in the silicon surface. The extra energy provided by high energy photons cannot be absorbed by the surface and is lost as heat in a process known as thermalisation (blue region). Thermalisation comprises the dominant loss mechanism in conventional silicon cells. Reproduced with permission from Rao and Friend (2017).

As described above SF involves the generation of two triplet states ($2T_1$) from one singlet excited state (S_1), and due to laws of energy conservation, the energy of the triplet states must be half that of the singlet excited state or less. Therefore, the absorption of one high energy photon to generate the S_1 state results in the formation of two lower energy states. The application of SF materials to conventional photovoltaic cells therefore proposes to harvest high energy photons and efficiently convert them to two lower energy states that can be more efficiently absorbed by the surface (Smith and Michl, 2013; Walker et al., 2013).

As outlined in a recent review (Rao and Friend, 2017) an ideal SF sensitizing material must satisfy a number of criteria. (1) Strong absorption coefficients in the visible and near UV regions. (2) Close energetic matching between the lowest triplet state and the bandgap of the surface (e.g. 1.1 eV for silicon). (3) Rapid singlet fission (< 10 ps) to outcompete other decay processes. (4) Long triplet lifetimes (> 1 ns) to maximise harvesting by the surface. (5) Sufficient stability for practical applications. As of the date of publication of this thesis, no such material has been reported that satisfies these criteria.

1.6.3 Singlet fission in tetracene

The acenes and their chemical derivatives constitute the most well-established SF materials; however, the exact mechanism in which SF proceeds still remains unclear. As with aggregates of carotenoids, the optical properties of tetracene in molecular crystals change dramatically through intermolecular interactions. Unlike the head-to-head and parallel configurations discussed so far, acenes appear to adopt a face-to-edge ‘herringbone’ structure. The intermolecular coupling is sufficient to permit the formation of triplets from the first, second and third excited states in tetracene (Zimmerman et al., 2011; Burdett and Bardeen, 2013).

Tetracene and its derivatives are an attractive prospect in generating a SF sensitizer as they have been shown to have long-lived triplet states (> 10 ns) that show close energetic matching to silicon, satisfying criteria 4 and 2 from section 1.6.2. However, the SF process in tetracene is relatively slow (~90 ps) and dependent on the morphology of the crystal or film generated (Wilson et al., 2013; Piland and Bardeen, 2015). Absorption in the visible region is also relatively weak and the material is highly sensitive to photooxidation.

1.6.4 Singlet fission in carotenoids

SF has been observed in natural carotenoproteins (Gradinaru et al., 2001; Yu et al., 2017) and more recently in carotenoid aggregates (Wang and Tauber, 2010; Musser et al., 2015). In astaxanthin, SF was shown to occur in both H-and J-aggregate types, proceeding directly from the $1B_u^+$ state (Musser et al., 2015). The exact mechanism for triplet formation from the $1B_u^+$ state is poorly understood, as are the potential implications of the $2A_g^-$ and $1B_u^-$ states.

Carotenoids provide an interesting alternative to the more established acenes, as they have ten-fold stronger absorption in the visible region and show ultrafast (< 100 fs), morphology-independent SF (Musser et al., 2015). Close energetic matching to silicon should be achievable by selecting an aggregate from the approximately 600 naturally occurring carotenoids. The triplet states in carotenoids, however, have been shown to decay rapidly with 90 % loss in 1 ns and photooxidative stability is also a major issue.

1.7. The twin-arginine translocase

Essential for the function of all living cells is the formation and regulation of a phospholipid bilayer that protects cytosolic components from the exterior environment. One of the key challenges faced by the cell is the translocation of specific proteins and small molecules across this membrane, through a mechanism that must also prevent the unwanted import or export of materials. In bacteria, proteins are translocated across the cytoplasmic membrane by either the general secretion (Sec) pathway or the twin-arginine translocation (Tat) pathway.

The Sec-pathway comprises the major route for protein export across the bacterial cytoplasmic membrane, using ATP hydrolysis and proton motive force (PMF) to translocate unfolded proteins into the periplasm or directly into the membrane. The structure and function of the various components of the Sec machinery have been determined and the mechanistic basis of its translocase activity is well established (reviewed in de Keyzer et al., 2003).

The Tat export system, on the other hand, selectively exports fully folded protein substrates and is driven solely by the PMF. Although atomic-resolution structures of the Tat translocase subunits have been determined, the basis for its activity remains contentious; particularly the mechanism by which it apparently senses the folded state of substrates. As the Tat apparatus and export mechanism have important ramifications for the experimental conclusions in chapter 6, these features will be briefly described below.

1.7.1 Occurrence and importance of the Tat system

The first suggestion of an additional bacterial export system, distinct from the Sec pathway, came from observations of the periplasmic hydrogenase of *Desulfovibrio vulgaris*. A putative, novel export signal peptide was suggested to attach to one of the two protein subunits, which was apparently cleaved following export (Prickril et al., 1986). It was subsequently shown that the presence of both subunits was required for export and removal of either subunit completely abolished translocation (van Dongen et al., 1988). Shortly after it was realised that many cofactor-containing periplasmic proteins, in a variety bacteria, contained similar signal peptides with a ubiquitously conserved twin-arginine motif (Berks, 1996). It was suggested that this novel bacterial export system was probably related to the ΔpH-dependent translocase that had recently been established in plant thylakoid membranes (Cline et al., 1992; Chaddock et al., 1995), with homology later shown between both systems (Settles et al., 1997).

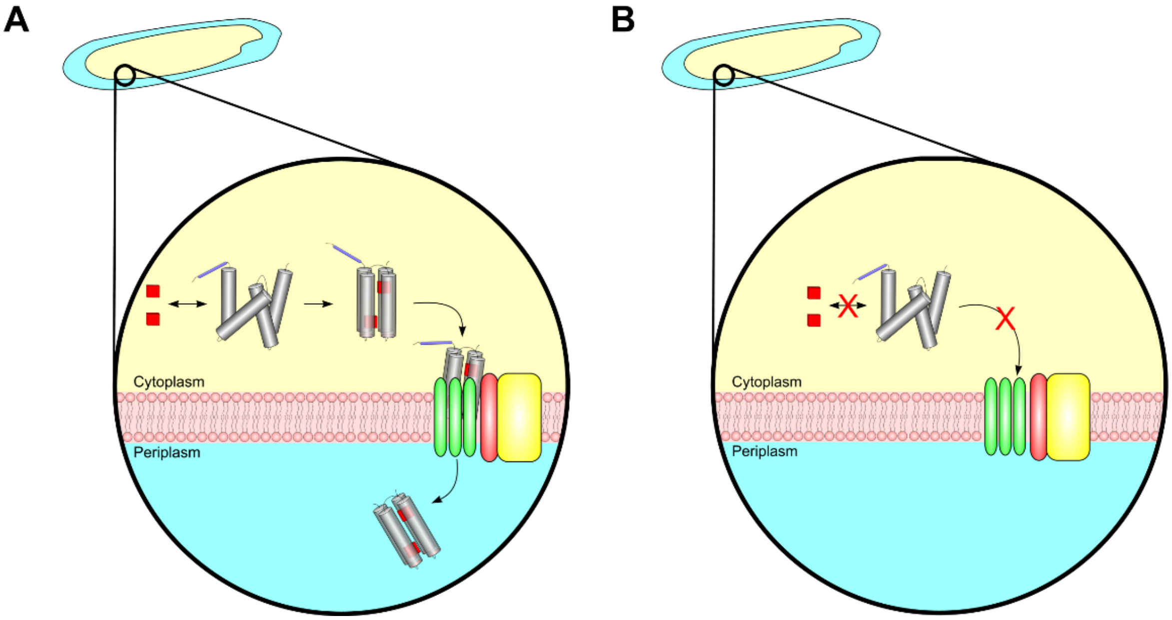


Figure 1.20 Schematic of Tat export system. The Tat system is responsible for transporting folded proteins from the cytoplasm (pale yellow) to the periplasm (light blue) of bacterial cells. Proteins (grey) are identified as Tat substrates by an N-terminal signal peptide (blue). The coordinated activity of the TatABC complex (shown in green, red and yellow, respectively) recognises the signal-peptide and folded state of the substrate, selectively exporting folded proteins. **(A)** An example of a Tat-substrate folding through incorporation of its cognate cofactor and being deemed suitable for export. **(B)** When the Tat substrate is unable to fold export is abolished.

Bacterial proteins translocated by the Tat pathway were originally thought to be cofactor-containing proteins, typically involved in energy metabolism (De Buck et al., 2008). However, it became apparent that substrates for the Tat system extended to some cofactor-free proteins (Stanley et al., 2000), proteins that fold too quickly for transport via the Sec system (Palmer and Berks, 2012) and oligomeric complexes that require assembly in the cytoplasm (Rodrigue et al., 1999).

More recently, the significance of the Tat pathway in bacterial pathogenesis is being realised (reviewed in De Buck et al., 2008). The Tat system is found in approximately 77 % of sequenced bacterial genomes (Simone et al., 2013) and is suggested to have a direct role in the virulence of many pathogens, including *E. coli*, *Neisseria meningitidis*, *Mycobacterium tuberculosis, Helicobacter pylori*, *Listeria monocytogenes*, *Salmonella enterica*, *Vibrio cholerae* and *Staphylococcus aureus*. The Tat system has been shown to play a role in the secretion of toxins (Pradel et al., 2003), the export of cell motility machinery (Ochsner et al., 2002; Pradel et al., 2003), biofilm formation (De Buck et al., 2005) and infection (Ochsner et al., 2002). As well as playing potentially a significant role in bacterial pathenogenesis, the Tat export pathway is used in a variety of important cellular processes, including photosynthesis, respiration, iron and phosphate nutrition, cell separation and nitrogen fixation (reviewed in Palmer and Berks, 2012; Berks, 2015).

1.7.2 The Tat-signal peptide

Further experimentation is required to understand the role of the Tat export system in biological processes from a wider variety of organisms. The classification of Tat substrates, which may contribute towards a particular biological process, can be done by the identification of the ubiquitously conserved Tat signal peptide sequence; which for bacteria follows the general Ser-Arg-Arg-Xaa-Phe-Leu-Lys motif (Stanley et al., 2000; Berks, 2015). Towards this end, sequence analysis tools have been generated to facilitate the identification of putative Tat substrates (Bendtsen et al., 2005).

Tat signal peptides are typically 30 residues or greater in length and are recognised at the cytoplasmic membrane, in bacteria, by the TatBC complex (discussed in section 1.7.4.) prior to export. The signal peptide is made up of three distinct regions: the polar n-region which contains the consensus twin-arginine motif, a hydrophobic h-region, and a second polar c-region containing the conserved peptidase cleavage site (figure 1.21).

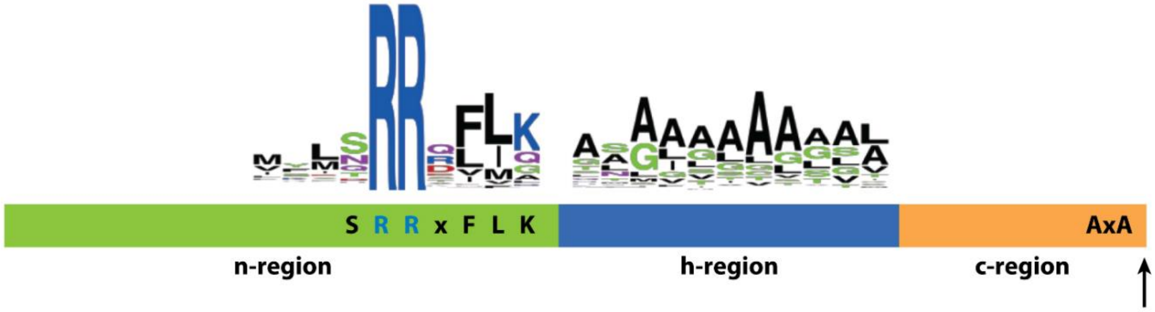


Figure 1.21 Schematic diagram of the tripartite Tat signal peptide. The relative positions of the n-region (green), h-region (blue) and c-region (orange) are shown. Consensus motifs that are present in greater than 50 % of bacterial Tat substrates are shown in lettering within the coloured regions. The relative size of the letters shown above the schematic corresponds to the frequency of the residue amongst sequenced bacterial genes that have been annotated as Tat substrates. The black arrow marks the peptidase cleavage site. Figure reproduced with permission from Berks (2015).

1.7.3 Subunits of the bacterial Tat system

The subunits comprising the Tat machinery are made up from just two structural families; the TatA family, denoted by a single transmembrane helix, and the TatC family, containing 6 transmembrane regions (Figure 1.22) (Berks, 2015). In *E*. *coli*, the Tat complex contains the 29 kDa TatC protein, the 9.7 kDa TatA protein and two other members of the TatA family known as TatB and TatE (18 kDa and 7.0 kDa, respectively). The TatA and TatE proteins are thought to be functionally interchangeable, with TatE expressed at lower concentrations in the cell (Sargent et al., 1998; Jack et al., 2001). For succinctness, only the structure and role of TatA will be considered here.

The atomic structures of the *E*. *coli* TatA and TatB proteins have been determined by solution NMR (figure 1.8) (Rodriguez et al., 2013; Zhang et al., 2014). The core structures of the TatA family proteins (TatA, TatB and TatE) are similar with an N-terminal transmembrane section arranged perpendicularly to an amphipathic helix oriented along the membrane surface, with the organisation of the helices defined by a short hydrophobic loop region between each section. Identifying universally conserved structural features to categorise TatA family proteins has proven difficult and Tat components are instead classified by their biological activity (Berks, 2015). TatB proteins are found in equimolar concentrations to TatC, whereas TatA proteins oligomerise in response to substrate recognition and have increased quantity relative to TatC (Jack et al., 2001).

The structure of the bacterial TatC protein from *Aquifex aeolicus* was determined by X-ray crystallography (Rollauer et al., 2012). The six transmembrane helices are arranged in a manner resembling a cupped hand forming a cavity known as the ‘palm’. The structure also showed that the polar side chain of a glutamate residue was oriented in the hydrophobic membrane bilayer, which is universally conserved as Glu or Gln and suggested to have role in signal peptide recognition (Ramasamy et al., 2013).

Native polyacrylamide gel electrophoresis revealed that the *E*. *coli* TatBC complex has an apparent molecular mass of over 400 kDa; suggesting the TatBC components are organised as multimeric complexes in the membrane (Orriss et al., 2007). The current understanding is that TatC components are able to self-assemble through specific interprotein interactions and form a complex to which TatB is then able to tightly associate (Oriss et al., 2007; Berks, 2015). The site of interaction between TatC components is unclear; no conserved contact points have been identified and the association is thought to be weak as detergent treatment results in fragmentation of TatBC subunits (Cline and Mori, 2001). The interaction site between TatB and TatC is better defined. As shown by cross-linking studies (reviewed in Berks, 2015), the transmembrane section of TatB has been shown to bind between helix 5 and 6 of TatC. Nascent crosslinking experiments have also suggested TatA binds between the docking site of TatB and the central cavity containing the conserved Glu/Gln residue.

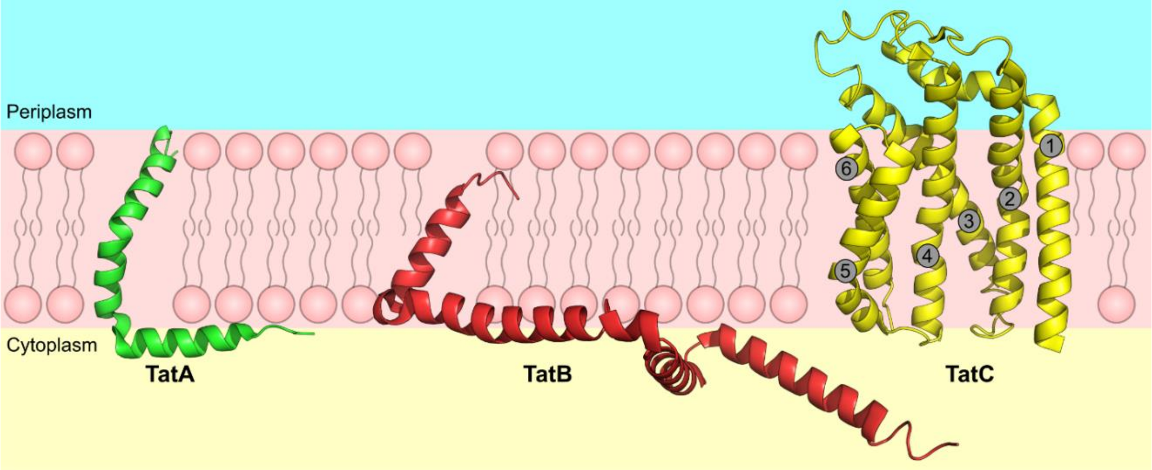


Figure 1.22 Atomic structures of the TatA (green), TatB (red) and TatC (yellow) relative to their position in the cytoplasmic membrane. The solution NMR structures of TatA (Rodriguez et al., 2013; PDB ID: 2LZS) and TatB (Zhang et al., 2014; PDB ID: 2MI2) revealed a single transmembrane region with an amphipathic helix arranged perpendicularly along the membrane surface. The crystal structure of the bacterial TatC protein (Rollauer et al., 2012; PDB ID: 4B4A) showed 6 transmembrane helices (numbered) arranged like a cupped hand.

1.7.4 Signal peptide recognition

Although the availability of atomic-level structures provides a model to predict how the structure of each component may relate its function, the exact basis of substrate recognition remains incompletely understood (reviewed in Berks et al., 2014 and Berks, 2015). Tat substrates are recognised at the cytoplasmic membrane by the TatBC complex with sub-micromolar affinities for their cognate N-terminal signal-sequence peptide (Whitaker et al., 2012). In pea (*Pisum sativum*) thylakoids, signal peptide recognition is suggested to occur in two binding modes; a weak peripheral interaction that leaves the signal peptide accessible to proteolysis and a strong interaction where the signal-peptide becomes buried within the TatBC complex (Gérard and Cline, 2007). Similar results have been observed in the *E*. *coli* Tat system where identical substrates have been found to have both weak (Bageshwar et al., 2009) and strong interactions (Tarry et al., 2009). The presence of the PMF across the membrane is thought to drive the conversion from a weak interaction to the strong deep-binding mode (Berks, 2015).

A recently proposed model of this interaction suggests that two conserved glutamate side chains on the cytoplasmic surface of TatC, between helix 2 and 3, directly interact with the twin-arginine residues of the signal peptide (Rollauer et al., 2012; Berks et al., 2014). Another model proposes that the n-region may bind further into the internal cavity of the TatC protein with the conserved Glu/Gln residue in the internal cavity of TatC suggested to bind a lysine residue at position 7 in the signal peptide of *E*. *coli* Tat substrates (Ramasamy et al., 2013). Although possible, this model remains contentious as many bacterial and plant substrates do not contain lysine at position 7 (Berks, 2015). The two binding models proposed here could perhaps account for the weak and strong binding modes, however, further experimentation is required to fully unravel the mechanism of signal peptide recognition.

Recent work has revealed that the TatC protein, in the absence of TatB and TatA, is able to translocate signal peptides across the membrane bilayer (Fröbel et al., 2012). The sole presence of TatC was sufficient to remove the signal sequence peptide from translocation-incompetent substrates. It can be concluded that the TatC protein provides a translocase activity as well as a substrate recognition role and the presence of TatB is required to identify suitable Tat substrates.

1.7.5 Mechanism of export

Following the substrate recognition event, the kinetics of membrane translocation can be described by a two-step mechanism; the first stage being the PMF dependent recruitment and oligomerisation of TatA, and the second stage being the coordination of export by the TatABC complex (Berks, 2015). Although known to be entirely dependent on the PMF (Alcock et al., 2013), the mechanism of TatA recruitment has not been fully elucidated. The oligomerisation state of TatA complexes is highly variable with pore sizes ranging between 30 and 70 Å (Gohlke et al., 2005). Interestingly, the range of pore sizes matches the range of diameters of known Tat substrates, leading to the conclusion that perhaps TatA monomers form a multimer of a specific size for each substrate.

Perhaps the most pressing challenge in understanding the Tat export system is to elucidate the mechanism in which folded proteins can be transported across the membrane, with the process unlikely to occur through a conventional system. Conventional export, as with the Sec system, proceeds through an aqueous pore that remains closed in the absence of substrate. When the pore opens for translocation, conformational changes result in constriction in the middle of the pore, preventing the leakage of small molecules (Park and Rapoport, 2012). Such a mode of transport is hard to envisage for the Tat system as creating a seal for the much larger and variable pore size would be mechanistically difficult to achieve (Berks, 2015).

It has been demonstrated that both oligomerisation of TatA and the presence of PMF are required for Tat-dependent export (Alcock et al., 2013), however, how these separate components combine to enable substrate translocation remains unclear. The so-called membrane-weakening model suggests that the recruitment of many TatA molecules to a specific area of the membrane reduces the integrity of the membrane in that region. The compromised membrane patch is then more susceptible to environmental stresses and the substrate, perhaps driven by the PMF, is then able to force its way through the weakened region. Another proposal is that the TatA multimer transduces the energy of the PMF into stored conformational energy which is then used following the arrival of substrate (Berks, 2015). The latter proposal, often referred to as the 'trap-door' mechanism, recognises that the entirely hydrophobic transmembrane helices of TatA oligomers are unlikely to be able to form an aqueous channel. Instead, the suggestion is that a conformational change could occur in TatA, in which the amphipathic helix moves position from along the membrane surface to fully integrate into the membrane in parallel with the other helical domain. This hairpin structure would allow the hydrophobic face of the amphiphilic helix to interact with the entirely hydrophobic N-terminal transmembrane region and allow the hydrophobic regions of other conformationally altered TatA monomers to form a polar pore through which proteins can move (Figure 1.23) (Walther et al., 2013). Currently, neither mechanism is supported by a wealth of evidence and questions surrounding the plausibility of either system remain (see Berks, 2015 for review).

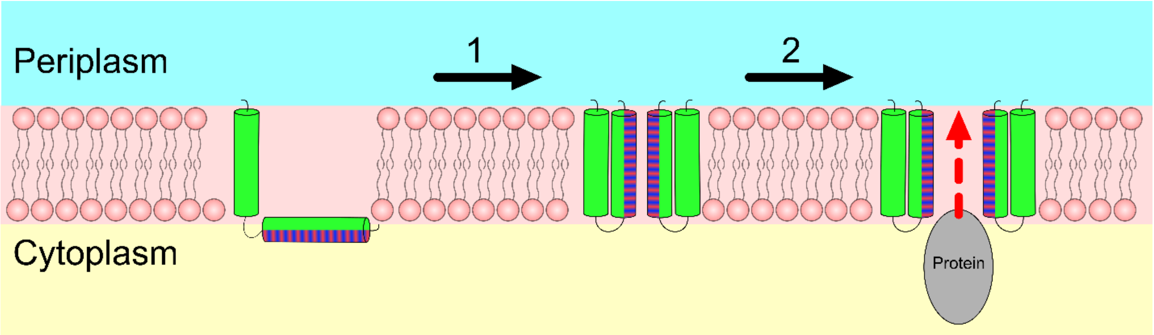


Figure 1.23 Schematic of the TatA trap-door export mechanism. (1) The energy of the PMF is used to drive the conformational change and oligomerisation of TatA (green). The amphipathic helical domain (one face green, one face red and blue) moves from an alignment along the membranes surface to a position parallel to the N-terminal hydrophobic transmembrane helix. The hydrophobic face of the amphiphilic helix aligns with the other hydrophobic helix, and the hydrophilic regions of multiple TatA monomers can oligomerise to form a channel. **(2)** The channel can then form an appropriate size to facilitate the export of Tat substrates.

1.7.6 The proof-reading mechanism of the Tat complex

Perhaps one of the most interesting questions surrounding Tat-mediated export is the apparent ability of the complex to determine the folded state of proteins and selectively export structured substrates (DeLisa et al., 2003). It has recently been realised that the basis of this selectivity cannot be due to an inherent inability of the complex to export unstructured proteins. Rocco et al., (2012) showed that minimal mutations (single or double amino acid substitutions) to Tat components were sufficient to permit the export of misfolded proteins whilst maintaining export of native substrates. It was also demonstrated that small unstructured peptides could be processed by the Tat system when components of the export complexes were produced at above-native levels. These studies show that there is no mechanistic reason for why the Tat system cannot export unfolded proteins and the basis for selectivity must derive from a type of proofreading mechanism that is inherent to the Tat machinery itself. The means by which the folded state of substrates is determined is unclear and could occur through the detection of hydrophobic patches on the proteins surface (Richter et al., 2007) or through sensing the conformational flexibility of the substrate.

1.7.7 Translocation of non-native substrates

Much work is needed to further understand the mechanistic basis of the Tat system, which is important not only for the biological roles discussed in section 1.7.1, but also in adapting the complex to perform roles in biotechnology. The use of recombinant biological products, such as antibody fragments and insulin, has seen a steady rise in pharmaceutical markets with an estimated sales revenue of approximately $140 billion (Walsh, 2014). The production of recombinant molecules must be met with a product release step where the molecule of interest is removed from the cell material. The process by which the molecule of interest is extracted will determine the amount of cell debris and smaller contaminants that require removal from the product. Therefore the means by which the product is obtained directly affects the cost of the downstream processing and the economic viability of the product created. A system in which the release of the product from the cell is selective against other molecules would greatly reduce the cost of processing and provide a useful tool for biotechnology (reviewed in Balasundaram et al., 2009).

The function of the Tat system therefore provides an interesting approach towards the aim of selective product release. The export of proteins allows extraction from the periplasm rather than from the cytoplasm, greatly reducing the number of contaminants, and the use of Tat pathway also ensures that the product will be in its correctly folded, cofactor-containing state (Browning et al., 2017). The export of recombinantly expressed proteins through the Tat system has been demonstrated for many molecules, including the green fluorescent protein (GFP) (Thomas et al., 2001), antibody fragments (DeLisa et al., 2003) and growth hormones (Alanen et al., 2015).

Browning et al., (2017) have recently reported that engineered mutants of *E*. *coli*, so-called ‘TatExpress’ strains, with enhanced expression of the Tat machinery were able to produce growth hormones to a greater extent than previously used systems.

1.8 Aims of this work

This thesis aims to utilise the simple design features of maquette proteins to develop novel biological tools to address some of the fundamental problems presently facing biochemistry and physical chemistry. In each case, the biotechnological roles for the functional maquettes are detailed and some of the future prospects for the work are outlined. Chapter 2 will detail the necessary methodology to describe the experimental work in chapters 3-6, while chapter 7 will form a critical analysis of the approaches used and discuss potential future prospects.

In chapters 3 and 4, the simple design of the BT6 maquette (Farid et al., 2013) is employed to develop variant structures with altered conformational flexibilities, to facilitate the hydrophobic partitioning of carotenoids and acenes into the core of the protein. This would constitute the first example of carotenoids and acenes being incorporated into artificial protein scaffolds and allow assessment of the optical properties of these molecules in aqueous solution. It was subsequently found that both astaxanthin and tetracene adopted aggregated conformations within the protein cavity and that the optical properties of the complexes were comparable to previously reported work. As the aggregated conformation of the chromophores was shown to enable the process of singlet fission, the feasibility of using protein-pigment complexes as a means of improving photovoltaic devices is discussed.

The work described in chapter 5 builds on the results from chapters 3 and 4, but aims to impart a specific interaction between the protein chassis and the excited state dynamics of the incorporated chromophores. Here, the aim is to generate a novel protein-based system to observe the external spin-orbital coupling interaction in tetracene and C8-BTBT. The effect of incorporating heavy-atoms, such as selenium and iodine, into maquettes structures in close proximity to the bound chromophores is investigated by transient absorption spectroscopy.

The experimental data outlined in chapter 6 uses minimal changes in maquette sequence to directly alter haem-binding capacity, with the aim of imparting differential folding on the protein variants. The twin-arginine translocase system of *E. coli* is then challenged to recognise the folded state of the artificial proteins and export each variant according to the extent of its folding. This not only provides a means with which to probe the proofreading mechanism of the Tat system, but also demonstrates the feasibility of exporting cofactor-containing maquettes to the periplasm prior to extraction.

Chapter 2

Materials and methods

2.1 Materials

All materials were purchased from Sigma-Aldrich unless stated otherwise.

2.2 Standard growth media, buffers and reagents

Cell culture media and buffers were prepared according to Sambrook et al. (1989). Details of the growth media used here can be found in appendix section 1. All solutions were prepared in Millipore (MilliQ) filtered water (QH_2O) (Barnstead Nanopure, ThermoFisher). Where required, growth media, buffers and reagents were sterilised by autoclaving (125 °C, 15 p.s.i., 20 min) or by filtration (0.2 µm pore diameter, Sartorious). Temperature sensitive reagents, such as antibiotics, were only added to solutions when they had cooled to 50 °C or below.

2.3 *Escherichia coli* cell culturing

For all the experiments described here, the desired strains of *Escherichia coli* (*E*. *coli*) were cultured as single colonies on agar plates containing the appropriate concentrations of lysogeny broth (LB) medium components. Liquid cell culturing was initiated by inoculating 5, 10 or 50 mL of LB broth, containing the appropriate antibiotics, with single colonies taken from the agar plates. Details of the *E*. *coli* strains and plasmid vectors used in this study can be found in appendix section 2. All strains of *E*. *coli* used here were grown semi-aerobically in the desired medium at 37 °C (unless stated otherwise) with 230 rpm agitation; 50 mL cultures were grown in 250 mL conical flasks and 500 mL cultures in 2 L conical flasks. Where required, antibiotics were included at the following concentrations: 100 µg mL^{-1} ampicillin, 30 µg mL^{-1} kanamycin.

2.4 Production of chemically competent *E*. *coli*

E. *coli* cell lines were cultured semi-aerobically in 50 mL LB broth until the OD_{600} (optical density at 600 nm) of the cultures had reached approximately 0.5. Cell pellets were generated by centrifugation (3900 x*g*, 30 min, 4 °C) and resuspended in 25 mL chilled 0.1 M $MgCl_2$. Cells were centrifuged again (3900 x*g*, 30 min, 4 °C) and resuspended in 25 mL chilled 0.1 M $CaCl_2$. Cells were centrifuged (3900 x*g*, 30 min, 4 °C) and resuspended in 1 mL chilled 0.1 M $CaCl_2$, 20 % (v/v) glycerol. All solutions were sterilised by filtration (0.2 µm pore size filter, Sartorius) before use and cells were resuspended by shaking. At all stages cells were kept on ice.

2.5 Transformation of *E. coli*

25 µL aliquots of chemically competent *E. coli* (section 2.4) were thawed on ice. 10-50 ng of pure plasmid DNA, diluted in QH_2O, was added to the aliquots and the solution incubated on ice for 30 min. Cells were then heat-shocked by transfer from ice to a 42 °C water bath for 45 seconds before immediately returning to ice for 5 min. Cells subjected to the heat-shock process were recovered by the addition of 1 mL LB broth followed by 1 hr incubation at 37 °C. Cells were then harvested by centrifugation (16000 x*g*, 2 min), resuspended in 100 µL LB broth and plated onto LB agar containing the appropriate antibiotic under aseptic conditions. The colonies on the agar plates were cultured for 16 hr at 37 °C and then stored at 4 °C.

2.6 Nucleic acid manipulation

2.6.1 Preparation of plasmid DNA (mini prep)

5 mL LB media, with the required antibiotics, was inoculated with JM109 *E. coli* containing the desired plasmid and cultured for 16 hr before centrifugation (16000 x*g*, 10 min). Plasmid DNA was prepared from JM109 cultures using FastGene™ (Nippon Genetics) plasmid purification kit, according to the manufacturer's instructions. The yield of pure plasmid was typically between 5-10 µg in a 30 µL elution. DNA was stored at -20 °C in QH_2O.

2.6.2 Polymerase chain reaction (PCR)

DNA was amplified by the polymerase chain reaction (PCR) using Q5® high-fidelity DNA polymerase (New England BioLabs) or MyTaq™ Red Mix (Bioline), according to the manufacturer's instructions. Briefly, 50 µL high-fidelity polymerase reactions (DNA polymerase 0.02 U μL^{-1}) were conducted using 0.5 µM of each primer, 200 µM dNTPs and 0.1-1 µg DNA template (1 % v/v dimethyl sulphoxide (DMSO) was added for GC rich DNA).

2.6.3 Site-directed mutagenesis

To generate single residue changes in protein sequences, site-directed mutagenesis of plasmid DNA was conducted using the QuikChange II Site-Directed Mutagenesis kit (Agilent), according to the manufacturer's instructions. Briefly, 50 µL reactions were set up, in which 5-50 ng of purified plasmid DNA was amplified by PCR with primers (125 ng both forward and reverse) designed to introduce specific base-pair changes. Following amplification, methylated and hemi-methylated parental plasmid DNA was digested with DpnI restriction enzyme (section 2.6.4) leaving only mutated, PCR-generated plasmid DNA. The vector containing the required mutations was then transformed (section 2.5) into XL1-Blue *E*. *coli*.

2.6.4 Restriction enzyme digestion

Restriction enzymes were purchased from Promega or ThermoFisher (FastDigest® restriction enzymes) and all reactions were conducted according to the manufacturer's instructions. Digested DNA fragments were purified directly from solution or from agarose gels (section 2.6.5), where necessary, using the FastGene™ (Nippon Genetics) Gel/PCR Extraction Kit, according to the supplier's instructions.

2.6.5 DNA gel electrophoresis

Analysis of DNA fragments from PCR and restriction enzyme digests was carried out by electrophoresis on 1 % (w/v) agarose gels made in TAE buffer (40 mM Tris-acetate, 1 mM ethylenediaminetetraacetic acid (EDTA)) with 0.5 mg mL^{-1} ethidium bromide. DNA samples were mixed with 6x Blue/Orange Loading Dye (Promega; 0.4 % orange G, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 15 % Ficoll® 400, 10 mM Tris-HCl (pH 7.5), and 50 mM EDTA) before loading in the agarose gel. Hyperladder™ 1kb (BIOLINE) was run alongside DNA samples to enable size estimates. All gels were visualised under ultraviolet light.

2.6.6 DNA ligation

Ligation of DNA fragments was achieved using T4 DNA ligase (New England Biolabs) with a 5:1 molar ratio of insert to vector typically used. 10 µL reactions were used; 1 µL T4 DNA ligase, 1 µL 10x T4 DNA ligase buffer (New England Biolabs), 8 µL insert-vector mixture at an appropriate ratio. Reactions were left at room temperature for at least 4 hr before transformation into JM109 *E*. *coli*.

2.6.7 DNA sequencing

DNA sequencing reactions were conducted by GATC Biotech (Germany). Purified plasmid DNA was sent with the appropriate primers for Sanger sequence analysis. Electronic sequencing data was analysed with CodonCode Aligner (version 3.7.1) software.

2.7 Protein production

All proteins described here were produced in BL21 (DE3) *E*. *coli* transformed with an expression vector, containing the desired gene, downstream of a T7 polymerase promoter under the control of the *lac* operator.

2.7.1 Standard protein production

Escherichia coli containing the desired plasmid was grown in 500 mL semi-aerobic cultures (section 2.3) until the optical density at 600 nm (OD_{600}) had reached 0.6, after which 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein production. Following induction, cells were cultured for 4 hr before harvesting the cells by centrifugation (4400 x*g*, 4 °C, 30 min). Proteins were purified by immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) (section 2.8), if necessary.

2.7.2 Protein overproduction

Where necessary, 500 mL *E. coli* cultures were grown as described previously (section 2.7.1) and induced with 1 mM IPTG at OD_{600} ~0.6 for 24 hr. After induction cells were harvested by centrifugation (4400 x*g*, 4 °C, 30 min). Overproduced proteins were purified through IMAC, anion exchange chromatography (AEC) and SEC (section 2.8).

2.7.3 Selenomethionine protein production

50 mL of LB broth was inoculated from single colonies of BL21 (DE3) *E. coli* containing the desired plasmid and cultured for 16 hr. Cell pellets were formed by centrifugation (4400 x*g*, 4 °C, 30 min) and resuspended by inversion at 4 °C in 40 mL SeM minimal medium (see appendix section 1). 5 mL of cell suspension was used to inoculate 500 mL SeM minimal medium, which was grown semi-aerobically at 37 °C. When the OD_{600} of the cultures reached ~0.4, the cells were induced with 1 mM IPTG for 24 hr. Cells were harvested by centrifugation (4400 x*g*, 4 °C, 30 min).

2.8 Protein purification

2.8.1 Immobilised metal affinity chromatography (IMAC)

Cell pellets obtained in section 2.7 were resuspended in buffer A (20 mM HEPES pH 7.4, 500 mM NaCl, 4 mM imidazole). Cells were lysed by sonication on ice and protease inhibitor added. Lysates were clarified by centrifugation (53000 x*g*, 4 °C, 30 min) and the supernatant collected.

Supernatants were passed through a filter (0.45 µm pore diameter, Sartorius) and loaded onto a 5 mL Chelating Sepharose Fast Flow column (GE Healthcare) pre-equilibrated with 10 mg mL^{-1} Ni sulphate. The column was washed with 20 mL of buffer A, 20 mL buffer B (20 mM HEPES pH 7.4, 500 mM NaCl, 50 mM imidazole) and 10 mL buffer C (20 mM HEPES pH 7.4, 100 mM NaCl, 400 mM imidazole). Column eluate was collected at all stages and analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (section 2.9.2), with fractions containing the desired protein collected and pooled. Where necessary, proteins were further purified by anion exchange chromatography (section 2.8.2) and/or size exclusion chromatography (section 2.8.3).

2.8.2 Anion exchange chromatography (AEC)

Protein samples were buffer exchanged into 50 mM HEPES (pH 7.4) prior to AEC and 10 mL solutions were loaded onto a FastFlow Q-sepharose column (GE Healthcare) connected to an ÄKTA Fast Protein Liquid Chromatography (FPLC) system (GE Healthcare). Samples were eluted using a linear gradient of 0-1 M NaCl in 50 mM HEPES (pH 7.4), with protein elution measured by monitoring 280 nm absorbance. Protein containing fractions were collected and analysed by SDS-PAGE.

2.8.3 Size exclusion chromatography (SEC)

Protein samples were buffer exchanged into SEC buffer (50 mM HEPES pH 7.4, 200 mM NaCl) and concentrated to 0.5-1 mL using Vivaspin centrifugal concentrators. Samples were then loaded onto a Superdex 200 Increase column (GE Healthcare) under the control of an ÄKTA FPLC. Protein elution was measured by monitoring the A_{280} signal and fractions were analysed by SDS-PAGE.

2.9 Protein analysis

2.9.1 Bradford protein assay

Estimates of protein concentration were conducted using the Bio-Rad protein assay kit (BIO-RAD), according to the manufacturer’s instructions. Briefly, a dilution series of bovine serum albumin ranging between 0.2-0.9 mg mL^{-1} was generated, with 100 µL of each dilution added to 4.9 mL of dye reagent. Samples were vortexed and incubated for 10 min at room temperature, after which absorbance was measured at 595 nm. The assay was run in triplicate and a standard curve was plotted between the averages of the 5 concentrations used (Bradford, 1976).

2.9.2 SDS-PAGE

Sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE) was used to analyse protein samples. Proteins were initially denatured in Laemmli sample buffer (4 % SDS, 20 % glycerol, 10 % 2-mercaptoethanol, 0.004 % bromophenol blue and 125 mM Tris HCl pH 6.8) with boiling for 5-10 min. Linearised sample was then loaded onto precast 12 % polyacrylamide gels (NuPAGE® Bis-Tris; Thermo Fisher Scientific). Gels were submerged in MES Running Buffer (50 mM MES, 50 mM Tris Base, 0.1 % SDS, 1 mM EDTA) and a constant voltage of 180 mV applied to separate proteins. Bio-Rad Precision Plus Protein™ standards (BIORAD) were run alongside proteins to enable estimation of protein size. Proteins were visualised by staining in Coomassie Brilliant Blue (BIORAD).

2.9.3 Immunoblot analysis

Proteins were separated by SDS-PAGE as described in section 2.9.2, with the exception of Coomassie staining. To enable the immunoblotting procedure a transfer 'sandwich' was created; 2 porous pads on the exterior containing 2 squares of filter paper with a polyacrylamide gel and a polyvinylidene fluoride (PVDF) membrane (Invitrolon™, 0.45 µm pore size, Life Technologies) at the centre of the 'sandwich'. The sandwich was submerged in transfer buffer (10 mM $NaHCO_3$, 3 mM Na_2CO_3, 10 % (v/v) methanol) and a voltage applied across the apparatus (constant amplitude 40 mA for 16 hr) such that proteins were transferred from the polyacrylamide gel to the PVDF membrane.

The membrane was removed from the apparatus and incubated in blocking buffer (5 % milk powder, 0.2 % TWEEN 20 (w/v), 50 mM Tris pH 7.6, 150 mM NaCl) for 1 hr. Blocking buffer was removed from the membrane and the relevant primary antibody was added for at least 4 hr. All antibodies were diluted, according to the supplier's instructions, in antibody buffer (0.05 % TWEEN 20 (w/v), 50 mM Tris pH 7.6, 150 mM NaCl). The membrane was then thoroughly washed in antibody buffer and the relevant HRP-conjugated secondary antibody was added for 60 min. The membrane was thoroughly washed again and imaged on an Amersham™ Imager 600 (GE Healthcare) after the addition of luminal peroxidase solution (WESTSTAR ηC, Cyanagen).

2.9.4 Direct determination of protein concentration

Protein concentrations were determined according to the method described in Gill and von Hippel (1989). Briefly, proteins were denatured by buffer exchange into denaturing buffer (6 M guanidinum chloride) using Zeba™ Spin Desalting Columns (ThermoFisher), according to the manufacturer's instructions. The absorbance of the denatured protein at 280 nm (A_{den}) was measured on a FLUOstar Omega absorbance spectra reader (BMG LABTECH). The molar extinction coefficient of denatured protein (εM_{den}) was calculated according to the following equation:

$$\varepsilon M_{den} = 5690 \cdot n_{Trp} + 1280 \cdot n_{Tyr} + 120 n_{Cys}$$

where n_{Trp}, n_{Tyr} and n_{Cys} correspond to the number of tryptophan, tyrosine and cysteine residues in the protein, respectively.

In order to determine the extinction coefficient of the native protein (εM_{nat}) in aqueous buffer the following equation was used:

$$\varepsilon M_{nat} = \frac{A_{nat} \cdot \varepsilon M_{den}}{A_{den}}$$

where A_{nat} and A_{den} correspond to the absorbance of the native and denatured protein at 280 nm, respectively. The Beer-Lambert equation can then be applied to the absorption of native protein at 280 nm, in a 1 cm path length cuvette, to determine the molar concentration of native protein (C_{nat}) in the sample, according to the following equation:

$$C_{nat} = \frac{A_{nat}}{\varepsilon M_{nat}}$$

2.10 Spectroscopy

2.10.1 Ultraviolet-visible (UV-Vis) absorbance spectroscopy

All UV-Vis absorbance spectroscopy was conducted on a Cary 60 spectrophotometer (Agilent), with a 1 cm path length, in either quartz or UV-permissive polystyrene cuvettes, unless stated otherwise. Samples were diluted such that absorbance units were below 1.5. All spectra were corrected to an appropriate blank.

2.10.2 Circular dichroism (CD) spectroscopy

Mean residue ellipticity (θ) of protein samples was measured by CD spectroscopy on a Jasco J810 spectropolarimeter with a Jasco PFD-425s peltier attachment enabling temperature control (Greenfield, 2006). Protein samples were buffer exchanged into enantiopure buffer (5 mM sodium phosphate buffer pH 7.4) and diluted to approximately 0.1 mg mL^{-1}. All measurements were taken in a 1 mm path length quartz cuvette. Temperature-dependent CD was conducted by monitoring the mean residue ellipticity at 222 nm (θ_{222}) across a reversible 15-90 °C gradient (1 °C per min) with θ_{222} recorded in 1 °C increments. Melting temperatures (T_m) were determined from a Boltzmann distribution (section 2.14.1).

2.10.3 One dimensional proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy

All data described here was recorded using a Bruker Avance 800 MHz spectrophotometer, with a 5 mm Quadruple Resonance Probe (Bruker) at 298 K. Protein samples were buffer exchanged into enantiopure buffer and diluted to 200-500 µM. 10 % (v/v) deuterated water (D_2O) was added as a spectrophotometer lock and 1 mM trimethysilyl propanate was added as a reference standard. Spectra were obtained as accumulations of 4096 transients, with a proton spectral width of 0-30.0 ppm, with respect to the reference standard. All data were processed using an exponent multiplication window function and 5 Hz line broadening without linear prediction in the TopSpin software v4.0 (Bruker).

2.10.4 Transient absorption spectroscopy

Sub-picosecond transient absorption (TA) measurements were recorded in a similar experimental system to that described by Rao et al. (2010). Briefly, excitation pulses of approximately 10 nm bandwidth were generated in a TOPAS optical parametric amplifier (Light Conversion Ltd) coupled to a 10 kHz regenerative amplifier (Solstice, Spectra-Physics), with pump fluences of 4 x 10^{18} photon per pulse per cm^2 typically used to prevent degradation of samples. The output from the regenerative amplifier was then passed through a sapphire crystal (Edmund Optics) to generate white light, which was used as a probe to determine the absorption spectra of the excited sample. The probe beam was then dispersed by a reflective grating and detected by a silicon detector (Ultrafast Systems). The differential transmission (ΔT/T) in the sample was recorded as a function of excitation photon energy and pulse-probe delay. The experimental set up enabled a temporal resolution of approximately 200 fs.

2.10.5 Ground-state absorbance decay

Samples generated according to the method described in sections 2.12 and 2.13 were sparged with nitrogen for 5 min, placed in either a 1 mm or 10 mm pathlength quartz cuvette and sealed under anoxygenic conditions. Samples were illuminated under constant white light (200 µmol photons m^{-2} s^{-1}) and absorption spectra gathered in appropriate time intervals.

2.11 Pigment reconstitutions

The reagent concentrations quoted for the pigment reconstitutions correspond to the final concentration of each compound in the reaction. For the carotenoid and tetracene reconstitutions, all buffers and solvent were sparged with nitrogen gas for 30 min prior to use, with the chromophores kept strictly in the dark at all times.

2.11.1 Carotenoids

Astaxanthin (AXT) stocks (1 mg mL^{-1}) were prepared in 100 % DMSO and exact molar concentrations were calculated using its extinction coefficient (ε482 = 1.25 × 105 L·mol-1·cm-1) (Régnier et al., 2015). Protein reconstitutions were conducted in a 10-fold molar excess of AXT to give the following reaction conditions: 336 µM AXT, 33.6 µM protein, 40 mM HEPES (pH 6-9), 0-800 mM NaCl, 20 % (v/v) DMSO. Reconstitution reactions were left for 30 min at 4-50 °C, diluted with chilled 50 mM HEPES (pH 6-9) and transferred to ice. Note that the temperature, pH and NaCl concentration of the reconstitutions varied according to the experimental strategy detailed in section 3.3.3. Samples were filtered (0.2 µm pore diameter; Sartorius) and applied to a 5 mL DEAE sepharose anion exchange column to remove unbound cofactor. Reconstitutions of echinenone and beta-carotene were conducted through an identical method to that described above, but with both carotenoid stocks prepared at 30 µg mL^{-1} in DMSO and protein concentrations diluted to maintain a 10-fold pigment excess.

2.11.2 Tetracene

The TIPS-tetracene derivative used here (eTet) was solubilised in DMSO at 0.2 mg mL^{-1} by sonication at 60 °C. Reconstitutions were conducted using an identical method to that used in section 2.11.1, with 100 µM eTet and 10 µM protein used in the reactions. Unbound pigment was removed by filtration (0.2 µm pore diameter) and AEC (5 mL DEAE sepharose column).

2.11.3 C8-BTBT

C8-BTBT was solubilised to give a solution of 125 µg mL^{-1} in 100 % DMSO by sonication and heating to 60 °C. The reconstitution of C8-BTBT into maquettes and subsequent purification was conducted in a similar manner to those previously described, but with a 20-fold molar excess of pigment relative to protein. C8-BTBT has also been reported to display stability to light and oxygen (Yuan et al., 2014), so the precautions taken in the previous experiments were deemed unnecessary in this work.

2.11.4 Haem

1 mg mL^{-1} hemin was solubilised in 100 % DMSO and reconstituted with maquette proteins in a 10-fold molar excess of pigment to protein. The reconstitution reaction was conducted in with 306 µM hemin, 30.6 µM protein, 40 mM HEPES (pH 7.5), 160 mM NaCl, 20 % (v/v) DMSO at room temperature and unbound pigment was removed as described previously.

2.11.5 Haem and astaxanthin reconstitution into the BT6 maquette

Haem-containing BT6 proteins were generated according to the protocol in section 2.11.4 and buffer exchanged into 50 mM HEPES (pH 8.5). The protein concentration of the BT6-haem sample was calculated assuming that there was no loss of protein during the aforementioned protocol. AXT was added to BT6-haem in a 10-fold molar excess in an identical method to that described in section 2.11.1.

2.12 Solution aggregate generation

Pigment aggregation was conducted strictly in the dark and all solvents were sparged with nitrogen gas for 30 min prior to use. Absorption spectra were gathered immediately after mixing.

2.12.1 Astaxanthin

Astaxanthin aggregate II (AXT agg II) was generated according to Musser et al., (2015). A solution of 1 mM AXT was prepared in 100 % DMSO and heated to 80 °C. The AXT stock was then mixed in 1:9 ratio with QH_2O pre-heated to 80 °C.

2.12.2 Tetracene

100 µM stock solutions of eTet were generated at 65 °C with sonication and added to QH_2O preheated to 4, 37, 50 or 80 °C in a 1:4 ratio. Solutions were mixed by repeated inversion.

2.13 Tat-mediated protein export

2.13.1 Cloning of maquette sequences

The maquette genes that included the N-terminal Tat signal peptide of trimethylamine-N-oxide reductase (TorA) were provided in the pJexpress414 plasmid by DNA2.0 (now ATUM). To limit the level of protein expression and avoid overload of the Tat export system, maquette sequences were cloned into the low copy number pEXT22 plasmid, through ligation of gene fragments into the vector’s XbaI/KpnI sites. The N-terminal Sec signal peptide from pectate lyase B (PelB) was added to the N-terminal of the BT6 variants by ligation of the maquette sequences into NcoI/XhoI sites of the pET22b(+) plasmid vector.

2.13.2 Growth and fractionation of *E. coli* for Tat-mediated protein export

BL21 (DE3) *E. coli* were transformed with pEXT22 plasmids, containing the desired maquette gene sequences, and an empty pET21a(+) vector. 50 mL cultures were grown semi-aerobically at 37 °C in LB broth to an OD_{600} of ~0.6, where gene expression was induced with 0.5 mM IPTG for 2 hr. Cells were harvested by centrifugation (4400 x*g*, 30 min, 4 °C) and processed to generate either cell-free extracts or periplasmic fractions as described below.

Cell-free extracts were generated by gentle resuspension of cell pellets to an OD_{600} ~2 with sucrose-free buffer (100 mM Tris-acetate pH 8.2, 5 mM EDTA) and sonication on ice. The lysate was clarified by centrifugation (16600 x*g*, 30 min, 4 °C) and the supernatant collected as the cell-free extract.

Periplamic fractions were generated using an osmotic shock method based on the protocol described in Randall and Hardy (1986). Cells were resuspended to an OD_{600} ~10 with high-sucrose buffer (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA) and 500 µL of cell suspension transferred to ice. 5 µL of 1 mg mL^{-1} lysozyme was added to the cell suspension, followed by 500 µL of chilled QH_2O. After 5 mins on ice, 20 µL of 1 M $MgSO_4$ was added and the solution mixed by gentle inversion. Spheroplasts and cell debris were removed by centrifugation (16600 x*g*, 30 min, 4 °C) and the supernatant collected as the periplasmic fraction. All solutions and labware were chilled before use.

The protein content of periplasmic fractions and cell-free extracts was analysed by immunoblotting (section 2.9.3), with primary antibodies against β-lactamase and GroEL used to assess the relative proportions of periplasmic and cytoplasmic proteins in each sample. Sample loadings were normalised to give approximately equal signal for β-lactamase, and the presence of GroEL was used to assess the extent of cytoplasmic contamination in the periplasmic fractions.

2.13.3 Purification of proteins from TatExpress E. coli

The *E. coli* W3110 TatExpress (Browning et al., 2017) cell line was transformed with pEXT22 vectors containing the gene sequences for the desired maquettes. 500 mL cultures were grown semi-aerobically at 30 °C and protein production was induced with 500 µM IPTG once the OD_{600} reached ~0.6. After 24 hr induction, cells were harvested by centrifugation (4000 x*g*, 30 min, 4 °C) and resuspended in 10 mL high-sucrose buffer. 10 mL chilled QH_2O was added, followed by 800 µL of 1 mg mL^{-1} lysozyme, and samples were incubated on ice for 10 min. 800 µL of 1 M $MgSO_4$ was added, with spheroplasts and cell debris removed by centrifugation (16600 x*g*, 30 min, 4 °C). Proteins were purified from the supernatant by IMAC (section 2.8.1).

2.14 Data normalisation and curve fitting

2.14.1 Boltzmann distribution

A sigmoidal Boltzmann distribution was used to describe the relationship between protein residue ellipticity and temperature according to the following equation:

$$\theta_{222} = \theta_{222}^{base} + \frac{\theta_{222}^{max} - \theta_{222}^{base}}{1 + exp\left(\frac{T_m - T_0}{\left(\frac{\delta T}{\delta \theta}\right)}\right)}$$

Where θ corresponds to the mean residue ellipticity, T corresponds to the temperature and the Tm refers to the transition temperature of melting or refolding.

2.14.2 One-phase exponential decay

One-phase exponential decays were fit to describe the absorbance decay of chromophores under high light according to the following equation:

$$A = A_{max} \cdot e^{-K \cdot T} + A_{min}$$

Where A is absorbance at a particular wavelength, K is a rate constant and T is time.

2.15 Molecular dynamics simulations

All molecular dynamics (MD) simulations were conducted by Dirk B. Auman (University of Pennsylvania, USA). Brief descriptions of the methods used are outlined below.

2.15.1 Computational prediction of an AXT dimer in solution and in the BT64A maquette

The conformation of an AXT dimer was generated using previously published structural parameters (Cerezo et al., 2013). An energy minimised state was generated through the Gromacs (v4.6) (Hess et al., 2008) molecular dynamics software, with water modelled using the TIP3P-CHARMM force field. A snapshot of the dimer structure was taken following a 50 ns trajectory and visualised in the PyMOL molecular graphics system (v1.7).

The amino acid sequence of BT64A was run through the PEP-FOLD structural prediction software (Shen et al., 2014) for a total of 100 simulations. The five best-scoring structures were analysed in PyMOL and one was taken forward for further experimentation. The structure of the AXT dimer was manually docked into the BT64A starting structure and the complex was energy-minimised using the Gromacs software. A snapshot of the BT64A-AXT structure was taken following 50 ns of the MD trajectory and visualised in PyMOL.

2.15.2 Computational prediction of haem binding into BT6 maquette variants

The predicted structure of each maquette was generated by analysing the peptide sequence in the PEP-FOLD software to generate 100 simulated structures, after which the five best-scoring were used for further experimentation. The starting structures were visualised using the PyMOL molecular graphics system (v1.7) and 2, 1 or 0 haem molecules were manually placed in near the histidine binding sites for the BT6, BT6M1 and BT6M0 variants, respectively. The structures were then relaxed to an energy-minimised state through the Gromacs molecular dynamics software.

Simulations were run using the CHARMM27 force field for the proteins and haem, with the positioning of water determined using the TIP3P-CHARMM force field. Covalent bonds were specified between the ferric iron centre of haem and the ε-nitrogen of the coordinating His-residues. Each structure was equilibrated for 100 ps to allow convergence of temperature, pressure and volume prior to the MD trajectory. Snapshots of the trajectories were taken after 50 ns and exported to .pdb files. The predicted molecular structures were visualised in PyMOL.

2.16 *In vitro* protein modification

2.16.1 Iodination of proteins

Iodine atoms were covalently attached to the tyrosyl moiety of purified proteins using Pierce Iodination Beads (Thermo Scientific) and sodium iodide (NaI) (Markwell, 1982). Iodination beads were washed thoroughly in an Eppendorf tube with reaction buffer (100 mM sodium phosphate buffer pH 7.2, 150 mM NaCl) and dried on filter paper. 3 iodination beads and 1 mg of protein were used in each 500 µL reaction. Protein and reaction buffer were added to a total volume of 250 µL in a round bottom 2 mL Eppendorf tube. 250 µL of iodination buffer (100 mM sodium phosphate buffer pH 7.2, 150 mM NaCl, 10 mM NaI) was then added, immediately followed by 3 washed iodination beads. The iodination reaction was conducted at room temperature for 10 min with gentle agitation. The reaction was terminated by removal of the iodination beads and the protein solution was buffer exchanged into 50 mM HEPES (pH 8.5) using Zeba™ Spin Desalting Columns (ThermoFisher).

2.17 Mass spectrometry (MS)

Note that all solutions for MS analysis were generated using LC-MS grade solvents and water.

2.17.1 Direct-infusion intact protein MS

Pierce C18 Spin Columns (Thermo Scientific) were used according to the manufacturer's instructions to remove any contaminants from the intact protein samples and solubilise the peptides in an appropriate solvent for MS analysis. Protein samples were prepared by adding 75 µL of protein (0.1 mg mL^{-1}) in 50 mM HEPES (pH 8.5) to 25 µL of sample solvent (2 % trifluoroacetic acid (TFA), 20 % acetonitrile (ACN)). The C18 resin was activated by methanol (50 %) and equilibrated in equilibration solvent (0.5 % TFA, 5 % ACN). Samples were loaded onto the column and washed in wash solvent (0.5 % TFA, 5 % ACN). The sample was then eluted with elution solvent (70 % ACN) and dried by evaporation under vacuum. The protein precipitate was gently resuspended in infusion solvent (60 % formic acid, 20 % ACN, 20 % H_2O) and directly applied to a Q Exactive HF quadrupole-Orbitrap mass spectrometer (Thermo Scientific) at 5 µL mL^{-1} via heated electrospray ionisation (ESI).

2.17.2 Tandem liquid chromatography mass spectrometry (LC-MS) of intact proteins

Tandem LC-MS was conducted on a 6530 Quadrupole-Time-of-flight (Q-TOF) mass spectrometer (Agilent) to determine the mass of intact protein samples. 20 µL of protein sample (1 mg mL^{-1}) was loaded onto a C18 reverse-phase column, pre-equilibrated in solvent A (0.1 % formic acid, 99.9 % H_2O). A linear gradient of solvent A to solvent B (0.1 % formic acid, 99.9 % ACN) was applied to the column to remove peptide samples, with the elution of protein monitored by tryptophan absorption. The eluted protein solution was then ionised by an ESI source and analysed by Q-TOF MS. The resulting charge series (m/z values) were then deconvoluted using the MassHunter charge profiling software (Agilent) to give the average molecular mass.

Chapter 3

A sterically-controlled, thermostable protein matrix to study the excited state dynamics of carotenoid aggregates

---

3.1 Motivation

The primary objective of the work detailed here was to develop a system that would permit the incorporation of carotenoids into artificial maquette proteins. Due to the importance of carotenoids in many biological processes (Britton et al., 1995), the generation of artificial carotenoproteins could enable new functions in synthetic biology and provide a tractable system for fundamental investigations into the excited state dynamics of carotenoids.

The first application that could be envisaged is the development of such complexes for roles in biotechnology. For example, artificial maquette structures have been demonstrated to have light-induced charge-separating functions through the incorporation of Zn-tetrapyrroles (discussed in section 1.3) (Farid et al., 2013; Moser et al., 2016; Ennist et al., 2017a). Although shown to be successful in initiating energy transfer processes, the sole presence of tetrapyrroles leads to the potential for highly damaging singlet oxygen ($^1O_2$) formation. In natural photosynthetic complexes the generation of singlet oxygen is either directly or indirectly quenched by the close physical arrangement of carotenoids and chlorophylls. Therefore, a means of integrating carotenoids into previously developed artificial charge-separating proteins would give improved photostability with the potential to impart useful operational lifetimes.

The second application of artificial carotenoproteins would involve their use as a controlled system for studying the photophysical properties of a range of carotenoids. As demonstrated for Fe- and Zn-tetrapyrroles, artificial protein constructs show reasonably promiscuous cofactor binding compared to naturally evolved proteins (Solomon et al., 2014). If such a property could also be realised for carotenoid incorporation, then the maquette approach would enable the comparison of the optical properties of many carotenoids in an identical environment. This system may then be used to determine the electronic transitions of a range of carotenoids and elucidate their role in biologically important processes, such as non-photochemical quenching (Muller et al., 2001; Ruban, 2016).

3.2 Summary

Assigning function to artificial peptides has largely been conducted through the 'maquette' approach to *de novo* protein design, pioneered by the work of the Dutton laboratory (Robertson et al., 1994; Dutton and Moser, 2011; Moser et al., 2016). As the majority of protein-catalysed reactions require the inclusion of cofactors (Koder and Dutton, 2006), maquette functionality has typically been achieved through site-specific ligand attachment or simple hydrophobic partitioning of cofactors (Farid et al., 2013; Mancini et al., 2018). The present study employs the latter binding method for several reasons: (1) carotenoid binding to native proteins remains poorly understood (García-Martín et al., 2008) and, at the time of writing this thesis, no known means for site-specific attachment exists; (2) as the motivation of this work was to generate a promiscuous system to enable the incorporation of many carotenoids, it was of interest to limit any specific interactions with a particular cofactor structure that may impede the attachment of other molecule types; (3) as carotenoids are highly hydrophobic, simple partitioning becomes largely favoured in aqueous environments.

To achieve the goals set out in section 3.1, the experimental strategy aimed to develop carotenoid binding in the BT6 maquette construct, first described by Farid et al., (2013). As detailed in section 1.3.2, the BT6 maquette is a single-polypeptide tetrahelical protein with a diverse range of potential cofactors. Here, the BT6 nomenclature will be used specifically to designate the maquette that has a di-haem binding function through four histidine residues (see table 3.1). Two variants of the BT6 maquette were generated that had progressively greater hydrophobicity in the core of the protein, resulting in concomitant differences in thermal stability, as assessed by temperature-dependent circular dichroism (CD) spectroscopy.

The ability of the maquette proteins to bind to the carotenoid astaxanthin (AXT) was assessed by ground state absorption of the purified carotenoproteins. It was found that all three maquettes bound AXT to similar extents and that the carotenoid was seemingly incorporated as a weakly-coupled H-aggregate (Musser et al., 2015). Note that the term aggregate will be used in this thesis to refer to any pigment complexes that contain two or more molecules. The purified AXT-maquettes could be concentrated to high optical densities (OD_{459nm} > 14, 1 cm pathlength) in aqueous solution, yielding carotenoid aggregates that were effectively-water soluble.

Due to intermolecular coupling, carotenoid aggregates display attenuated optical properties compared to monomers (see sections 1.4 and 1.5), with the interaction permitting transitions that are optically forbidden in the monomeric state (i.e. transfer between singlet and triplet states).

Transition to the optically forbidden lowest triplet energy level (T_1) was observed in carotenoids through transient absorption spectroscopy. TA lifetime analysis confirmed the kinetics of this transition were rapid and largely similar to previously reported data on AXT solution aggregates.

The observation of intermolecular coupling in the AXT-maquette not only changed the electronic transitions of the incorporated chromophore but also extended the potential applications of the protein-pigment complex as a whole. Electronically coupled carotenoid aggregates can be described as an organic semi-conducting material due to their high electron-hole mobility. Organic semi-conducting materials have seen considerable interest in the past three decades as an alternative to traditional inorganic thin-film transistors and possess many potential device applications, such as organic light-emitting diodes (reviewed in Dimitrakopoulos and Mascaro, 2001; Reese et al., 2004; Mei et al., 2015). In the interests of succinctness, the development of the carotenoid maquettes in this section will only be discussed in relation to their application as a charge multiplication material for use in photovoltaic devices.

The formation of triplets in carotenoid aggregates occurs rapidly through a spin-permitted quantum mechanical process known as singlet fission (SF), in which two independent triplet states are formed from a singlet exciton (Smith and Michl, 2013). As discussed in section 1.6, there has been great interest in the potential of SF to overcome efficiency limitations in conventional photovoltaic devices. As the mechanism of SF in AXT does not appear to be greatly affected by the protein chassis, the carotenoid-maquettes provide an attractive approach in the fundamental investigation of this process and its future device applications.

3.3 Experimental work

3.3.1 Maquette design and production

All the maquette genes discussed in this chapter were purchased from DNA 2.0 (now ATUM) and provided in the high copy-number pJexpress414 plasmid. The details of the three maquette protein variants are displayed in table 3.1. The BT6 maquette was modified by substituting the four histidine residues for alanine or phenylalanine residues, generating variants named BT64A and BT64F, respectively. As the histidine residues are believed to be located in or near the hydrophobic cavity of the protein, these changes were predicted to progressively increase the hydrophobicity of the core.

All the maquette proteins in this chapter were produced and purified from *E. coli* BL21 (DE3) according to the method set out in section 2.7.1. Briefly, 500 mL cell cultures were grown at 37 °C to an optical density at 600 nm (OD_{600}) of ~0.6, after which production of the maquette was initiated through IPTG-induction of the bacterial T7 promoter. Following induction, cells were cultured for a further 4 h at 37 °C before lysis by sonication. Following clarification of the cell lysate, maquettes were initially purified by immobilized metal affinity chromatography (IMAC), with purity assessed by SDS-PAGE (section 2.9.2). Where necessary, maquettes were further purified by size exclusion chromatography (SEC) (section 2.8.3). An example of the purity of the proteins obtained through this procedure is displayed in figure 3.1.

3.3.2 Maquette stability

The stability of each of the maquette constructs was measured by temperature dependent CD to determine what affect the alanine and phenylalanine substitutions had on the BT6 structure. Residue ellipticity (θ) was measured between 190 and 260 nm for each variant in 50 mM HEPES (pH 6-9) at 15 °C with the spectra displaying typically α-helical character for each maquette (figure 3.2, black lines). Spectra were also recorded at 90 °C (figure 3.2, red lines) and following cooling to 15 °C (figure 3.2, blue lines).

The thermal stability of the maquette constructs was then determined by monitoring residue ellipticity at 222 nm (θ_{222}) across a reversible temperature gradient (Greenfield, 2006). Where possible the datapoints for melting and refolding were described by a Boltzmann distribution (see section 2.14.1) and the midpoint of the melting curve (T_m) was determined from this non-linear fit (figure 3.3 and table 3.1). The inclusion of increasingly hydrophobic residues into the core of the maquette stabilised the protein structures, raising the T_m from 34 °C in the BT6 maquette to 49 °C for BT64A, when measured at pH 9. A Boltzmann distribution could not be accurately fit to describe the data for the phenylalanine maquette and as this construct retained typically α-helical character at 90 °C (figure 3.2, red line), the melting transition is assumed to be greater than this temperature. Note that due to the strong buffering conditions used, the accuracy of the measurements at wavelengths shorter than ~215 nm is poor; however, the buffering conditions were selected to accurately record the stability of the proteins in the conditions used for the subsequent experiments in this chapter. Full unfolding spectra of BT6 variants in weak buffering conditions are displayed in section 6.3.3.

It was shown that the pH of the buffering solution greatly affected the stability of the proteins. At pH 7.5, only minor shifts in protein stability were observed, with just a 3 °C increase in the T_m of BT6 and BT64A. However, at pH 6, a dramatic stabilization was observed, increasing the T_m for BT6 by 45 °C (figure 3.3A). Similarly to BT64F, a Boltzmann distribution could no longer be fit to the BT64A data at pH 6, indicating that the T_m was in excess of 90 °C, and T_m values could not be fit at any pH value for the BT64F maquette. All melting curves were shown to be reversible with no significant aggregation (figure 3.3).

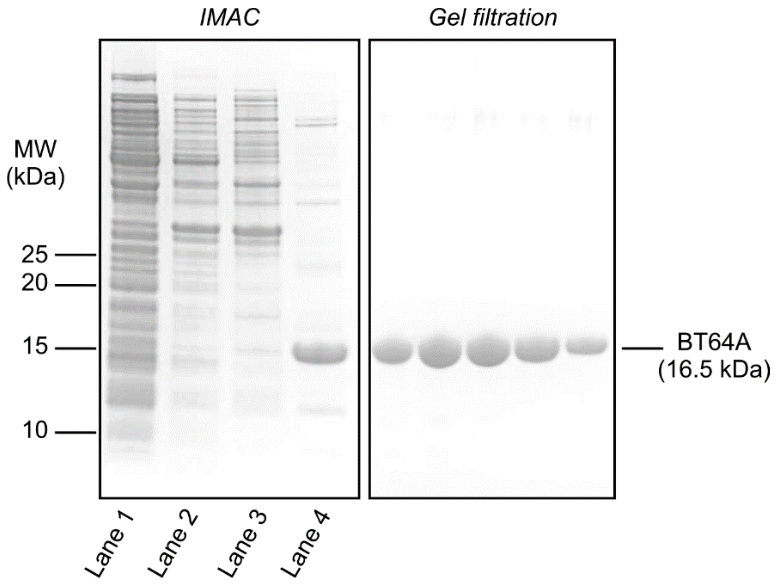


Figure 3.1 SDS-PAGE analysis of BT64A maquette purity. The flow-through (lane 1), wash steps (lane 2 and 3) and elution (lane 4) from an immobilised metal affinity chromatography (IMAC) column (left). Elution fractions from a gel filtration column (right) loaded with the sample from lane 4.

3.3.3 Optimisation of AXT binding

Several variables were considered to find optimal conditions for the partitioning of carotenoids into the hydrophobic cavity of the maquette proteins, including the pH of the buffering solution, the temperature of the reconstitution reaction and the effect of the salt (NaCl) concentration. These first two considerations were chosen as they affect the conformational flexibility of the protein (as described in section 3.3.2), with the hypothesis that the propensity for cofactor incorporation would depend on the availability of the hydrophobic cavity. In this section, the incorporation of AXT is assessed by quantifying the total ground state absorption in the visible region, relative to the concentration of protein (assessed by tryptophan absorbance at 280 nm). At this stage, little consideration was made of the overall character of the absorption spectra with discussion on spectral features deferred until section 3.3.5.

All work in this chapter that involved the use of chromophores was conducted strictly in the dark and, where necessary, all buffers and solvents were depleted of oxygen by sparging with nitrogen. AXT was mixed with the BT64A protein in largely aqueous solution (40 mM HEPES, 180 mM NaCl, 20 % (v/v) DMSO), with a ten-fold molar excess of the carotenoid, according to the method described in section 2.11.2. Briefly, 350 µM stocks of AXT were made up in 100 % DMSO. This stock was then added to the BT64A protein in aqueous buffer (50 mM HEPES, 200 mM NaCl) at room-temperature, with the solution at various pH values (pH 6, 7, 7.5, 8, 8.5 and 9). Note that the pH of the solutions, along with the concentrations of buffer and salt, are quoted before the addition of the carotenoid stock. The protein-carotenoid mixture was left for 30 mins at room temperature, after which large unbound aggregates were removed by filtration (0.2 µm pore diameter) and AXT-maquette complexes purified by anion exchange chromatography (AEC). The ground-state absorption spectra of the resulting eluates showed that the pH of the solution greatly affected the incorporation of AXT into the maquette structure, with greater binding occurring at higher pH values and little or no incorporation at pH 6 or 7 (figure 3.4A). Taken together with the information on protein stability, it could be concluded that the conformational flexibility of the hydrophobic cavity is an important determinant for the incorporation of AXT. All subsequent carotenoid reconstitutions in this chapter were conducted in 50 mM HEPES at pH 8.5 unless stated otherwise.

Next, the effect of temperature on carotenoid incorporation was assessed. Pigment incorporation was conducted essentially as described above, but now the temperature of the solution was varied prior to and following the addition of the carotenoid. Protein-containing solutions (50 mM HEPES, 200 mM NaCl, pH 8.5) were incubated in water baths at 4, 37 and 50 °C for 15 mins. Carotenoid was then added and the reactions were incubated for a further 30 mins at the corresponding temperatures. Following filtration and AEC, the normalised absorption spectra of the complexes revealed that carotenoid incorporation occurred more efficiently at 37 °C than at 4 or 50 °C (figure 3.4B). The difference in carotenoid binding at 4 and 37 °C can be adequately explained by the conformational flexibility hypothesis previously discussed; however, this suggestion cannot rationalise the difference between the incorporation at 37 and 50 °C. At 50 °C the protein has greater flexibility, which would be expected to make the protein cavity more accessible to cofactors. It could be argued that discrepancy arises from degradation of the protein at 50 °C, but this seems unlikely due to the ability of the proteins to melt and refold following heating to 90 °C. As these results cannot be explained by the conformational flexibility or degradation of the protein, a more complicated mechanism must underlie carotenoid incorporation (discussed in section 3.3.5). All subsequent AXT-binding experiments were conducted at 37 °C unless stated otherwise.

The final consideration in these AXT binding optimisations was the effect of NaCl concentration. The salt concentrations used, before the addition of carotenoid, were 0, 200, 500 and 1000 mM. AXT incorporation was tested as described above and absorption spectra were recorded. It was found that increased salt concentration concomitantly reduced the efficiency of carotenoid incorporation, with the most intense absorption occurring in the absence of any salt (figure 3.4C).

Table 3.1. Amino acid sequences and physical properties of peptides used in this study.

Protein	Sequence	T_m (°C)		
		pH 6	pH 7.5	pH 9
BT6	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLHHHHHH	79	37	34
BT64A	MGGDGENLYFQG EIWKQ**A**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**A**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**A**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**A**EDALQKFEEALNQFEDLKQLHHHHHH	>90*	52	49
BT64F	MGGDGENLYFQG EIWKQ**F**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**F**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**F**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**F**EDALQKFEEALNQFEDLKQLHHHHHH	>90*	>90*	>90*

The positions of residue substitutions are shown in bold. *The midpoint of the melting curve could not be accurately determined within the temperature range tested; all proteins displayed a typically helical signal at 90 °C.

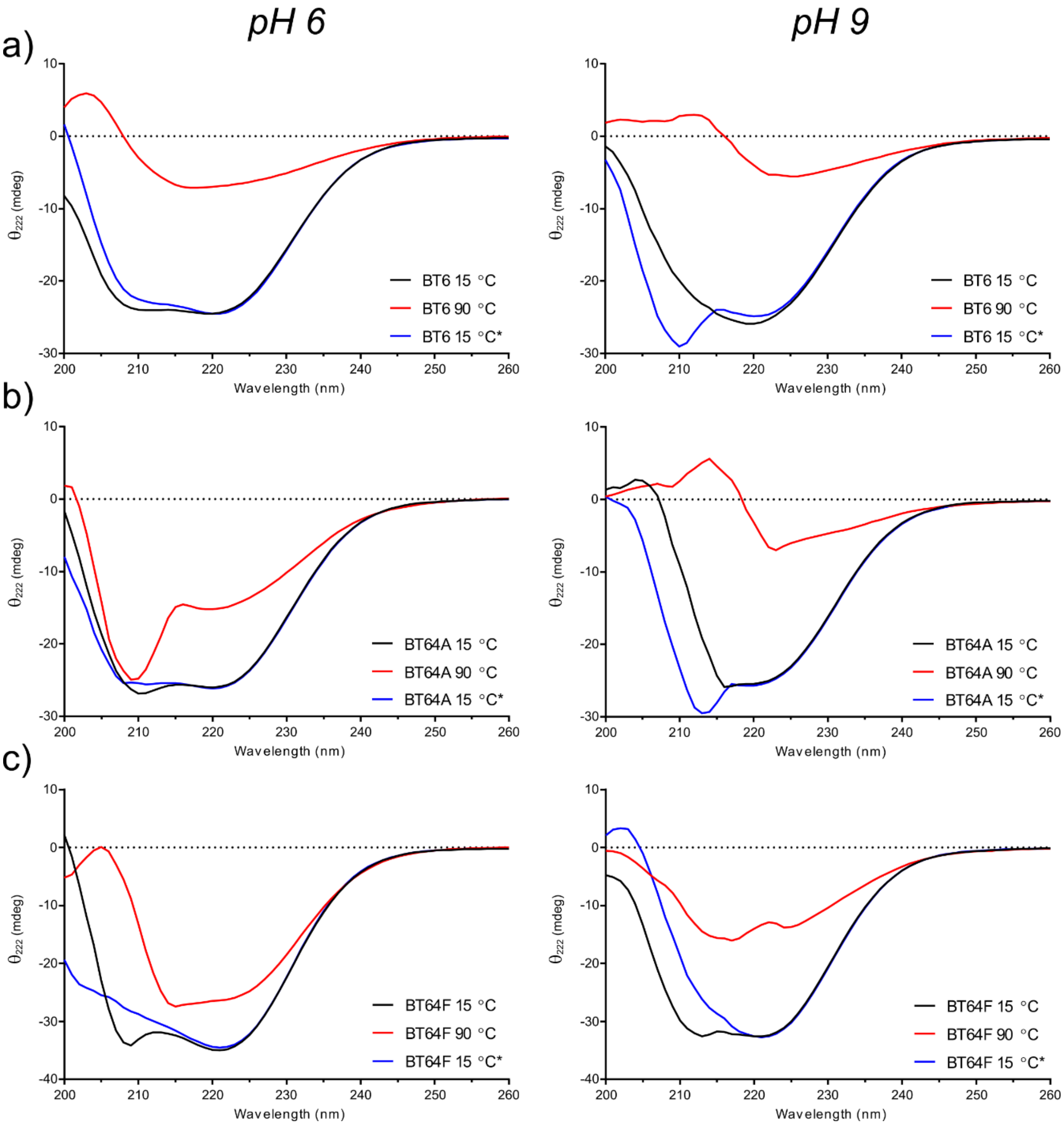


Figure 3.2 Residue ellipticity spectra of BT6 maquettes at pH 6 (left) and pH 9 (right). The residue ellipticity (θ) of BT6 **(A)** BT64A **(B)** and BT64F **(C)** was recorded at 15 °C (black), 90 °C (red) and following the return from 90 to 15 °C (blue). Note that due to the strong buffering conditions used the ellipticity values for wavelengths shorter than ~215 nm contain high error.

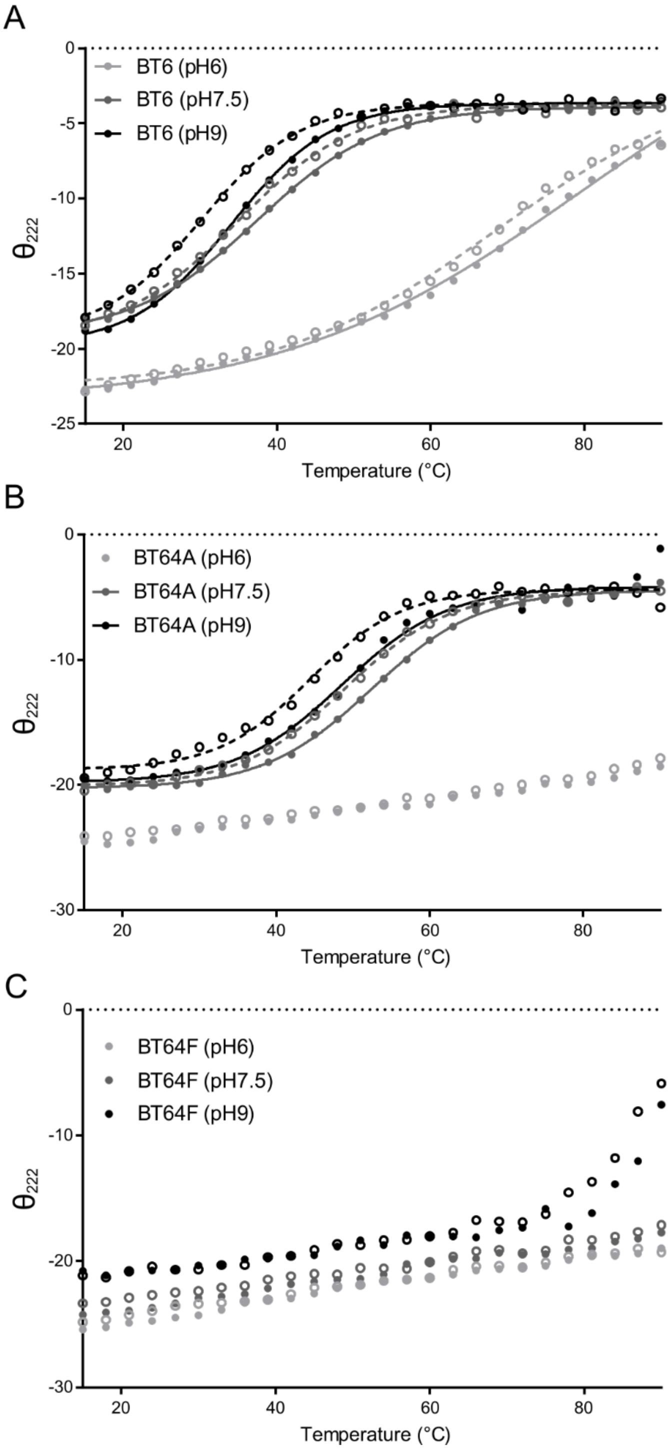


Figure 3.3 The thermal stability of maquette variants at various pH values. Melting and refolding data for BT6 **(A)**, BT64A **(B)** and BT64F **(C)**. Residue ellipticity at 222 nm (θ_{222}) is shown across a temperature gradient. Solid markers indicate values for melting and open markers for refolding. Ellipticity was measured in 1 °C intervals but only every third datapoint is shown for clarity. Boltzmann distributions are shown as solid (melting) or dashed lines (refolding), where possible.

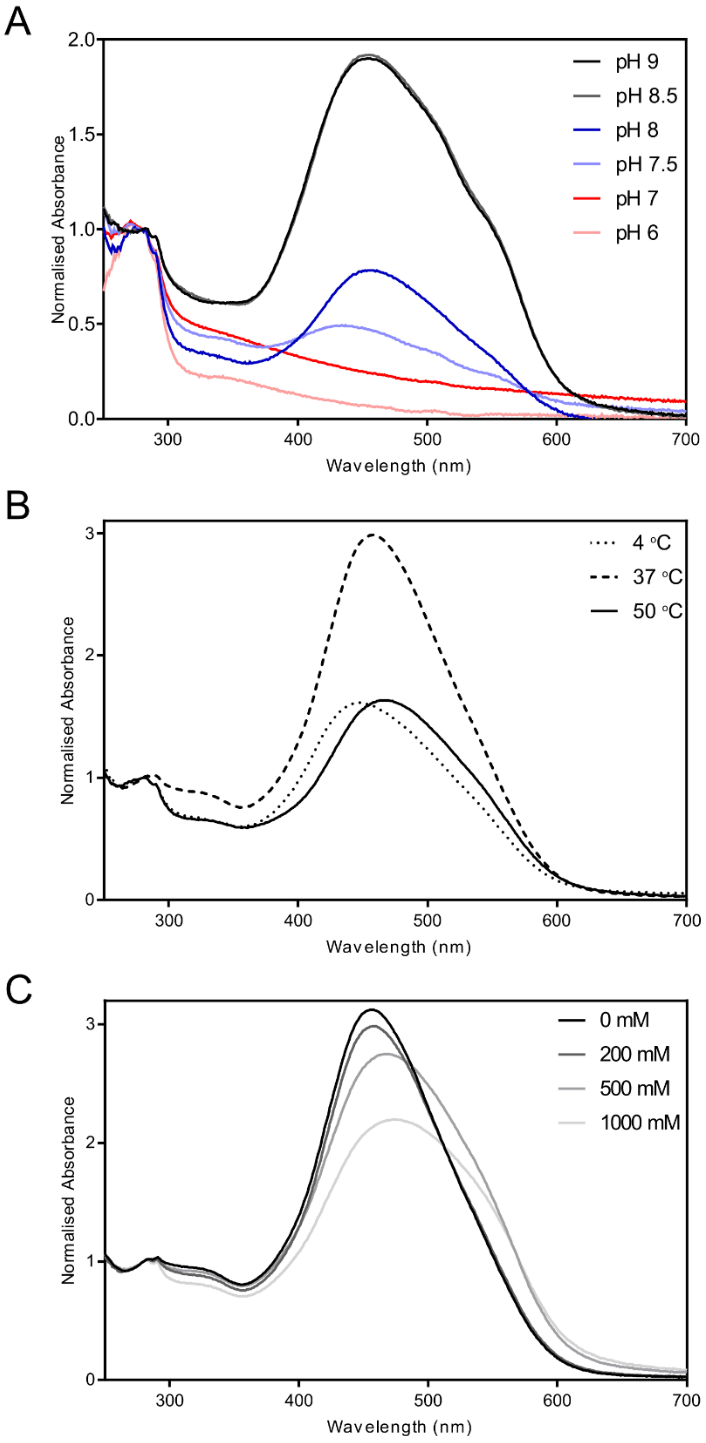


Figure 3.4 Optimisation of AXT binding to BT64A. The incorporation of AXT into the BT64A maquette was assessed by ground state absorption spectroscopy, with the spectra normalised on the basis of maquette absorption at 280 nm. **(A)** Variation of AXT incorporation into maquette complexes with pH; the pH value corresponding to each absorption spectrum is indicated in the figure legend inset. **(B)** AXT was added to BT64A in solution at 4 (dotted line), 37 (dashed line) and 50 °C (solid line). The absorption spectra of samples are shown following filtration and AEC, with the solution at room temperature. **(C)** The effects of salt concentration on AXT incorporation.

3.3.4 Comparison of AXT binding to maquette variants

AXT was incorporated into the BT6 and BT64F maquettes using an identical method to that described in section 3.3.3 to identify any differences in binding between the three variants. Controls were performed in the absence of protein to elucidate the spectral characteristics of AXT in solution under the conditions used for the incorporations and to demonstrate the aptitude of filtration and AEC in removing unbound carotenoids.

The absorption spectra of each sample, prior to and following filtration, are displayed in figure 3.5 and normalised absorption spectra following elution from AEC are shown in figure 3.6A. Precipitation of AXT could be observed by eye in the no-protein control, but not in any of the maquette samples. This precipitation is manifested in the absorption spectra shown in figure 3.5D, which is characteristic of disordered pigment aggregation. Filtration was sufficient to remove all pigment in the no-protein control, whereas, the maquette samples largely retained the full amount of carotenoid (figure 3.5).

Following AEC, all three maquette constructs retained similar carotenoid levels. Although the relative heights of these peaks were found to be slightly variable on repetition, the normalised spectra displayed in figure 3.6A provide a representative example of AXT binding. The fact that AXT binding is largely similar for the three proteins calls into question the hypothesis that conformational flexibility determines the efficiency of incorporation.

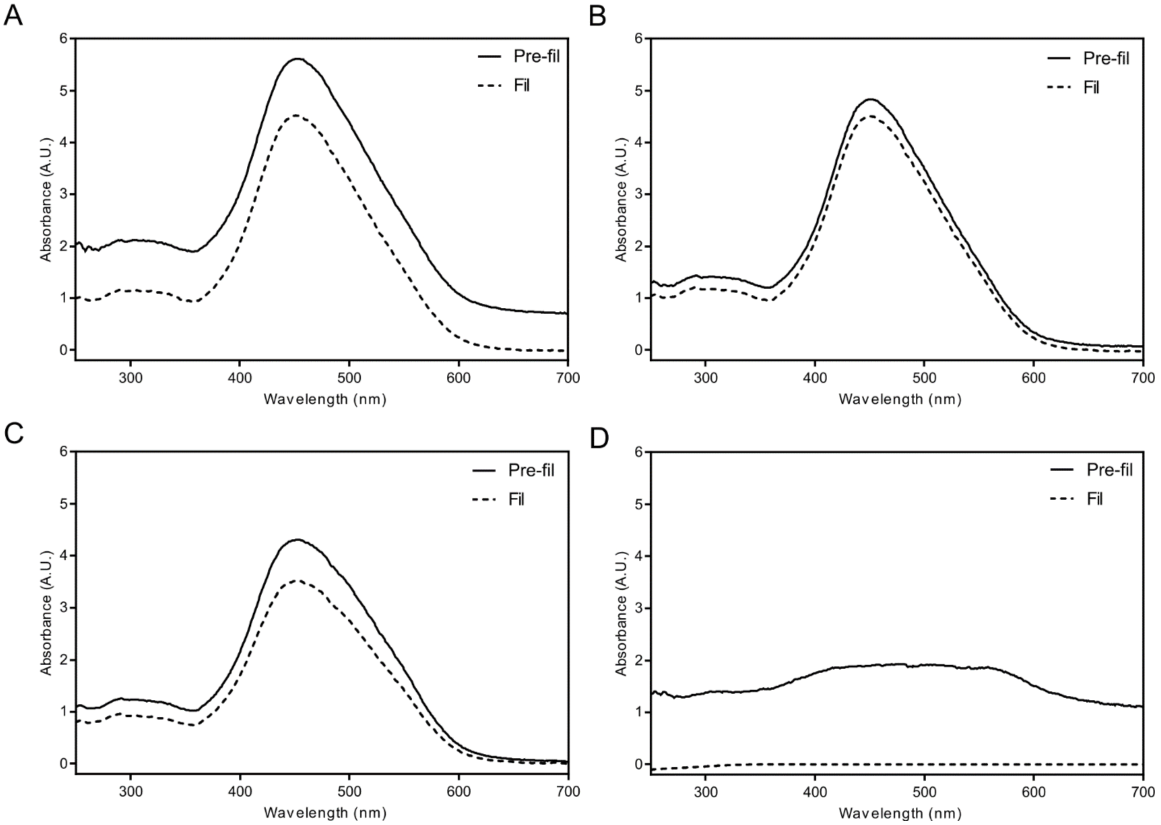


Figure 3.5 Absorption spectra of AXT in BT6 (A), BT64A (B), BT64F (C) and in solution (D). Absorption spectra are shown before (Pre-fil; solid line) and after (Fil; dashed line) filtration.

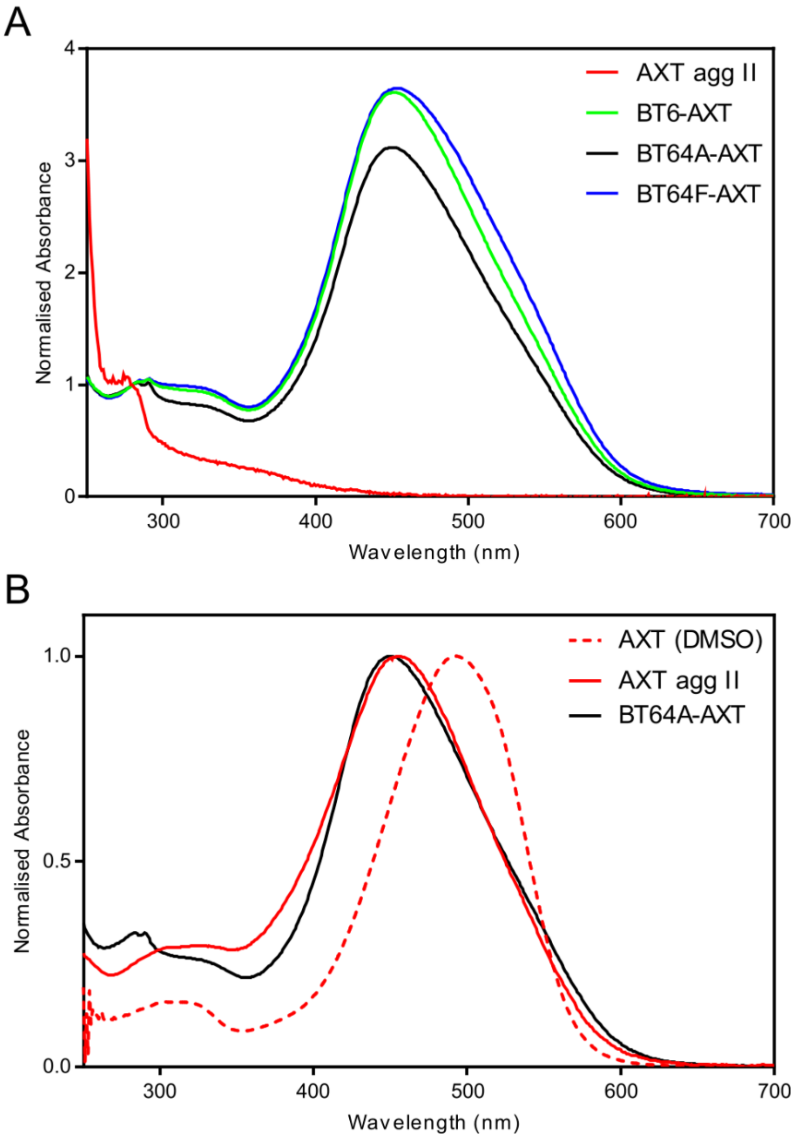


Figure 3.6 Absorption spectra of AXT in protein and in solution. (A) Normalised absorption of AXT in BT6 (green), BT64A (black), BT64F (blue) and in solution (red), following filtration and AEC. All samples normalised to absorbance at 280 nm. **(B)** Normalised absorption spectra of monomeric AXT (red, dashed), AXT agg II (red, solid) and BT64A-AXT. All samples were normalised to the maximum absorbance value.

3.3.5 Mechanism of AXT binding

As described in section 1.5.2, ground-state absorption provides a useful indication of any modifications in carotenoid structure or environment (Schiedt and Liaaen-Jensen, 1995). A particularly important observation is the blue-shift in the absorption maximum of the AXT-maquette samples compared to monomeric AXT in DMSO (figure 3.6B, dashed line). The absorption of the purified AXT-maquettes is similar to a previously reported weakly-coupled AXT solution H-aggregate, herein called aggregate II (AXT agg II) (figure 3.6B, solid line) (Musser et al., 2015). AXT agg II is one of five unique, stable AXT aggregation states and is formed by the mixing of water and monomeric AXT in DMSO at 80 °C (see section 2.12.1).

The similar absorption characteristics of AXT agg II and the AXT-maquette complexes leads to the hypothesis that two or more carotenoids are incorporated into the hydrophobic cavity of the protein, presumably in a parallel alignment (see sections 1.4.8 and 1.5.2). It was shown by Musser et al., (2015) that the type of AXT aggregate formed could be dramatically changed by relatively subtle alterations to the reaction conditions affecting features such as solvation, temperature and pigment concentration. This demonstrates the fragile nature of carotenoid aggregate formation, which is believed to occur largely through hydrogen bonding and hydrophobic interactions (Billsten et al., 2005; Musser, 2013).

If the formation of aggregates is considered as the key determinant of incorporation, rather than the makeup of the protein, the results obtained so far can be rationalised more easily. This explanation works on the assumption that only very small aggregates are permitted into the hydrophobic cavity (which will be addressed in section 3.3.6) and that the aggregate size is the limiting factor. Figure 3.7 depicts the mechanisms in which AXT aggregates may be sequestered within protein scaffolds, showing the insertion of AXT into the scaffold (horizontal arrows) and the formation of aggregates as separate processes (vertical arrows). Hypothetically, if the rate of aggregate formation is slower than the binding rate, then monomeric carotenoid would be incorporated into the protein cavity (as in scheme 1). If the protein cavity sterically permits more than one carotenoid, then monomers may be added sequentially until the maximum aggregate size is reached. An alternative pathway could also be envisaged, where the rate of binding and aggregation are similar. In such a scenario, aggregates that are sufficiently small to be permitted into the internal protein cavity could be pre-formed in solution (scheme 2). In scheme 3 the rate of aggregation is faster than the rate of carotenoid binding (scheme 3) and, under such conditions, the rapid formation of large aggregates would sterically prevent their insertion into the maquette interior.

These modes of binding are consistent with the optimisations detailed in section 3.3.3 concerning pH, temperature and salt concentration, all of which would affect the hydrogen bonding and hydrophobic interactions in carotenoid aggregates. Therefore, the conditions of the solution must permit the formation of aggregates of the appropriate size and shape to permit incorporation into the protein. The presence of AXT in a weakly-coupled H-aggregate conformation in all three maquette variants suggests that the optimised conditions follow scheme 1 or 2 in figure 3.7.

The optical properties of AXT in the no-protein control were characteristic of absorption scatter and are attributable to the formation of a variety of different and very large aggregation states. As filtration through pores of 200 nm diameter was sufficient to remove all carotenoid from the solution aggregates, it can be assumed that all complexes must be of greater size, following the 30 min incubation period. As the H-aggregation state of all AXT-maquettes was maintained through filtration and AEC, the proteins must limit the size of the aggregates formed, most likely through encapsulation of the chromophore in the hydrophobic cavity.

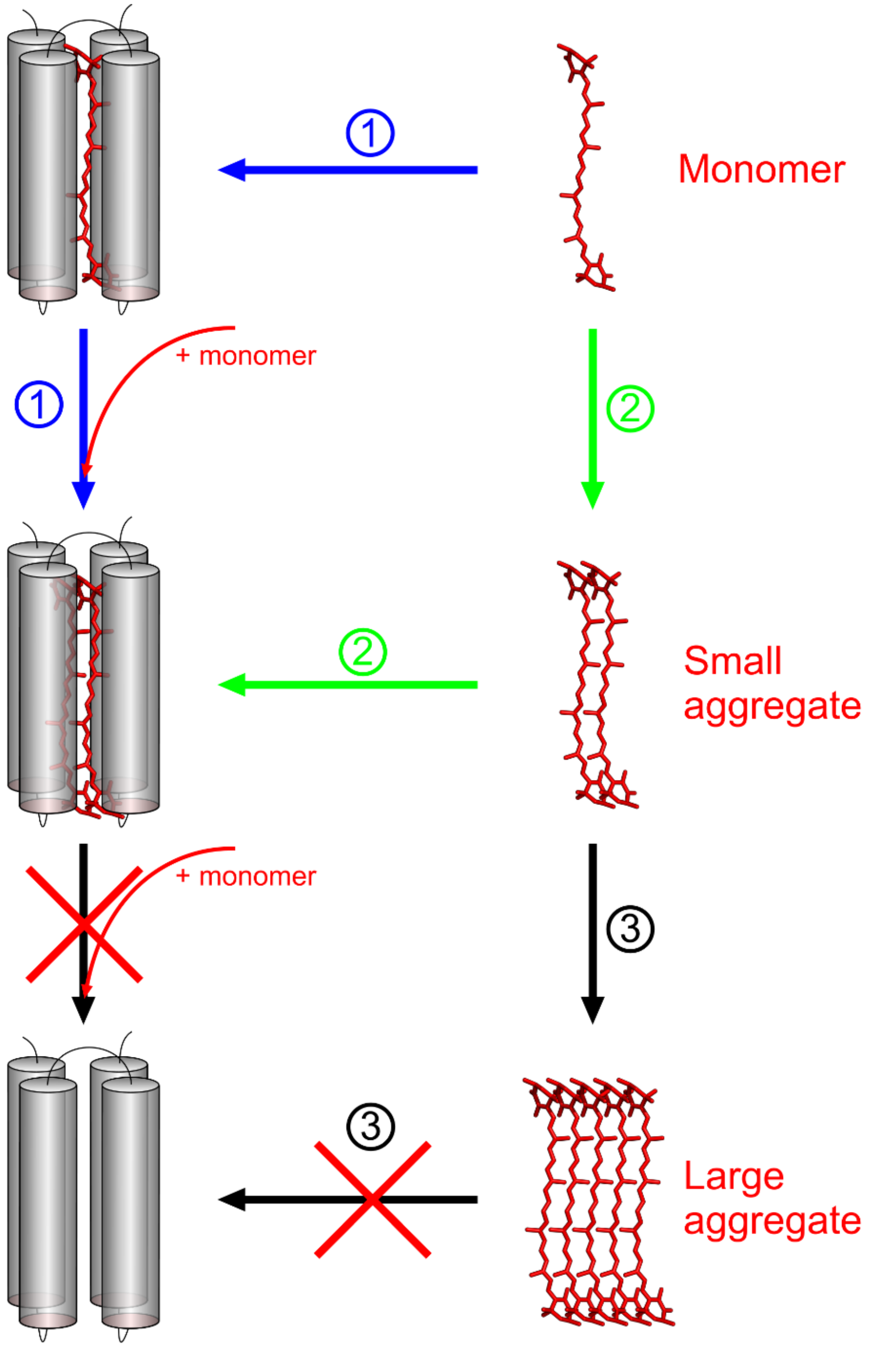


Figure 3.7 Schematic to demonstrate the potential binding modes of AXT to a maquette protein. The maquette protein is represented by the grey cylinders and AXT by the red structures. If the incorporation of AXT occurs more rapidly than aggregation, then monomeric carotenoid will be added sequentially to the protein core until a maximum size is reached (scheme 1, blue arrows). If the rate of aggregation and binding are similar then a scenario may arise in which small aggregates form and subsequently partition into the maquette structure (scheme 2, green arrows). If aggregation is very rapid, then structures could form that are too large to be permitted into the protein (scheme 3, black arrows). The aggregate size and composition in this illustration are entirely arbitrary and may not reflect the true reaction conditions. All arrows are shown as irreversible for simplicity, but this is unlikely to reflect the true nature of carotenoid incorporation, which would be somewhat reversible under aqueous conditions.

3.3.6 Location of AXT binding

To examine the notion that the AXT aggregates are binding within the protein core, haem was ligated into the BT6 maquette (section 2.11.5). As described in Farid et al., (2013) and section 1.3, the BT6 maquette binds two molecules of haem *b* in the hydrophobic protein cavity via axial ligation of the iron centre by four coordinating histidine residues. Following ligation of haem, the BT6 maquette was mixed with AXT as described above then purified by filtration and AEC. Little or no trace of the carotenoid was observed in the ground-state absorption spectra (figure 3.8), suggesting that AXT could not partition into maquettes when haem is already present. Although the ligation of haem has been shown to stabilise the maquette structure (discussed at length in chapter 6), the data in section 3.3.4 shows that the stability of the protein has little effect on carotenoid incorporation. Therefore, the inhibition of aggregate binding by prior ligation of haem is likely to arise from steric clashes within the hydrophobic cavity rather than any restriction caused by the reduction in conformational flexibility. This leads to the conclusion that the carotenoid must be sequestered within the hydrophobic cavity, rather than somehow attaching to the polar surface of the BT6 proteins. If the latter process occurred, the prior attachment of haem within the cavity would not affect carotenoid binding.

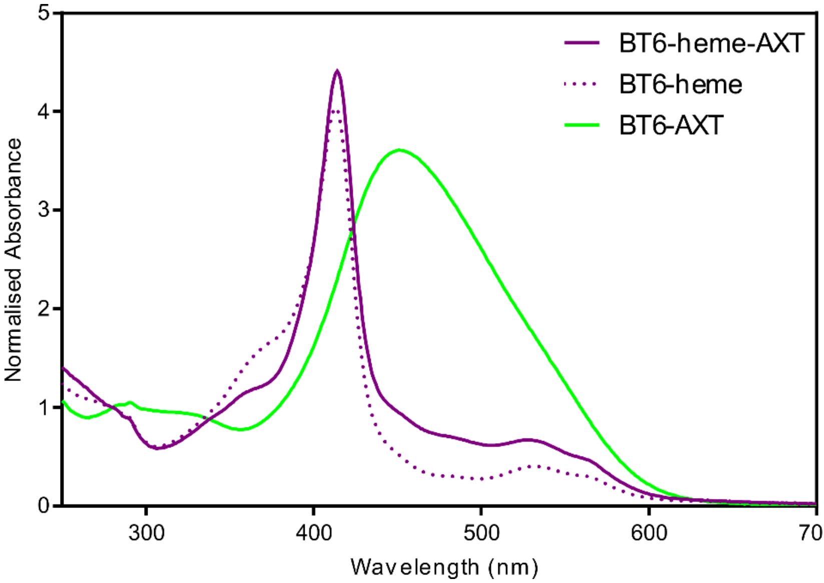


Figure 3.8 Absorption spectra of the BT6 maquette with haem and AXT. The BT6 maquette is shown following the insertion of haem (purple, dotted line). AXT was then added to BT6-haem and spectra shown after filtration and AEC (purple, solid line). BT6 with AXT is shown for comparison. All spectra are normalised to absorbance at 280 nm.

3.3.7 Molecular dynamic simulations of AXT maquettes

To complement the experimental findings, molecular dynamics (MD) simulated trajectories were used to indicate whether an AXT H-aggregate would be sterically permitted in the cavity of the BT64A maquette and give a likely conformation for the protein structure. In the interests of simplicity, only a dimeric pigment conformation was used in the simulations. MD trajectories were conducted by Dirk B. Auman (University of Pennsylvania, USA).

The conformation of an AXT dimer was generated using the Gromacs molecular dynamics software (v4.6) (Hess et al., 2008) according to previously published structural parameters (Cerezo et al., 2013). An energy minimized *apo*-protein structural model was generated for the BT64A protein using Gromacs. The AXT dimer was then docked into the *apo*-protein structure and MD simulations were performed as described in section 2.15.1. The atomised coordinates of the 60 ns trajectories are visualised in PyMOL format in figure 3.9.

The resulting structure of the AXT dimer was similar in configuration to that observed previously (Olsina et al., 2012). On the timescales of the simulation, the carotenoid dimer was sterically permitted into the hydrophobic cavity of the protein. Although the structural models shown here seemingly agree with the empirically derived data, the theoretical structures are included for illustrative purposes and are not intended as a replacement for atomically accurate experimentally-determined structures.

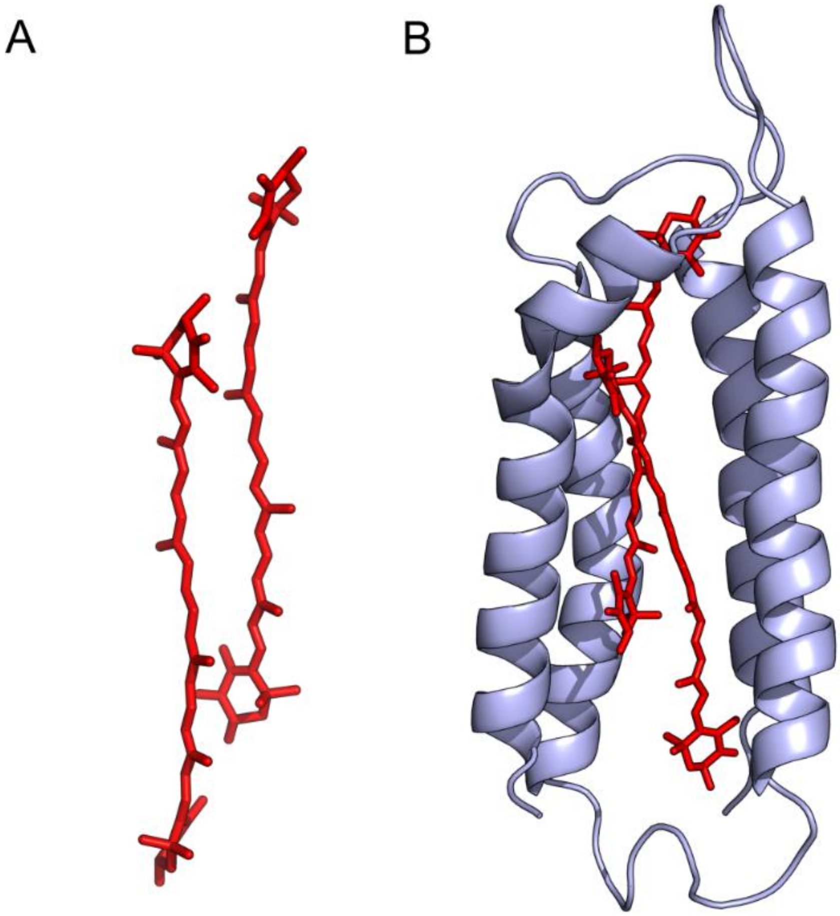


Figure 3.9 Structural models of AXT agg II and BT64A-AXT from simulated MD trajectories. (A) Structure of an astaxanthin H-aggregate. **(B)** The tetrahelical BT64A protein (purple) with a dimeric H-aggregate of astaxanthin bound (red). Both structures were derived from a 60 ns MD trajectory and visualized in PyMol (v1.7).

3.3.8 Transient absorption spectroscopy of AXT maquettes

The ground-state absorption spectra of AXT-maquette complexes show that the aggregated conformation of the carotenoids results in substantial changes in optical properties compared to the monomer. The intermolecular coupling responsible for these changes also results in different excited state dynamics, where optically forbidden transitions become permissible (see sections 1.4.8, 1.5.2 and 1.6.3 for more detail). Of particular interest here is the process of singlet fission (SF), which has been shown to occur in aggregates of zeaxanthin and AXT (Wang and Tauber, 2010; Musser et al., 2015).

As described in detail in section 1.6, SF is a quantum mechanical process that permits the generation of two independent triplet excited states from one singlet excited state. In previously reported work, ultrafast transient absorption (TA) spectroscopy revealed the photo-induced absorption (PIA) spectra of the $1B_u$, $2A_g$ and lowest (T_1) triplet states in monomeric and aggregated AXT, as well as the rate of the formation and decay of these excited states (Musser et al., 2015). The identification of these PIA features in the TA spectrum of AXT-maquettes, along with subsequent investigation of the excited state kinetics, would provide a diagnostic means of assessing the intermolecular interactions within the protein core and demonstrate the feasibility of using artificial proteins to study electronic transitions in carotenoid aggregates.

Ultrafast-TA spectroscopy on AXT-maquettes was carried out in collaboration with Daniel Polak (University of Sheffield, UK), as described in section 2.10.4. Briefly, carotenoprotein samples in aqueous buffer (50 mM HEPES, pH 6) were concentrated to obtain maximum absorbance values in excess of ten absorbance units in a 1 cm pathlength cuvette (A_{459} > 10 A.U.). Normalised TA spectra of BT64A-AXT, BT64F-AXT and AXT agg II were measured following laser excitation at 400 nm, 500 Hz 120 fs pulse length. The spectra of the PIA peak corresponding to the T_1 excited state are shown for all samples after 3 and 30 ps probe delay (figure 3.10A,B), with the spectra of the two AXT-maquettes comparable to AXT agg II at both timepoints. As the formation of triplets is unlikely to occur through single photon transitions, the presence of the T_1 state confirms that the carotenoids must undergo SF in the protein. As the PIA features appear to overlay at the timepoints shown it also suggests that the kinetics of T_1 formation and decay are consistent between all samples. To further confirm this, the emergence and decay of the 600 nm PIA feature were monitored over time (figure 3.10C); the similar excited state dynamics of all of the AXT samples suggest that the protein has little or no effect on the electronic transitions of the incorporated material.

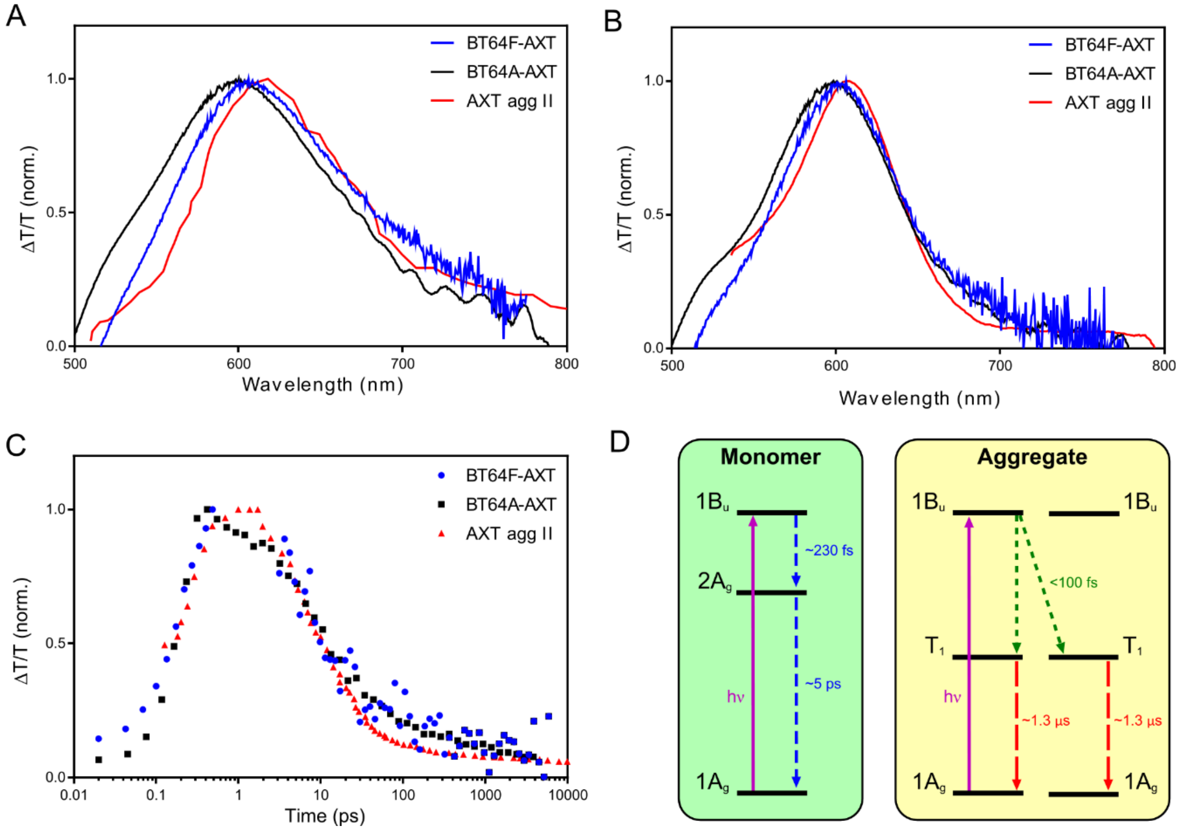


Figure 3.10 Transient absorption spectroscopy of AXT aggregates in solution and in maquettes. Normalised TA spectra of AXT agg II (red), BT64A-AXT (black) and BT64F-AXT (blue) in solution, taken after a 3 ps **(A)** or 30 ps **(B)** delay following excitation at 400 nm. **(C)** Transient absorption kinetics of the AXT T_1 state, shown as formation and decay of the 600 nm PIA signal over time, following photo-induction. **(D)** Excitation scheme for monomeric and aggregated AXT with approximate lifetime kinetics indicated. Photon absorption is shown in purple, internal conversion in blue, singlet fission in green and intersystem crossing in red. The data on AXT agg II and the lifetimes in panel D were reproduced with permission from Musser et al. (2015).

3.3.9 Stability of AXT maquettes

The primary aim of the work in this section was to determine the practical advantages the maquette approach may offer over aggregate based systems for investigating SF. Maquettes overcome some of the limitations that arise from the instability and unpredictability of the carotenoid aggregates generated in mixed solvent systems, with the lifetime of samples often short due to the susceptibility of photooxidative decay and incompatibility with long-term storage (Schiedt and Liaaen-Jensen, 1995). As shown in Musser et al. (2015), the temperature of the reaction solution dramatically changes the aggregation state of AXT and, as such, solution-based samples cannot be cryogenically stored. Therefore, the measurement of carotenoid aggregates requires the frequent regeneration of samples, which can lead to difficulties in reproducibility due to the highly sensitive nature of aggregation.

The maquette approach can also address a second, more ambitious aim, of generating carotenoid aggregates with sufficient stability for applications in photovoltaic devices. This concept is discussed at length in section 1.6, but the essential objective is to generate a material capable of being used in conjunction with conventional solar cells, which generates multiple excited states from a single photon. Such a system is suggested to help overcome some of the efficiency limitations of conventional solar cells (Smith and Michl, 2013; Rao and Friend, 2017). In practical applications SF-sensitising materials would be exposed to constant solar illumination and therefore require high thermo- and photo-stability for operational lifetimes. At the time of submission of this thesis, no such material has been described in the literature.

Following sample generation, a BT64A-AXT sample in aqueous buffer (50 mM HEPES, pH 6) was submerged in liquid nitrogen and stored at -80 °C under atmospheric conditions. After 6 months the sample was thawed and the absorption spectra of the samples before and after storage showed negligible differences in optical properties (figure 3.11A). In a separate experiment, a BT64A-AXT sample was treated with successive rounds of freezing in liquid nitrogen and thawing in a 30 °C water bath, with ground state absorption spectra taken after each cycle. The normalised absorption spectra showed no change in the overall character of the sample following 10 cycles (figure 3.11B) and little change in the absorbance intensity at 459 nm (A_{459}) (figure 3.11C). These results demonstrate that the carotenoid aggregates, encapsulated by the surrounding protein, are amenable to long-term cryogenic storage. This feature greatly improves the practicality of studying such carotenoid aggregates, as the same sample can be measured multiple times several months apart and also raises the possibility of transporting such materials by conventional means to a range of locations for spectroscopic or structural studies.

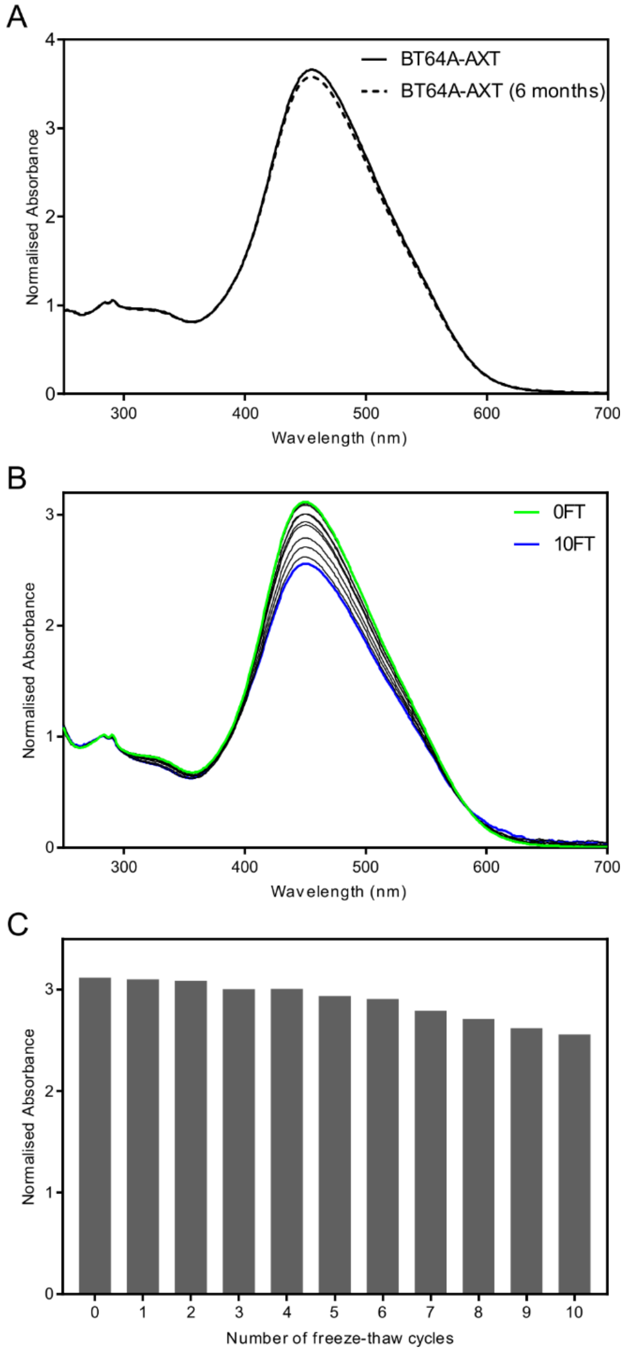


Figure 3.11 Solution stability of BT64A-AXT constructs. (A) Absorption spectra of a BT64A-AXT sample before and after cryogenic storage (solid and dashed lines, respectively). **(B)** The absorption spectra of a BT64A-AXT following successive cycles of cryogenic freezing and thawing. The sample prior to freezing is shown in green and following 10 cycles in blue. Intervening cycles are shown in black. **(C)** The absorbance at 459 nm for the samples in (B) is displayed for each cycle. All values were normalised to absorbance at 280 nm.

The photostability of AXT agg II in solution and in the protein matrix was compared by monitoring the ground state absorption profile over time under constant white light illumination (200 µmol photons m^{-2} s^{-1}) (section 2.10.5). Photooxidative decay was determined by measuring the reduction in A_{459} signal at 30 min intervals for 100 hours. A non-linear exponential decay was fit to the data as described in section 2.14.2. AXT aggregates incorporated in the protein maquettes showed an approximately three-fold increase in half-life compared to AXT agg II in solution alone (31 hours and 11 hours, respectively). Although the photostability is far below that necessary for practical use, the data show that maquettes provide a means of improving the lifetime of the SF sensitiser and form a promising prototype that warrants further development.

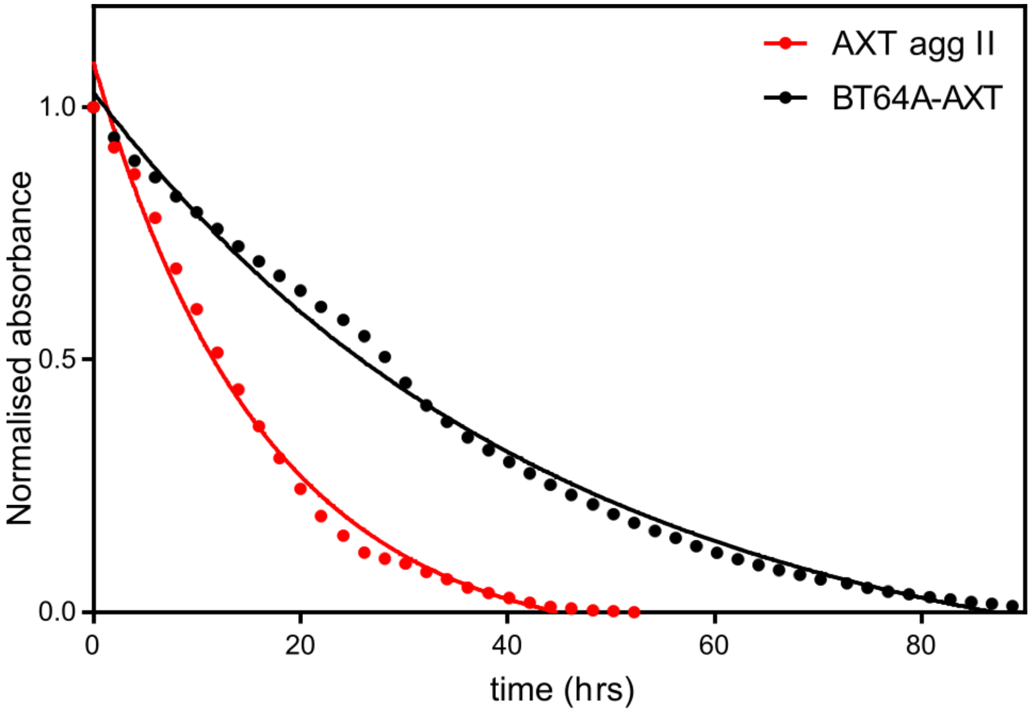


Figure 3.12 Photostability of AXT in solution and in BT64A. Normalised absorbance at 459 nm is shown as a function of time for AXT aggregates in solution (red) and in BT64A (black), with both samples under constant white light illumination. Measurements were taken every 30 mins but only every third data point is shown (closed dots). A non-linear decay was fit to describe each dataset (solid lines).

3.3.10 Incorporation of other carotenoids

As outlined in section 3.1, one of the aims of the work undertaken here was to create a promiscuous system capable of binding different carotenoid structures. As the experimental work so far has only concerned AXT, the range of available carotenoids was expanded by incorporating β-carotene (βC) or echinenone (ECH) into the BT64A maquette, essentially as described for AXT (detailed in section 2.11.1), with pigment added in ten-fold molar excess to protein, and unbound pigment removed by filtration and AEC. The resulting protein-pigment complexes were measured by ground state absorption spectroscopy (figure 3.13). A clear blue-shifted absorption was observed in BT64A-ECH compared to the monomer, suggesting that, similarly to AXT, the carotenoids were being bound in a conformation similar to a weakly coupled H-aggregate. The absorption shift is less clear in βC, but the spectral characteristics look similar to a previously reported H-aggregate (Lu et al., 2016). As little or no published TA data are available for aggregates of these molecules, the excited state transitions in the carotenoid-maquettes await further characterisation.

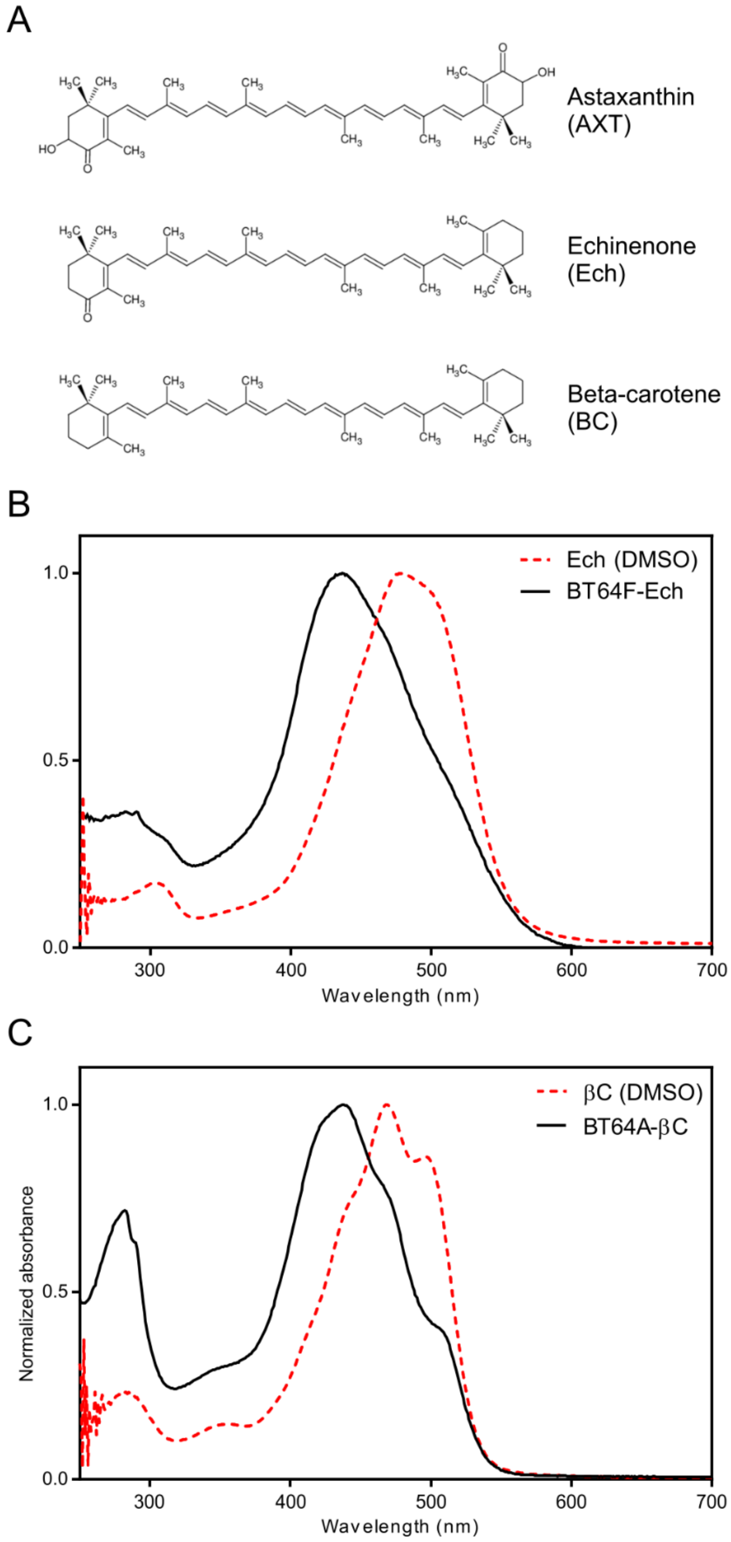


Figure 3.13 Incorporation of echinenone and β-carotene into BT64A. (A) Chemical structures of the carotenoids used in this chapter. **(B)** Absorption spectra of monomeric ECH in DMSO and in the BT64A maquette. **(C)** Absorption spectra of monomeric βC in DMSO and in the BT64A maquette. All absorption profiles are normalised to the maximum absorbance value of the sample.

3.4 Discussion

The system developed here provides a means of measuring the optical properties of defined aggregates of carotenoids in aqueous solution. The consistency in the excited state dynamics between previously reported work and the data described here suggests that the protein chassis does not alter the photophysical properties of the incorporated chromophores. The maquette proteins therefore provide a practical, modular system to further investigate the fundamental electronic processes in a range of carotenoids and also provide a unique material to enable biotechnological applications in the field of photovoltaics. To my knowledge, the data described here represent the first example of such a system, although many questions remain surrounding the exact mechanism of carotenoid incorporation, as well as the potential future applications of the complexes in primary research and biotechnological processes. Potential further developments in relation to these points are discussed below.

3.4.1 Further protein development

The three maquette variants described here differ by minor alterations in peptide sequence, which were largely immaterial for the desired carotenoid binding function. However, a more extensive protein redesign strategy may be used to better understand the current properties of the proteins and impart new functions on the carotenoid-maquettes.

Although the data described here suggests that aggregated AXT binds within the protein core, a number of questions remain surrounding the homogeneity of the samples and the exact orientation of the carotenoids within the hydrophobic cavity. These questions can be addressed through the determination of the carotenoid-maquette structure at a resolution sufficient to identify the orientation of the carotenoids, which would assign an accurate structural conformation to the AXT H-aggregate. Existing spectroscopic studies on solvent-derived AXT aggregates, to my knowledge, have only been modelled *in silico* (Olsina et al., 2012; Zajac et al., 2018), so no empirically derived correlation can be described between excited state dynamics and structurally defined carotenoid-carotenoid interactions. Elucidating such correlations is essential for understanding energy transfer processes, for example, where the atomic structures of light-harvesting complexes have been essential in modelling their activity (Roszak et al., 2003; Polivka and Frank, 2010). Structural determination of carotenoid-maquette complexes would not be possible through electron microscopy methods due to the small size of the protein (~17 kDa), so would require X-ray crystallography or NMR. However, NMR assignment is particularly challenging in maquette constructs due to the repetitive composition of the amino acid sequence. Both techniques would also require a high level of conformational homogeneity, which cannot be confirmed by the current data.

Addressing the issue of conformational heterogeneity requires optimising both the incorporation of the pigment, where the reaction conditions must ensure each complex has equal stoichiometry, and the structural uniqueness of the BT6 proteins themselves, which show a degree of disorder in the presence and absence of cofactor (personal communication, Prof. P. Leslie Dutton FRS). A redesign strategy, similar to that outlined in section 1.3.2, could be used to impart greater structural uniqueness on the BT6 scaffold which would improve the likelihood of crystallisation and diffraction quality (Gibney et al., 1999; Huang et al., 2003). Further experimentation could be used to identify conditions in which AXT is incorporated equally into each scaffold or to show how a more specific carotenoid binding interaction could be engineered into the maquettes.

The optical properties of the AXT-maquettes shown here are typical of carotenoids in a weakly-coupled H-aggregated state. As this was observed consistently for all three maquette structures, it raises the question of whether changes could be made to the protein that enforce only the inclusion of one carotenoid or facilitates the binding of other aggregate types. The inclusion of one carotenoid molecule would require reaction conditions that favoured partitioning into the maquette over aggregation and a protein structure that would sterically permit only a single cofactor (see figure 3.7). As many naturally occurring proteins bind single carotenoid molecules (Kerfeld et al., 2003), this function must be inherent to native peptide structures and, thus, achieving such a property in maquettes should be possible through structural redesign to impart stricter protein conformations or other helical arrangements (Woolfson, 2015).

In the work demonstrated here, carotenoids were incorporated into the maquette constructs without the attachment of any additional cofactors. It was shown in section 3.3.6 that the ligation of haem at either end of the cavity within the maquette complex was sufficient to inhibit carotenoid incorporation, most likely through steric restrictions. However, this finding does not disprove the ability of maquette constructs to bind other cofactors in conjunction with carotenoids. For example, the removal of one haem binding site in the BT6 structure may permit the incorporation of carotenoid at one end of the molecule and the site-specific coordination of a metallated chlorin at the other. Elucidating the requirements for such design features could see the inclusion of carotenoids in more established photo-active maquette structures (Farid et al., 2013; Moser et al., 2016), with the potential offering some degree of photoprotection to the incorporated tetrapyrroles.

3.4.2 Future fundamental investigations

The proof-of-concept work described here demonstrates the feasibility of the maquette approach in the study of excited state transitions in carotenoid aggregates. Although only AXT has been described in detail here, the incorporation of βC and ECH aggregates indicates the potential for an extensive array of chromophores and, in theory, any carotenoid may be permitted into the maquette with the potential to analyse the cognate optical properties. Generating the precise individual conditions to allow the formation of stable aggregation states in multiple carotenoids would be incredibly time consuming. Although the maquette approach seemingly requires H-aggregates to be formed at a rate appropriate to facilitate partitioning, the conditions necessary appear to be far less stringent than solution aggregates alone. Therefore, the electronic transitions of many carotenoids may be analysed in aqueous solution in maquettes, allowing comparison in identical environments. For example, this may allow relationships to be identified between the length and symmetry of conjugation in the carotenoid structure and its photophysical properties.

As explained at length in section 3.3.5, the incorporation of AXT is thought to depend on the rate and type of aggregate formed, which is apparently perturbed by conditions such as temperature, pH and salt concentration. The notion that AXT aggregation is greatly affected by the solution conditions gives a reasonable explanation of the observed maquette-binding characteristics, but can only be inferred from the absorption spectra following protein incorporation. A more insightful analysis of carotenoid aggregation kinetics could be conducted through time-resolved ground-state absorption spectroscopy to reveal the rate and type of the AXT aggregates formed under different conditions. This would allow the reaction conditions to be optimised to ensure the slow formation of H-aggregates in many different carotenoids, prior to maquette incorporation.

One potential drawback of the work here is that the carotenoids seem to reproducibly incorporate only one type of aggregated state. Although the attachment of J-aggregated carotenoids seems unlikely to occur given the current structural designs, the inclusion of other parallel H-type aggregates, such as the strongly-coupled AXT H-aggregate described by Musser et al. (2015) (figure 1.16), may be possible. Investigating the incorporation of other parallel aggregate types will give an indication of whether the protein is dictating a particular aggregate type onto the carotenoids or if aggregates are simply generated in solution and partition without any influence from the protein.

The maquette approach was originally conceived to generate a highly controlled system in which to assess the redox properties of bound haems (section 1.3). A similar approach may therefore be utilised in studying the transfer of excited states between carotenoids and other bound chromophores, such as Zn-tetrapyrroles. Assuming the incorporation of carotenoids and other site-specific cofactors would be achievable with the current experimental designs, the attachment of a variety of Zn-tetrapyrroles can be conducted with different carotenoids and energy transfer rates between each chromophore measured by TA spectroscopy or time-resolved fluorescence emission. This approach would elucidate the transfer kinetics between carotenoids and tetrapyrroles and delineate the fundamental rules governing this process, which would be of general interest in photosynthesis research.

3.4.3 Future applications

The semi-conducting properties of coupled organic polymers, such as carotenoids, has seen considerable interest in the last thirty years for a multitude of different applications (Mei et al., 2015). One potential application is the use of carotenoid aggregates as a SF-sensitising material to enhance the power-conversion efficiency of conventional photovoltaic devices (see section 1.6). In a recent review, Rao and Friend (2017) called for a more collaborative effort from a wider variety of research areas to address the challenges currently facing solar energy generation, with a particular desire to expand the number of available chromophores in the identification of an ideal SF-sensitising material.

Although in the nascent stages of development, the work presented here demonstrates the potential of carotenoid-maquettes in addressing these issues. As the maquettes appear to be able to bind several different carotenoids and hold them in a defined aggregation state, the number of available materials could, in theory, encompass far more naturally occurring carotenoids, of which there are approximately 600. The solubility of carotenoid-maquettes in aqueous solution also raises the possibility of using combinations of pigments in order to harvest more solar radiation.

Operational photovoltaic devices generate significant amounts of heat, so any additional material used would require high thermostability. As the melting transitions of maquettes is in excess of 90 °C, it may suggest that they could withstand the temperatures generated, although more extensive experimentation would be required to confirm this. The prototype maquettes discussed here improve the photostability of carotenoid aggregates to some degree, but drastic improvements would be required to give lifetimes appropriate for operational use. Such improvements may be achievable through protein redesign, optimisation of the buffering conditions, or encapsulation of maquette arrays in polymers.

Chapter 4

A thermostable protein matrix to study the excited state dynamics of tetracene

---

4.1 Motivation

The results of the previous chapter demonstrated that the incorporation of carotenoids into maquette proteins occurred through simple partitioning, due to the inherent hydrophobicity of the maquette interior and the chromophores used. As such binding does not rely on any specific interactions between the protein and cofactor, it was hypothesised that other hydrophobic molecules may also be able partition into the maquette cavity. This section investigates the incorporation of tetracene into maquette proteins.

Since the discovery of singlet fission (SF) in crystals of anthracene, the study and development of this process has largely focussed on the polyacenes and their derivatives. As detailed in chapter 3, the incorporation of carotenoid aggregates into maquette structures did not affect the photophysical properties of the bound chromophores and resulted in a more stable, practical system for optical analysis. The work in this chapter therefore aimed to impart such characteristics onto maquette-bound aggregates of a tetracene derivative, in order to determine the excited state dynamics of this material in aqueous solution for the first time. Elucidation of the excited state dynamics would allow comparison of the SF process between the maquette-bound chromophores and previously reported tetracene crystals. As interest in polyacenes is largely based on potential applications in photovoltaics, the use of tetracene-maquettes towards this aim will also be discussed.

4.2 Summary

The mechanism of SF was first proposed by Singh et al., (1965) to explain an unusual observation in crystalline anthracene and was later discovered in tetracene (Merrifield et al., 1969), carotenoids (Rademaker et al., 1980) and polyenes (Austin et al., 1989); with the mechanism theoretically described by Merrifield (1971). Interest in the process has recently been revived due to the potential applications of SF-sensitisers in photovoltaic (PV) devices (Dexter, 1979; Hanna and Nozik, 2006). Advances in the past decade have seen huge increases in the number of available SF materials and a more comprehensive understanding of the electronic transitions underlying the SF mechanism (reviewed in Smith and Michl, 2013; Rao and Friend, 2017; Casanova, 2018).

The most well-established SF materials derive from the polyacene structural family, including anthracene, tetracene and pentacene. Particular interest has been given to this class of molecules due to their long-lived triplet states and close energetic matching to silicon, making them an attractive material for PV applications (Rao and Friend, 2017). However, the polyacenes display three disadvantages that currently preclude their use in operational devices: (1) slow triplet formation (>10 ps), (2) morphology-dependent SF and (3) poor stability (Wilson et al., 2013; Piland and Bardeen, 2015). The experimental work in this chapter aimed to address the latter two disadvantages as outlined below.

The hydrophobic composition of the polyacenes results in poor solubility in most solvent systems and complete insolubility in water, with the π-conjugated structure also highly susceptible to photooxidative damage. As described by Anthony (2008), the addition of 'functional' chemical groups, such as triisopropylsilyl (TIPS), to the acene chromophore may be used to improve solubility. Here, a single TIPS functional group was covalently attached to tetracene to impart reasonable solubility in organic solvents, with a chemical structure hypothesised to be of appropriate size for incorporation into a maquette cavity (figure 4.1). It was shown that the addition of water to the tetracene-derivative in organic solvent was sufficient to impart an aggregated state, with optical properties similar to previously reported data on crystals of conventional tetracene (Kim et al., 2003). The inclusion of maquettes in the aqueous mixture resulted in the partitioning of the aggregated tetracene-derivative into the hydrophobic protein cavity, through a mechanism identical to that described for the incorporation of carotenoid aggregates (chapter 3). The resulting tetracene-protein complexes showed improved stability, compared to aggregates in solution alone, enabling optical characterisation by ground-state and transient absorption (TA) spectroscopy. A more ambitious aim was to determine whether the incorporation into maquettes would confer improved photostability on the chromophores. The absorption decay of tetracene aggregates in solution and in maquettes was recorded under constant illumination, with three- and five-fold increases in decay half-life observed for the maquette-bound samples.

As described by Piland and Bardeen (2015), the morphology of tetracene in the solid state (i.e. films or crystals) greatly affects the excited state dynamics of the chromophore, with the observed lifetime of the lowest singlet state ranging from 30 to 90 ps in films and 200 to 300 ps in crystals depending on the sample quality. A comprehensive description of the mechanism and kinetics of SF in polyacenes will require the use of a variety of different instruments, particularly in the analysis of the spin-behaviour of the system that cannot be described by optical techniques. As such, these measurements will require the production of multiple samples, which must be generated in a reproducible manner to obtain identical morphology and comparable kinetic lifetimes. Here, we aimed to develop a strategy that would result in the reproducible formation of tetracene aggregates with a defined structural conformation and size through maquette incorporation.

4.3 Experimental work

4.3.1 Synthesis of a soluble tetracene derivative

Considering the incorporation of tetracene into maquette proteins, two requirements were identified: (1) suitable solubility in an appropriate, water-miscible solvent that would not damage the maquettes and (2) a structural design to limit the potential for steric clashes within the protein core. As described above, functional TIPS groups may be covalently linked to acene moieties to improve solubility, typically in a perpendicular orientation with respect to the long axis of the molecule (figure 4.1A) (Payne et al., 2004). It was hypothesised that the perpendicular arrangement of the functional groups in bis-TIPS-tetracene would cause steric clashes within the protein cavity, so instead a tetracene derivative, herein referred to as elongated tetracene (eTet), was synthesised that contained a single TIPS group at one end of the molecule (figure 4.1B). The eTet molecule showed reasonable solubility in DMSO (0.2 mg mL^{-1}) following heating to 60 °C and sonication (1-2 h). Synthesis and purification of the eTet compound was conducted by the laboratory of Prof. John Anthony (University of Kentucky, USA).

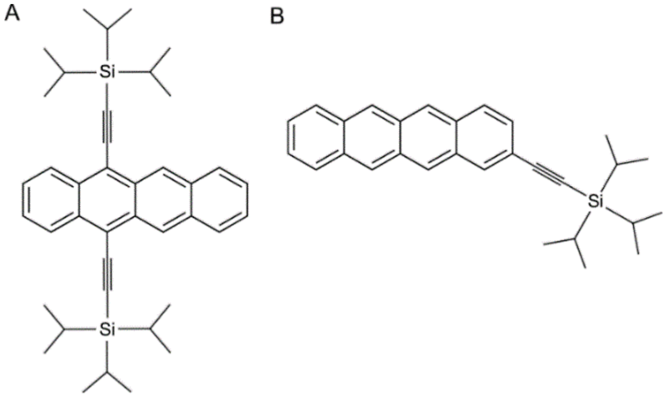


Figure 4.1 Chemical structures of functionalised tetracenes. (A) Structure of bis-TIPS-tetracene. **(B)** The ‘elongated’ tetracene (eTet) derivative used in this study.

The absorption spectrum of the eTet monomer in 100 % DMSO (figure 4.2B) was similar in character to conventional tetracene (figure 4.2C), but displayed red-shifted absorption maxima. Note that, unlike carotenoids, tetracene does not display pronounced solvatochrosim (Müller et al., 2007), so this absorption shift is attributable to the inclusion of a functional TIPS group (figure 4.2A). All buffers, solvents and labware used in this chapter were sparged with nitrogen prior to use, and eTet was kept strictly in the dark unless stated otherwise. All absorption spectra in this section are shown to highlight features within the visible range (350-700 nm); for description of the UV region, see appendix section 3.

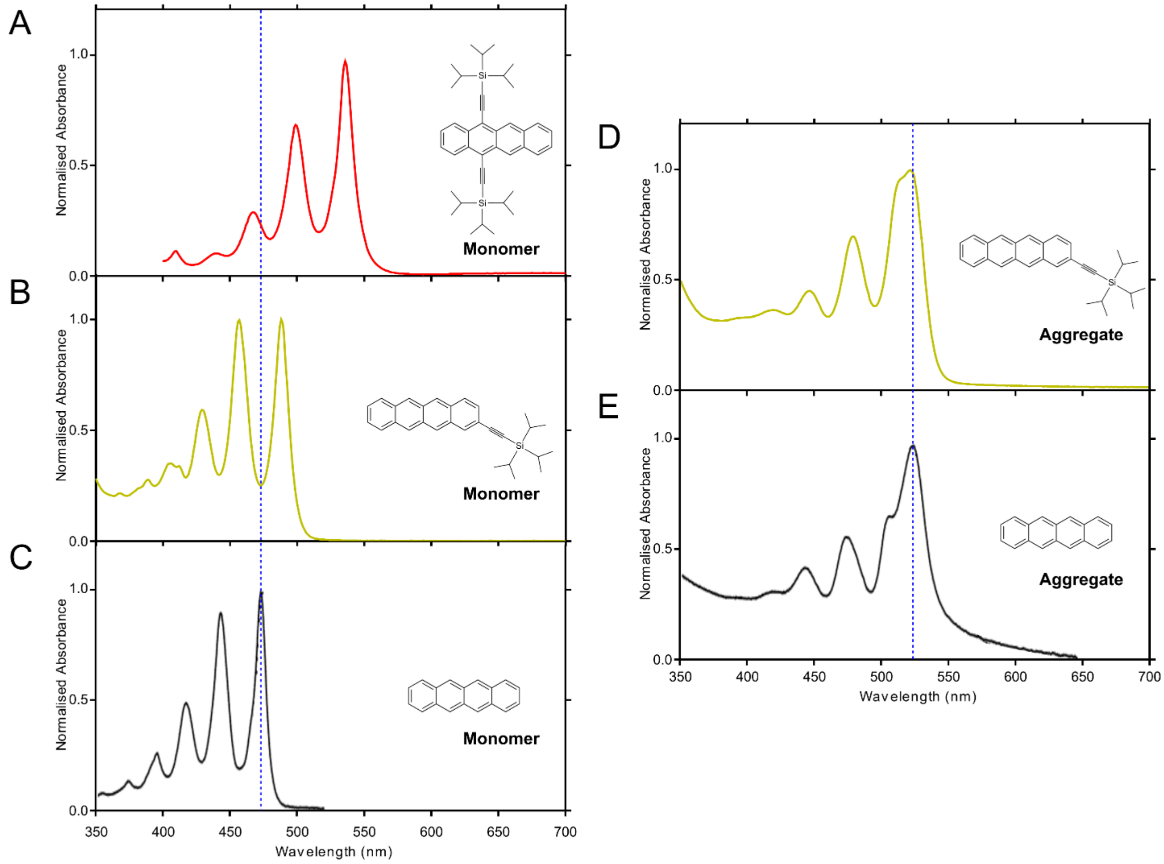


Figure 4.2 Absorption spectra of monomeric and aggregated tetracene samples. Absorption spectra of monomeric bis-TIPS-tetracene in chloroform **(A)**, eTet in DMSO **(B)** and tetracene in THF **(C)**. Absorption spectra of aggregated eTet **(D)** and tetracene **(E)** in largely aqueous solution. The spectrum displayed in panel A was reproduced with permission from Stern et al., (2015); the data in panels C and E were reproduced with permission from Kim et al., (2003). The dashed vertical lines indicate the wavelength of the lowest energy absorption feature in monomeric and aggregated tetracene. Note that tetracene is not greatly affected by solvent polarity (Müller et al., 2007) and the absorption shifts in panels A-C result from the addition of functional TIPS groups.

4.3.2 Generation of eTet solution aggregates

In chapter 3, the aggregation of carotenoids was shown to be susceptible to seemingly minor changes in the conditions of the solution, which in turn affected the propensity of the molecules to partition into the maquettes. It was also suggested that analysis of aggregation kinetics in solution could be used to pre-determine the optimal conditions for generating pigment complexes of appropriate size and conformation for insertion into the maquettes. This notion was investigated by looking at the formation of eTet aggregates in solution at different temperatures. As tetracene molecules have only been reported to adopt a 'herringbone' conformation in the solid state (figure 4.3A), kinetic analysis of aggregate formation would only require consideration of the aggregation rate and size, rather than the formation of any other aggregate types. A simple analysis of the aggregation rate was conducted by measuring the ground-state absorption spectra of eTet aggregates in solution after 30 min, comparing the intensity of the absorbance at 525 nm (A_{525}) with the amount of light-scatter, observed as undefined absorption features in the 600-700 nm region.

Aggregate formation was conducted at 4, 37, 50 and 80 °C, using a method similar to that described previously (Kim et al., 2003) (section 2.12.2). Briefly, stock concentrations of eTet were made up to 500 µM in 100 % DMSO and added in a 1:4 ratio (v/v) with pre-heated aqueous buffer (50 mM HEPES pH 8.5) in a volume of 4 mL. Solutions were mixed by inversion and incubated for 30 mins, after which absorption spectra were recorded (figure 4.3A). The results revealed that more light-scattering occurred at higher temperatures, with the greatest A_{525} intensity and lowest scatter occurring at 4 °C. This data suggests that larger eTet aggregates were present in solution at the higher temperatures and that the rate of aggregate formation occurs more rapidly with increasing temperature. Interestingly, the spectral characteristics of the eTet solution aggregates were largely similar to previously reported solution nano-crystals of non-functionalised tetracene (Kim et al., 2003) and do not show the same red-shifted absorption features observed in the monomeric samples (figure 4.2D,E). All eTet solution aggregates were filtered (0.2 µm pore size), with absorption spectroscopy confirming the complete removal of the chromophore in each sample (figure 4.3B).

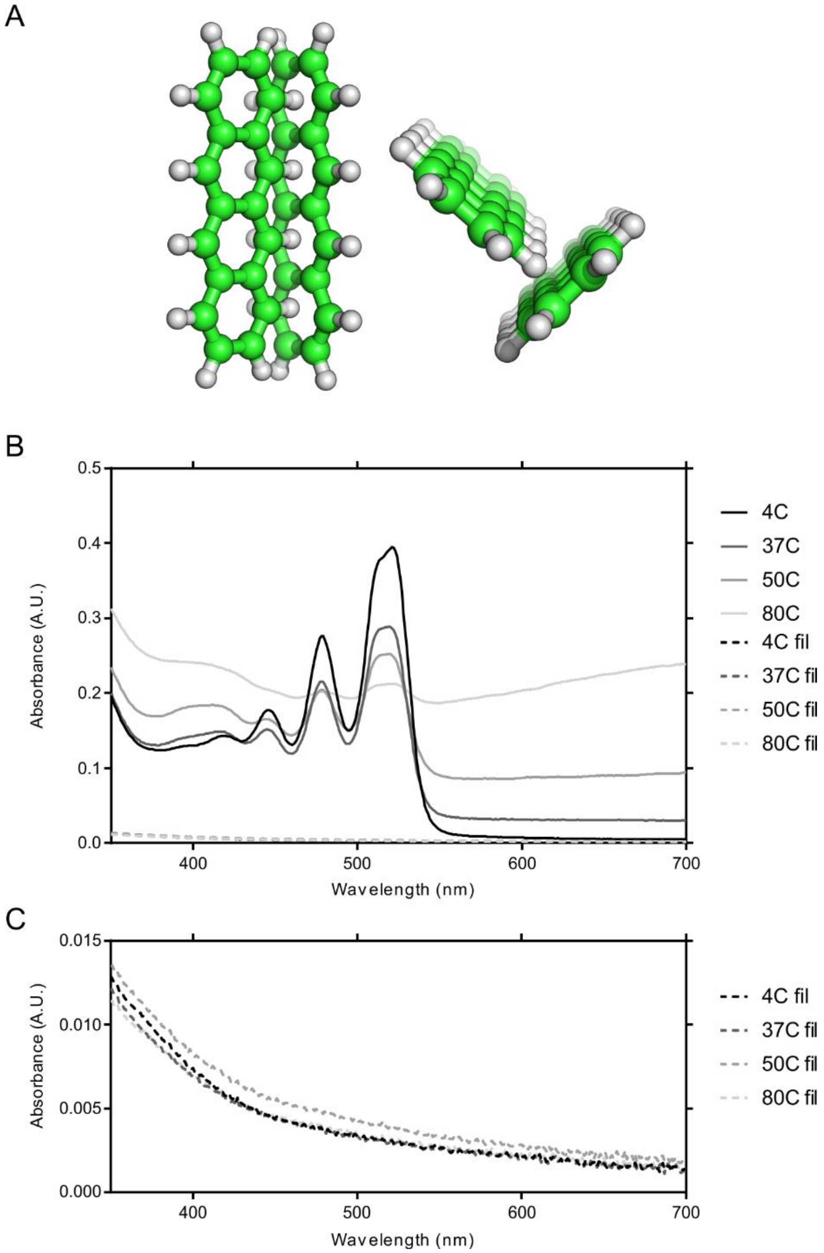


Figure 4.3 Temperature-dependence of eTet aggregate generation. (A) Structural representation of a tetracene dimer in a 'herringbone' conformation. **(B)** Ground-state absorption spectra of eTet aggregates generated at various temperatures before (solid lines) and after filtration (dashed lines). **(C)** Magnified display of data shown in (B).

4.3.3 Optimisation of eTet partitioning into maquette protein scaffolds

The experiment in section 4.3.2 was repeated with each of the BT6, BT64A and BT64F maquettes (previously described in chapter 3) in the aqueous solution, to determine if a correlation could be observed between the pre-determined rate of eTet aggregation and the efficiency of protein incorporation. The incorporation method involved the mixing of each maquette with a ten-fold excess of eTet at the various temperatures. Following a 30 min incubation, the solutions were filtered and diluted five-fold in aqueous buffer (50 mM HEPES pH 8.5) before anion exchange chromatography (AEC) on a DEAE Sepharose column. The proteins were eluted from the column by the addition of elution buffer (50 mM HEPES pH 8.5, 500 mM NaCl) and concentrated to exactly 1 mL. Normalisation to tryptophan absorbance (A_{280}) is not feasible for the work presented in this chapter as the UV absorption of eTet results in values that are often in excess of the sensitivity of the spectrophotometer and spectra that are highly variable in A_{280} intensity depending on the presence of relatively small amounts of light scattering (figure A3).

The relationship between the apparent rate of aggregation and maquette incorporation was not coherent for all samples; while the efficiency of partitioning into the BT64A maquette followed the predicted trend (4 > 37 > 50 > 80 °C), this was not upheld by either the BT6 or BT64F variant (figure 4.4 A-C). However, consistency was observed at 80 °C, with each maquette displaying poor eTet binding at this temperature. The incorporation of eTet into BT64A was also higher at 4, 37 and 50 °C than at any temperature in either of the other two maquettes. Interestingly, this observation is unlike the results of chapter 3, where carotenoids were incorporated equally across each variant. However, the observation that eTet aggregates were maintained in the protein samples following filtration and AEC suggests that the interaction is similar to that observed for carotenoid partitioning, in that the size of the eTet aggregate must be limited by the protein in order to pass through filtration, most likely by encapsulation within the maquette cavity.

So far, the incorporation of eTet aggregates has been conducted at pH 8.5 as this was found to be optimal for carotenoid binding; however, investigation of the optimal pH conditions for eTet partitioning was required to determine if more efficient chromophore binding was achievable. The protocol described above was repeated at various pH values (pH 6, 7, 8 and 9) and absorption spectra were recorded following filtration and AEC. The results showed that optimal eTet incorporation occurred at pH 8, with more absorption scatter observed at pH 9 and no observable binding at pH 6 or pH 7 (figure 4.4D). All subsequent experiments were conducted at pH 8 unless stated otherwise. The incorporation of eTet into each maquette was conducted at 4 °C with the solution at pH 8 and the absorption spectra of each sample was recorded to enable comparison of each variant under the optimised conditions (figure 4.4E).

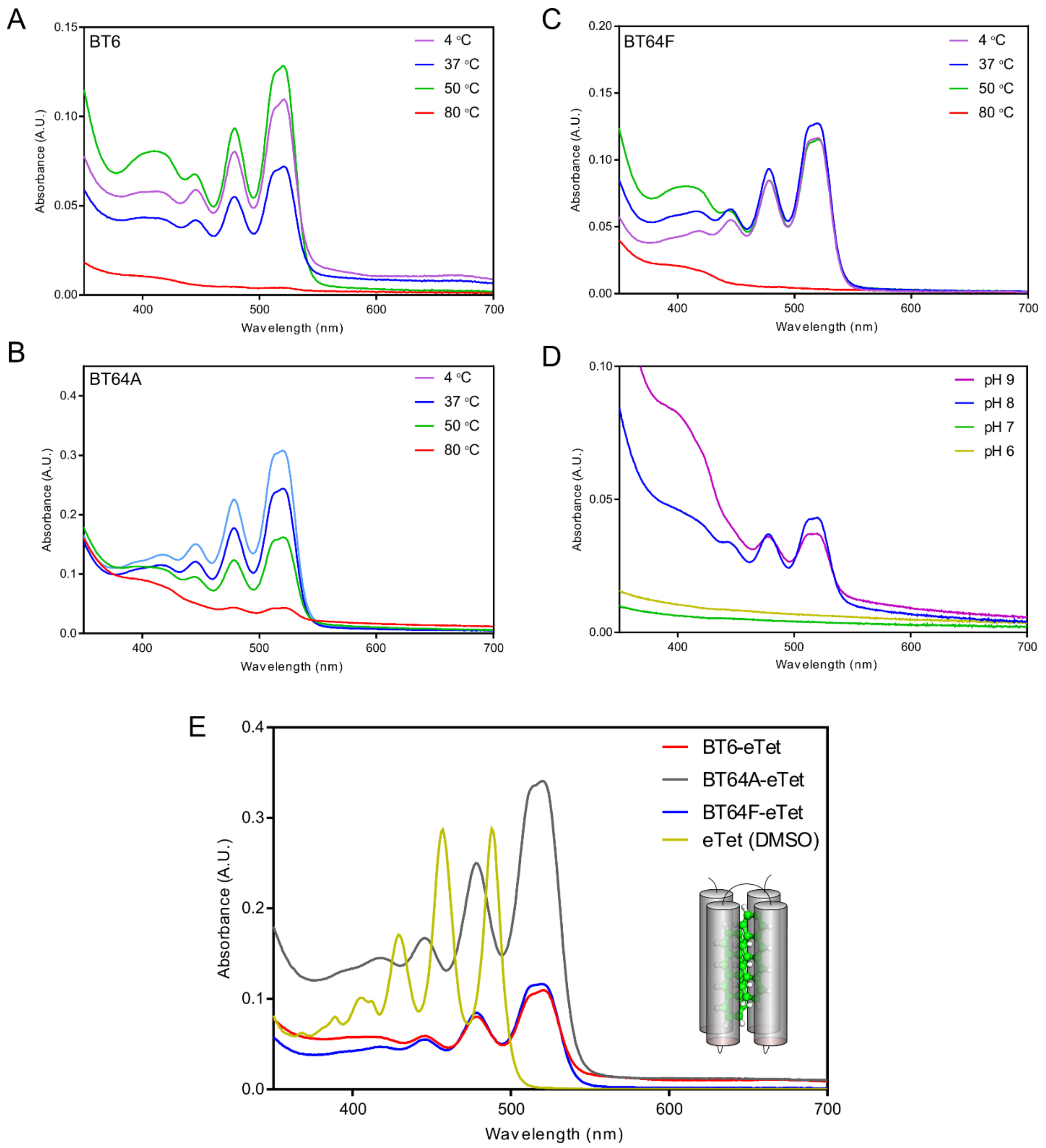


Figure 4.4 Optimisation of eTet partitioning. Absorption spectra of BT6 **(A)**, BT64A **(B)** and **(C)** BT64F following incorporation of eTet at various temperatures. **(D)** Absorption spectra of BT64A-eTet with the incorporation of pigment conducted at various different solution pH values. **(E)** Absorption spectra of monomeric eTet (yellow); BT6 (red), BT64A (grey) and BT64F (blue) following incorporation of eTet. A schematic diagram of a tetracene-maquette is shown (inset). All samples are shown at a volume of exactly 1 mL, following filtration and AEC.

4.3.4 Photostability of eTet aggregates in maquettes

Although the polyacenes constitute the most well-established materials for the study of singlet fission, the instability of the chromophore causes significant issues in the curation of data and precludes their use in functional devices (Anthony, 2008). In order to assess the effect maquette proteins had on the photostability of the incorporated eTet aggregates, photo-induced decay was measured as the reduction in absorbance over time, according to the experimental procedure in section 2.10.5. Briefly, samples were generated to give A_{525} values of approximately 0.5, then sparged with nitrogen gas and sealed. After an initial measurement in the dark, the samples were illuminated with white-light (200 µM photons m^{-2} s^{-1}) and full absorption spectra were measured in 2.5 min intervals.

The light-induced absorption changes of eTet in solution revealed that the decay of the spectral characteristics was coupled to the emergence of light scattering (figure 4.5A), which could be observed by eye as a very pale yellow precipitate after 90 min. To enable comparison to other samples, the absorption spectra was scatter-corrected by normalisation to the intensity of the A_{600} signal. Interestingly, after repeating the same experiment with the maquette-bound samples, no significant absorption scatter was observed (figure 4.5B,C). This observation provides further evidence for the notion that pigments are sequestered within the maquette cavity; in solution alone the eTet aggregate size increases to a level where they are visible by eye and as this does not happen in either of the maquette samples, it suggests that the maquettes prevent this precipitate formation by encapsulation. It is worth noting that the absorption decay of the BT64F-eTet sample began from a lower starting value with A_{525} less than 0.15 (figure 4.5C) due to issues in generating a sample with sufficient absorbance. Despite using four-fold more total reagent and concentrating to the same volume as BT64A, the absorbance of the BT64F-eTet sample remained lower than the 0.5 value required; further experimentation is required to explain this issue.

Analysis of the A_{525} intensity over time was use to compare the photo-stability of eTet within each sample; for all decay plots each data point was normalised by subtraction of the lowest absorbance value and division by the maximum absorbance value. For the eTet solution aggregates, obtaining such data required the use of the scatter-corrected absorbance values, however, the necessity for such normalisation in the eTet-maquette sample was questionable as neither sample displayed any significant light-scattering. The scatter-corrected and non-normalised A_{525} values for BT64A and BT64F are shown in figure 4.6A and figure 4.6B, respectively. For the BT64A-eTet sample, both the corrected and non-corrected data series resulted in equivalent decay plots. However, the scatter-corrected data for the BT64F-eTet maquette showed far larger variation compared with the non-corrected series, which is most likely due to the low A_{525} intensity relative to the background A_{600} signal of the sample. For this reason, only non-corrected data were used for the maquette sample comparison. The photo-induced absorbance decay of the corrected eTet solution aggregate and the two non-corrected eTet-maquettes are displayed in figure 4.6C. A one-phase exponential decay was fit to all datapoints according to the protocol in section 2.14.2. Following incorporation into both proteins, eTet showed significantly improved photostability compared to aggregates in solution alone, with approximately three-fold and 5-fold increases in the decay half-life for the BT64A and BT64F variants, respectively.

The difference in the photostability between the two eTet-maquette constructs could arise from the greater thermal stability of the BT64F maquette (see section 3.3.2) or due to the fact that the BT64F-eTet sample began the decay from a lower absorbance intensity. As described in chapter 3, the pH of the buffering solution greatly affected the thermal stability of the maquette proteins, with the BT64A maquette showing a melting transition midpoint (T_m) of 49 °C at pH 9 and a T_m in excess of 90 °C at pH 6. It was hypothesised that if the thermal stability of the maquette structure had any effect on the photostability of the incorporated eTet aggregate, then the BT64A-eTet sample would show enhanced photostability at lower pH values. An important consideration in generating such a sample was to determine at what stage during the incorporation and purification process to reduce the pH in order to maintain optimal eTet binding. The initial reconstitution process must occur at pH 8 as incorporation was inhibited at pH 6 (figure 4.4D), however, the reduction of the pH could occur in the initial dilution of the sample prior to AEC (method 1), in the elution from the column (method 2) or by buffer exchange following elution (method 3). All pH exchange methods are described in table 4.1 and the results of the incorporations shown by ground-state absorption in figure 4.7A, with optimal binding observed via method 2. The low pH sample was subsequently subjected to the photostability assessment described above, with no change observed in the photostability of the sample (figure 4.7B).

Taken together, the results of these experiments suggest that the thermal stability of the protein structures has little bearing on the photostability of the incorporated chromophore and that the differences in the half-life of the BT64A and BT64F samples was likely to be caused by the differences in the intensity of their starting absorption values; however, from the current data it is unclear why the concentration of bound chromophore would have any effect on the rate of decay. As no significant aggregation formed in the maquette samples, the decay process of eTet in maquettes must be different than in solution, but perhaps the maquettes simply act to prevent the formation of aggregates and the photostability of the molecules remains the same.

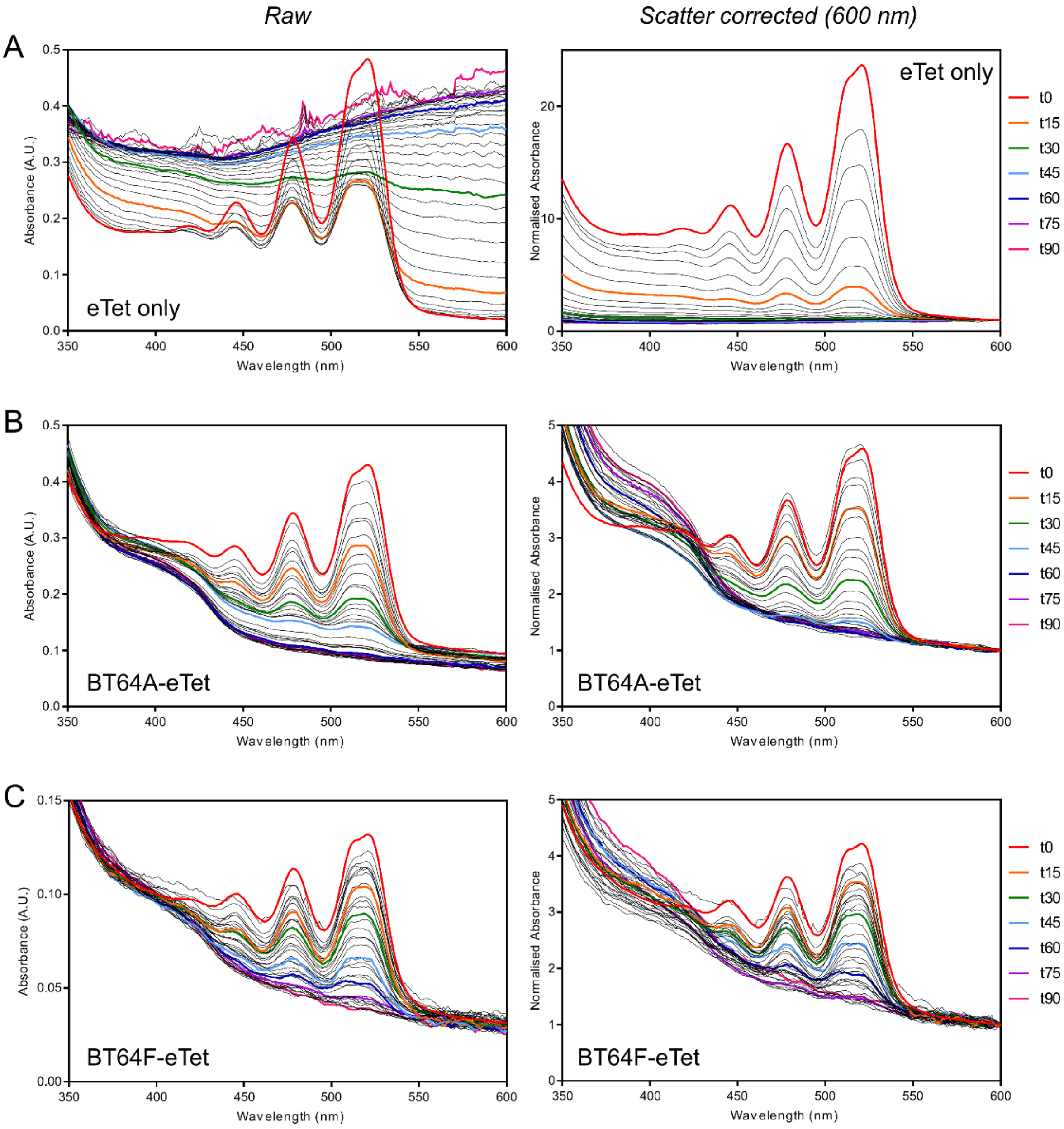


Figure 4.5 Absorption spectra of eTet aggregates during light-induced decay. The change in the absorption characteristics of eTet aggregates in solution **(A)** is compared with BT64A-eTet **(B)** and BT64F-eTet **(C)**. Panels on the left show the uncorrected recorded data and the panels on the right show the same data after normalisation to the absorbance at 600 nm. Sample absorption was recorded in 2.5 min intervals with each timepoint shown in thin black lines; every sixth timepoint is highlighted in thick coloured lines.

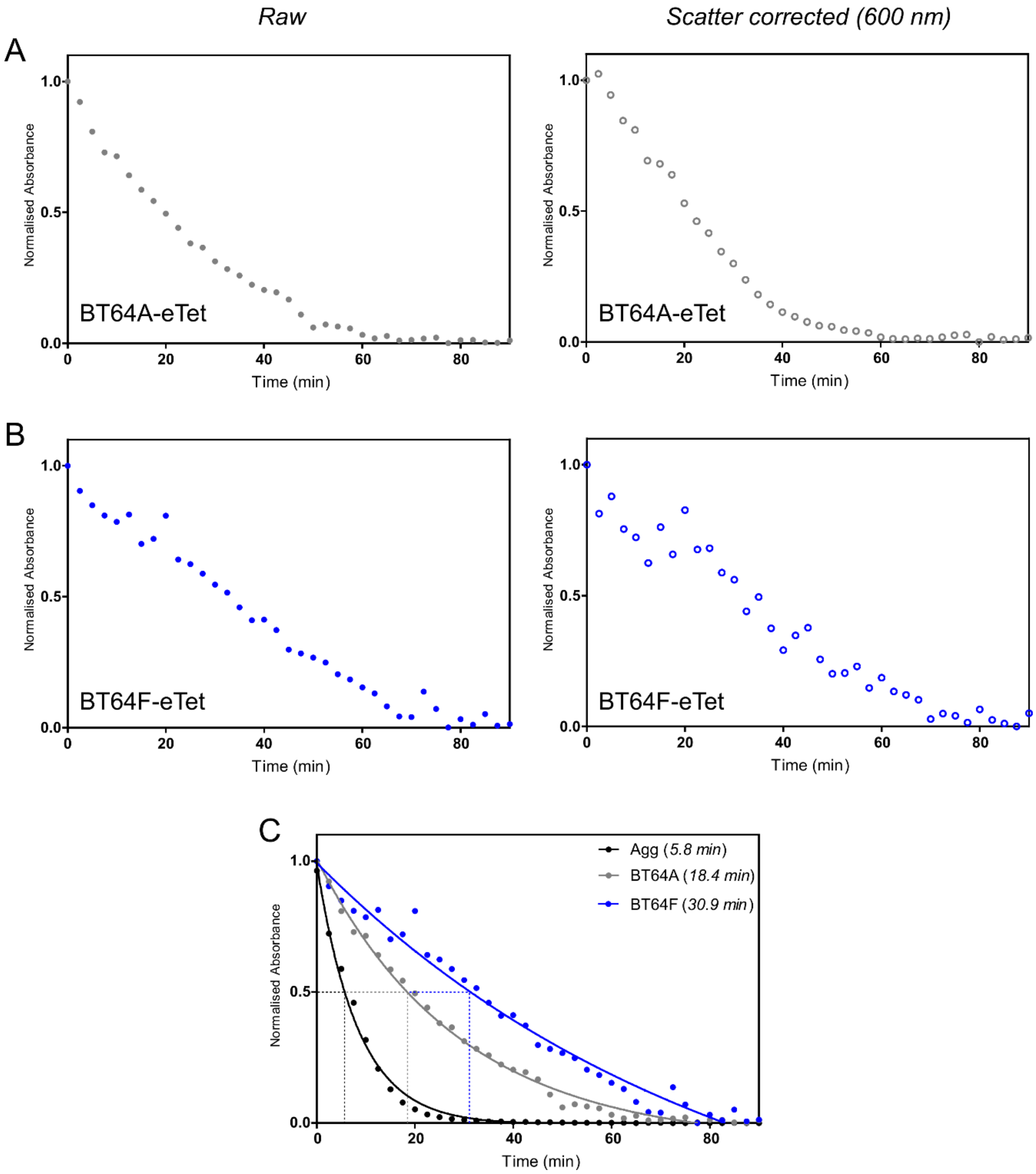


Figure 4.6 Photo-induced decay of eTet aggregates. The photostability of eTet was assessed by monitoring the intensity of absorbance at 525 nm. The A_{525} values for BT64A-eTet **(A)** and BT64F-eTet **(B)** are shown prior to (left, closed circles) and following normalisation to A_{600} (right, open circles). **(C)** The absorbance decay of eTet in solution (black), BT64A (grey) and BT64F (blue). Due to the high amount of light scattering in the solution aggregate (figure 4.5A), only the scatter corrected absorbance values are shown for this sample; for BT64A and BT64F, absorbance values are shown without correction to A_{600}. All absorbance values were described by a non-linear decay (coloured lines) and the half-life of each sample is indicated in parentheses.

Table 4.1. Buffer exchange methodologies.

	Buffer pH		
	Method 1	Method 2	Method 3
Reconstitution	8.0	8.0	8.0
Dilution	8.0	8.0	6.0
Elution	8.0	6.0	6.0
Buffer exchange	6.0	6.0	6.0

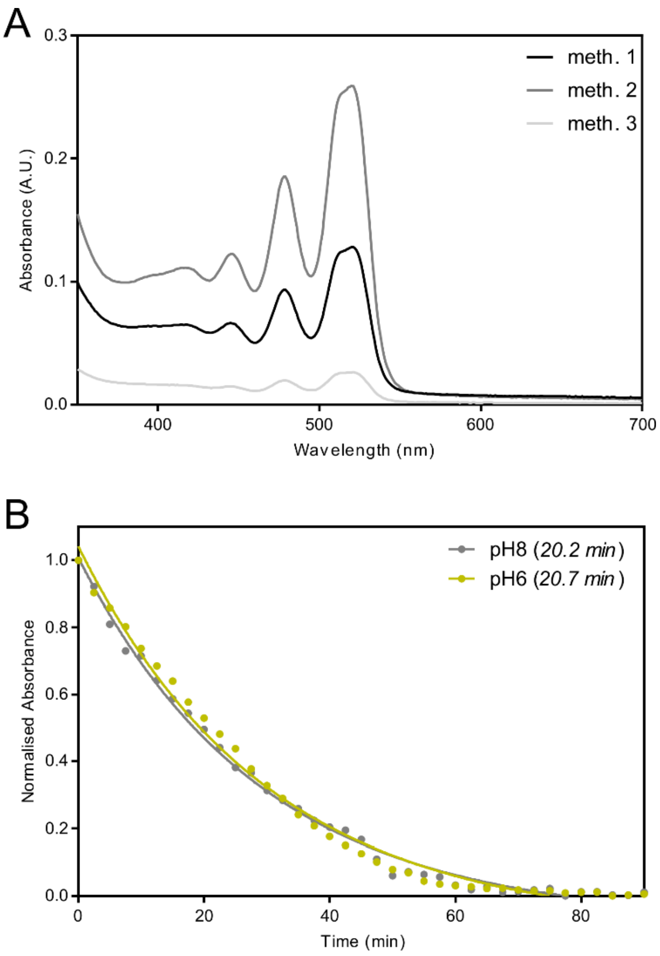


Figure 4.7 Comparison of BT64A-eTet photostability in different solution pH values. (A) Absorption spectra of BT64A-eTet samples generated according to the different methodologies described in table 4.1. **(B)** The decay of A_{525} signal in BT64A-eTet samples at pH 8 (grey) and pH 6 (yellow).

4.3.5 Stability of tetracene maquettes in aqueous solution

SF in polyacenes has typically been investigated in the solid state, though the dynamics of the electronic transitions have been shown to be dependent on the morphology of the film or crystal generated (Piland and Bardeen, 2015). Thus, for accurate determination of excited state kinetics, measurements would have to be conducted on samples with identical morphologies. One way this may be achieved is to generate a system where the stability of the chromophore aggregate is sufficient to maintain morphology following long-term storage and repeated measurements. Here, the absorption of BT64A-eTet was measured following cycles of cryogenic freezing (-196 °C) and thawing (FT) under normal atmospheric conditions in aqueous buffer (50 mM HEPES pH 6), with little change in spectral character following seven FT cycles (figure 4.8). Due to time-constraints, the absorption of BT64A-eTet following long-term storage could not be measured, however, as the spectral characteristics did not greatly alter during the seven FT cycles it may be assumed that the eTet-maquettes behave in a similar manner to the carotenoid-maquettes in section 3.3.9, where samples showed little change in absorption following six months of storage at -80 °C (figure 3.11).

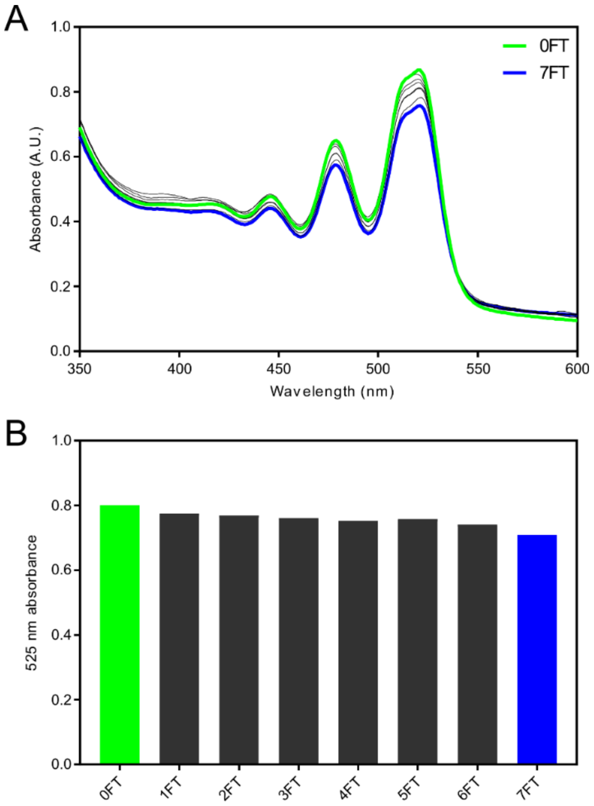


Figure 4.8 Cryogenic freezing of BT64A-eTet. (A) Absorption spectra of BT64A-eTet prior to cryogenic freezing (green) and following successive rounds of freeze thaw (black), up to seven cycles (7FT; blue). **(B)** A_{525} intensity of samples from panel (A).

4.3.6 Transient absorption spectroscopy of eTet aggregates in a protein matrix

Ultrafast transient absorption (TA) spectroscopy was used to elucidate the optical characteristics and transition kinetics of the excited states in eTet. Unlike the AXT aggregates in chapter 3, the optical properties of aggregated tetracene in solution remain poorly characterised, most likely due to issues associated with achieving homogenous sample morphology. An additional complication is that the covalent inclusion of the functional TIPS-group in eTet may alter the excited state transitions in the monomer, compared with conventional tetracene. Thus, the work in this section aimed to identify the optical characteristics of the excited states in eTet and offer a kinetic description of the electronic transitions in a monomeric and aggregated morphology.

Ultrafast TA spectroscopy was conducted in collaboration with Daniel Polak (University of Sheffield, UK) and performed essentially as described in section 2.10.4. Samples were excited with either 400 or 500 nm laser illumination and absorption spectra were gathered, with the instrument response estimated to be 120 fs. The absorption spectra of monomeric and aggregated eTet is shown after a 1, 100, 1000 and 2000 ps delay following excitation, in figure 4.9 (left panels). The lifetime kinetics of several spectral features are shown (figure 4.9, right panels) as the average absorption of 10 nm intervals, with each interval indicated by the coloured marks above each TA spectrum (figure 4.9, left panels). Note that the intervals selected were not necessarily at the maxima of each absorption peak, but instead selected to give minimal interference from other features. From the available data, the identity of the excited states and an absolute description of the electronic transitions cannot be categorically determined. However, the information that may be inferred from each dataset and a proposed model are described below, with the recognition that the data is insufficiently complete to confirm the proposition.

Excited-state absorption spectra were recorded following 500 nm excitation, which is at the band-edge of monomeric eTet absorption in toluene. The results showed intense absorption in the 400-500 and 1100-1300 nm regions, as well as spectral features at 600 and 950 nm (figure 4.9A). Note that transient absorption measurements in the infra-red region were of limited sensitivity due to issues with the experimental set-up. Kinetic analysis of the average absorption from each of these regions showed that each spectral feature decayed within approximately the same time frame, suggesting that all of these characteristics arise from the same state. Based on theoretical calculations, transient absorption spectroscopy and time-resolved photo luminescence, Burdett et al., (2010) assign the 1200 nm feature to the S_1 energy state (figure 4.9A). Thus, we tentatively assign the 1200 nm spectral feature, as well as all other features with similar decay kinetics to the S_1 state.

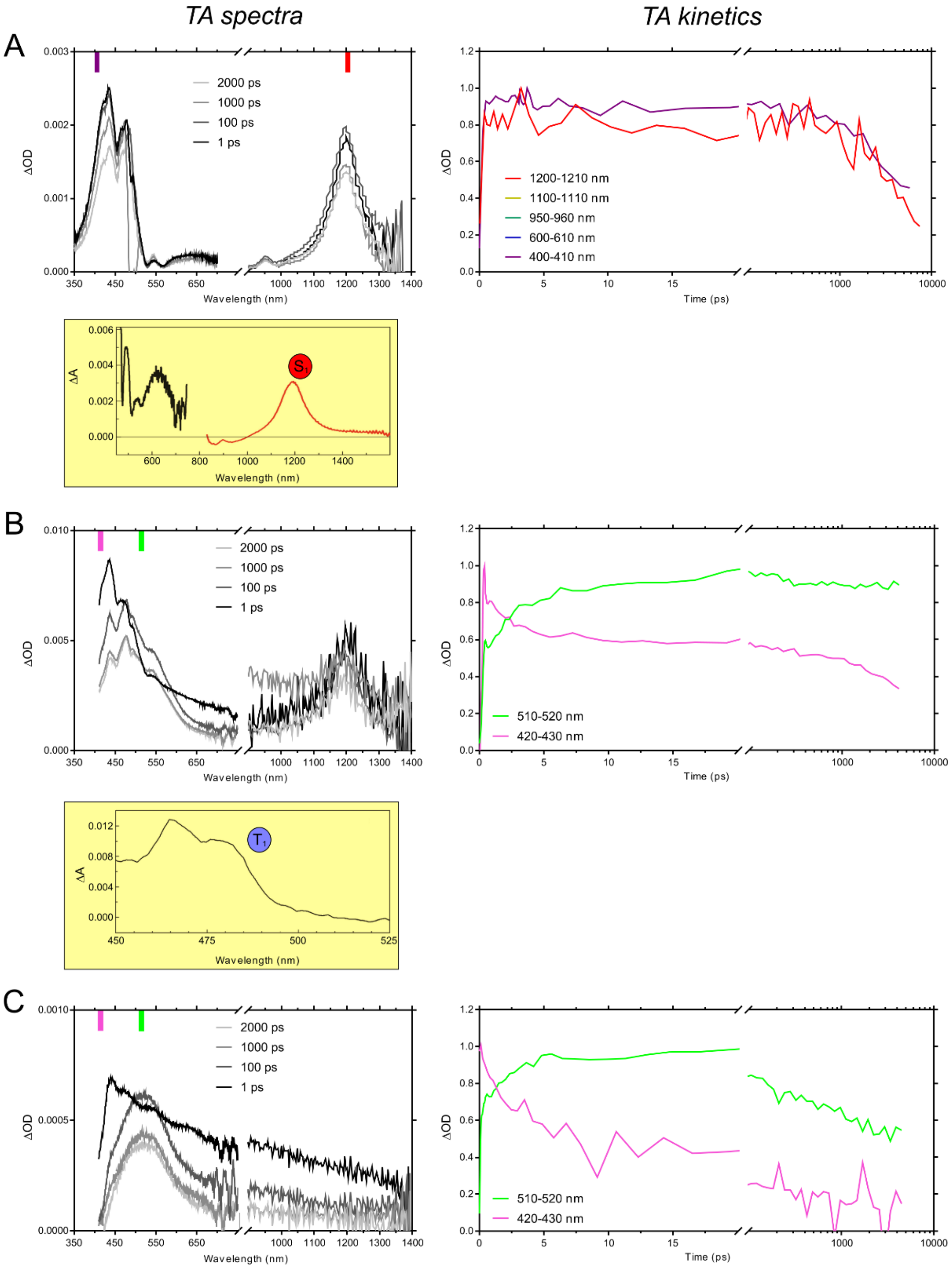


Figure 4.9 Transient absorption spectroscopy of eTet. Monomeric eTet in toluene after excitation at 500 nm **(A)** and 400 nm **(B)**. **(C)** BT64A-eTet in 50 mM HEPES (pH 6). For each sample, full PIA spectra are shown at various timepoints (left) and the lifetime kinetics of specific spectral regions are analysed (right). The spectral region used for the lifetime is indicated in colour at the top of the TA spectra. Transient absorption spectra that were reproduced, with permission, from Burdett et al., (2010) are highlighted in yellow. Note that absorption detection in the infra-red region for B and C was of limited sensitivity due to instrument constraints.

Excitation higher into the absorption band of eTet at 400 nm resulted in different TA characteristics compared with those observed at 500 nm (figure4.9B), displaying shifts in the relative intensity of the peaks in the 400-550 nm region over time. Additional spectral changes were observed for the two different excitation wavelengths; however, as the reasons for such differences are unclear from the current data, these will not be discussed further here. Following excitation at 400 nm, the predominant absorption peak after a 1 ps delay is at ~440 nm with a smaller 'shoulder' at 460 nm (figure 4.9B). After 100 ps the relative intensities of these two peaks alters, so that the most intense absorption comes from the 460 nm peak. Note that a similar PIA feature is assigned to the T_1 state by Burdett et al., (2010) (figure 4.9B).

Kinetic analysis of the 440 nm feature was taken between 420-430 nm and the 460 nm feature between 510-520 nm. The absorption of the 420-430 nm feature decays to approximately 60 % of its highest intensity within 10 ps, then follows similar kinetics to those seen in the putative S_1 state. Conversely, the 510-520 nm feature increases in absorption, reaching ~90 % of its highest intensity within 10 ps, and is maintained within the measurement lifetime of the instrument. The emergence of absorption over time may possibly be due to an excited state transition in the chromophore; in this case, the second excited state that is has a similar spectrum to the putative T_1 state also appears to have a longer lifetime compared to the S_1 state.

The current model for this data proposes that the 500 nm excitation results in a transition to the lowest vibrational level in the S_1 state, which is maintained for 1 ns and begins to decay within 10 ns (figure 4.10A). The higher-energy 400 nm excitation is suggested to cause a transition to a high vibrational level in the S_1 state, which may relax to the lowest vibrational level and decay in an identical manner to the 500 nm excitation, but may also undergo some intersystem crossing to the triplet manifold (figure 4.10B). From the current data it is not clear whether the high-energy excitation causes transition to a high vibrational level in the S_1 state or to another state within the singlet manifold (e.g. S_2); for simplicity, only transitions involving the S_1 state will be considered.

The justification for the current hypothesis comes from the fact that the 440 nm feature decays rapidly by ~40 % and then behaves identically to the lifetimes observed after 500 nm excitation, possibly resulting from 60 % of the population remaining in the S_1 state and the remainder transitioning to another state. The emergence of another excited state that does not follow the kinetics of the S_1 state, instead displaying an extended lifetime, may be indicative of triplet character, as transition to the ground state is optically forbidden from the triplet manifold. The absorption characteristics arising from the putative eTet triplet state here are also in agreement with the work of Burdett et al., (2010) that reported a triplet state following a 20 ns delay.

BT64A-eTet samples in 50 mM HEPES (pH 6) were concentrated to 500 μL giving A_{525} values of approximately 1, sparged with nitrogen gas for 10 min and sealed within a 1 mm pathlength quartz cuvette for laser excitation at 400 nm. The resulting TA spectra showed broadening of the absorption peaks, which is characteristic of data gathered in maquettes (see chapter 3), as well as increased noise, arising from the reduced absorption intensity of the samples (figure 4.9C). The absorption peak at 440 nm corresponding to the S_1 state is observed following a 1 ps delay, but completely decays after 100 ps (figure 4.9C, left panel); importantly, this is unlike the decay observed for the eTet monomer after 400 nm excitation, where the S_1 peak remained present after 2 ns. In the eTet-maquette sample, the decay of the 420-430 nm feature is coupled the emergence of an excited state with a broad absorption character in a similar region (500-550 nm) to the putative triplet state in the monomer. Interestingly, the 510-520 nm absorption state decays more rapidly than the triplet state in the monomer (discussed in section 4.4.4). Note that the average absorption of the 420-430 nm feature drops to approximately 40 % within 10 ps and then follows a decay similar to the 510-520 nm feature. This suggests that the absorption of the 420-430 nm feature, after ~5 ps, results from the overlap of the broad triplet peak, with the overall S_1 character completely depleted in this time frame. This can be confirmed on analysis of the full TA spectra, where the peak at 425 nm is completely lost after 100 ps (figure 4.9C).

In summary, the eTet monomer displays altered electronic transitions depending on the energy of excitation. Lower-energy excitation causes a transition to the lowest vibrational level in the S_1 state, which decays directly back to the ground state (the mechanism of this will be discussed in section 4.4.4), while higher-energy excitation induces a transition to a high vibrational level in the S_1 state, which may undergo intersystem crossing to the triplet manifold or vibrationally relax and decay through an identical mechanism to the putative S_1 state in the monomer. High-energy excitation of BT64A-eTet resulted in complete conversion of the S_1 state into the putative triplet state that decayed more rapidly than in the monomer. Such kinetic behaviour is suggestive of singlet fission in tetracene where complete transition into the triplet manifold has been shown to occur on the timescale of 30-90 ps in thin films (Piland and Bardeen, 2015). As mentioned above, the results displayed in this section cannot fully describe the electronic transitions of eTet in either the monomeric or aggregated state, requiring the further experimentation discussed in section 4.4.4.

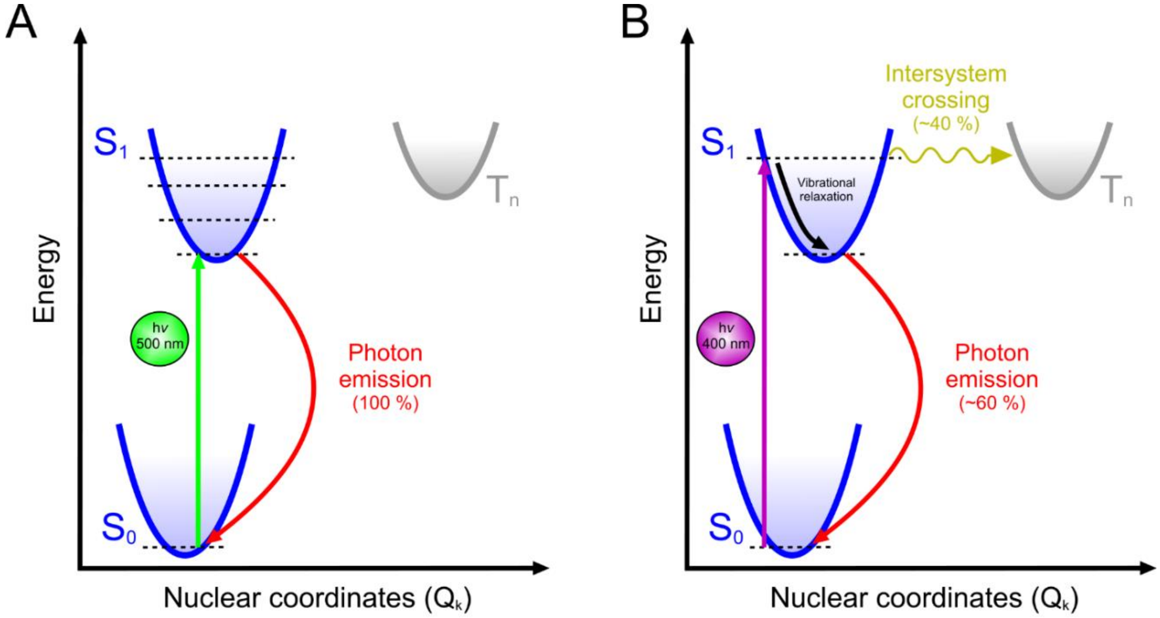


Figure 4.10 Schematic to show the proposed energy transitions in monomeric eTet. (A) Excitation at 500 nm (green arrow) causes a transition from the ground state (S_0) to the lowest vibrational level in the S_1 state, which returns to the ground state by photon emission. **(B)** Excitation at 400 nm (purple arrow) induces a transition to a high vibrational level in the S_1 state, with sufficient energy to permit some intersystem crossing (orange arrow) to the triplet manifold (T_n). The energy and nuclear coordinates of the states shown here are arbitrary and shown for clarity. The proportion of energy states entering each decay pathway is estimated from the TA kinetic data, but may not reflect the true energy-loss mechanism.

4.4 Discussion

The experimental work in this chapter was founded on the results of chapter 3, utilising the same three artificial protein variants and chromophore incorporation method to explore the potential use of maquettes to analyse polyacenes in solution. As the data described above is largely in line with the work on carotenoid-maquettes, many of the same issues previously discussed also arise here, particularly those concerning stoichiometry, structural conformation and developments required for device applications. To avoid repetition, the discussion below will largely focus on issues specific to acene incorporation; for information on issues common to the incorporation of both chromophores, the reader is directed to section 3.4.

4.4.1 Incorporation of other polyacenes

To determine whether a polyacene derivative would partition into the hydrophobic cavity of a maquette protein, a number of chemical and structural factors were considered from the outset to give the highest probability of incorporation. These considerations ensured that the candidate molecule, eTet, had sufficient solubility to be of practical use and a chemical structure that minimised the likelihood of steric clashes within the protein. Given the eTet incorporation demonstrated above, it would be of interest to determine if other tetracene derivatives with different structural conformations, such as bis-TIPS-tetracene (figure 4.1A), would be incorporated into the protein cavity or whether the functional groups sterically prevent insertion as predicted. Identification of chemical structures that are not permitted into the protein, would allow some general rules surrounding cofactor conformation to be delineated.

Functional groups may be added in either perpendicular or parallel arrangements to other polyacenes, such as anthracene, pentacene and hexacene (Anthony, 2008), with the conditions required for incorporation identifiable through the experimental protocol described here. This would allow acenes of various lengths with different functional additions to be sequestered in an identical environment, enabling accurate comparison of the photophysical properties.

An important consideration for the data discussed above is the effect of the functional group on the excited state dynamics of the chromophore. Therefore, for accurate comparison of functionalised and conventional acenes, the insertion of conventional tetracene into the maquette cavity is required. As the solubilisation of non-functionalised polyacenes generally requires the use of strong solvents, combination with proteins would likely cause damage to the maquette structures; although, this assumption would be worth investigating given the unusual composition of the artificial peptides. If strong, water-miscible solvents are shown to irreversibly damage the maquette proteins then a two-phase solvation system could be envisaged; maquette proteins could be solubilised at high concentration in a buffered aqueous lower-phase with acenes in a solvent upper phase, which can subsequently be evaporated, potentially forcing the acene molecules into the lower phase and permitting partitioning into the maquettes.

4.4.2 Aggregation and decay kinetics

From the observations made in chapter 3, it was suggested that the rate of aggregate formation in solution would pre-determine the ability of cofactors to partition into the maquettes (figure 3.7). This notion was investigated in the incorporation of eTet, where the absorption characteristics of the pigment in solution alone were used as an estimate of the aggregation rate after a defined time period. A correlation was observed between the temperature and the rate of aggregation, with a reduced aggregation rate leading to greater BT64A incorporation. This experiment only considered the temperature of the reaction but could be employed to analyse other factors such as pH, buffer conditions and pigment concentrations, possibly in a 96-well plate format, allowing many different conditions to be assessed. More frequent absorption recordings over longer periods of time would also enable more detailed description of the aggregation kinetics. Such developments may identify conditions in which eTet more readily partitions into the maquettes and could be applied to other polyacenes before trialling incorporation into the proteins.

A significant complication in the photo-induced decay of eTet in solution was the formation of precipitate during the experiment. From the current data it is not possible to say whether the observed rate of aggregate formation was solely due to the temperature of the solution or whether the incident light increased the aggregation rate. This is further complicated given the photochromic nature of tetracene, where incident light can induce covalent dimerization, forming a so-called ‘butterfly’ compound (figure 4.11) (Bouas-Laurent and Dürr, 2001; Anthony, 2008). Disentangling the independent processes of photooxidative decay and aggregation may be conducted by comparing the absorption decay of eTet solution aggregates at constant temperature, under illumination and in the dark. For example, comparison of solution aggregates at 4 °C, under illumination and in the dark, would allow the rate of photooxidative decay to be separated from the rate of aggregation. This information may then be used to compare the decay mechanism of eTet in the maquettes and to determine whether the increased half-life of eTet in the maquettes was due to protection from photooxidative damage or if the proteins function simply to limit the rate of pigment aggregation.

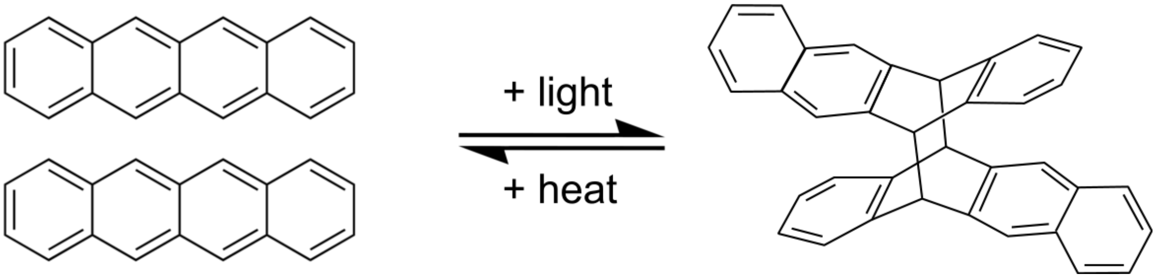


Figure 4.11 Photochromism in tetracene. Absorption of light leads to the dimerisation tetracene, which is reversible with the application of heat. Adapted from Anthony, 2008.

4.4.3 Binding and Stoichiometry

One notable difference between the results of this chapter and those of chapter 3, is that the partitioning of eTet was dependent on the composition of the inner protein cavity, unlike AXT. This discrepancy cannot be explained from current data, but may be due to the presence of polar histidine residues in the BT6 protein and steric clashes from the bulky phenylalanine residues in BT64F. The generation of other maquette variants with altered core residue size or polarity may identify the cause of the discrepancy.

As with the data on carotenoid-maquettes in chapter 3, the stoichiometery of eTet binding remains unknown, however, the absolute concentrations of eTet and protein may be determined by quantitative H^1 NMR (Pauli et al., 2013). This procedure would not be applicable in the quantification of carotenoids as the chemical shift data of protons in AXT significantly overlapped with those of the protein (data not shown). However, the methyl protons in the TIPS functional group of the eTet molecule give chemical shifts in the 0.1-0.09 ppm region (Sasi Kumar et al., 2015), which does not significantly overlap with any of the signals from the BT6 proteins (see appendix section 4).

4.4.4 Singlet fission

As described above, the results of the TA spectroscopy were insufficient to describe the electronic transitions of eTet in the monomeric and aggregated state, but did reveal some interesting characteristics that warrant further investigation. The proposed model to describe the data suggests that some triplet character is formed in the monomer following excitation at 400 nm, but not at 500 nm, with the intersystem crossing proposed to occur from a high vibrational level in the S_1 state. A putative triplet state is also formed in BT64A-eTet following 400 nm excitation, which leads to the complete decay of the S_1 state within ~10 ps, possibly through a SF-mechanism. TA analysis of eTet-maquettes after 500 nm excitation will be critical in determining whether the formation triplets in monomeric and aggregated eTet arise from distinct decay mechanisms. As SF in tetracene films outcompetes all other decay mechanisms (Wilson et al., 2013), excitation at 500 nm should result in the complete conversion of the S_1 state to the same triplet species identified at 400 nm, confirming the SF process in the eTet-maquettes

Although the transient absorption of the putative triplet state in the eTet monomer is largely in agreement with previously reported work on conventional tetracene (Burdett et al., 2010), further analysis should be conducted to confirm this. As the transition from the putative S_1 state to the ground state would likely be emissive in eTet (Wilson et al., 2013), information about excitation transitions may be determined from time-resolved photoluminescence spectroscopy, as triplet decay is not emissive in such organic systems. Thus, assuming that 100 % of the energy from the S_1 state decays emissively, the proportion of energy lost by ISC following 400 nm excitation may be deduced from the loss of fluorescence signal, compared to excitation at 500 nm. Transient electron paramagnetic resonance (EPR) may also be used to describe the spin-state of monomeric and aggregated eTet following excitation, thus confirming the generation of triplets (Schweiger and Jeschke, 2001).

Interestingly, the observed decay rate of the putative triplet state was accelerated in the maquettes with respect to the monomer. This finding is in agreement with previous work on covalently linked tetracene dimers (Korovina et al., 2015; Korovina et al., 2018), where shorter triplet lifetimes are generally observed compared to larger solid-state structures, due to triplet-triplet annihilation (TTA) or failure of the two triplets to dissociate from the entangled TT state (see section 1.6) (Casanova, 2018). The short triplet lifetimes observed provide further evidence for the presence of small aggregates within the protein cavity, most likely arranged as a molecular dimer.

Providing the results of the aforementioned experiments are in agreement with the proposed model for the excited state transitions of eTet, the rate of triplet formation in the BT64A-eTet samples becomes interesting. Piland and Bardeen (2015) demonstrate that the rate of singlet fission in solid state tetracene is highly dependent on the morphology of the crystal or film, with singlet lifetimes ranging from 30-300 ps. More recently, Stern et al., (2017) demonstrated the formation of the entangled TT state on the timescale of 300 fs. As the singlet lifetime is less than 10 ps in the BT64A-eTet sample, SF in the protein-bound aggregates is among the most rapid observed. As discussed in section 3.4.1, the atomic structure of the eTet complexes would permit an empirically derived assessment of the excited state dynamics with respect to the structural configuration of the incorporated chromophores and perhaps explain the rapid SF process observed in the maquettes.

Chapter 5

Design and production of a protein matrix for tuning the photophysical properties of organic semi-conductors through spin-orbit coupling

---

5.1 Motivation

The experimental strategy detailed thus far has aimed to develop a protein-based system to facilitate the measurement of organic semi-conductors without perturbing the underlying electronic transitions of the incorporated material. Conversely, the objective of this chapter was to alter the photophysical properties of the incorporated chromophores by increasing the rate of intersystem crossing though a protein-mediated spin-orbit coupling (SOC) mechanism.

SOC constitutes the best established mechanism for achieving rapid intersystem crossing – the transition from singlet to triplet spin states – and is heavily reliant on the atomic mass of the nuclei involved. As the atomic masses of the nuclei in organic chromophores are small they do not play a significant role in SOC, and intersystem crossing is not commonly observed. The so-called 'external heavy-atom effect' provides a mechanism to impart a SOC interaction in a chromophore without modification to its chemical structure. In this chapter, selenium and iodine were attached to maquettes to alter the electronic transitions of incorporated tetracene and 2,7-dioctyl[1]benzothieno[3,2-b][1]benzothiophene (C8-BTBT). The development of a protein-based system to control the external heavy-atom affect would provide a unique tool to probe the SOC process in a variety of organic molecules, allowing empirically derived correlations to be generated between the optical properties of the chromophores and the predicted distance of the heavy atom.

As the decay of triplets in organic chromophores typically occurs in a non-radiative manner, harvesting the energy of such states is difficult and provides a significant challenge for singlet fission (SF)-based photovoltaic devices. Although several materials have recently been reported to extract the triplet energy of SF-chromophores, these mechanisms rely solely on Dexter energy transfer, which is highly dependent on the close spatial arrangement of donor and acceptor. If the protein-mediated SOC interaction was of sufficient strength, the decay of triplets could occur emissively, allowing the energy to be extracted by resonance energy transfer or photon emission over longer distances.

5.2 Summary

Spin-orbit coupling (SOC) describes the relativistic interaction between the spin of an electron and its orbital angular momentum in a magnetic field, with important connotations in the properties of photonic devices. To consider the importance of SOC in such applications, intersystem crossing is discussed in relation to the well-established system of organic light-emitting diodes (OLEDs) (Burroughes et al., 1990). Electronic excitation of an OLED may result in the formation of four distinct electron spin-states, one in the singlet manifold and three in the triplet (section 1.4.3). In organic systems, the relatively weak SOC interactions means that excitation emission may only occur from the singlet state, as emissive decay from the triplet manifold is optically-forbidden. Therefore the energy of the emission ($S_1 \rightarrow S_0$) may only account for 25 % of the electrical energy input in an ideal system. OLED efficiencies have been improved through the inclusion of phosphorescent dyes, such as octaethyl-21H,23H-porphine platinum(II), allowing light-emission efficiencies of 90% to be reached (Blado et al., 1998; Baldo et al., 2000). Such improvements in electroluminescent efficiencies rely on energy transfer between the OLED material and the cognate dye, where the 'heavy-atom' effect exerted by the coordinated Pt centre imparts a strong SOC interaction on the chromophore macrocycle, permitting intersystem crossing and phosphorescent emission from the triplet manifold. Note that the term 'heavy-atom' will be used herein to describe nuclei of large atomic number and the *external* heavy-atom effect will refer to atoms that do not form part of the chromophore's chemical structure.

Recent studies have reported energy transfer from triplet states in conventional tetracene and pentacene to lead (Pb) quantum dots, primarily mediated through Dexter transfer (Tabachnyk et al., 2013; Thompson et al., 2014). However, as Dexter transfer requires the close physical arrangement of donor and acceptor, employing such energy transfer mechanisms may limit the applications of the material. The external heavy-atom effect permits optical-transitions that are forbidden under normal circumstances, without alteration to the chemical structure of the chromophore. Development of this energy-harvesting mechanism is desirable as transfer from triplet states could occur via photon emission or FRET, permitting transfer over longer distances. This leads to the suggestion that the process may be applied to SF-sensitised photovoltaic devices (see section 1.6), where generating a mechanism to extract the energy of triplet states has proved challenging (Bardeen, 2014; Rao and Friend, 2017). The experimental strategy below aimed to utilise the external heavy-atom effect to tune transitions to and from the triplet manifold in eTet (see chapter 4) and C8-BTBT (figure 5.6A, inset).

The two chromophores used here were selected for their entirely different photophysical properties, as described below. Given that the results of chapter 4 demonstrated that SF was likely to occur in the eTet-maquettes, the transient absorption (TA) spectra of the complexes could be monitored to determine what effect the placement of heavy atoms had on the formation and decay of the putative triplet state. Any evidence of interaction would demonstrate the capacity of maquette proteins to alter the photophysical properties of a highly-established SF material, with potential applications in the devices previously discussed. C8-BTBT was selected as triplets are readily formed in the monomeric state, which may also decay emissively (personal communication, Dr. Jenny Clark, University of Sheffield, UK), allowing the assessment of triplet lifetimes in the maquettes by both TA and time-resolved photoluminescence (PL) analysis. Note that the unusual ability of the organic chromophore, C8-BTBT, to readily transition to and emit from the triplet manifold is believed to occur through light-induced puckering of the thiophene rings (Schnappinger et al., 2017). The experimental strategy here aimed to introduce selenium and iodine into maquette structures in order to observe the external heavy-atom effect in the incorporated chromophores, with a brief justification for the approach described below.

In biochemistry, protein iodination is typically conducted using the ^{131}I or ^{125}I isotopes for use in radioimmunoassays and, as such, many commercially available reagents exist to facilitate the covalent attachment of iodine to tyrosyl moieties in protein structures (Markwell, 1982; Jakubowski et al., 2008). The external heavy-atom effect of the halides was demonstrated through the pioneering work of Michael Kasha in the late 1940s and early 1950s, where the combination of colourless ethyl iodide and colourless 1-chloronaphthalene resulted in a yellow coloured solution despite no chemical reaction taking place (Kasha, 1952). This was shown to be caused by a direct S_0 to T_1 transition, enabled by the close contact of the naphthalene chromophore and iodine.

The *in vivo* incorporation of selenomethionine into protein structures has been used for many years in X-ray crystallography as an answer to the phase problem in structural biology (Hendrickson et al., 1990; Hendrickson, 1991). As selenium-labelled proteins can be purified directly from *E. coli* this approach would be considered advantageous over *in vitro* protein iodination due to the time saved in generating samples. As the process of SOC is dependent on the fourth power of the atomic mass of the nuclei used (see section 1.4.7), the effect of selenium (standard atomic mass (A_r) 79.0) was hypothesised to be weaker than that of iodine (A_r 126.9). However, recent work by Schott et al., (2017) describes the increase in SOC in C8-BTBT through an internal heavy atom-effect, following the substitution of sulphur atoms with selenium.

Based on the BT64A scaffold, three maquette variants were designed for the heavy-atom incorporation. The BT64A_A47Y and BT64A_Q49Y maquettes were generated to position a single tyrosine residue within the protein core and into the aqueous surroundings, respectively. The BT64A6M maquette was generated to position six methionine residues within the hydrophobic cavity of the protein. Tandem liquid-chromatography mass spectrometry (LC-MS) was used to determine the efficiency of heavy-atom incorporation in the protein scaffolds, showing >99 % selenomethionine identity in BT64A6M and a 2:1 iodine to tyrosine binding stoichiometry in both BT64A_A47Y and BT64A_Q49Y.

The incorporation of eTet into selenium-labelled and non-labelled BT64A6M resulted in ground-state absorption characteristics similar to those described in chapter 4. The TA characteristics of the selenium-containing complex were recorded and are compared with BT64A in section 5.3.4. The partitioning of C8-BTBT into a protein structure is demonstrated for the first time into the BT64A maquette, with optical properties closely resembling those of thin films (Keum et al., 2018). The TA characteristics of monomeric and protein-bound C8-BTBT are discussed in section 5.3.6, with a proposed model for the energetic transitions in both the monomeric and oligomeric states. Although the partitioning of C8-BTBT was demonstrated to occur in BT64A_A47Y and BT64A_Q49Y, the formation of white precipitate was visible by eye after ~30 mins at room temperature and further analysis of these constructs was not conducted.

5.3 Experimental work

5.3.1 Heavy atom derivative protein redesign strategy

The four different maquette genes discussed in this chapter (see table 5.1) were all ordered from DNA 2.0 (now ATUM) and provided in the pJexpress414 plasmid. All variants derive from the BT64A maquette discussed in section 3.3.1 and have been designed to incorporate selenium or iodine as follows.

Methionine residues were substituted into the *a* and *d* positions of the α-helical heptad repeats (see section 1.3.1), which are predicted to project into the protein interior, generating a BT64A-methionine variant containing 6 methionine residues, herein named 4A6M. Selenium was incorporated into these positions *in vivo* by culturing BL21 (DE3) *E. coli* in minimal medium containing selenomethionine (SeM+ minimal medium) according to the experimental method in section 2.7.3. The purified protein was analysed by SDS-PAGE analysis with the purity estimated to be in excess of 95 % (figure 5.1).

The BT64A_Q49Y and BT64A_A47Y maquettes (herein called Q49Y and A47Y) were generated to position a single tyrosine residue towards the exterior and interior of the protein, respectively. The maquette proteins were purified from BL21 (DE3) *E. coli* as described previously (section 2.7.1) and covalent attachment of iodine to the tyrosyl moiety of the purified maquette variants was conducted *in vitro* using sodium iodide (NaI) and Pierce Iodination beads, according to the manufacturer's instructions (see section 2.16.1). SDS-PAGE analysis of the A47Y and Q49Y maquettes gave estimated purities in excess of 95 % (figure 5.1).

5.3.2 Analysis of heavy atom incorporation by mass spectrometry

The stoichiometry of tyrosine iodination in the A47Y and Q49Y maquettes was determined by direct infusion mass spectrometry (MS). The iodinated proteins generated as described in the previous section were precipitated using a 2D Gel Clean-Up Kit (GE Healthcare) according to the protocol in section 2.17.1, with the precipitate re-solubilised in infusion solvent (60 % formic acid, 20 % acetonitrile, 20 % H_2O). Samples were then directly infused into a Q Exactive HF quadrupole-Orbitrap mass spectrometer (Thermo Scientific) at 5 µL min^{-1} via a heated electrospray ionisation (ESI) source (section 2.17.2). MS operation and data analysis of the iodinated maquettes was conducted by Dr. Philip J. Jackson (University of Sheffield, UK).

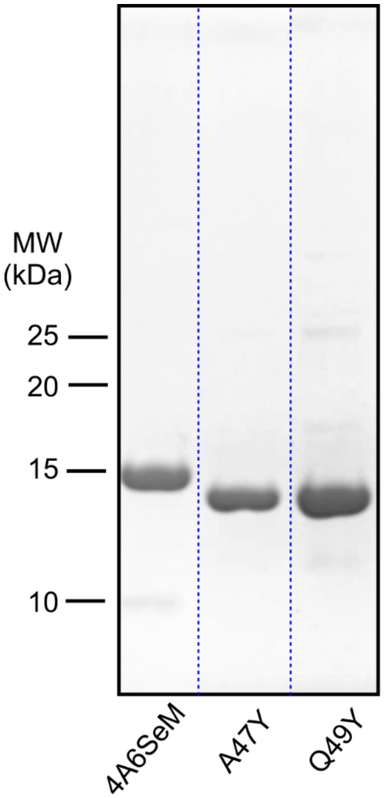


Figure 5.1 SDS-PAGE analysis of heavy-atom maquette variants. The purity of 4A6(Se)M grown in SeM+ minimal media, A47Y and Q49Y was estimated to be greater than 95 % by SDS-PAGE analysis. The positions of molecular weight markers are displayed (right) and the sample identity is indicated below each lane. The blue dashed lines indicate splice points in the image.

The theoretical charge series for the maquette variants were calculated from the predicted average molecular weight of each protein, modelling the mass to charge ratio (m/z) based on the incorporation of 12-19 protons (figure 5.2, pink circles) and the attachment of zero, one, two or three iodine atoms (I_{0-3}) per tyrosine (figure 5.2, green vertical lines). For both maquettes, the observed charge states were of higher mass than predicted, most likely due to the attachment of sodium or potassium ions to glutamate residues in the protein, rather than the protons considered in the modelling. As the apex of almost all charge states is between the predicted mass of the I_2 and I_3 iodination states the stoichiometry of iodine attachment is likely to be 2:1 with respect to each tyrosine, confirming the success of the labelling method used in section 5.3.1. Iodinated proteins will be referred to as A47Yi and Q49Yi.

Tandem liquid chromatography mass spectrometry (LC-MS) was used to determine the extent of selenomethionine (SeM) incorporation into the 4A6M maquette. LC-MS data acquisition of 4A6M was conducted in collaboration with Mr. Simon Thorpe (University of Sheffield, UK) according to the method detailed in section 2.17.3. Briefly, approximately 20 µg of protein in 50 mM HEPES (pH 8.5) was loaded onto a C18 reverse-phase column (Agilent), with a linear gradient of solvent A (0.1 % formic acid in H_2O) and solvent B (0.1 % formic acid in 99.9 % acetonitrile) used to elute the protein. The protein elution was subsequently ionised by heated ESI and analysed by quadrupole-TOF (time-of-flight) MS (6530 Q-TOF, Agilent). The MassHunter (Agilent) software was used to deconvolute the isotopic patterns and charge states of the recorded species to give the average isotopic mass of the molecules in the sample (figure 5.3).

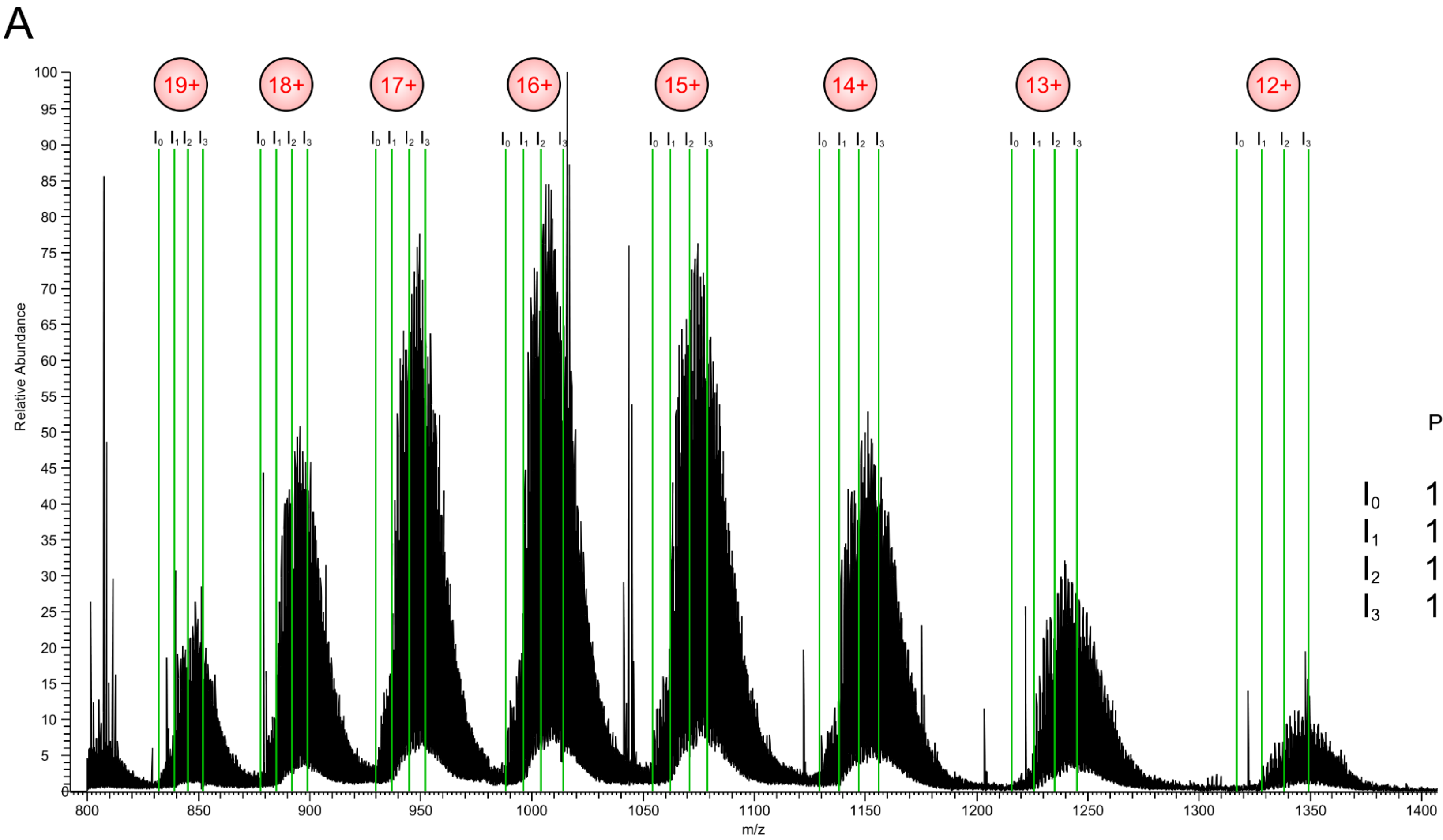
As the N-terminal methionine (NTMet) residue may be cleaved in proteins produced from *E. coli*, the average theoretical mass of the 4A6M protein was calculated with and without NTMet (-NTMet), giving predicted values of 16699.22 and 16568.03 Da, respectively. Assuming 100 % incorporation of SeM, the predicted mass of 4A6M (+7Se) was 17027.597 Da and 16849.496 Da for 4A6M(-NTMet) (+6Se). The predominant average isotopic mass species observed in the LC-MS analysis were 16850.4 Da and 17028.3 Da, corresponding to 4A6M(-NTMet) and 4A6M with full incorporation of SeM. The mass error between the observed and theoretical masses were 54 and 42 ppm for 4A6M(-NTMet) and 4A6M, respectively, which is within the error range expected for the analysis. The theoretical mass of the 4A6M and 4A6M(-NTMet) proteins following the incorporation of 0-7 SeM residues are displayed in table 5.2.

The MS results show that the overwhelming majority of signals in the sample can be attributed to maquettes solely containing SeM, comprising a mixed population of proteins with and without an N-terminal SeM. As the N-terminal SeM residue is unlikely to affect the chromophores bound in the maquette interior it will not be considered in the work below and the mixed population of selenium-labelled proteins will be referred to as 4A6SeM.

Table 5.1. Amino acid sequences of peptides used in this chapter.

Protein	Abbreviation	Sequence	Details
BT64A	-	MGGDGENLYFQG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLHHHHHH	Original maquette used in chapters 3 and 4
BT64A6M	4A6M or 4A6SeM	M‡GGDGENLYFQG EIWKQAEDA**M**QKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**M**EDALQKFEEA**M**NQFEDLKQLGGSGSGSGG EIWKQAEDA**M**QKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**M**EDALQKFEEA**M**NQFEDLKQLHHHHHH	6 methionine residues, all predicted to position within hydrophobic cavity
BT64A_Q49Y	Q49Y or Q49Yi	MFQG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDAL**Y**KFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLHHHHHH	Tyrosine residue, predicted to project to protein exterior
BT64A_A47Y	A47Y or A47Yi	MFQG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAED**Y**LQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLHHHHHH	Tyrosine residue, predicted to position within hydrophobic cavity

The positions of residue substitutions are shown in bold, with (seleno)methionine residues in red and tyrosine in blue. ‡ Note that protein production in *E. coli* may result in cleavage of the N-terminal methionine residue.



	Predicted MW
I_0	15793
I_1	15919
I_2	16045
I_3	16171

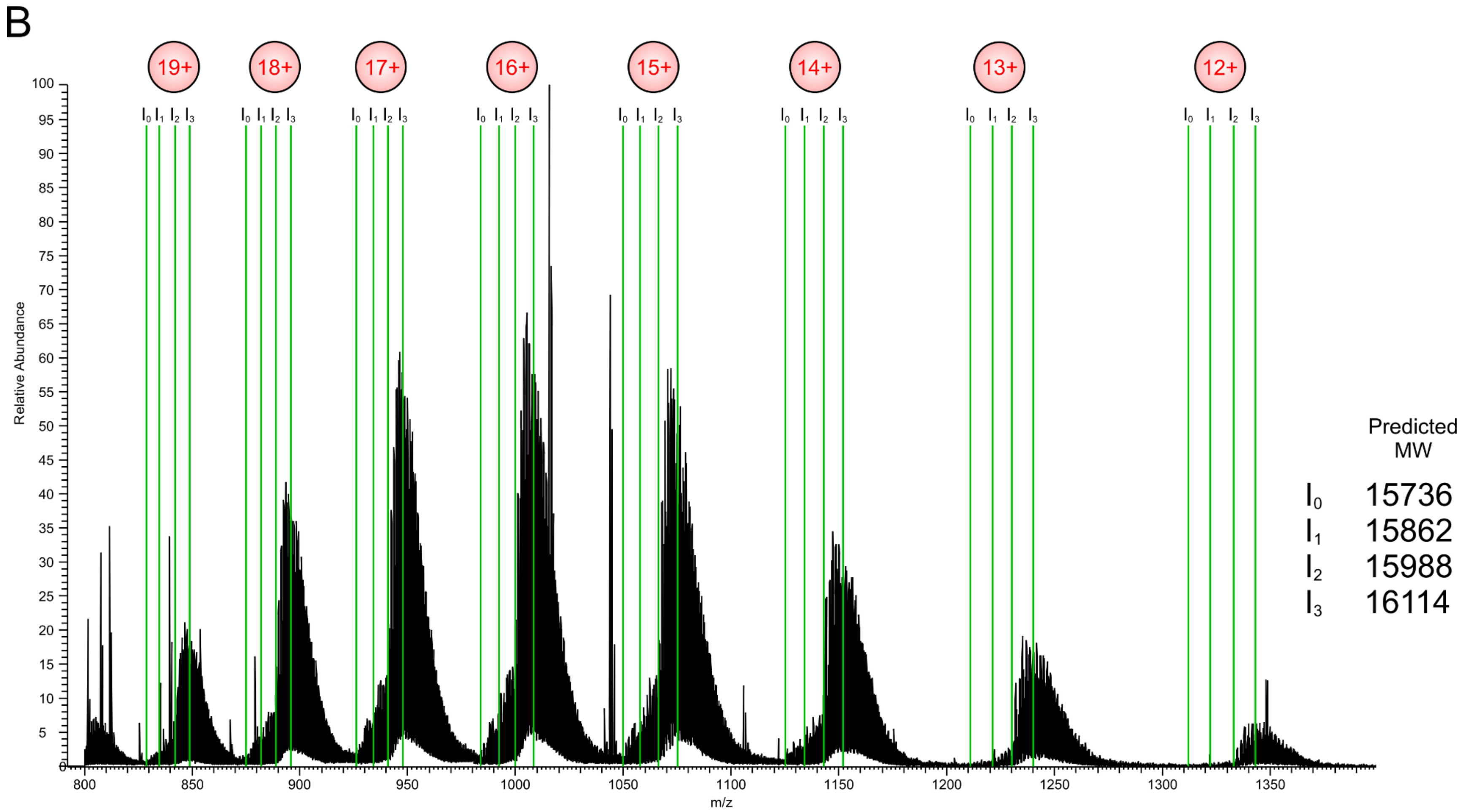
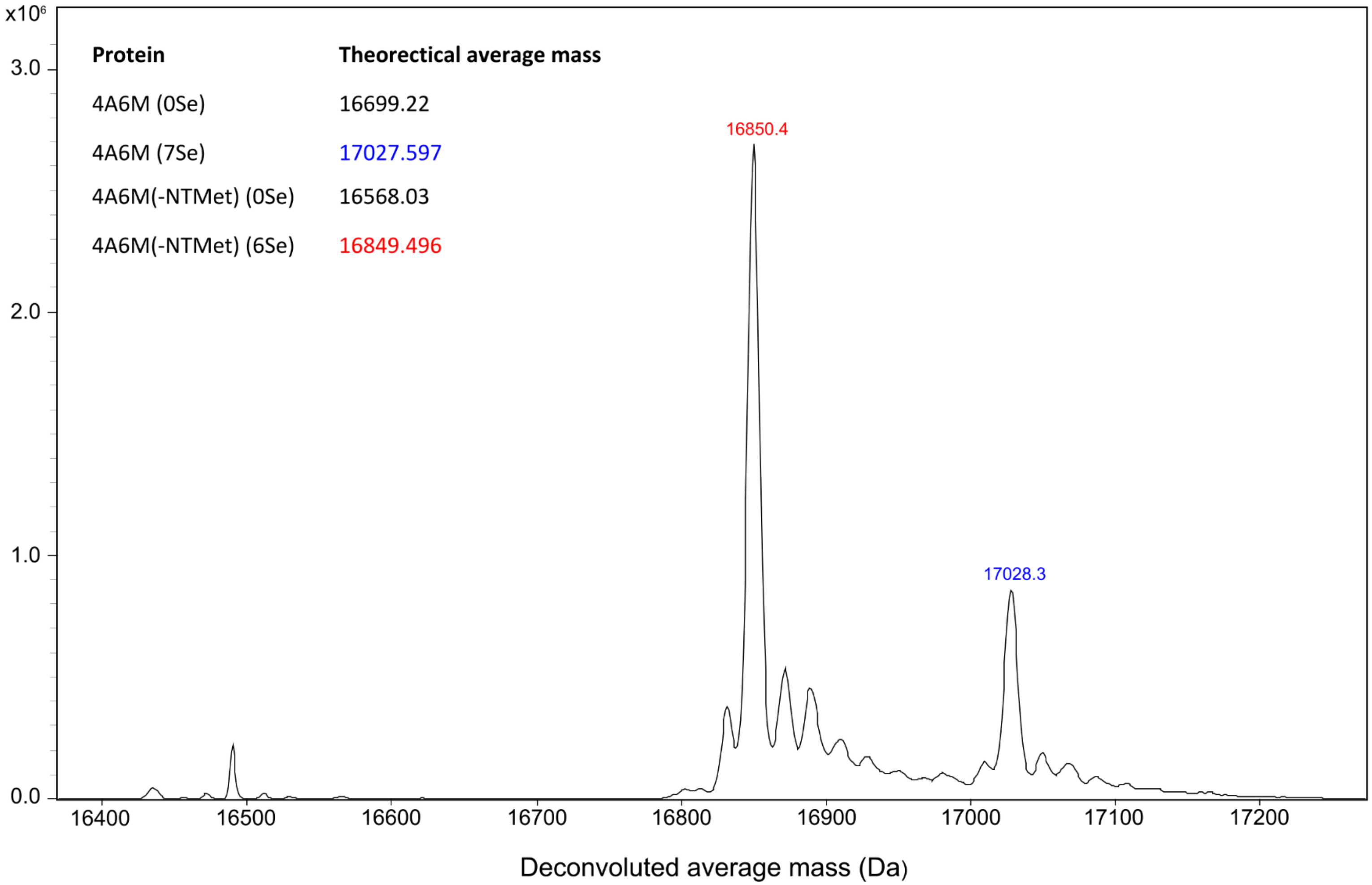


Figure 5.2 MS analysis of iodinated maquette proteins. The relative abundance of charge species are shown for A47Y **(A)** and Q49Y **(B)** following the attachment of iodine. For each data set, a charge series is shown for the addition of 12-19 protons (pink circles). The predicted m/z that corresponds to the attachment of zero, one, two and three iodine atoms ($I_{0\text{-}3}$) is indicated by the vertical green lines. The predicted molecular weights of the I_0-I_3 states are shown in the table inset.



Protein	**Theorectical average mass**
4A6M (0Se)	16699.22
4A6M (7Se)	17027.597
4A6M(-NTMet) (0Se)	16568.03
4A6M(-NTMet) (6Se)	16849.496

Figure 5.3 LC-MS analysis of selenium labelled maquettes. The average molecular mass for each of the two predominant species is indicated above each peak, with the average theoretical mass of the maquettes in the table inset.

Table 5.2. Theoretical mass of selenium-labelled proteins.

Protein	Theoretical average mass*	Observed mass	Δmass
4A6M (0Se)	16699.22	17028.3	329.08
4A6M (1Se)	16746.131	17028.3	282.169
4A6M (2Se)	16793.042	17028.3	235.258
4A6M (3Se)	16839.953	17028.3	188.347
4A6M (4Se)	16886.864	17028.3	141.436
4A6M (5Se)	16933.775	17028.3	94.525
4A6M (6Se)	16980.686	17028.3	47.614
4A6M (7Se)	17027.597	17028.3	0.703 (41 ppm)
4A6M(-NTMet) (0Se)	16568.03	16850.4	282.37
4A6M(-NTMet) (1Se)	16614.941	16850.4	235.459
4A6M(-NTMet) (2Se)	16661.852	16850.4	188.548
4A6M(-NTMet) (3Se)	16708.763	16850.4	141.637
4A6M(-NTMet) (4Se)	16755.674	16850.4	94.726
4A6M(-NTMet) (5Se)	16802.585	16850.4	47.815
4A6M(-NTMet) (6Se)	16849.496	16850.4	0.904 (54 ppm)

*Note that the additional theoretical mass of selenomethionine incorporation was calculated from the difference in the average atomic mass of S and Se, which is 46.911 Da. The mass error is quoted in parentheses for the protein species that resulted in the lowest Δmass.

5.3.3 Incorporation of eTet into labelled maquette variants

The incorporation of eTet into the 4A6M and 4A6SeM maquettes was conducted using a method identical to that described previously (section 2.11.2), to determine if any differences in binding could be observed between the variants and BT64A. Following the incorporation of eTet, samples were filtered (0.2 µm pore diameter) and unbound pigment was removed by anion exchange chromatography (AEC). All samples were concentrated to exactly 1 mL and the ground-state absorption of the 4A6M and 4A6SeM variants confirmed that eTet was incorporated in an aggregation state similar to that observed in chapter 4 (figure 5.4A). Although significantly lower chromophore incorporation was observed in 4A6M and 4A6SeM compared to the BT64A protein, the absorption spectra of both variants showed less scatter than that of BT64A (Figure 5.4B), which may be indicative of a tighter binding interaction.

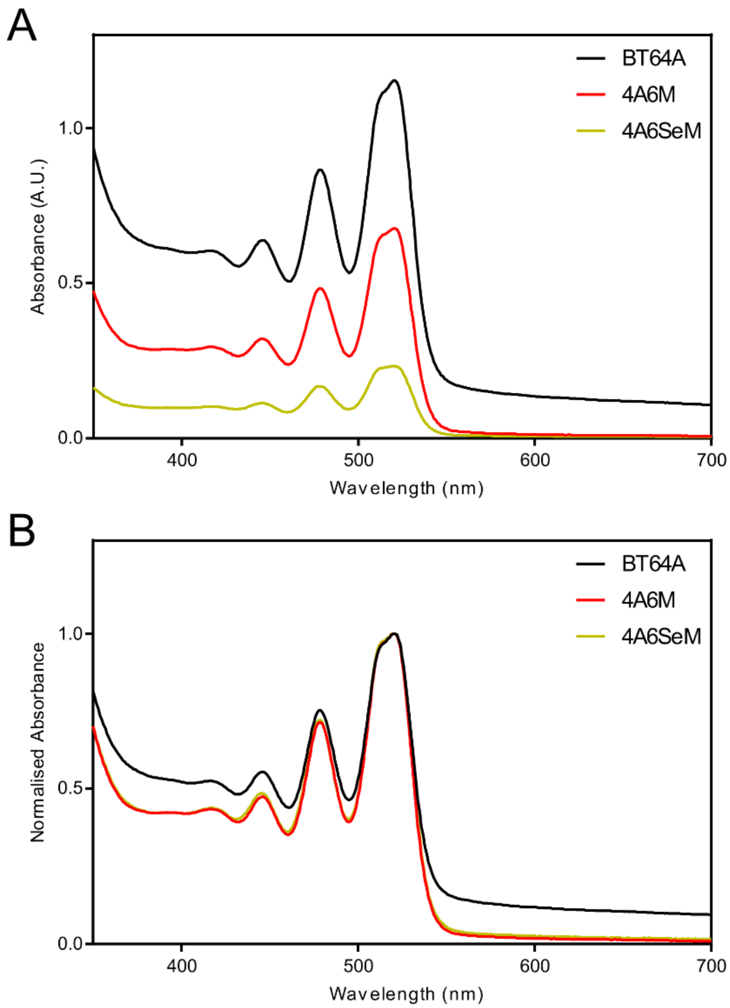


Figure 5.4 Incorporation of eTet into selenomethionine labelled maquettes. (A) Absorption spectra of eTet in the BT64A (black), BT64A6M (red) and BT64A6SeM (orange); all samples were recorded at equal volumes. **(B)** Data in panel A, normalised to absorbance at 525 nm.

5.3.4 Assessment of the SOC interaction in eTet-maquettes using transient absorption spectroscopy

The 4A6SeM-eTet maquette sample generated in the previous section was buffer exchanged into no-salt buffer (50 mM HEPES pH 6) and concentrated to give an optical density at 525 nm (OD_{525}) of approximately 1. The sample was then sparged with nitrogen for 5 mins and sealed in a 1 mm pathlength quartz cuvette under anoxygenic conditions. Photoinduced absorption (PIA) spectra of the 4A6SeM-eTet sample were recorded at discrete time intervals following excitation with a 400 nm laser, using methodology identical to that described in chapter 4 (see sections 2.11.4 and 4.3.6). Due to limitations in instrument availability, the equipment configuration for the TA measurements here could only accurately record between 450-800 nm, so the putative S_1 peak at ~425 nm could not be observed in the 4A6SeM-eTet sample (figure 5.5A). All TA measurements and analysis in this section were conducted in collaboration with Daniel Polak (University of Sheffield, UK).

The results of the TA spectroscopy on 4A6SeM-eTet were inconclusive. As the absorption range of the instrument did not permit the identification of the S_1 state, the potential population of this state is unknown. However, the 520 nm PIA feature in the BT64A-eTet sample (figure 5.5B), which was assigned to the T_1 state in chapter 4, could not be observed at any time point in the 4A6SeM-eTet maquette. Given that the absorption measurements were recorded within 1 ps of excitation it is highly unlikely that the heavy-atom effect was responsible for the formation and decay of the triplet state within this timeframe. The absence of the putative T_1 state in the 4A6SeM-eTet sample remains unclear and is discussed in section 5.4.2.

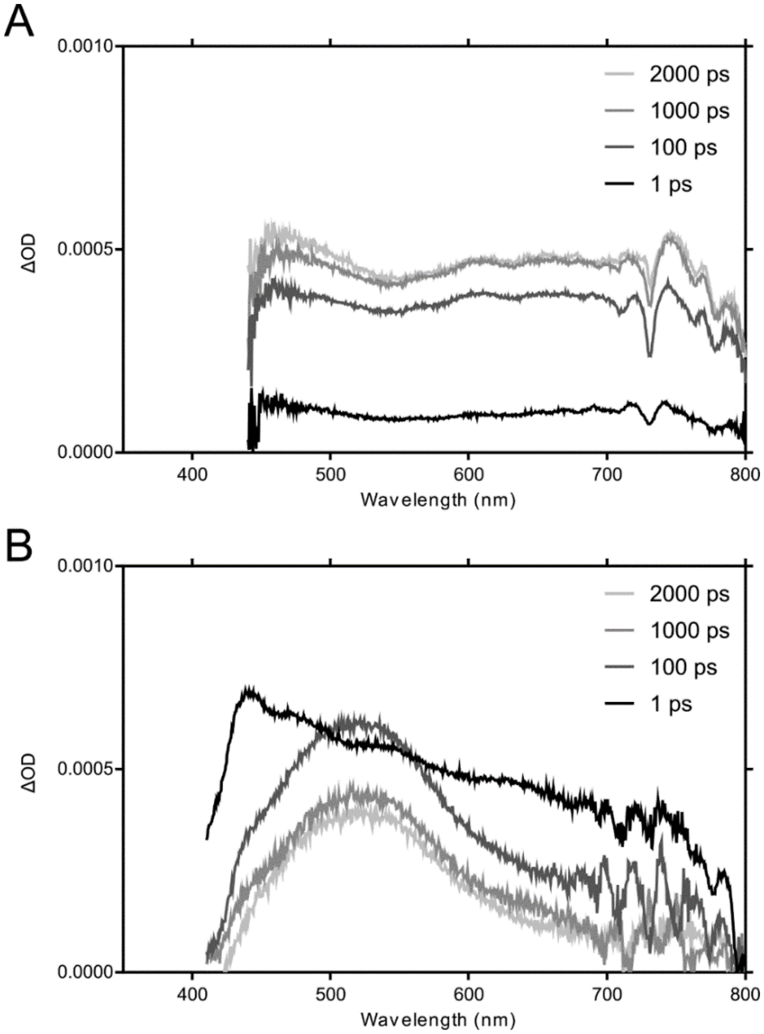


Figure 5.5 Transient absorption spectroscopy of eTet in selenium labelled maquettes. TA spectra of 4A6SeM-eTet **(A)** and BT64A-eTet **(B)** recorded after 1-2000 ps delay following 400 nm excitation.

5.3.5 Incorporation of C8-BTBT into maquette structures

C8-BTBT was solubilised in 100 % DMSO at 60 °C with sonication to give stock solutions of 125 µg mL^{-1}, with the ground-state absorption recorded after dilution to 25 µg mL^{-1} (figure 5.6A). Incorporation of C8-BTBT into the BT64A maquette was conducted according to the method in section 2.11.3. Briefly, the stock C8-BTBT solution was added in a 1:4 (v/v) ratio to a solution of 50 mM HEPES (pH 8.5) containing BT64A, with the concentration of protein 20-fold less than that of the chromophore. The incorporation reaction was incubated for 30 min at 4, 37 and 50 °C, before four-fold dilution with chilled 50 mM HEPES and filtration (0.2 µm pore diameter). The filtered solution was then applied to AEC column to remove unbound pigment, concentrated to exactly 1 mL and analysed by ground-state absorption. The absorption spectra revealed that the chromophore was incorporated into the BT64A maquette at all temperatures tested, with the greatest levels of partitioning occurring at 50 °C. The absorption characteristics of C8-BTBT in BT64A were significantly different to those of the monomer; importantly the lowest energy absorbance peak shifted from 330 to 355 nm (see section 5.3.6). Interestingly, the absorption profile of this feature in BT64A closely matches previously reported spectra of C8-BTBT thin films (figure 5.6B) (Keum et al., 2018), suggesting that the pigment forms an aggregated state in the maquette.

The propensity for C8-BTBT to partition into the A47Y and Q49Y maquettes was assessed using a method identical to that described above. The absorption spectra revealed that C8-BTBT partitioned into both the A47Y and Q49Y maquettes to some extent, with the level of pigment in the Q49Y maquette higher than that of BT64A (figure 5.6C). Note that at wavelengths shorter than 300 nm, the absorption observed in the A47Y and Q49Y maquettes was considerably higher than that of BT64A, which is thought to be due to degradation of the pigment stock following storage.

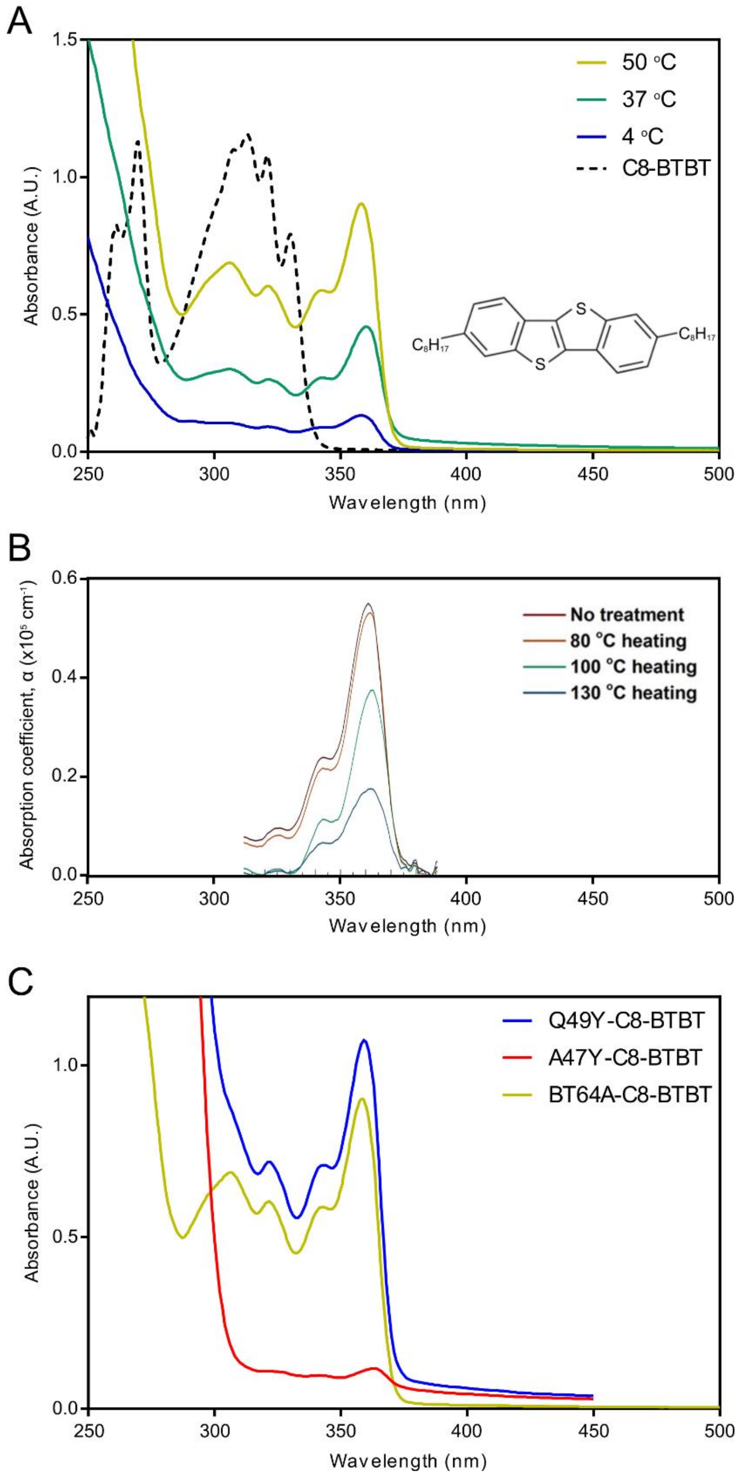


Figure 5.6 Absorption spectra of C8-BTBT. (A) The absorption spectra of monomeric C8-BTBT in DMSO (dashed, black line) and in BT64A following incorporation at various temperatures (solid, coloured lines); all measurements were taken from samples of the same volume (1 mL). **(B)** The absorption spectra of C8-BTBT films generated under different temperatures; reproduced with permission from Keum et al., (2018). **(C)** Absorption spectra of C8-BTBT in BT64A (yellow), A47Y (red) and Q49Y (blue); all measurements were taken from samples of the same volume (1 mL). Note that differences in absorption intensity may be due to loss of protein during sample purification.

5.3.6 Assessment of the SOC interaction in C8-BTBT-maquettes using transient absorption spectroscopy

TA spectroscopy of monomeric and maquette-bound C8-BTBT was performed by Jozra Garrido Velasco (University of Sheffield, UK). Extensive information concerning the nature of the electronic transitions of monomeric C8-BTBT will discussed elsewhere, with the information provided in this section limited to understanding the differences between the monomer and the maquette. While the information available for the maquette samples is insufficient to fully describe the electronic transitions of the materials, a proposal is suggested to explain the observations.

In monomeric C8-BTBT, excitation at 310 nm resulted in the formation of the S_1 state, which can be observed after a 100 ps delay as a PIA peak at 535 nm (figure 5.7A). The absorbance of the S_1 feature decays within ~500 ps to give rise to another PIA characteristic at 445 nm, which is maintained for longer than 1.2 ns; this second feature is firmly assigned to the T_1 state (data not shown). In the BT64A-C8-BTBT sample, excitation at 310 nm results in the formation of a broad absorption feature after a 2 ps delay, with maxima at 510 and 540 nm (figure 5.7B). This broad absorption is tentatively assigned to the S_1 state as it occupies a similar spectral region to the S_1 state of the monomer. The putative S_1 feature in the maquette-bound sample then decays until a flat, featureless absorption spectrum is reached at approximately 6 ns. Note that the decay of the S_1 state in the maquette was not coupled to the emergence of the T_1 state as was observed in the monomer. The TA of the BT64A-C8-BTBT sample also displayed stimulated emission in the 385 nm region after 2 ps that decreased over time. This stimulated emission is assigned to a S_1 to S_0 transition; however, as the emission spectrum significantly overlaps with the absorption of the S_1 state, valid kinetic data could not be obtained.

In summary, the excitation of C8-BTBT leads to a S_1 to T_1 transition in the monomer (figure 5.8A), which is seemingly inhibited by incorporation into the BT64A maquette. The reasons for this change are not fully understood, though two proposals are offered to explain the observations. (1) The formation of triplets in the monomeric state is thought to arise from puckering of the thiophene ring in the C8-BTBT chromophore, which may be prevented by pigment aggregation or steric hindrances provided by the protein chassis. (2) The ground-state absorption of C8-BTBT shows that the lowest energy transition shifts from ~330 nm in the monomer to 360 nm in the aggregated state, suggesting that the energy of the S_1 state is lowered relative to S_1 of the monomer. It is proposed that the lower energy of the aggregate S_1 state prevents intersystem crossing to the triplet manifold, which is presumed to be of higher energy than the lowest vibrational level of S_1, thus making the conversion energetically unfavourable (figure 5.8B).

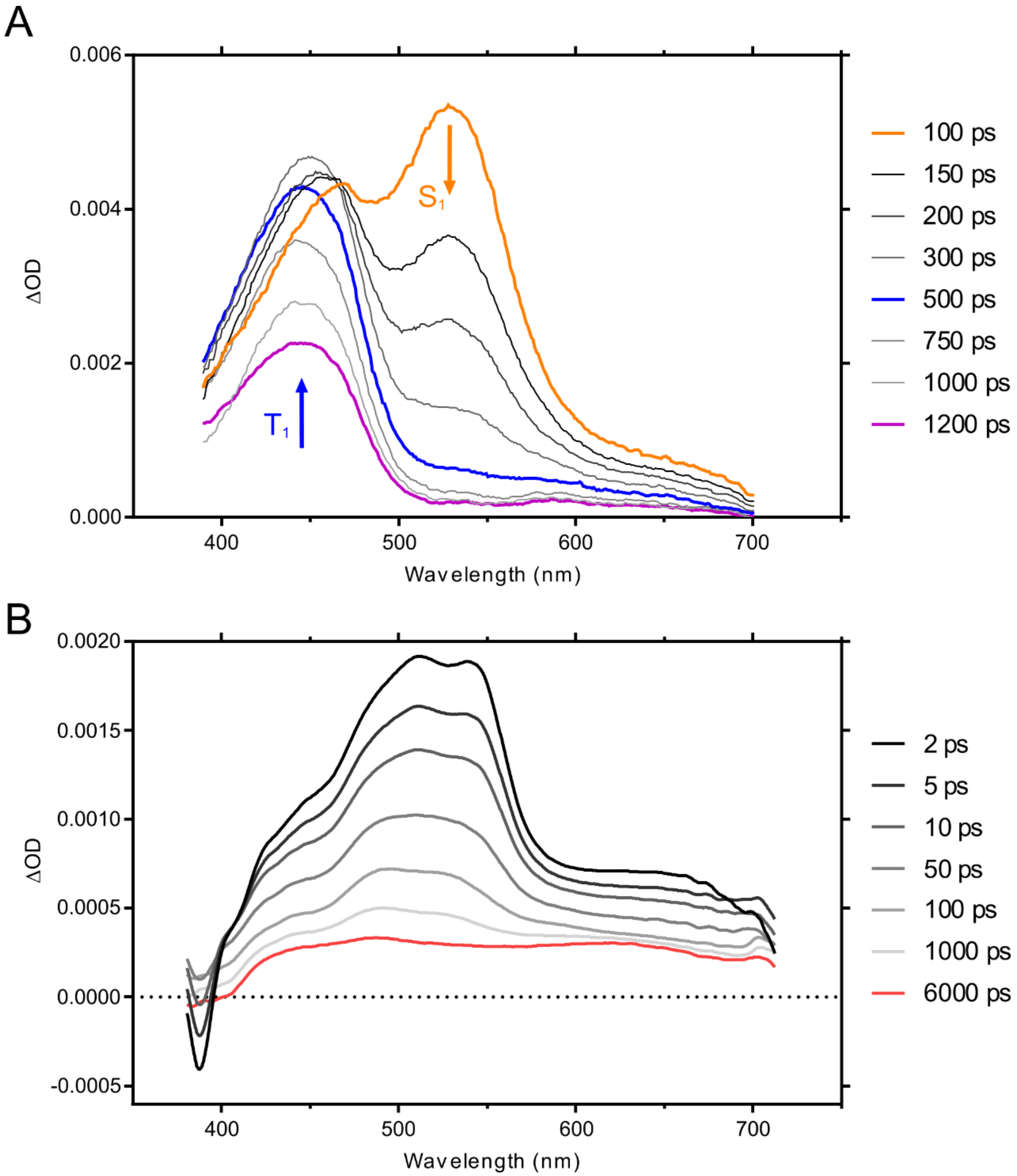


Figure 5.7 Transient absorption spectroscopy of C8-BTBT. (A) TA spectra of monomeric C8-BTBT (3 mM) in 100 % toluene after 100-1200 ps delay following excitation at 310 nm. Key timepoints are highlighted in colour, with intermediate spectra shown in greyscale. **(B)** TA spectra of BT64A-C8BTBT in aqueous buffer (50 mM HEPES, 500 mM NaCl) after 1-6000 ps delay following excitation at 310 nm. A broad absorption peak forms in the 450-570 nm region within 2 ps (black line) and decays (greyscale lines) to form a broad, featureless absorption feature after 6 ns (red).

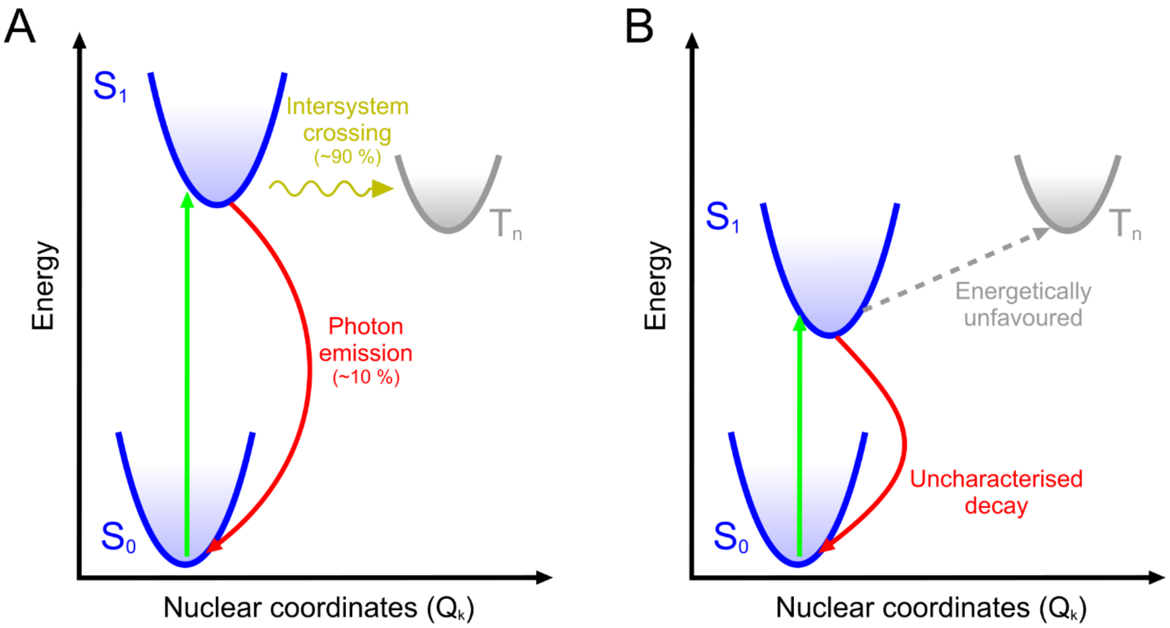


Figure 5.8 Proposed model for energy transitions in C8-BTBT. (A) In monomeric C8-BTBT the absorption of light energy (green arrow) causes a transition to the S_1 state, which is of sufficient energy to permit intersystem crossing (orange arrow) to the triplet manifold (T_n). **(B)** The arrangement of aggregated C8-BTBT in the maquettes is thought to reduce the energy of the S_1 state, such that transition to the triplet manifold becomes energetically unfavoured. The relative positioning of the nuclear coordinates is entirely arbitrary and shown for clarity of argument. While the absorption of light is shown to occur directly to the S_1 state, the excitation may occur to a higher energy state within the singlet manifold.

5.4 Discussion

The work in this chapter aimed to generate three variants of the BT64A maquette to enable the close physical arrangement of heavy atoms and chromophores to observe an enhanced SOC interaction in eTet and C8-BTBT. The proteins were produced in large quantities in *E. coli*, with the *in vitro* and *in vivo* incorporation of iodine and selenium confirmed by mass spectrometry. Due to the time constraints in the publication of this thesis and the limited availability of instruments for ultrafast TA analysis, the results of the spectroscopy in this section remain largely unclear and are discussed at length in the sections below. Although the results of the biochemical manipulation of the maquette proteins were successful, several additional experiments are required to explain the observed activity.

5.4.1 Further biochemical characterisation

As the incorporation of eTet into 4A6M and 4A6SeM resulted in less light-scatter than in BT64A, relative to 525 nm absorbance, it may be suggested that the chromophores are more tightly bound in these variants. As the dissociation of pigment from the protein may have important implications in the photophysical properties of the material (see section 5.4.2), the determination of the binding kinetics in each variant will be essential for forming accurate conclusions. Experimental strategies for determining the strength of binding are discussed in section 7.3.1.

Although the incorporation of C8-BTBT was shown to occur in all three maquette variants tested, the inconsistencies in the absorption in the near UV region warrant further investigation, as does the propensity for precipitation to form in the A47Y and Q49Y samples. The aggregation of C8-BTBT in solution and subsequent incorporation into proteins requires more rigorous investigation, akin to the work conducted in chapters 3 and 4.

As described in section 3.3.2, temperature-dependent CD can be used to indicate the stability of protein structure, which should be used here to determine what impact the peptide sequence changes had on the folding of the maquette variants. Determination of the conformational flexibility of the maquettes may offer some insight into the observed binding characteristics, particularly for the A47Y and Q49Y variants, where dramatically different C8-BTBT incorporation was observed.

5.4.2 Transient absorption of eTet

As the TA observations in 4A6SeM did not cover wavelengths shorter than 450 nm (figure 5.5A), the population of the putative S_1 state could not be determined and as such, more complete measurements are required for any firm conclusions to be drawn. Although the results detailed in section 5.3.4 suggest that the T_1 state is not formed, without confirmation that the S_1 state is initially populated, this observation is inconsequential. If the putative S_1 state is not populated at any stage following excitation, this would suggest a more dramatic alteration to the photophysical properties of the incorporated material than expected and would require much further experimentation. In addition to repeating the observations following 400 nm excitation to allow the presence of the S_1 state to be detected, TA analysis of 4A6SeM-eTet should be conducted following 500 nm excitation as this provides further information of the kinetics of SF, as described in section 4.4.4.

5.4.3 Analysis of the optical properties of C8-BTBT

TA spectroscopy of C8-BTBT revealed that the formation of triplet states, as seen in the monomer in 100% toluene, was seemingly inhibited by insertion into maquettes through an unknown mechanism. One of the main issues in characterising the electronic transitions of C8-BTBT was that few examples of the optical properties are available in the literature, with the majority of the work described by electron paramagnetic resonance (EPR). Characterisation of C8-BTBT in the solid state (i.e. films or crystals) by TA spectroscopy may explain whether the inhibition of triplet formation is an effect caused by the protein or due to the aggregation of the chromophore.

One explanation offered for the lack of triplet formation in C8-BTBT claimed that the energy of the S_1 state was lowered by aggregation and that the energy triplet manifold was sufficiently high to make the transition unfavourable. To investigate this notion the energy of the T_1 state in the protein-bound aggregate may be determined by doping the solution with a triplet donor, such as biphenyl. Triplet energy transfer from the biphenyl donor to the protein-bound aggregate would then allow the transient absorption characteristics of the T_1 state to be determined (Cogdell et al., 1983). The energy of the T_1 and S_1 states may then be inferred from TA analysis and ground state absorption, respectively.

The observed precipitation in the A47Y and Q49Y samples prevented the assessment of the external heavy-atom effect in C8-BTBT, which requires further investigation before use. However, given that the formation of the triplet state is prevented in the BT64A maquette, the effect of the heavy-atom maquettes would result in a readily apparent indication of SOC, by the reformation of the T_1 absorption that was identified in the monomer. Another possibility that would permit a more simple assessment of the heavy-atom effect would be to redesign the maquettes to incorporate only a single C8-BTBT pigment, allowing direct kinetic assessment of the T_1 state. Any perturbations to the rate of triplet formation or decay would thus provide evidence of a protein mediated SOC interaction.

5.4.4 Further improvements to the SOC interaction

The results of this chapter do not provide evidence for or against the induction of a protein-mediated external heavy-atom effect, and further experimental approaches should be used to better observe a SOC interaction through protein redesign or alternative detection techniques. The protein variants generated here consist of reasonably conservative changes to the BT64A precursor and further alterations may be required to observe a SOC interaction. For example, the A47Yi and 4A6SeM variants aimed to introduce one and six heavy atoms into the internal cavity of the maquette, respectively. As the incorporation of pigments occurs through simple partitioning, greater changes to protein sequence may be possible without perturbing the attachment of chromophores so, in theory, the amino acid composition of the inner cavity could consist entirely of iodinated tyrosine or selenomethionine residues.

As the SOC interaction is dependent on the fourth power of the atomic mass of the nuclei involves, greater effects may be observed though the incorporation of larger elements. For example, the incorporation of telluromethionine rather than selenomethionine would result in the mass of the ‘heavy-atom’ changing from 78.971 to 127.60 Da (Budisa et al., 1997) and protein labelling with europium (151.96 Da) would also result in a significant mass increase in comparison with iodine (126.90 Da) (Patel et al., 2008).

Although TA spectroscopy has proved sufficient in identifying the excited states of eTet and C8-BTBT, techniques such as pulsed electron paramagnetic resonance (EPR) may provide more informative evidence to answer some of the questions raised above. EPR spectroscopy has strong similarities with the principles of nuclear magnetic resonance (NMR) spectroscopy, describing the spin of electrons in a magnetic field rather than nuclei (Schweiger and Jeschke, 2001). Pulsed EPR may then be used as a direct indication of the spin-state of a chromophore, providing a sensitive probe to observe the SOC interaction (Schott et al., 2017).

Chapter 6

Developing the interface between biology and artificial haem proteins*

6.1 Motivation

The experimental work detailed so far has utilised artificial proteins to improve or modify the stability or photophysical properties of materials that could be considered of interest to specific areas of physical chemistry. As discussed at length already, the results of this experimental work provide an interesting system with promising future potential. However, it could be argued that by using purified apo-proteins and incorporating molecules in vitro, the proteins essentially become a material and the justification for their use would become more difficult if other materials, such as chemically synthesised polymers, were able to fulfil the same role. Therefore, it is of interest to expand the potential interactions between the maquettes and the biological activities of the cell prior to extraction. For example, the incorporation of molecules from native or engineered biosynthetic pathways in vivo would simplify the production of cofactor containing proteins. The covalent attachment of haem c to a maquette scaffold provides a recent example of this approach (Anderson et al., 2014), but many other post-transcriptional modifications to artificial proteins could also be envisaged.

Developing the interface between maquettes and native biological processes takes advantage of the exquisite specificity and selectivity of native and non-native biological catalysts, and could be used to impart specific functions on protein structures that would not be possible in other materials. With this motivation in mind, the following work hoped to address a number of criteria: (1) Could an entirely artificial protein sequester haem from the native biosynthetic pathway of E. coli through simple bis-histidine ligation (detailed in section 1.3.2)? (2) Could such proteins be recognised and processed by a periplasmic export system; in this case, the twin-arginine translocase. (3) Would the twin-arginine translocase show substrate specificity and more efficiently export folded proteins? (4) Provided the aforementioned criteria are met, can a protein product be purified from the cell periplasm in quantities applicable for use in biotechnology?

---

* A version of this work was published in the Journal of Biological Chemistry (Sutherland et al., 2018). All rights and permissions for all visual data and artwork have been granted. Where appropriate work conducted by other authors is acknowledged.

6.2 Summary

The tightly regulated transport of materials across membranes provides a significant challenge for the cell. In bacteria, the translocation of proteins across the cytoplasmic membrane is conducted by the general secretory (Sec) system (De Keyser et al., 2003) or the twin-arginine translocase (Tat) (Berks, 2015). In native systems the Tat machinery is responsible for the export of folded, often cofactor-containing, proteins. The mechanistic basis for substrate selection and export remains poorly understood, although the current consensus is that the Tat machinery offers a 'proofreading' function and is solely responsible for determining the suitability of proteins for export (see section 1.7.6). How proteins are identified as suitable for export remains contentious and may occur through sensing the conformational flexibility of the substrate or through the recognition of hydrophobic surface patches (Berks, 2015). For extensive information on the Tat system the reader is directed to section 1.5.

The experimental strategy undertaken here was largely founded on the biophysical data detailed in Farid et al., (2013), which showed that the BT6 maquette became gradually more structured following the addition of one and two molar equivalents of haem. Rather than rely on cofactor concentration, the primary aim of this work was to generate maquette variants with intrinsically altered haem binding capacities, with the aim of imposing differential stability and folding onto the protein scaffolds.

Three different maquettes were generated, each varying by a single amino acid substitution predicted to directly alter the haem binding capacity of the protein. The stability and folded state of artificial proteins can be assayed with reasonable accuracy by temperature-dependent circular dichroism (CD) and one-dimensional proton nuclear magnetic resonance (1D ^{1}H NMR) spectroscopy (described in Farid et al., 2013 and Koder et al., 2006, respectively). All three protein constructs were analysed using the aforementioned biophysical techniques in the presence and absence of haem. The results of these experiments revealed that the ligation of haem greatly stabilised structures, where binding was possible, and that the removal of haem binding sites caused concomitant reduction in the extent of protein folding.

As described in Thomas et al., (2001) the inclusion of the well-characterised Tat signal peptide of the trimethylamine-N-oxide reductase (TorA) to the N-terminus of the green fluorescent protein was sufficient to enable Tat-dependent processing. The TorA signal peptide was fused to the N-terminus of each maquette variant and the *E. coli* Tat machinery was challenged to recognise and export the artificial proteins. *In vivo* export assays confirmed the maquette proteins could be processed by the Tat system and that the degree of export correlated with the extent of the protein folding.

As this was the first published example of a pigment containing artificial protein being exported by Tat-machinery, it was of interest to see whether the haem-containing proteins could be purified in quantities suitable for use in biotechnology (Balasundaram et al., 2009). Each of the maquette proteins was produced from the W3110 ‘TatExpress’ cell line in large culture volumes (Browning et al., 2017) and following purification from the periplasm, visibly pigmented maquette protein was obtained.

The results of this proof-of-concept work therefore satisfy all the criteria set out in section 6.1. The experimental approach undertaken here also provides a framework that is applicable for other molecules or complexes that may be of interest in biotechnology, particularly for molecules that may cause toxicity to the host while in the cytoplasm. In addition, these *de novo* designed structures provide a useful tool for probing the quality control mechanism of the Tat system, as modifications to protein structure that alter folding and surface hydrophobicity are entirely intuitive.

6.3 Experimental work

6.3.1 Design and production of maquette proteins

The gene encoding the di-haem binding BT6 maquette was synthesised and provided in the high copy number pJexpress414 plasmid by DNA2.0 (now ATUM). The removal of haem ligating residues was achieved through single histidine to alanine substitutions using QuikChange II site directed mutagenesis (section 2.6.3). This resulted in the one-haem binding BT6M1 maquette, which contained a single His-Ala substitution (H53A), and the zero-haem binding BT6M0, containing two His-Ala substitutions (H53A/H88A). The peptide sequences used here are detailed in table 6.1.

Unlike the proteins discussed in chapters 3-5, the maquettes here did not contain His-tags and required longer production timeframe with a more extensive purification protocol (section 2.7.2). *Escherichia coli* BL21 (DE3) cells were transformed with the pJexpress414 plasmid containing the desired gene under the control of the plasmid's IPTG-inducible T7 promoter. Following the addition of IPTG, cells were incubated for 24 hours at 37 °C and subsequently lysed by sonication. The resulting cell lysates were applied to an immobilised metal affinity chromatography (IMAC) which was used to bind some of the contaminating protein and cell debris from the maquette sample which was able to pass through the column. The flow-through from the column was further purified (>95 % purity, as assessed by SDS-PAGE) by anion exchange chromatography (AEC) and size exclusion chromatography (SEC). The overproduction and purification protocol yielded high quantities of the maquette proteins with no significant amounts of haem retained (figure 6.2A).

Molecular dynamics simulations were run using the methodology set out in section 2.15.2, in order to visualise the structures and conformations of the three haem-maquette constructs. As outlined in section 1.3, haem incorporation into the BT6 maquette structure requires a 55° rotation of the α-helices to enable axial bis-histidine ligation of the iron centre (Farid et al., 2013). This theoretical prediction is in apparent agreement with the structures shown in figure 6.1, where minor changes in helix orientation can be observed between the different constructs. The structures generated are included for illustrative purposes and are not intended as a substitute for experimentally determined atomic coordinates. Molecular dynamics simulations and structural visualisations were conducted by Dirk B. Auman (University of Pennsylvania, USA).

Table 6.1. Amino acid sequences of maquettes and signal peptides used in this study.

Name	Sequence*	Details
BT6	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLK	Di-haem binding
BT6M1	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLK	Single haem binding due to removal of a coordinating histidine residue
BT6M0	MGGDGENLYFQG EIWKQ**H**EDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQAEDALQKFEEALNQFEDLKQLGGSGSGSGG EIWKQ**H**EDALQKFEEALNQFEDLK	No haem binding due to removal of 2 coordinating histidine residues
TorA	MNNNDLFQAS*RR*RFLAQLGGLTVAGMLGPSLLTPRRATAAQAA	TorA signal peptide for periplasmic localization by the Tat system
TorA R12/13K	MNNNDLFQAS*KK*RFLAQLGGLTVAGMLGPSLLTPRRATAAQAA	R12K/R13K TorA signal peptide
PelB	MKYLLPTAAAGLLLLAAQPAMA	PelB signal peptide for periplasmic localization by Sec system

*Haem coordinating histidine residues are shown in bold. The twin arginine motif of the Tat signal peptide is shown in italics.

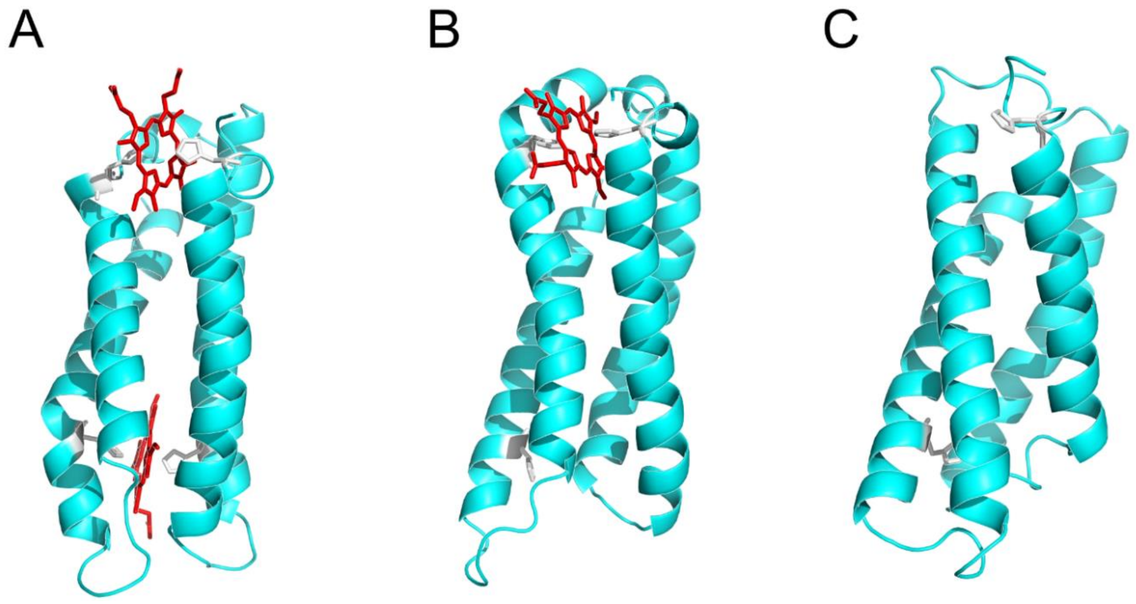


Figure 6.1 Structural models of the BT6 (A), BT6M1 (B) and BT6M0 (C) maquette proteins. Coordinating histidine residues are shown in white and the macrocycle of the haem in red. All images were generated in PyMol and taken from 50 ns trajectories.

6.3.2 *In vitro* haem incorporation

Purified *apo*-proteins were reconstituted with haem *in vitro* using a 10-fold molar excess of pigment (section 2.11.4). In a methodology similar to that used by Farid et al., (2013) unbound pigment was removed by filtration (0.2 µm pore size) and AEC (section 2.8.2), and absorption spectra of the purified haemoproteins were recorded. Following normalisation to tryptophan absorbance (280 nm), the relative intensity of the Soret band absorbance at ~400 nm progressively decreased between the BT6, BT6M1 and BT6M0 maquettes (figure 6.2B). A minor shift in the Soret band absorption maximum was observed between the BT6 and BT6M1 proteins (413 and 412 nm, respectively). A larger blue-shift was observed for the BT6M0 maquette (Soret absorption of 396 nm) which is likely to be due to a low level of adventitiously bound haem, possibly in an aggregated state. The absorption of protein-bound haem was compared to solvated states in 100 % DMSO and 50 mM HEPES (pH 7.4), with Soret absorbance maxima at 404 and 384 nm, respectively (figure 6.2C).

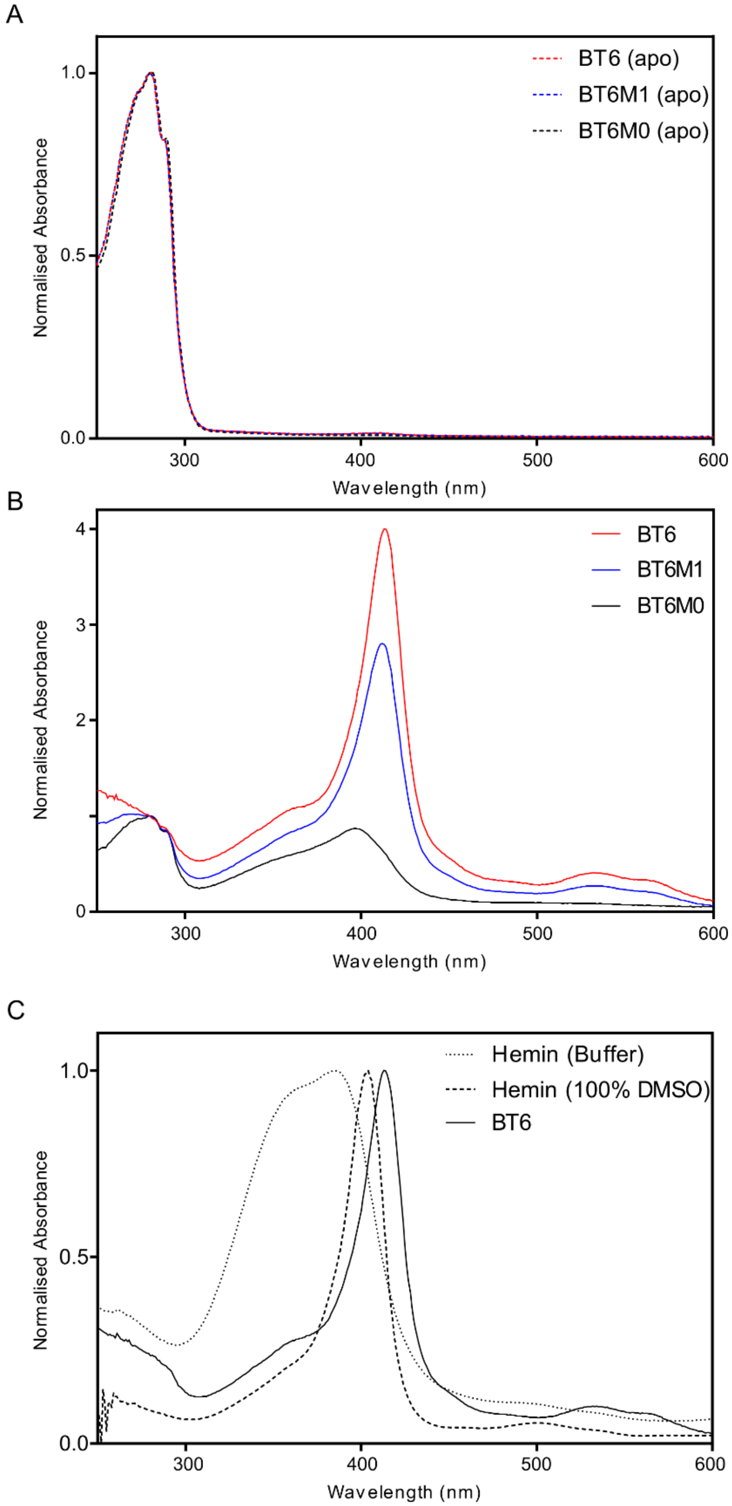


Figure 6.2 Absorption spectra of *apo*-maquettes and haem in various conditions. (A) Absorption spectra of maquette proteins, purified according to section 2.8.2; normalised to absorbance at 280 nm **(B)** Absorption spectra of *in vitro* haem-reconstituted maquette proteins; normalised to absorbance at 280 nm. (**C)** Absorption spectra of haem in aqueous buffer (dotted line), in 100 % DMSO and bound to a maquette protein. All spectra are normalised to the maximum absorbance value.

6.3.3 Thermostability of maquette variants

Temperature-dependent CD spectroscopy was used to assess what impact the removal of haem binding had on the thermal stability of the protein scaffolds (section 2.10.2) (Greenfield, 2006). Protein residue ellipticity was measured for all *apo*- and haem-bound maquette variants between 190 to 250 nm at 15 °C, with all samples displaying typically α-helical character at this temperature (figure 6.3). The temperature of the sample was raised at a rate of 1 °C min^{-1} and full CD spectra were measured at 5 °C intervals; at 80 °C all samples displayed spectra of a typically unfolded protein.

Thermal denaturation and refolding was determined by monitoring the reduction of ellipticity at 222 nm during temperature cycling from 15 to 80 to 15 °C at a rate of 1 °C min^{-1}. Boltzmann distributions were fit to the data for all samples according to section 2.14.1. For BT6, the ligation of haem dramatically increased the stabilisation of the protein, shifting the midpoint of the melting curve (T_m) from 22 to 56 °C, a 34 °C difference (figure 6.4A). Less of a shift was observed in the BT6M1 variant following haem insertion (figure 6.4B), with the T_m increasing by just 16 °C (25 to 41 °C), and no significant change was imparted on the BT6M0 protein in the presence of haem (32 to 35 °C) (figure 6.4C).

6.3.4 Assessment of haem-maquette folding by one-dimensional proton NMR spectroscopy

With evidence that haem ligation resulted in an increase in protein stability, one-dimensional protein NMR (1D ^{1}H NMR) spectroscopy was used to investigate the conformational effects of cofactor binding by comparing the folding of the *apo*-and *holo*- states. Protein samples were prepared in 5 mM sodium phosphate buffer (pH 7.4) and measured on a Bruker Avance 800 MHz spectrophotometer (see section 2.10.3 for details). Data acquisition and analysis were carried out with the help of Angus J. Robertson (University of Sheffield, UK).

In the absence of haem all maquette scaffolds displayed a characteristic ^{1}H NMR spectrum of a protein with limited tertiary structure, with undispersed, sharp resonance signals (figure 6.5A). The addition of haem to the BT6 maquette resulted in far greater dispersion of signals, most clearly observed in the amide region (5-11 ppm), consistent with a folding event following cofactor ligation (figure 6.5B). The addition of haem to the BT6M1 and BT6M0 proteins did not have a large effect on the dispersion of resonances, suggesting that the presence of cofactor had only a minor influence on the folding of these structures (figure 6.5C,D). The results described here are consistent with previously reported two-dimensional NMR data on the BT6 maquette (Farid et al., 2013).

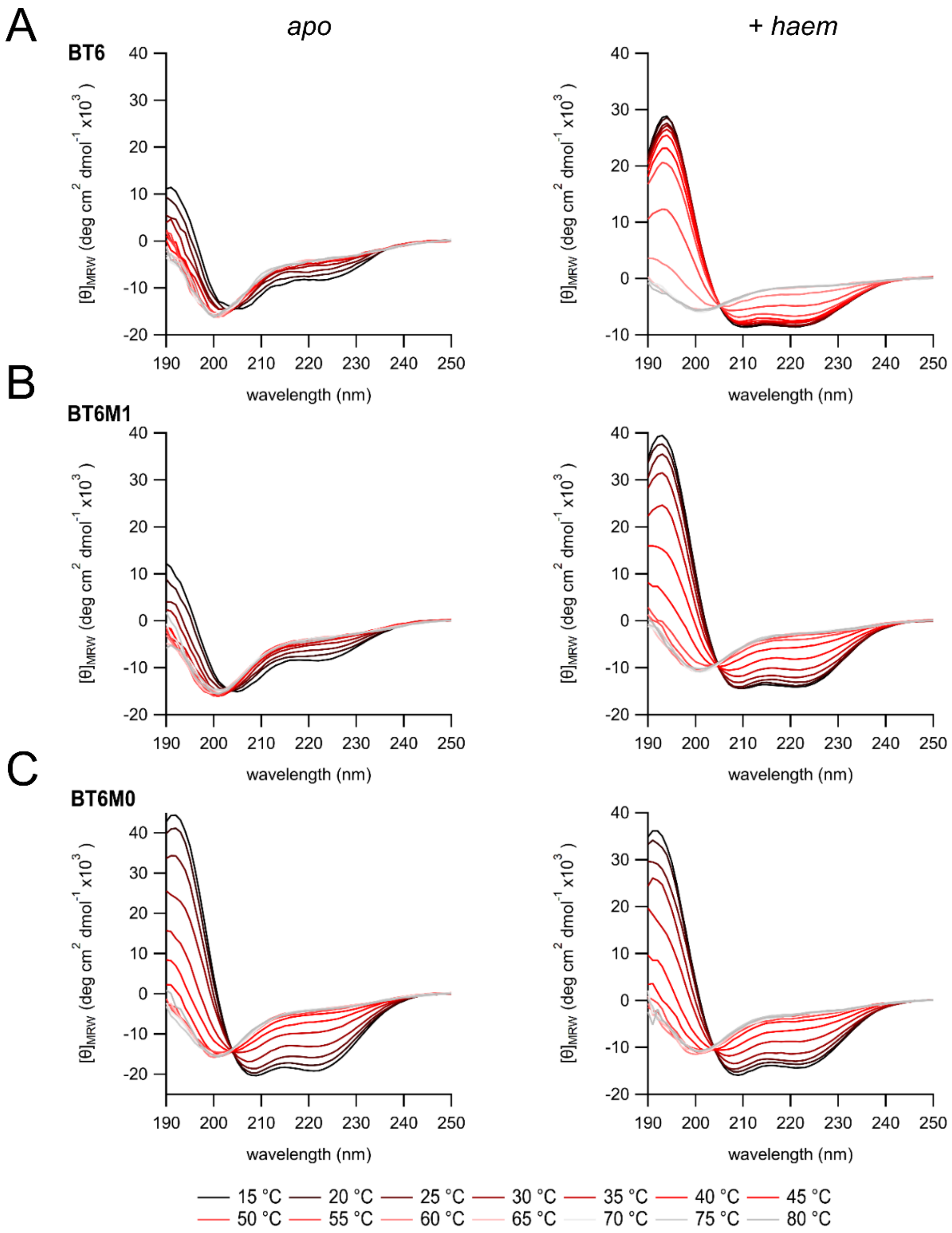


Figure 6.3 Mean residue ellipticity spectra of *apo*- and haem-bound maquette proteins. All samples show typically α-helical character at 15 °C (black lines) and spectra of an unfolded protein at 80 °C (light pink).

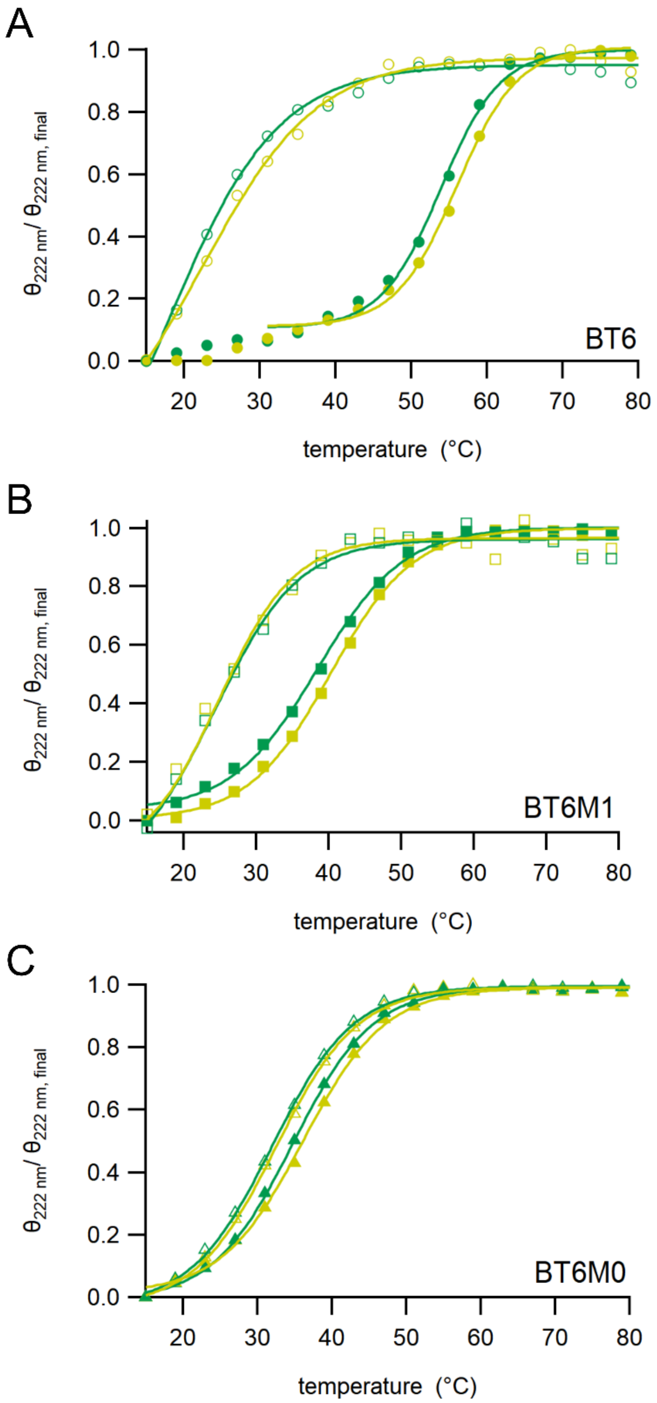


Figure 6.4 Temperature-dependent CD spectroscopy of *apo*- and haem-maquettes. Thermal denaturation (yellow) and refolding (green) of protein secondary structure, measured as the reduction in CD ellipticity at 222 nm across a temperature gradient. Open markers indicate the thermal transition of *apo*-proteins, while closed markers correspond to haem-bound proteins. Ellipticity was recorded every 1 °C but only every third data point is displayed for clarity. A Boltzmann distribution was fit for each dataset (solid lines). CD data were normalised by subtracting the minimum 222 nm ellipticity value for each dataset and dividing by the maximum ellipticity for each dataset.

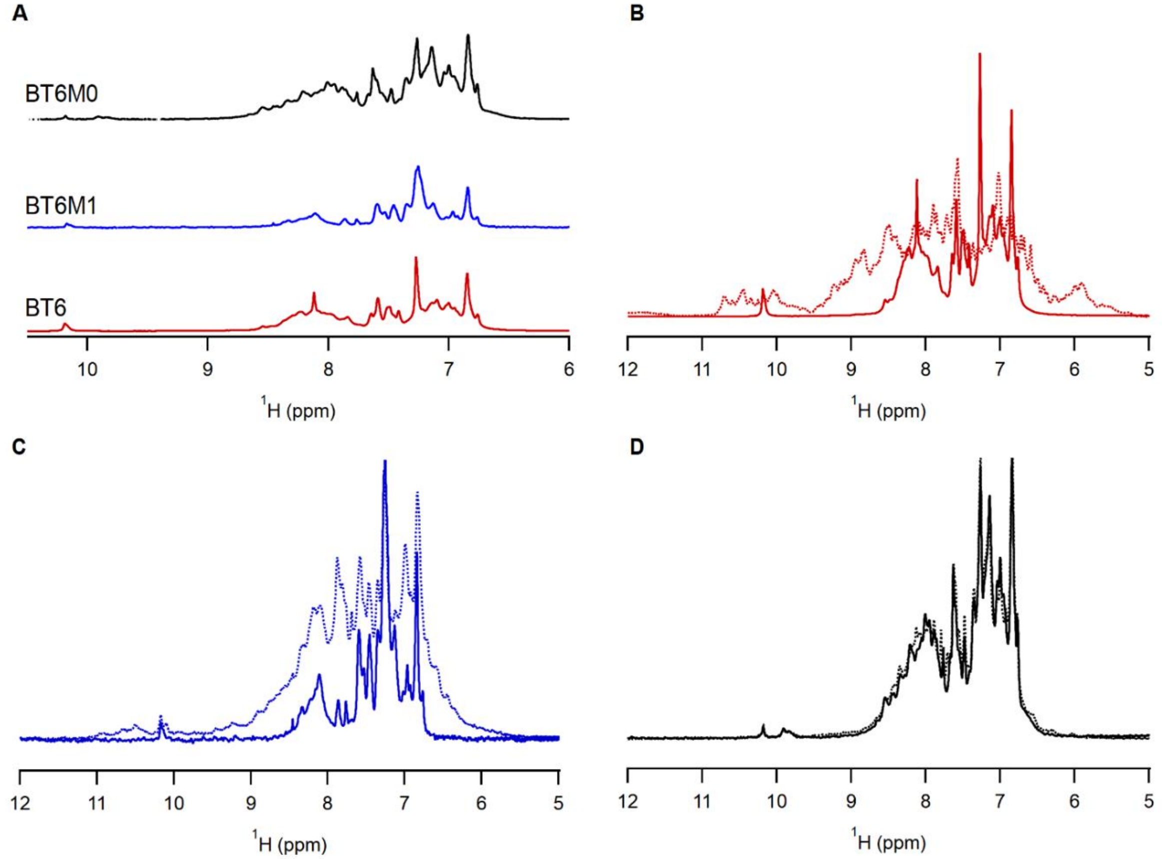


Figure 6.5 ^{1}H NMR spectra of *apo* and *holo* maquettes. (A) Spectra of *apo* BT6 (red), BT6M1 (blue) and BT6M0 (black). **(B-D)** Overlay of NMR resonances for *apo* (solid lines) and haem-bound maquettes (dashed lines). Only amide resonance signals are shown for clarity, for full resonance spectra see appendix section 4.

6.3.5 Tat-dependent export of maquette variants

Thomas et al., (2001) showed that the in-frame fusion of the TorA signal peptide to the green fluorescent protein (GFP) was sufficient for Tat mediated transport. Here, the same approach was used for the three maquette constructs according to the method described in section 2.13.1. The constructs generated consisted of the 39 residue TorA signal peptide followed by the first 4 amino acids of the mature TorA protein, which were fused in-frame to the N-terminus of each of the maquette constructs. A C-terminal 6xHis-tag was also included to enable immunoblot detection (figure 6.6). The resulting genetic fusion was cloned into the low copy number (1 to 1.5 copies per chromosome) pEXT22 plasmid vector, allowing IPTG-inducible expression under the tightly regulated control of the *tac* promoter (Liu et al., 1983; Dykxhoorn et al., 1996). The use of the pEXT22 expression vector resulted in weaker production of the maquette constructs so that the Tat system was not overwhelmed.

Escherichia coli BL21 (DE3) cells were co-transformed with the pEXT22 vector containing the gene for the desired TorA-maquette variant and an empty pET21a(+) vector. The pET21a(+) vector was included to allow immunodetection of the periplasmic β-lactamase protein, which would enable normalisation of periplasmic content between samples (figure 6.7B). The native *E*. *coli* GroEL protein was also detected by immunoblotting, allowing the relative quantities of cytoplasmic proteins to be determined (figure 6.7A). Following growth and protein production, cells were lysed to obtain cell-free extracts or subjected to osmotic shock to obtain periplasmic fractions (section 2.13.2) (Randall and Hardy, 1986).

As the signal peptide of Tat substrates is cleaved following translocation, the predicted molecular mass of cytoplasmic 'precursor' protein (~21 kDa) is larger than that of the processed 'mature' protein (~17 kDa). Thus, the relative quantities of mature and precursor protein, qualitatively determined here by the intensity of the anti-His immunoblot signal (figure 6.7C) report the efficiency of Tat export. For the TorA-BT6 variant the cell free extract displayed a strong signal for mature protein, with little or no detection of the precursor protein. Under the same growth conditions, periplasmic fractions were obtained with little evidence of cytoplasmic contamination, as assessed by the GroEL immunoblot signal. As the only signal observed in the periplasm corresponded to mature BT6 protein, it could be assumed that the presence of this lower molecular mass band was due to Tat translocation rather than cytoplasmic degradation of the signal peptide.

To confirm the mechanism of export was Tat-dependent, a signal peptide mutant was generated in which each of the conserved twin-arginine residues were substituted with lysine (TorA-KK-BT6). Immunoblots of TorA-KK-BT6 cell free extracts revealed the presence of precursor protein with no signal for mature protein. No signal was observed in the periplasmic fraction of the same cells, suggesting that the signal peptide mutant accumulates in the cytoplasm and is not translocated by the Tat machinery (or any other export mechanism). Taken together, these results suggest that the TorA signal peptide is able to direct the non-native BT6 maquette to the periplasm in a Tat-dependent manner.

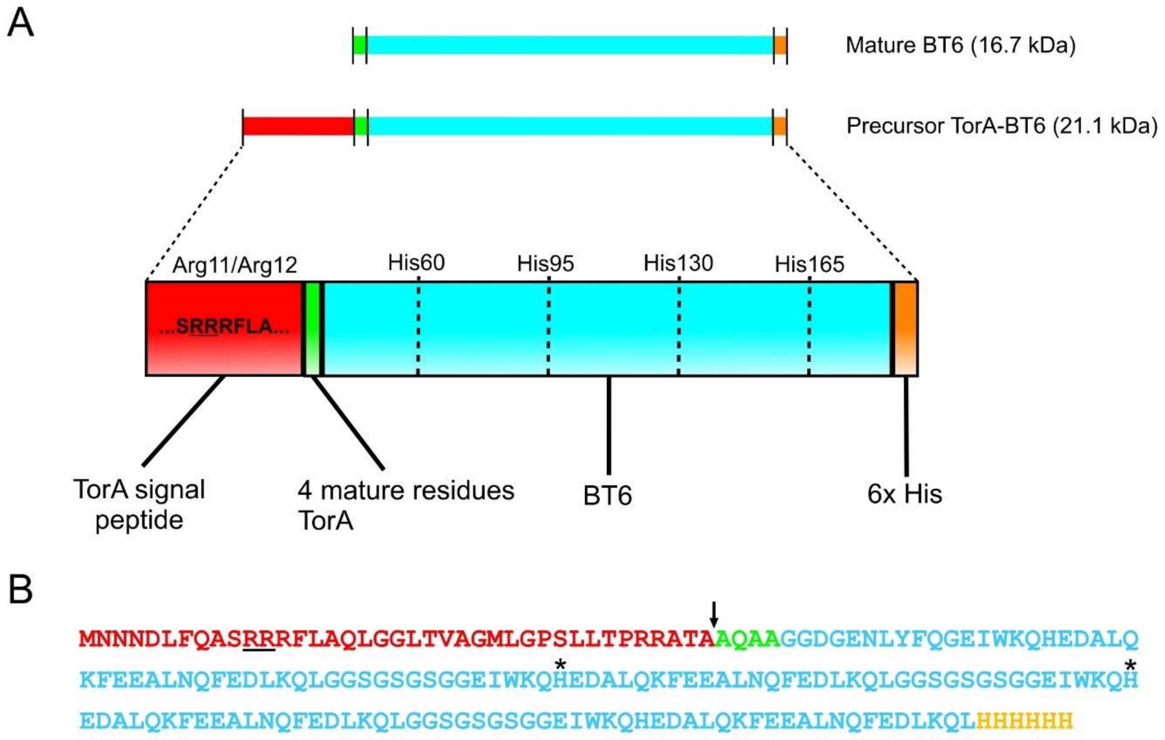


Figure 6.6 Schematic representation of the TorA-BT6 fusion. (A) The predicted molecular masses of the precursor and mature BT6 are shown in parentheses. The relative size and position of different motifs within the genetic fusion are shown in block colours. The twin-arginine residues of the signal peptide are indicated, as are the relative positions of the four haem-ligating histidine residues. **(B)** The amino acid sequence of the TorA-BT6 genetic fusion is shown, with the twin arginine motif underlined, the signal peptidase cleavage site indicated by a downward arrow and the His-Ala substitution sites shown by asterisks.

To determine whether the Tat system would also be able to recognise the folded state of the artificial proteins, the TorA signal peptide was also added to the BT6M1 and BT6M0 maquettes. This hypothesis assumes that the maquettes would bind haem *in vivo*, according to the previously discussed *in vitro* work, and that the extent of their folding would be recognised by the Tat system, with the efficiency of export reduced accordingly. Immunoblots of the TorA-BT6M0 protein showed only weak signal for mature protein in the cell free extract and periplasmic fraction, while a relatively strong precursor band was observed in cell free extracts. The one-haem binding TorA-BT6M1 maquette showed a signal corresponding to the mature protein in both the cell free extract and the periplasmic fraction, with precursor protein also present in the cell free extract.

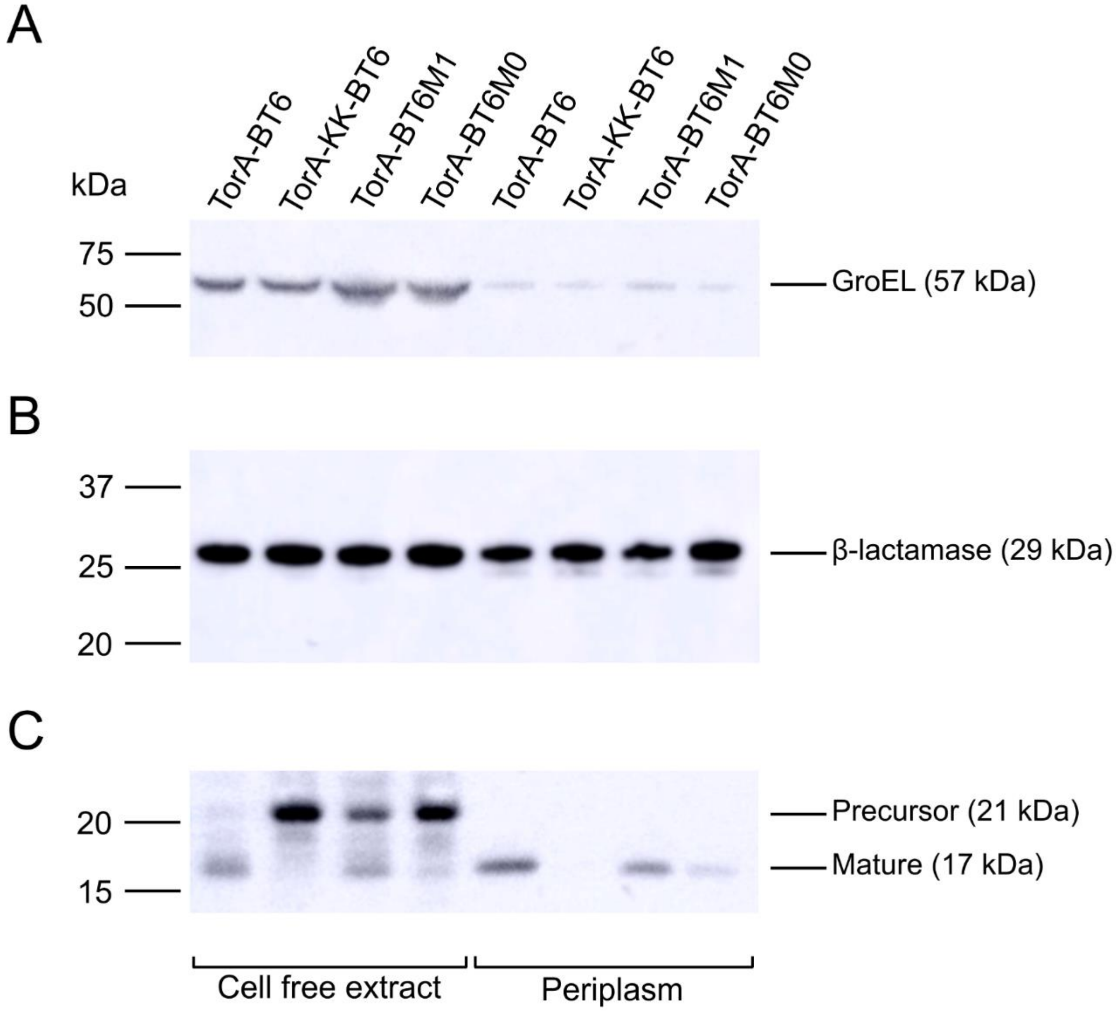


Figure 6.7 Immunodetection of TorA-BT6 variants in *E. coli* cell free extract and periplasm. (A) The relative intensity of the GroEL signal was used to assess the amount of cytoplasmic protein present in protein samples. **(B)** The amount of sample loaded was normalised to the intensity of the periplasmic β-lactamase protein. **(C)** Anti-6xHis antibodies were used to determine mass, quantity and localisation of the TorA-BT6 variants. For all panels, the positions of molecular mass markers are indicated (left) as is the expected location of each protein (right). The identity of the samples is noted above and the origin of each sample is indicated below. Each blot is representative of at least 3 repeats.

Two additional controls were run to determine whether the differential cellular localisation and abundance of the maquette variants were entirely due to Tat-selective export. The first involved the quantification of signal peptide-free maquettes under the same growth conditions used for TorA-BT6, to allow the relative levels of protein production and cytoplasmic degradation to be compared. The second control involved the genetic fusion of the *Erwinia carotovora* pectate lyase B (PelB) signal peptide to each of the BT6 maquettes (section 2.13.1) to direct the proteins to the Sec export system (Lei et al., 1987; Yoon et al., 2010). As the Sec system transports proteins in an unfolded state, the translocation of each PelB-maquette does not depend on the structural characteristics of the substrate and export for each variant should be equally efficient, so any variation in periplasmic abundance would likely be due to differential degradation in the periplasm. The results of both control experiments showed that when total protein loadings were normalised, the signals for all maquette variants were approximately equal (figure 6.8). This leads to the conclusion that the periplasmic targeting observed in figure 6.7 cannot be due to differences in expression, production or degradation between the maquettes and must result from the varying conformational flexibility of the proteins, which is apparently recognised by the Tat components.

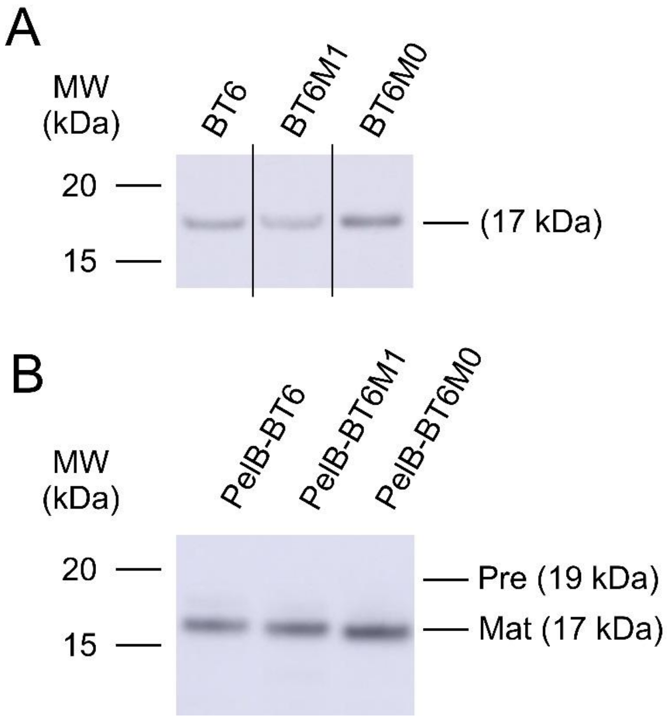


Figure 6.8 Production and degradation of maquettes, assessed by anti-His immunoblot analysis. (A) Cellular levels of signal peptide-free maquettes are shown. **(B)** Immunodetection of cell-free extracts revealed that Sec-exported PelB-fusion maquettes are equally produced and only found in their mature, processed states (molecular mass ~ 17 kDa). All samples were taken from cell-free extracts and the total amount of protein in each lane, as assessed by Bradford assay (section 2.9.1), was equal. The positions of the molecular mass markers are indicated (left) as is the predicted size of each protein (right). The vertical lines in panel A indicate where the original image has been spliced for clarity; all samples are from the same image and have not been altered.

6.3.6 Purification of haem-containing maquettes from the *E. coli* periplasm

The results described here demonstrate that the efficiency of Tat-mediated export increased with greater stability of the artificial proteins. This would only occur if the BT6 and BT6M1 were able to sequester haem from the native biosynthetic pathway, as the *apo* forms of these two proteins are less stable than BT6M0 (figure 6.3 and figure 6.5). To confirm this assertion and to determine whether the artificial proteins would retain the cofactor following export, the pEXT22 plasmid was used to produce the TorA-BT6 maquettes in the W3110 'Tat-Express' cell line (Browning et al., 2017). As described in section 1.5.7, the Tat-Express cell line has enhanced expression of Tat components due to the inclusion of the inducible pTac promoter upstream of the *tatABCD* operon.

Cell cultures were grown according to section 2.13.3 and periplasmic fractions were prepared by an osmotic lysis protocol similar to that described previously. Maquettes were subsequently purified from periplasmic fractions by IMAC and the maquette-containing elution was spin-concentrated (Vivaspin centrifugal concentrators, Sartorius) to exactly 1 mL. Striking differences in pigmentation could be observed by eye, with deep red pigmentation in the purified BT6 maquette, weaker colouring in the BT6M1 maquette and no obvious pigmentation in the BT6M0 protein (figure 6.9A inset). The absorbance spectra of the protein eluates (figure 6.9A) were similar to the spectra obtained from the *in vitro* reconstitutions (figure 6.3A). One notable difference between the *in vitro* and *in vivo* experiments is the incorporation of haem into the BT6M0 protein; while low quantities of haem seem to bind to the BT6M0 maquette *in vitro*, no haem is observable following protein purification from the cell. As the concentration of haem in the *in vitro* reconstitutions is significantly higher than that in the cytoplasm of *E. coli*, this discrepancy may be explained by the non-specific attachment of haem to BT6M0 at higher concentrations.

Interestingly, however, when the normalised absorption spectra of BT6-haem from the *in vitro* reconstitutions and from the purification from the TatExpress strain are overlaid, the traces are almost identical. This suggests that the amount of haem loaded into the BT6 scaffold is identical for both samples. In the BT6 maquette, sub-2 nM dissociation constants have been reported for haem *b* at both binding sites (Farid et al., 2013). As the concentration of haem in the *in vitro* reconstitutions was approximately 300 µM, it can be assumed that both the bis-His binding sites will be saturated with haem under these conditions. It may then be inferred that the cytoplasmic concentration of haem was also sufficient to completely saturate the maquette binding sites and the resulting purified product had its full complement of haem.

To determine whether the observed pigmentation was due to the presence of maquettes and not from other contaminants, the purity of the eluted maquettes was determined by SDS-PAGE. No visible sign of any protein species other than mature processed maquette was detected, with the quantity of each protein in the 1 mL eluates following the general trend observed previously; higher amounts of the BT6 protein compared to BT6M1, and even less BT6M0.

Small amounts of processed BT6M0 can be observed both in SDS-PAGE analysis of proteins purified from TatExpress *E*. *coli* (fig. 6.9B) and in the immunodetection assays discussed in section 6.3.6. Tat processing of the BT6M0 protein suggests that the maquette must be deemed as suitable for limited levels of export under cytoplasmic conditions, so the trend observed (BT6>M1>M0) for temperature-dependent CD and 1D ^{1}H NMR is retained but the extent of export of BT6M0 is higher than expected. This discrepancy can be explained by considering that the environment within the cytoplasm is significantly different from the weak buffering conditions required for NMR and CD analysis. The richer buffering conditions inside the cell, as well as hydrophobic or van der Waals interactions with other cellular components may stabilise the BT6M0 protein sufficiently to allow limited export through the Tat system.

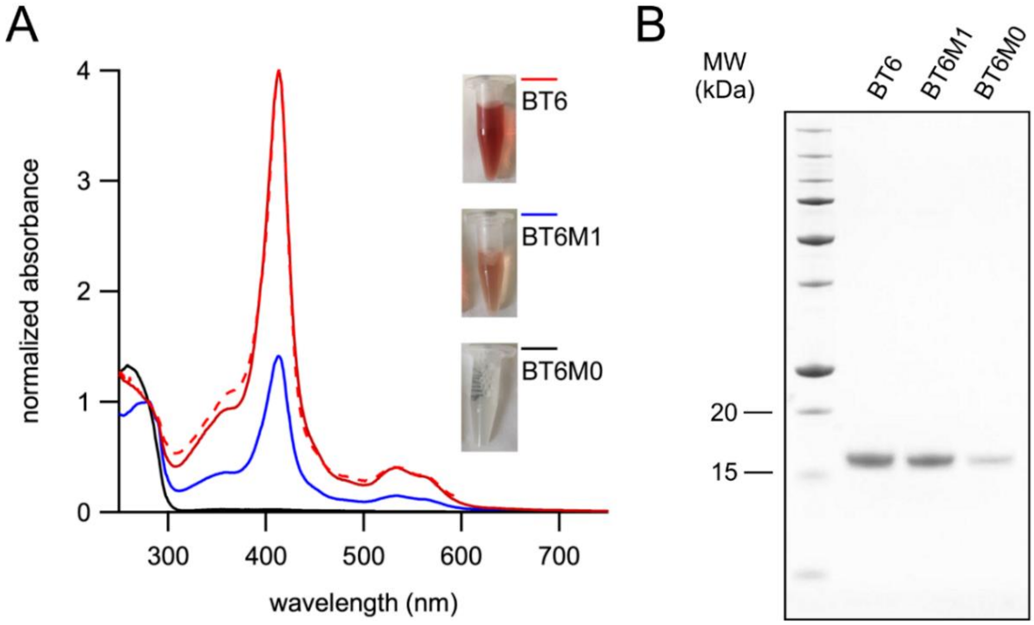


Figure 6.9 Purification of maquettes from *E. coli* TatExpress. (A) The absorption spectra of BT6 (red), BT6M1 (blue) and BT6M0 (black) maquette proteins purified from the TatExpress cell line (solid lines). The absorption spectra of haem-bound BT6 generated *in vitro* are displayed for comparison (dashed). The inset panels show the purified proteins. **(B)** SDS-PAGE analysis of IMAC eluates; the size of the molecular mass markers are indicated (left).

6.3 Discussion

The results presented in this chapter provide insights into the fundamental function of the twin-arginine translocase and the use of artificial proteins in biotechnology. The conclusions and potential future prospects from the experimental work described here will be discussed in relation to both these research areas.

6.3.1 Impact on Tat research

As described at length in section 1.5, the Tat system is of global significance due to its role in exporting proteins of importance to wide ranging biological activities, including those with roles in pathogenesis (Palmer and Berks, 2012; De Buck et at., 2008). As the mechanistic basis of substrate recognition and translocation remains poorly understood (Berks et al., 2015), new biochemical tools with which to probe the Tat system are desirable; artificially designed maquette proteins provide such a tool. As the presence of each amino acid in the protein structure is understood, the permutations of the removal or addition of residues becomes much more predictable than in natural proteins. Modifications affecting protein folding, surface hydrophobicity or cofactor binding, which would be more difficult to predict in natural protein scaffolds, can therefore be rationally incorporated into maquettes (see section 1.2).

In the work described here, artificial proteins were designed to have different cofactor binding capacities, imparted through single point mutations in residues that coordinate haem. This in-turn affected the stability of the complexes, as shown by temperature dependent CD and ^{1}H NMR. Where possible, these proteins were shown to sequester haem from the biosynthetic pathway of *E*. *coli* and fold accordingly, with the extent of this folding seemingly recognised by the native Tat machinery. Although the correlation between the folded state of a substrate and the rate of Tat-dependent export is not a new finding (DeLisa et al., 2003; Rocco et al., 2012), the determination of stability and folding in the maquette variants described here, along with the structural modelling of the protein, provides researchers with a well characterised tool to enable future experimentation.

One way in which these artificial proteins could be used to further investigate the Tat export mechanism would be to categorically determine the basis for the substrate 'proofreading' function. Currently, the mechanism by which Tat substrates are deemed suitable for export remains contentious, and could occur through sensing conformational flexibility or the recognition of hydrophobic patches. In the work presented here the maquette structures vary only by a single point mutation in the hydrophobic core, but it is likely that there are substantial conformational changes following haem ligation. As described in Farid et al., (2013) and Huang et al., (2004), the axial ligation of the haem iron centre requires substantial rotation (~ 55°) of the helical sections, in agreement with the structural models shown in figure 6.1. The design of the BT6 maquettes takes this rotational change into account in the binary positioning of hydrophobic and hydrophilic residues, so in the absence of haem the helices will not rotate and hydrophobic residues will likely be exposed to the solvent. Therefore, given the present data, the Tat recognition mechanism cannot be determined and may involve sensing of hydrophobic patches and conformational mobility in the substrate.

The use of more thermostable protein scaffolds, such as the BT64F maquette used in chapters 3 and 4, could provide further insights into the proofreading mechanism. The substitution of all four of the histidine residues in the BT6 maquette to phenylalanine residues (BT64F) significantly increased the thermostability of the protein constructs ($T_m > 90$ °C) (figure 3.3). Molecular dynamics simulations of BT64F and any other derivatives may then be used to reveal the orientation of hydrophobic amino acids in the protein structure. Such thermostable proteins provide a unique tool that is highly structured in the absence of ligand under normal cell culture conditions, with a surface likely to be made up entirely of polar residues. Variants of the maquette that have progressively larger hydrophobic regions on the surface of the protein could be generated with little effect on the stability of the structure. This experimental approach essentially eliminates the possibility of variable conformational flexibility and any differential translocation must occur through the recognition of hydrophobic patches.

6.3.2 Impact on biotechnology

To my knowledge, the work presented here is the first example of an artificial protein binding to a natively-derived cofactor and subsequently being translocated across a membrane by a host export system. The significance of this finding may have implications in biotechnology in two ways. The first application to consider is the acquisition of the cofactor without regard to the protein chassis used to translocate it to the periplasm. In this scenario the cofactor will always be separated from the protein at the final stage of processing and the export protein can be thought of solely as a means to accumulate the cofactor. In this study the accumulation and translocation of cofactor has been demonstrated for haem, but it is conceivable that such a methodology may also be applied to other molecules within the cell and/or used in alternative host organisms. This may prove particularly useful for the accumulation of cofactors that are not stable in the cytoplasm or for molecules that may cause cytoplasmic toxicity.

An example of where such small molecule binding may prove useful to synthetic biology comes from a recently reported study by Hitchcock et al., (2016). In this work, the native bacteriochlorophyll *a* biosynthetic pathway from a purple bacterium was genetically altered to produce chlorophyll *a*, which is typically found in higher plants. Significant accumulation of chlorophyll *a* was only observed following the recombinant production of the cytoplasmic water-soluble chlorophyll protein from Brussels sprout (*Brassica oleracea*), likely because chlorophyll *a* is degraded in the cytoplasm if it is not able to bind to a protein complex. Therefore, maquettes that have been appropriately redesigned for chlorophyll binding, could be used to enhance the heterologous production of chlorophyll pigments.

The second application would address the growing demand in biotechnology for folded, active recombinant proteins to be exported prior to extraction, as described in Balasundaram et al., (2009). Translocation of the protein from the cytoplasm greatly simplifies downstream processing and reduces the costs of purification. Artificial proteins purified from the periplasm, with the full complement of their required cofactors attached, would therefore provide a simpler means of product generation and greater economic viability. The oxygen-transport maquette protein described in section 1.3.6 (Koder et al., 2009) is currently showing promising results in clinical trials as a means of generating artificial blood (personal communication, Prof. P. Leslie Dutton FRS, University of Pennsylvania, USA). The similarity of the BT6 proteins described here to the oxygen transport maquettes raises the possibility of using a similar Tat-export approach to simplify the incorporation of cofactor and downstream purification of these biotechnologically important molecules.

Chapter 7

Concluding remarks and future work

7.1 Overview

The work in this thesis aimed to increase the potential functions of maquettes to address some of the current disputes in fundamental research and develop these artificial protein complexes for applications in biotechnology.

The work in chapters 3 and 4 aimed to optimise a system that would enable the incorporation of hydrophobic pigments for optical analysis. The results of these chapters suggested that the process of singlet fission (SF) was occurring in protein-bound aggregates of astaxanthin (AXT) and functionalised tetracene (eTet). The maquettes therefore provide a system to study the photophysical properties of chromophores in an identical environment and show promise towards generating a material for applications in photovoltaic devices. The work of chapter 5 showed that the maquettes were amenable to heavy-atom labelling, but the effects of selenium and iodine incorporation on the optical properties of the incorporated chromophores could not be determined here. The results of chapter 6 demonstrated that maquette proteins could effectively interface with the native export systems of E. coli and provide a valuable tool to probe the activity of the twin-arginine translocase.

As the results of each experimental chapter are reviewed within the corresponding section, the discussion here will look to avoid repetition of these specific conclusions, instead forming a broader analysis of the work undertaken introducing some possible further work.

7.2 Structural information

One of the major challenges to the work undertaken here was the lack of experimentally derived structural information on the maquette proteins. Although the atomic structure of the H10H24 maquette has been published (Huang et al., 2003), this constitutes the only experimentally derived maquette structure available to date. As the sequence of H10H24 is significantly different to the BT6 variants used here, the structures of the two maquettes may differ and, as such, molecular modelling of the BT6 variants was required (as described in chapters 3 and 6). As structural prediction software derives parameters from published protein scaffolds, the accuracy of the maquette modelling is questionable as the peptide sequences are unlike the naturally occurring proteins that constitute the overwhelming majority of published coordinates. Therefore, experimental determination of artificial protein structures is of paramount importance for future development of the maquette approach; strategies towards structural determination are outlined in section 7.2.1.

Arguably of greater interest would be to determine the structural arrangement of the maquette-bound pigment aggregates discussed in chapters 3-5. For example, the structural determination of light-harvesting complexes, such as the LH2 antenna complex in *Rhodoblastus acidophila* (formerly *Rhodopseudomonas acidophila*) (figure 7.2) (Cogdell et al., 1997), enabled accurate correlations to be derived between energy-transfer processes and the arrangement of chromophores. Such discoveries were critical to understanding light absorption and energy transfer in photosynthesis, tying together theoretical predictions with empirical data. Experimentally derived correlations between the photophysical properties and structure of AXT agg II, aggregated C8-BTBT and aggregated tetracene have yet to be described. The maquette approach therefore provides a means to solve the structure of such aggregates, allowing conclusions to be drawn between the arrangement of the chromophores and the extensive theoretical and spectroscopic data available in the literature.

7.2.1 Structural redesign

As mentioned in chapter 3, the structural determination of BT6 maquettes been unsuccessful due to issues with crystal formation and poor X-ray scattering (personal communication, Prof. P. Leslie Dutton, University of Pennsylvania, USA). This is believed to be due to structural heterogeneity in the BT6 proteins, where the positioning of the helices may change relative to one another (illustrated in figure 7.1). The binary patterning in maquettes usually ensures that the *a* and *d* positions of the α-helical heptad repeat face in towards the core, but in the BT6 variants no additional structural features are present to specify the relative positioning of each helical section. Structural homogeneity may be imparted through the introduction of electrostatic interactions between the charged surface residues of each helical section (Moser et al., 2016). For example, by ensuring that the *c* position in the heptad is glutamate (negatively charged at neutral pH) in helices 1 and 3, and lysine (positively charged at neutral pH) in helices 2 and 4, the opposing charges will form a salt bridge, forcing the protein into a homogeneous conformation. The introduction of charge patterns on the surface of the BT6 protein may result in sufficient homogeneity to permit appropriate crystal formation and structural determination.

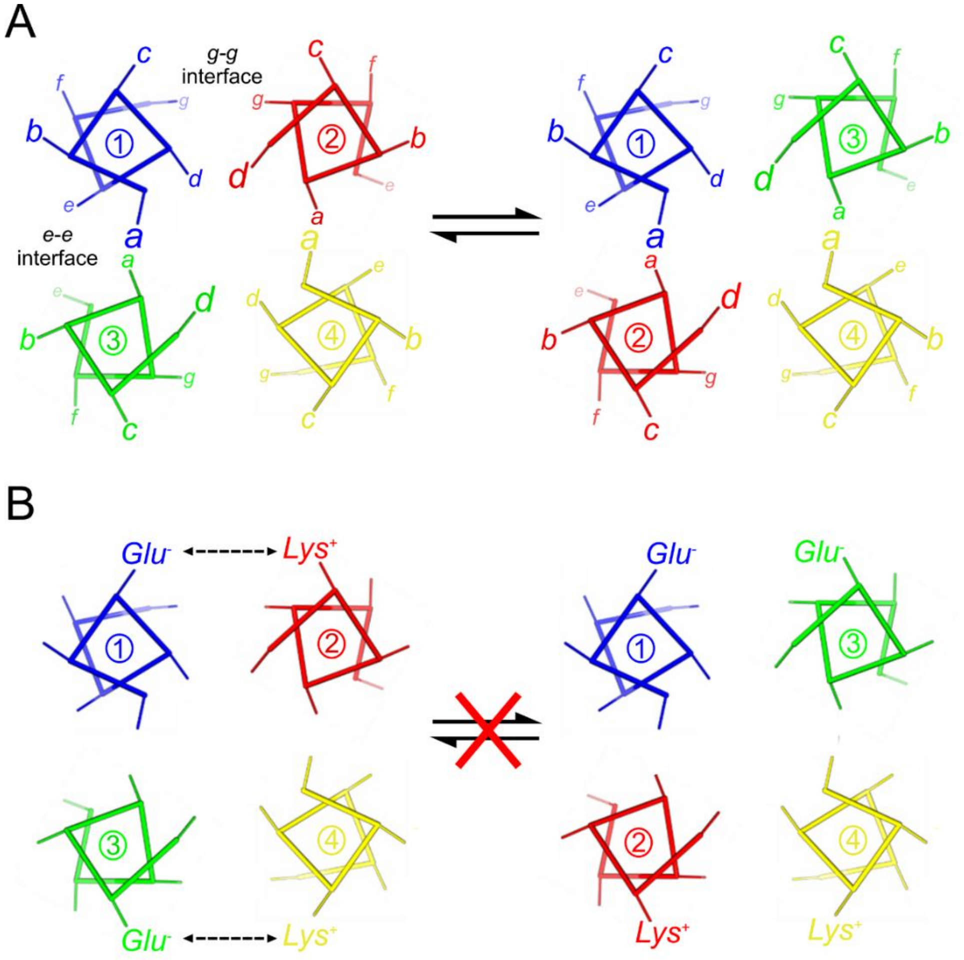


Figure 7.1 Helical redesign strategy. (A) In the BT6 maquettes the binary patterning does not specify any helix-helix interactions other than the hydrophobic packing of the *a* and *d* positions. Therefore, the relative positions of the helices may alter; for example, the *g* interface of helix 1 may coordinate with the *g* interface of either helix 2 or 3. **(B)** By specifying the charge of the amino acid in the *c* positions it is possible to force a homogeneous conformation onto the protein scaffolds through the formation of salt bridges. Adapted with permission from Ennist et al., (2017a).

7.2.2 Simplification of maquette designs for hydrophobic pigment insertion

As described in chapter 1, part of the justification for the maquette approach to protein engineering was that the 'bottom-up' design circumvented the use of natural protein sequences, which contain high levels of complexity due to the relics of evolutionary processes. As the BT6 maquettes were originally designed to investigate the redox properties of coordinated tetrapyrroles, it could be argued that the protein designs in chapters 3-5 contain the relics of previous experiments and are more complex than required for simple hydrophobic partitioning. While it was prudent for such proof-of-concept work to be established in a well-characterised system, the simplification of the maquette sequences may be necessary to address some of the issues raised in those chapters.

The seminal work on *de novo* protein design was conducted by generating a tetrahelical bundle protein containing only three amino acids; leucine, lysine and glutamate (Regan and DeGrado, 1988; DeGrado et al., 1989). Beginning with this simple structure, analysis of hydrophobic pigment incorporation may be conducted as described in chapters 3-5, allowing the requirements for pigment incorporation to be determined through a bottom-up approach. This may identify the structural requirements necessary for the incorporation of only one hydrophobic molecule or more simply generate a scaffold with appropriate homogeneity for structural determination.

One significant complication in the BT6 maquette designs, for the purposes of hydrophobic pigment incorporation, are the alterations in heptad patterning that account for the helical rotation caused by haem ligation. As described in section 1.3.2, the ligation of haem into the maquette structure causes ~50° rotation of the helices relative to one another, with the binary patterning of hydrophobic and polar residues adjusted to compensate for this change. While the modelling of the BT64A maquette showed that the substitution of all four polar histidine residues for hydrophobic alanine induced some helical rotation (data not shown), permanent or transient exposure of hydrophobic residues may occur and could provide a binding site for hydrophobic pigments on the protein surface. While this is considered an unlikely binding mechanism from the data shown in this thesis, the incorporation of hydrophobic molecules into a maquette with more strictly ordered patterning may be necessary to elucidate whether such molecules are binding to exposed residues on the surface or within the protein.

7.3 Pigment photostability

One of the primary aims for the work undertaken in chapters 3 and 4 was to improve the photooxidative stability of the maquette-bound pigments. While some improvement was granted in the stability of both AXT and eTet, the observations made in the latter (section 4.4) suggested that the maquette proteins may simply function to prevent the formation of large featureless aggregates. Regardless of the mechanism of improved photostability, the extended decay lifetimes of both materials remain far shorter than required for applications in photovoltaic (PV) devices. In this section, two possible methods of improving the photooxidative stability of the maquette-bound chromophores will be discussed.

7.3.1 Improved binding

Provided that the strength of cofactor binding proves to be an important determinant for the stability of maquette-bound chromophores, improvements may be made through alterations to protein sequence or the composition of the aqueous buffer. This latter point assumes that changes to the buffering conditions may reduce the dissociation constant (K_d), by 'forcing' the hydrophobic pigments into the maquettes and reducing the propensity for the chromophores to dissociate and form larger aggregates in solution.

For eTet, the rate of aggregate formation was determined through the presence of light-scatter in the absorption spectra of the chromophore in 20 % DMSO, following a 30 min incubation at different temperatures. As described in section 4.4.2, such analysis could also be translated into a 96-well plate format, allowing many different conditions to be assessed. This simple readout for the aggregation rate in solution may also be used to determine the strength of pigment binding in the maquettes. For example, the protein-pigment complex may be placed into each well of a 96-well plate with conditions such as the protein sequence, pH, salt concentration, glycerol concentration and the buffering system altered. The absorption spectra of each well could be measured over time to determine the presence of light-scattering, with the strongest binding resulting in the least scatter.

7.3.2 Trehalose glass

Another technique that may improve the photooxidative stability of the materials is to encase the protein pigment complexes in a trehalose glass. Trehalose is used by biological systems to protect proteins under dehydrating conditions and has been used in spectroscopy as a tool to create optically clear glass structures that encapsulate proteins in their native state (Hagen et al., 1995;

Wright et al., 2002). If the glass is formed under anoxic conditions the chromophores will be completely restricted of oxygen as this cannot penetrate the glass. Trehalose glass is also purported to rigidly maintain protein tertiary structure meaning that the incorporated molecules would not be able to dissociate following glass formation and presumably retain their aggregated state.

7.4 Alternative functions

The maquettes described here display a narrow range of functions compared to the abundance of roles facilitated by naturally occurring proteins. While the premise of the maquette approach asserts that any naturally occurring function may also be replicated through *de novo* designed proteins, the scope of the discussion in this section will be limited to alternative functions that readily extend from the experimental work conducted in this thesis.

7.4.1 Artificial light harvesting functions

The results of this thesis and those in Farid et al., (2013) demonstrate that maquette proteins are able to facilitate the insertion of light-active Zn-chlorins and carotenoids. The potential combination of such molecules into maquette structures raises the possibility of generating artificial light-harvesting proteins reminiscent of the $\alpha\beta$ subunits of the LH2 antenna protein in purple bacteria (figure 7.2, inset).

Artificial systems allow greater control of the positioning of chlorins and may be used to access wavelengths of light that are not available to native proteins. However, several issues are readily apparent in the generation these complexes given the current data. One obvious issue is that carotenoids bind to the maquettes in an aggregated state, but as a single molecule in the $\alpha\beta$ monomer. As the carotenoids in LH2 participate in both light-harvesting and photoprotection, it is not clear how the aggregated state may affect either process, particularly in relation to the rapid SF observed in chapter 3. Bis-His ligation of two haem molecules into the BT6 structure prevented the insertion of AXT, most likely through steric clashes within the protein core (section 3.3.6) and thus, it is unclear whether the attachment of Zn-chlorins would result in the same effect. Although the insertion of Zn-chlorins only requires a single histidine residue for attachment, the presence of two macrocycles at either end of the molecule, as with haem in BT6, may block carotenoid partitioning. A maquette may be designed to attach one Zn-chlorin at one end of the cavity with the other end able to sequester carotenoids; this however raises the question of how close the carotenoids and tetrapyrroles may be positioned, which would also affect the desired energy-transfer functions of the complex.

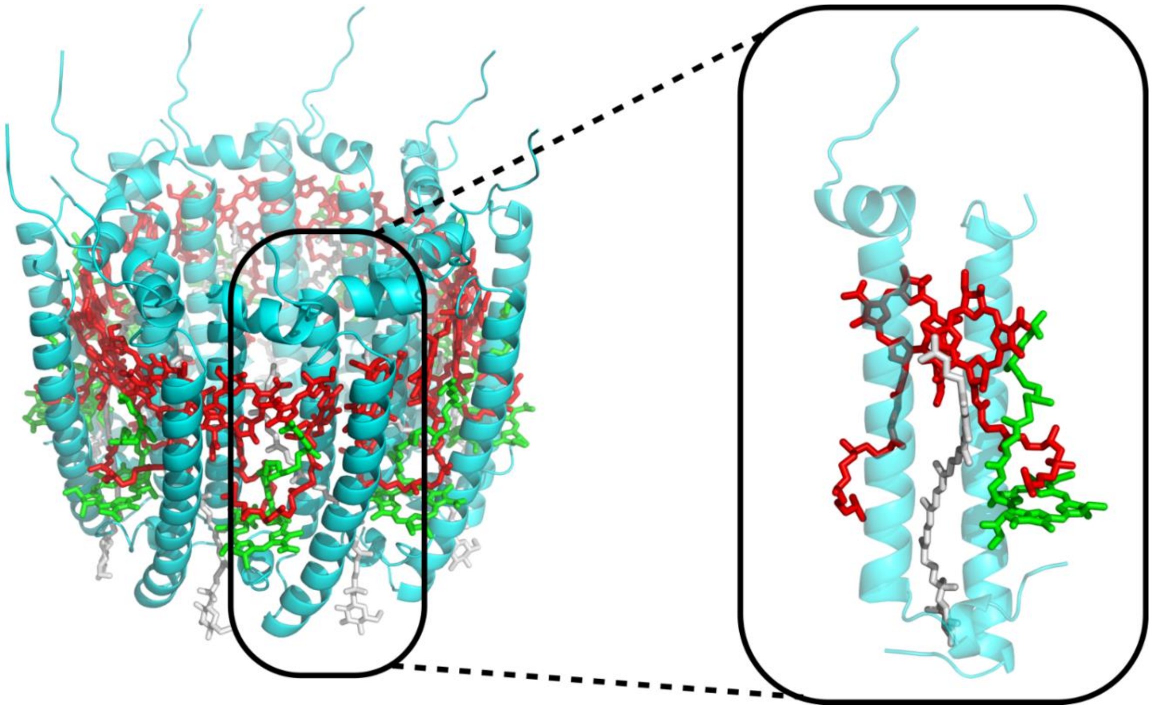


Figure 7.2 Structure of the light-harvesting 2 (LH2) antenna protein from *Rhodoblastus acidophilus* (formerly *Rhodopseudomonas acidophila*). The atomic structure of the LH2 protein is shown in cyan, carotenoids shown in white, with the B850 and B800 bacteriochlorophylls shown in red and green, respectively. A representation of a monomeric αβ-subunit is shown inset. Structural representations were generated in PyMol (PDB ID: 1NKZ).

7.4.2 Insertion of other hydrophobic pigments

In the past 10 years, the number of materials shown to undergo SF has increased dramatically (Rao and Friend, 2017; Casanova, 2018). The molecules assessed in this thesis therefore capture the properties of only a fraction of the available materials (as highlighted in figure 7.3). It is conceivable that many of these structures may be incorporated into the maquette proteins described here, possibly used in combinations to permit broad light-harvesting capabilities in PV devices.

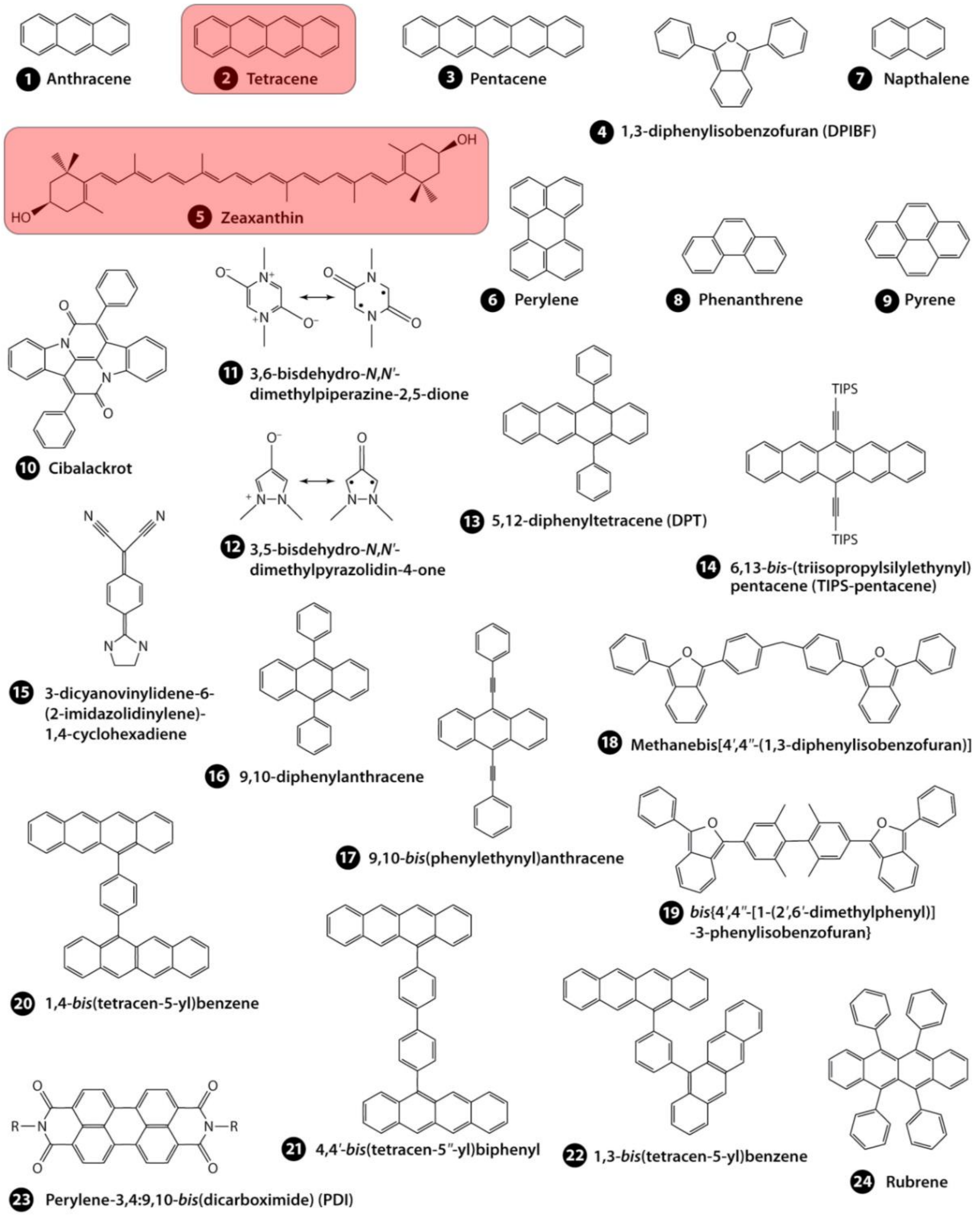


Figure 7.3 The chemical structure of some singlet fission materials. A selection of SF materials with the molecule classes used in this study highlighted in red. Diagram adapted with permission from Smith and Michl, 2013.

7.4.3 Tat-mediated export of carotenoid-maquettes

The results described in chapter 6 demonstrate that the production of maquettes in *E. coli* could lead to the *in vivo* incorporation of cofactors and purification the *holo*-protein complex following export to the periplasm. This capability could be explored further to permit the incorporation of carotenoids and other tetrapyrroles from the biosynthetic pathways of the host organism. For example, this notion could be employed in a strains of *E. coli* that have been engineered to accumulate a variety of different carotenoids (Lee et al., 2003) or through heterologous expression in bacteria that accumulate carotenoids as part of their native metabolism, such as *Rhodobacter sphaeroides* (Chi et al., 2015).

7.4.4 Attachment to surfaces

The site-specific attachment of proteins to a range of surfaces is now routine in biochemistry (Vasilev et al., 2014) and has recently been demonstrated for light-active BT6 maquettes and gold surfaces (Lishchuk et al., 2018). The two-dimensional arrangement of light-active redox proteins gives a highly controlled tool to investigate energy transfer processes and may be used here to investigate how native or native-like complexes may interact with the proteins described in chapters 3-5. Another possibility is the two- and three-dimensional arrangement of various maquette complexes on glass or silicon surfaces (unpublished work, Dr. Cvetelin Vasilev, University of Sheffield, UK) for applications in photovoltaic devices. This would enable the absorption of a range of wavelengths with any intermolecular energy-transfer activities controlled by the arrangement of the maquette proteins.

7.5 Conclusions

- The maquette approach to protein design provides a system to predictably alter the structural and functional properties of a peptide.
- Maquettes may be employed to investigate the fundamental mechanisms underlying many important biological processes or used to confer entirely novel protein functions that are not observed in nature.
- The incorporation of AXT into maquettes was demonstrated in chapter 3, with transient absorption spectroscopy revealing that the excited state dynamics were similar to a previously reported AXT solution aggregate.
- The incorporation of other carotenoids was also demonstrated, which will enable assessment of the electronic character of these molecules under identical conditions, allowing firm comparisons to be drawn between the chemical structure of the molecules and their optical properties.
- This may be used to explain the roles of individual carotenoids in relation to their biological functions, but also to describe the process of SF in a variety of chromophores, perhaps allowing identification of a carotenoid aggregate with appropriate energetics for applications in photovoltaic devices.
- The partitioning of eTet into maquettes demonstrated that the incorporation of hydrophobic pigments is not limited to carotenoids, resulting in the first reported tetracene-protein complex.
- The eTet maquette was highly soluble in aqueous buffer, which enabled the characterisation of well-established SF aggregates in solution for the first time.
- The results of chapter 5 demonstrated that the maquettes were amenable to heavy-atom labelling, with the incorporation of iodine and selenium confirmed by mass spectrometry. Such proteins may be used to impart a protein mediated heavy-atom effect on the bound chromophores, but the results presented here could not confirm this hypothesis.
- The final experimental chapter demonstrated that maquette proteins could interface with native biological export systems and form an effective tool to probe the activity of the twin-arginine translocase.
- The *in vivo* attachment of haem into the BT6 protein demonstrates the ability of maquettes to form cofactor-containing complexes within the host organism, which may potentially be applied to facilitate the incorporation of other biomolecules.

Chapter 8

References

Alanen, H.I., Walker, K.L., Lourdes Velez Suberbie, M., Matos, C.F.R.O., Bönisch, S., Freedman, R.B., Keshavarz-Moore, E., Ruddock, L.W., and Robinson, C. (2015). Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the Escherichia coli Tat pathway in the absence of prior disulfide bond formation. Biochim. Biophys. Acta - Mol. Cell Res. *1853*, 756–763.

Alcock, F., Baker, M.A.B., Greene, N.P., Palmer, T., Wallace, M.I., and Berks, B.C. (2013). Live cell imaging shows reversible assembly of the TatA component of the twin-arginine protein transport system. Proc. Natl. Acad. Sci. *110*, E3650–E3659.

Anderson, J.L.R., Armstrong, C.T., Kodali, G., Lichtenstein, B.R., Watkins, D.W., Mancini, J. A, Boyle, A.L., Farid, T.A, Crump, M.P., Moser, C.C., et al. (2014). Constructing a man-made c-type cytochrome maquette in vivo: electron transfer, oxygen transport and conversion to a photoactive light harvesting maquette. Chem. Sci. *5*, 507–514.

Anthony, J.E. (2008). The larger acenes: Versatile organic semiconductors. Angew. Chemie Int. Ed. *47*, 452–483.

Arnold, F.H. (2015). The nature of chemical innovation: New enzymes by evolution. Q. Rev. Biophys. *48*, 404–410.

Austin, R.H., Baker, G.L., Etemad, S., and Thompson, R. (1989). Magnetic field effects on triplet exciton fission and fusion in a polydiacetylene. J. Chem. Phys. *90*, 6642–6646.

Bageshwar, U.K., Whitaker, N., Liang, F.C., and Musser, S.M. (2009). Interconvertibility of lipid- and translocon-bound forms of the bacterial Tat precursor pre-Sufl. Mol. Microbiol. *74*, 209–226.

Balasundaram, B., Harrison, S., and Bracewell, D.G. (2009). Advances in product release strategies and impact on bioprocess design. Trends Biotechnol. *27*, 477–485.

Baldo, M.A., O'Brien, D.F., You, Y., Shoustikov, A., Sibley, S., Thompson, M.E., and Forrest, S.R. (1998). Highly efficient phosphorescent emission from organic electroluminescent devices. Nature *395*, 151–154.

Baldo, M.A., Thompson, M.E., and Forrest, S.R. (2000). High-efficiency fluorescent organic light-emitting devices using a phosphorescent sensitizer. Nature *403*, 750–753.

Bale, J.B., Gonen, S., Liu, Y., Sheffler, W., Ellis, D., Thomas, C., Cascio, D., Yeates, T.O., Gonen, T., King, N.P., et al. (2016). Accurate design of megadalton-scale two-component icosahedral protein complexes. Science. *353*, 389–394.

Balevičius, V., Abramavicius, D., Polívka, T., Galestian Pour, A., and Hauer, J. (2016). A unified picture of S∗ in carotenoids. J. Phys. Chem. Lett. *7*, 3347–3352.

Barber, J. (2006). Photosystem II: an enzyme of global significance. Biochem. Soc. Trans. *34*, 619–631.

Bardeen, C.J. (2014). Bringing dark states to light. Nat. Mater. *13*, 1001–1003.

Barford, W. (2005). Electronic and Optical Properties of Conjugated Polymers. (New York, USA: Oxford University Press).

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks to transfer and process information. Nature *387*, 913–917.

Beljonne, D., Shuai, Z., Pourtois, G., and Bredas, J.L. (2001). Intersystem crossing in conjugated polymers: A configuration interaction description. J. Phys. Chem. A *105*, 3899–3907.

Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005). Prediction of twin-arginine signal peptides. BMC Bioinformatics *6*, 1–9.

Berks, B.C. (1996). A common export pathway for proteins binding complex redox cofactors? Mol. Microbiol. *22*, 393–404.

Berks, B.C. (2015). The twin-arginine protein translocation pathway. Annu. Rev. Biochem. *84*, 843–864.

Berks, B.C., Lea, S.M., and Stansfeld, P.J. (2014). Structural biology of tat protein transport. Curr. Opin. Struct. Biol. *27*, 32–37.

Bierzynski, A, Kim, P.S., and Baldwin, R.L. (1982). A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A. Proc. Natl. Acad. Sci. U. S. A. *79*, 2470–2474.

Billsten, H.H., Sundström, V., and Polívka, T. (2005). Self-assembled aggregates of the carotenoid zeaxanthin: Time-resolved study of excited states. J. Phys. Chem. A *109*, 1521–1529.

Blankenship, R.E. (2014). Molecular Mechanisms of Photosynthesis. (West Sussex, UK: John Wiley and Sons, Ltd)

Bouas-Laurent, H., and Dürr, H. (2001). Organic photochromism (IUPAC Technical Report). Pure Appl. Chem. *73*, 639–665.

Bradford, M.M. (1976). Rapid and sensitive method for quantitation of microgram quantities of protein utilizing principle of protein-dye binding. Anal. Biochem. *72*, 248–254.

Brédas, J.L., Beljonne, D., Coropceanu, V., and Cornil, J. (2004). Charge-transfer and energy-transfer processes in π-conjugated oligomers and polymers: A molecular picture. Chem. Rev. *104*, 4971–5003.

Britton, G., Liaaen-Jensen, S., and Pfander, H. (1995) Carotenoids today and challenges for the future. In Carotenoids, Britton, G., Liaaen-Jensen, S., Pfander, H., eds. (Boston, USA: Birkhauser) pp. 13-20.

Browning, D.F., Richards, K.L., Peswani, A.R., Roobol, J., Busby, S.J.W., and Robinson, C. (2017). *Escherichia coli* 'TatExpress' strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. Biotechnol. Bioeng. *114*, 1–9.

Budisa, N., Karnbrock, W., Steinbacher, S., Humm, A, Prade, L., Neuefeind, T., Moroder, L., and Huber, R. (1997). Bioincorporation of telluromethionine into proteins: a promising new approach for X-ray structure analysis of proteins. J. Mol. Biol. *270*, 616–623.

Burdett, J.J., Müller, A.M., Gosztola, D., and Bardeen, C.J. (2010). Excited state dynamics in solid and monomeric tetracene: The roles of superradiance and exciton fission. J. Chem. Phys. *133*.

Burdett, J.J., and Bardeen, C.J. (2013). The dynamics of singlet fission in crystalline tetracene and covalent analogs. Acc. Chem. Res. *46*, 1312–1320.

Burroughes, J.H., Bradley, D.D.C., Brown, A.R., Marks, R.N., Mackay, K., Friend, R.H., Burns, P.L., and Holmes, A.B. (1990). Light-emitting diodes based on conjugated polymers. Nature *347*, 539–541.

Busby, E., Xia, J., Wu, Q., Low, J.Z., Song, R., Miller, J.R., Zhu, X.Y., Campos, L.M., and Sfeir, M.Y. (2015). A design strategy for intramolecular singlet fission mediated by charge-transfer states in donor-acceptor organic materials. Nat. Mater. *14*, 426–433.

Caffrey, M., and Cusanovich, M. (1994). Site-specific mutagenesis studies of cytochromes c. Biochim. Biophys. Acta *1187*, 277–288.

Casanova, D. (2018). Theoretical modeling of singlet fission. Chem. Rev. *118*, 7164–7207.

Cerezo, J., Zúñiga, J., Bastida, A., Requena, A., and Pedro Cerón-Carrasco, J. (2013). Conformational changes of β-carotene and zeaxanthin immersed in a model membrane through atomistic molecular dynamics simulations. Phys. Chem. Chem. Phys. *15*, 6527–6538.

Cerullo, G., Polli, D., Lanzani, D., De Silvestri, S., Hashimoto, H., and Cogdell, R.J. (2002). Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State. Science *298*, 2395–2398.

Chaddock, A.M., Mant, A., Karnauchov, I., Brink, S., Herrmann, R.G., Klösgen, R.B., and Robinson, C. (1995). A new type of signal peptide: central role of a twin-arginine motif in transfer signals for the delta pH-dependent thylakoidal protein translocase. EMBO J. *14*, 2715–2722.

Chen, G.E., Canniffe, D.P., Barnett, S.F.H., Hollingshead, S., Brindley, A.A., Vasilev, C., Bryant, D.A., and Neil Hunter, C. (2018). Complete enzyme set for chlorophyll biosynthesis in Escherichia coli. Sci. Adv. *4*, 1–9.

Chi, S.C., Mothersole, D.J., Dilbeck, P., Niedzwiedzki, D.M., Zhang, H., Qian, P., Vasilev, C., Grayson, K.J., Jackson, P.J., Martin, E.C., et al. (2015). Assembly of functional photosystem complexes in Rhodobacter sphaeroides incorporating carotenoids from the spirilloxanthin pathway. Biochim. Biophys. Acta *1847*, 189–201.

Choma, C.T., Lear, J.D., Nelson, M.J., Dutton, P.L., Robertson, D.E., and DeGrado, W.F. (1994). Design of a heme binding four helix bundle. J. Am. Chem. Soc. *116*, 856–865.

Christensson, N., Žídek, K., Magdaong, N.C.M., Lafountain, A.M., Frank, H.A., and Zigmantas, D. (2013). Origin of the bathochromic shift of astaxanthin in lobster protein: 2D electronic spectroscopy investigation of β-crustacyanin. J. Phys. Chem. B *117*, 11209–11219.

Claydon, J., Greeves, N., Warren, S. (2012) Organic chemistry. (New York, USA: Oxford University Press Inc.)

Cline, K., Ettinger, W.F., and Theg, S.M. (1992). Protein-specific energy requirements for protein transport across or into thylakoid membranes. Two lumenal proteins are transported in the absence of ATP. J. Biol. Chem. *267*, 2688–2696.

Cline, K., and Mori, H. (2001). Thylakoid ΔpH-dependent precursor proteins bind to a cpTatC-Hcf106 complex before Tha4-dependent transport. J. Cell Biol. *154*, 719–729.

Cogdell, R.J., Land, E.J., and Truscott, T.G. (1983). The triplet extinction coefficients of some bacterial carotenoids. Photochem. Photobiol. *38*, 723–725.

Cogdell, R.J., Isaacs, N.W., Freer, A.A., Arrelano, J., Howard, T.D., Papiz, M.Z., Hawthornthwaite-Lawless, A.M., and Prince, S. (1997). The structure and function of the LH2 (B800-B850) complex from the purple bacterium Rhodopseudomonas acidophila strain 10050. Prog Biophys Mol Biol. *68*, 1–27.

Cohen-Ofri, I., Van Gastel, M., Grzyb, J., Brandis, A., Pinkas, I., Lubitz, W., and Noy, D. (2011). Zinc-bacteriochlorophyllide dimers in de novo designed four-helix bundle proteins. A model system for natural light energy harvesting and dissipation. J. Am. Chem. Soc. *133*, 9526–9535.

Collman, J.P., Boulatov, R., Sunderland, C.J., and Fu, L. (2004). Functional analogues of cytochrome c oxidase, myoglobin, and hemoglobin. Chem. Rev. *104*, 561–588.

Conlan, B., Cox, N., Su, J.H., Hillier, W., Messinger, J., Lubitz, W., Dutton, P.L., and Wydrzynski, T. (2009). Photo-catalytic oxidation of a di-nuclear manganese centre in an engineered bacterioferritin “reaction centre.” Biochim. Biophys. Acta - Bioenerg. *1787*, 1112–1121.

Csete, M., and Doyle, J. (2002). Reverse engineering of biological complexity. Science *295*, 1664–1669.

Dahiyat, B.I., Sarisky, C.A., and Mayo, S.L. (1997). De novo protein design: towards fully automated sequence selection. J. Mol. Biol. *273*, 789–796.

De Buck, E., Lammertyn, E., and Anné, J. (2008). The importance of the twin-arginine translocation pathway for bacterial virulence. Trends Microbiol. *16*, 442–453.

De Buck, E., Maes, L., Meyen, E., Van Mellaert, L., Geukens, N., Anné, J., and Lammertyn, E. (2005). Legionella pneumophila Philadelphia-1 tatB and tatC affect intracellular replication and biofilm formation. Biochem. Biophys. Res. Commun. *331*, 1413–1420.

De Keyzer, J., van der Does, C., and Driessen, A.J.M. (2003). The bacterial translocase: A dynamic protein channel complex. Cell. Mol. Life Sci. *60*, 2034–2052.

DeGrado, W.F., Summa, C.M., Pavone, V., Nastri, F., and Lombardi, A. (1999). De novo design and structural characterization of proteins and metalloproteins. Annu. Rev. Biochem. *68*, 779–819.

Degrado, W.F., Wasserman, Z.R., and Lear, J.D. (1989). Protein design, a minimalist approach. Science *243*, 622–628.

DeLisa, M.P., Tullman, D., and Georgiou, G. (2003). Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc. Natl. Acad. Sci. U. S. A. *100*, 6115–6120.

Dexter, D.L. (1953). A theory of sensitized luminescence in solids. J. Chem. Phys. *21*, 836–850.

Dexter, D.L. (1979). Two ideas on energy transfer phenomena: ion-pair effects involving the OH stretching mode, and sensitization of photovoltaic cells. J. Lumin. *19*, 779–784.

Dimitrakopoulos, C.D., and Mascaro, D.J. (2001). Organic thin-film transistors: A review of recent advances. IBM J. Res. Dev. *45*, 11–27.

Discher, B.M., Noy, D., Strzalka, J., Ye, S., Moser, C.C., Lear, J.D., Blasie, J.K., and Dutton, P.L. (2005). Design of amphiphilic protein maquettes: Controlling assembly, membrane insertion, and cofactor interactions. Biochemistry *44*, 12329–12343.

Doyle, L., Hallinan, J., Bolduc, J., Parmeggiani, F., Baker, D., Stoddard, B.L., and Bradley, P. (2015). Rational design of α-helical tandem repeat proteins with closed architectures. Nature *528*, 585–588.

Dutton, P.L., and Moser, C.C. (2011). Engineering enzymes. Faraday Discuss. *148*, 443–448.

Dykxhoorn, D.M., Pierre, R.S., and Linn, T. (1996). A set of compartible tac promoter expression vectors. Gene *177*, 133–136.

Ebata, H., Izawa, T., Miyazaki, E., Takimiya, K., Ikeda, M., Kuwabara, H., and Yui, T. (2007). Highly soluble [1]benzothieno[3,2-b]benzothiophene (BTBT) derivatives for high-performance, solution-processed organic field-effect transistors. J. Am. Chem. Soc. *129*, 15732–15733.

Edelman, G.M., and Gally, J.A. (2001). Degeneracy and complexity in biological systems. Proc. Natl. Acad. Sci. U. S. A. *98*, 13763–13768.

Ennist, N.M., Mancini, J.A., Auman, D.B., Bialas, C., Iwanicki, M.J., Esipova, T. V, Discher, B.M., Moser, C.C., and Dutton, P.L. (2017a). Maquette strategy for creation of light-and redox-active proteins. In Photosynthesis and Bioenergetics, J. Barber, A.V. Ruban, Eds. (London, UK: World Scientific Publishing Co.) pp. 1–33.

Ennist, N.M. (2017b) Design, structure, and action of an artificial photosynthetic reaction center. PhD thesis University of Pennsylvania, USA. *Dissertations available from ProQuest*. AAI10264634. https://repository.upenn.edu/dissertations/AAI10264634

Espiritu, E., Olson, T.L., Williams, J.C., and Allen, J.P. (2017). Binding and energetics of electron transfer between an artificial four-helix mn-protein and reaction centers from rhodobacter sphaeroides. Biochemistry *56*, 6460–6469.

Farid, T.A., Kodali, G., Solomon, L.A., Lichtenstein, B.R., Sheehan, M.M., Fry, B.A., Bialas, C., Ennist, N.M., Siedlecki, J.A., Zhao, Z., et al. (2013). Elementary tetrahelical protein design for diverse oxidoreductase functions. Nat. Chem. Biol. *9*, 826–833.

Fletcher, J.M., Boyle, A.L., Bruning, M., Bartlett, G.J., Vincent, T.L., Zaccai, N.R., Armstrong, C.T., Bromley, E.H.C., Booth, P.J., Brady, R.L., et al. (2012). A basis set of *de novo* coiled-coil peptide oligomers for rational protein design and synthetic biology. ACS Synth. Biol. *1*, 240–250.

Förster, T. (1965). Delocalized excitation and excitation transfer. In modern quantum chemistry. O. Sinanoglu, ed. (New York, USA: Academic Press Inc.) pp. 93–137.

Fröbel, J., Rose, P., Lausberg, F., Blümmel, A.S., Freudl, R., and Müller, M. (2012). Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. Nat. Commun. *3*, 1–10.

Fry, B.A., Solomon, L.A., Leslie Dutton, P., and Moser, C.C. (2016). Design and engineering of a man-made diffusive electron-transport protein. Biochim. Biophys. Acta - Bioenerg. *1857*, 513–521.

García-Martín, A., Pazur, A., Wilhelm, B., Silber, M., Robert, B., and Braun, P. (2008). The role of aromatic phenylalanine residues in binding carotenoid to light-harvesting model and wild-type complexes. J. Mol. Biol. *382*, 154–166.

Gellman, S.H. (1998). Minimal model systems for beta sheet secondary structure in proteins. Curr Opin Chem Biol *2*, 717–725.

Gérard, F., and Cline, K. (2007). The thylakoid proton gradient promotes an advanced stage of signal peptide binding deep within the Tat pathway receptor complex. J. Biol. Chem. *282*, 5263–5272.

Ghadiri, R.M., and Case, M.A. (1993). De novo design of a novel heterodinuclear three-helix bundle metalloprotein. Angeir. Chem. Int. Ed. Engl. *32*, 1594–1597.

Gibney, B.R., Isogai, Y., Rabanal, F., Reddy, K.S., Grosset, A.M., Moser, C.C., and Dutton, P.L. (2000). Self-assembly of heme A and heme B in a designed four-helix bundle: Implications for a cytochrome c oxidase maquette. Biochemistry *39*, 11041–11049.

Gibney, B.R., Rabanal, F., Skalicky, J.J., Wand, A.J., and Dutton, P.L. (1997). Design of a unique protein scaffold for maquettes. J. Am. Chem. Soc. *119*, 2323–2324.

Gibney, B.R., Rabanal, F., Skalicky, J.J., Wand, A.J., and Dutton, P.L. (1999). Iterative protein redesign. J. Am. Chem. Soc. *121*, 4952–4960.

Gibney, B.R., Mulholland, S.E., Rabanal, F., and Leslie Dutton, P. (1996). Ferredoxin and ferredoxin–heme maquettes. Biochemistry *93*, 15041–15046.

Gill, S., and Hippel, P. Von (1989). Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. *182*, 319-326

Gillbro, T., and Cogdell, R.J. (1989). Carotenoid fluorescence. Chem. Phys. Lett. *158*, 312–316.

Gohlke, U., Pullan, L., McDevitt, C.A., Porcelli, I., de Leeuw, E., Palmer, T., Saibil, H.R., and Berks, B.C. (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. Proc. Natl. Acad. Sci. U. S. A. *102*, 10482–10486.

Goldsmith, M., and Tawfik, D.S. (2012). Directed enzyme evolution: Beyond the low-hanging fruit. Curr. Opin. Struct. Biol. *22*, 406–412.

Gouterman, M. (1961). Spectra of Porphyrins. J. Mol. Spectrosc. *6*, 138–163.

Gouterman, M. (1978). Optical spectra and electronic structure of porphyrins and related rings. In The Porphyrins, D. Dolphin, ed. (New York, USA: Academic Press, Inc). pp. 1-156.

Gradinaru, C.C., Kennis, J.T.M., Papagiannakis, E., van Stokkum, I.H.M., Cogdell, R.J., Fleming, G.R., Niederman, R.A., and van Grondelle, R. (2001). An unusual pathway of excitation energy deactivation in carotenoids: Singlet-to-triplet conversion on an ultrafast timescale in a photosynthetic antenna. Proc. Natl. Acad. Sci. U. S. A. *98*, 2364–2369.

Greenfield, N.J. (2006). Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat. Protoc. *1*, 2527–2535.

Hagen, S.J., Hofrichter, J., and Eaton, W.A. (1995). Protein reaction kinetics in a room temperature glass. Science *269*, 959–962.

Hanna, M.C., and Nozik, A.J. (2006). Solar conversion efficiency of photovoltaic and photoelectrolysis cells with carrier multiplication absorbers. J. Appl. Phys. *100*.

Hedstrom, L., Szilagyi, L., and Rutter, W.J. (1992). Converting trypsin to chymotrypsin: the role of surface loops. Science *255*, 1249–1253.

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. EMBO J. *9*, 1665–1672.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science *254*, 51–58.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load balanced, and scalable molecular simulations. J. Chem. Theory Comput. *4*, 435–447.

Hill, R.B., Raleigh, D.P., Lombardi, A., and DeGrado, W.F. (2000). De novo design of helical bundles as models for understanding protein folding and function. Acc. Chem. Res. *33*, 745–754.

Hitchcock, A., Jackson, P.J., Chidgey, J.W., Dickman, M.J., Hunter, C.N., and Canniffe, D.P. (2016). Biosynthesis of chlorophyll a in a purple bacterial phototroph and assembly into a plant chlorophyll-protein complex. ACS Synth. Biol. *5*, 948–954.

Huang, P.-S., Boyken, S.E., and Baker, D. (2016a). The coming of age of de novo protein design. Nature *537*, 320–327.

Huang, P.S., Feldmeier, K., Parmeggiani, F., Velasco, D.F., Hocker, B., and Baker, D. (2016b). De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. *12*, 29–34.

Huang, S.S., Koder, R.L., Lewis, M., Wand, A.J., and Dutton, P.L. (2004). The HP-1 maquette: From an apoprotein structure to a structured hemoprotein designed to promote redox-coupled proton exchange. Proc. Natl. Acad. Sci. *101*, 5536–5541.

Huang, S.S., Gibney, B.R., Stayrook, S.E., Leslie Dutton, P., and Lewis, M. (2003). X-ray structure of a maquette scaffold. J. Mol. Biol. *326*, 1219–1225.

Iwata, S., Lee, J.W., Okada, K., Lee, J.K., Iwata, M., Rasmussen, B., Link, T.A., Ramaswamy, S., and Jap, B.K. (1998). Complete structure of the 11-subunit bovine mitochondrial cytochrome bc 1 complex. Science *281*, 64–71.

Jack, R.L., Sargent, F., Berks, B.C., Sawers, G., and Palmer, T. (2001). Constitutive expression of Escherichia coli tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth. J. Bacteriol. *183*, 1801–1804.

Jakubowski, N., Messerschmidt, J., Añorbe, M.G., Waentig, L., Hayen, H., and Roos, P.H. (2008). Labelling of proteins by use of iodination and detection by ICP-MS. J. Anal. At. Spectrom. *23*, 1487–1496.

Jaschke, P.R., Hardjasa, A., Digby, E.L., Hunter, C.N., and Beatty, J.T. (2011). A bchD (magnesium chelatase) mutant of Rhodobacter sphaeroides synthesizes zinc bacteriochlorophyll through novel zinc-containing intermediates. J. Biol. Chem. *286*, 20313–20322.

Kadish, K.M., and Van Caemelbecke, E. (2003). Electrochemistry of porphyrins and related macrocycles. J. Solid State Electrochem. *7*, 254–258.

Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M., and Hecht, M.H. (1993). Protein design by binary patterning of polar and nonpolar amino acids. Science *262*, 1680-1685.

Kasha, B.Y.M. (1950). Characterisation of electronic transitions in complex molecules. Discuss. Faraday Soc. *9*, 14–19.

Kasha, M. (1952). Collisional perturbation of spin-orbital coupling and the mechanism of fluorescence quenching. a visual demonstration of the perturbation. J. Chem. Phys. *20*, 71–74.

Kasha, M. (1963). Energy transfer mechanisms and the molecular exciton model for molecular aggregates. Radiat. Res. *20*, 55–70.

Kerfeld, C.A., Sawaya, M.R., Brahmandam, V., Cascio, D., Ho, K.K., Trevithick-Sutton, C.C., Krogmann, D.W., and Yeates, T.O. (2003). The crystal structure of a cyanobacterial water-soluble carotenoid binding protein. Structure *11*, 55–65.

Keum, C.-M., Liu, S., Al-Shadeedi, A., Kaphle, V., Callens, M.K., Han, L., Neyts, K., Zhao, H., Gather, M.C., Bunge, S.D., et al. (2018). Tuning charge carrier transport and optical birefringence in liquid-crystalline thin films: A new design space for organic light-emitting diodes. Sci. Rep. *8*, 699.

Khakshoor, O., and Nowick, J.S. (2008). Artificial β-sheets: chemical models of β-sheets. Curr. Opin. Chem. Biol. *12*, 722–729.

Khalil, A.S., and Collins, J.J. (2010). Synthetic biology: Applications come of age. Nat. Rev. Genet. *11*, 367–379.

Khoury, G.A., Smadbeck, J., Kieslich, C.A., and Floudas, C.A. (2014). Protein folding and de novo protein design for biotechnological applications. Trends Biotechnol. *32*, 99–109.

Kim, H.Y., Bjorklund, T.G., Lim, S.H., and Bardeen, C.J. (2003). Spectroscopic and photocatalytic properties of organic tetracene nanoparticles in aqueous solution. Langmuir *19*, 3941–3946.

Kodali, G., Mancini, J.A., Solomon, L.A., Episova, T. V, Roach, N., Hobbs, C.J., Wagner, P., Mass, O.A., Aravindu, K., Barnsley, J.E., et al. (2017). Design and engineering of water-soluble light-harvesting protein maquettes. Chem. Sci. *8*, 316–324.

Koder, R.L., Anderson, J.L.R., Solomon, L.A., Reddy, K.S., Moser, C.C., and Dutton, P.L. (2009). Design and engineering of an O2 transport protein. Nature *458*, 305–309.

Koder, R.L., and Dutton, P.L. (2006). Intelligent design: the de novo engineering of proteins with specified functions. Dalt. Trans. 3045-3051.

Koder, R.L., Valentine, K.G., Cerda, J., Noy, D., Smith, K.M., Wand, A.J., and Dutton, P.L. (2006). Nativelike structure in designed four α-helix bundles driven by buried polar interactions. J. Am. Chem. Soc. *128*, 14450–14451.

Köhler, A., and Bässler, H. (2009). Triplet states in organic semiconductors. Mater. Sci. Eng. R Reports *66*, 71–109.

Korendovych, I. V., and DeGrado, W.F. (2014). Catalytic efficiency of designed catalytic proteins. Curr. Opin. Struct. Biol. 247–253.

Korovina, N. V., Das, S., Nett, Z., Feng, X., Joy, J., Haiges, R., Krylov, A.I., Bradforth, S.E., and Thompson, M.E. (2016). Singlet fission in a covalently linked cofacial alkynyltetracene dimer. J. Am. Chem. Soc. *138*, 617–627.

Korovina, N. V., Joy, J., Feng, X., Feltenberger, C., Krylov, A.I., Bradforth, S.E., and Thompson, M.E. (2018). Linker-dependent singlet fission in tetracene dimers. J. Am. Chem. Soc. *140*, 10179–10190.

Kraut, D.A., Carroll, K.S., and Herschlag, D. (2003). Challenges in enzyme mechanism and energetics. Annu. Rev. Biochem. *72*, 517–571.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. Science *302*, 1364–1368.

Laquai, F., Park, Y.-S., Kim, J.-J., and Basché, T. (2009). Excitation energy transfer in organic materials: from fundamentals to optoelectronic devices. Macromol. Rapid Commun. *30*, 1203–1231.

Lax, M. (1952). The franck-condon principle and its application to crystals. J. Chem. Phys. *20*, 1752–1760.

Lee, P.C., Momen, A.Z.R., Mijts, B.N., and Schmidt-Dannert, C. (2003). Biosynthesis of structurally novel carotenoids in Escherichia coli. Chem. Biol. *10*, 453–462.

Lei, S.P., Lin, H.C., Wang, S.S., Callaway, J., and Wilcox, G. (1987). Characterization of the Erwinia-Carotovora Pel-B gene and its product pectate lyase. J. Bacteriol. *169*, 4379–4383.

Lewis, N.S. (2016). Research opportunities to advance solar energy utilization. Science *351*.

Lewis, N.S., and Nocera, D.G. (2007). Powering the planet: Chemical challenges in solar energy utilization. Proc. Natl. Acad. Sci. U. S. A. *104*, 15729–15735.

Lichtenstein, B.R., Bialas, C., Cerda, J.F., Fry, B.A., Dutton, P.L., and Moser, C.C. (2015). Designing light-activated charge-separating proteins with a naphthoquinone amino acid. Angew. Chemie - Int. Ed. *54*, 13626–13629.

Lichtenstein, B.R., Farid, T.A., Kodali, G., Solomon, L.A., Anderson, J.L.R., Sheehan, M.M., Ennist, N.M., Fry, B.A., Chobot, S.E., Bialas, C., et al. (2012). Engineering oxidoreductases: maquette proteins designed from scratch. Biochem. Soc. Trans. *40*, 561–566.

Lin, C., Robertson, D., Ahmad, M., Raibekas, A., Jorns, M., Dutton, P., and Cashmore, A. (1995). Association of flavin adenine dinucleotide with the Arabidopsis blue light receptor CRY1. Science *269*, 968–970.

Liu, C., Churchward, G., and Caro, L. (1983). The repA2 gene of the plasmid R100.1 encodes a repressor of plasmid replication. Plasmid *10*, 148–155.

Lishchuk, A., Kodali, G., Mancini, J.A., Broadbent, M., Darroch, B., Mass, O.A., Nabok, A., Dutton, P.L., Hunter, C.N., Törmä, P., et al. (2018). A synthetic biological quantum optical system. Nanoscale *10*, 13064–13073.

Lu, L., Wu, J., Wei, L., and Fang, W. (2016). Temperature dependence of aggregated structure of β-carotene by absorption spectral experiment and simulation. Spectrochim. Acta - Part A Mol. Biomol. Spectrosc. *169*, 116–121.

Macpherson, A.N., and Gillbro, T. (1998). Solvent Dependence of the Ultrafast S_2 $-S_1$ Internal Conversion Rate of β-Carotene. J. Phys. Chem. A *102*, 5049–5058.

Maeda, Y., Makhlynets, O. V., Matsui, H., and Korendovych, I. V. (2016). Design of catalytic peptides and proteins through rational and combinatorial approaches. Annu. Rev. Biomed. Eng. *18*, 311–328.

Mancini, J.A., Kodali, G., Jiang, J., Reddy, R., Lindsey, J.S., Bryant, D.A., Dutton, P.L., and Moser, C.C. (2017). Multi-step excitation energy transfer engineered in genetic fusions of natural and synthetic light-harvesting proteins. J. R. Soc. Interface *14*

Mancini, J.A., Sheehan, M., Kodali, G., Chow, B.Y., Bryant, D.A., Dutton, P.L., and Moser, C.C. (2018). De novo synthetic biliprotein design, assembly and excitation energy transfer. J. R. Soc. Interface *15*.

Markwell, M.A.K. (1982). A new solid state reagent to iodinate proteins. Anal. Biochem. *125*, 427–432.

Mei, J., Leung, N.L.C., Kwok, R.T.K., Lam, J.W.Y., and Tang, B.Z. (2015). Aggregation-induced emission: together we shine, united we soar! Chem. Rev. *115*, 11718–11940.

Merrifield, R.E., Avakian, P., and Groff, R.P. (1969). Fission of singlet excitons into pairs of triplet excitons in tetracene crystals. Chem. Phys. Lett. *3*, 155–157.

Merrifield, R.E. (1971). Magnetic effects on triplet exciton interactions. Pure Appl. Chem. *27*, 481–498.

Mocny, C.S., and Pecoraro, V.L. (2015). De novo protein design as a methodology for synthetic bioinorganic chemistry. Acc. Chem. Res. *48*, 2388–2396.

Monera, O.D., Sereda, T.J., Zhou, N.E., Kay, C.M., and Hodges, R.S. (1995). Relationship of sidechain hydrophobicity and α-helical propensity on the stability of the single-stranded amphipathic α-helix. J. Pept. Sci. *1*, 319–329.

Moser, C.C., Sheehan, M.M., Ennist, N.M., Kodali, G., Bialas, C., Englander, M.T., Discher, B.M., and Dutton, P.L. (2016). De novo construction of redox active proteins. In Methods in Enzymol. *580*, 365–388.

Moser, C.C., Keske, J.M., Warncke, K., Farid, R.S., and Dutton, P.L. (1992). Nature of biological electron transfer. Nature *355*, 796–802.

Moser, C.C., Ennist, N.M., Mancini, J.A., and Dutton, P.L. (2018). Making maquette models of bioenergetic structures. In Mechanisms of Primary Energy Transduction in Biology, M. Wikström ed. (London, UK: The Royal Society of Chemistry) pp. 1–24.

Muller, H.J. (1964). The Relation of recombination to mutational advance. Mutat. Res. *106*, 2–9.

Muller, P. (2001). Non-photochemical quenching. A response to excess light energy. Plant Physiol. *125*, 1558–1566.

Müller, A.M., Avlasevich, Y.S., Schoeller, W.W., Müllen, K., and Bardeen, C.J. (2007). Exciton fission and fusion in bis(tetracene) molecules with different covalent linker structures. J. Am. Chem. Soc. *129*, 14240–14250.

Musser, A.J. (2013) Singlet exciton fission in unconventional systems. PhD thesis. University of Cambridge, UK.

Musser, A.J., Maiuri, M., Brida, D., Cerullo, G., Friend, R.H., and Clark, J. (2015). The nature of singlet exciton fission in carotenoid aggregates. J. Am. Chem. Soc. *137*, 5130–5139.

Narayan, A.R.H., Jiménez-Osés, G., Liu, P., Negretti, S., Zhao, W., Gilbert, M.M., Ramabhadran, R.O., Yang, Y.F., Furan, L.R., Li, Z., et al. (2015). Enzymatic hydroxylation of an unactivated methylene C-H bond guided by molecular dynamics simulations. Nat. Chem. *7*, 653–660.

Nelson, N., and Junge, W. (2015). Structure and energy transfer in photosystems of oxygenic photosynthesis. Annu. Rev. Biochem. *84*, 659–683.

Niedzwiedzki, D.M., Hunter, C.N., and Blankenship, R.E. (2016). Evaluating the Nature of So-Called S∗-State Feature in Transient Absorption of Carotenoids in Light-Harvesting Complex 2 (LH2) from Purple Photosynthetic Bacteria. J. Phys. Chem. B *120*, 11123–11131.

Niedzwiedzki, D.M., Swainsbury, D.J.K., Martin, E.C., Hunter, C.N., and Blankenship, R.E. (2017). Origin of the S* excited state feature of carotenoids in light-harvesting complex 1 from purple photosynthetic bacteria. J. Phys. Chem. B *121*, 7571–7585.

Ochsner, U.A., Snyder, A., Vasil, A.I., and Vasil, M.L. (2002). Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis. Proc. Natl. Acad. Sci. U. S. A. *99*, 8312–8317.

Olsina, J., Durchan, M., Minofar, B., Polivka, T., and Mancal, T. (2012). Absorption spectra of astaxanthin aggregates. Arxiv *1208*, No. 4958

Orriss, G.L., Tarry, M.J., Ize, B., Sargent, F., Lea, S.M., Palmer, T., and Berks, B.C. (2007). TatBC, TatB, and TatC form structurally autonomous units within the twin arginine protein transport system of Escherichia coli. FEBS Lett. *581*, 4091–4097.

Palmer, T., and Berks, B.C. (2012). The twin-arginine translocation (Tat) protein export pathway. Nat. Rev. Microbiol. *10*, 483–496.

Park, E., and Rapoport, T.A. (2012). Mechanisms of sec61/secy-mediated protein translocation across membranes. Annu. Rev. Biophys. *41*, 21–40.

Patel, P., Jones, P., Handy, R., Harrington, C., Marshall, P., and Evans, E.H. (2008). Isotopic labelling of peptides and isotope ratio analysis using LC-ICP-MS: A preliminary study. Anal. Bioanal. Chem. *390*, 61–65.

Pauli, G.F., Gödecke, T., Jaki, B.U., and Lankin, D.C. (2012). Quantitative ^{1}H NMR. Development and potential of an analytical method: An update. J. Nat. Prod. *75*, 834–851.

Payne, M.M., Odom, S.A., Parkin, S.R., and Anthony, J.E. (2004). Stable, crystalline acenedithiophenes with up to seven linearly fused rings. Org. Lett. *6*, 3325–3328.

Paz, R. (1989). Feynman’s office: the last blackboards. Phys. Today, 42, p. 88.

Piland, G.B., and Bardeen, C.J. (2015). How morphology affects singlet fission in crystalline tetracene. J. Phys. Chem. Lett. *6*, 1841–1846.

Polívka, T., and Frank, H.A. (2010). Light harvesting by carotenoids. Acc. Chem. Res. *43*, 1125–1134.

Polívka, T., and Sundström, V. (2009). Dark excited states of carotenoids: Consensus and controversy. Chem. Phys. Lett. *477*, 1–11.

Polizzi, N.F., Wu, Y., Lemmin, T., Maxwell, A.M., Zhang, S.Q., Rawson, J., Beratan, D.N., Therien, M.J., and DeGrado, W.F. (2017). De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. Nat. Chem. *9*, 1157–1164.

Pradel, N., Ye, C., Livrelli, V., Xu, J., Joly, B., and Wu, L.F. (2003). Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic Escherichia coli O157:H7. Infect. Immun. *71*, 4908–4916.

Price, N.C., Dwek, R.A., Ratcliffe, R.G., and Wormald, M.R. (2001) Principles and Problems in Physical Chemistry for Biochemists. (New York, USA: Oxford University Press)

Prickril, B.C., Czechowski, M.H., Przybyla, A.E., Peck, H.D., and LeGall, J. (1986). Putative signal peptide on the small subunit of the periplasmic hydrogenase from Desulfovibrio vulgaris. J. Bacteriol. *167*, 722–725.

Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: From modules to systems. Nat. Rev. Mol. Cell Biol. *10*, 410–422.

Rademaker, H., Hoff, A.J., Van Grondelle, R., and Duysens, L.N.M. (1980). Carotenoid triplet yields in normal and deuterated Rhodospirillum rubrum. Biochim. Biophys. Acta - Bioenerg. *592*, 240–257.

Ramasamy, S., Abrol, R., Suloway, C.J., and Clemons Jr, W.M. (2013). The glove-like structure of the conserved membrane protein TatC provides insight into signal sequence recognition in twin-arginine translocation. Structure *21*, 777–788.

Randall, L.L., and Hardy, S.J.S. (1986). Correlation of competence for export with lack of tertiary structure of the mature species: a study in vivo of maltose-binding protein in E. coli. Cell *46*, 921–928.

Rao, A., and Friend, R.H. (2017). Harnessing singlet exciton fission to break the Shockley–Queisser limit. Nat. Rev. Mater. *2*.

Rao, A., Wilson, M.W.B., Hodgkiss, J.M., Albert-Seifried, S., Bassler, H., and Friend, R.H. (2010). Exciton fission and charge generation via triplet excitons in pentacene/C60 bilayers. J. Am. Chem. Soc *132*, 12698–12703.

Reese, C., Roberts, M., Ling, M.M., and Bao, Z. (2004). Organic thin film transistors. Mater. Today *7*, 20–27.

Regan, L., and DeGrado, W. (1988). Characterization of a helical protein designed from first principles. Science *241*, 976–978.

Regan, L., Caballero, D., Hinrichsen, M.R., Virrueta, A., Williams, D.M., and O'Hern, C.S. (2015). Protein design: Past, present, and future. Biopolymers *104*, 334–350.

Reichardt, C., (2003). Solvents and solvent effects in organic chemistry. (Weinheim, Germany: Wiley-VCH)

Richter, S., Lindenstrauss, U., Lücke, C., Bayliss, R., and Brüser, T. (2007). Functional tat transport of unstructured, small, hydrophilic proteins. J. Biol. Chem. *282*, 33257–33264.

Robertson, D.E., Farid, R.S., Moser, C.C., Urbauer, J.L., Mulholland, S.E., Pidikiti, R., Lear, J.D., Wand, A.J., DeGrado, W.F., and Dutton, P.L. (1994). Design and synthesis of multi-haem proteins. Nature *368*, 425–432.

Rocco, M.A., Waraho-Zhmayev, D., and DeLisa, M.P. (2012). Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. Proc. Natl. Acad. Sci. U. S. A. *109*, 13392–13397.

Rodrigue, A., Chanal, A., Beck, K., Müller, M., and Wu, L. (1999). Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. J. Biol. Chem. *274*, 13223–13228.

Rodriguez, F., Rouse, S.L., Tait, C.E., Harmer, J., De Riso, A., Timmel, C.R., Sansom, M.S.P., Berks, B.C., and Schnell, J.R. (2013). Structural model for the protein-translocating element of the twin-arginine transport system. Proc. Natl. Acad. Sci. U. S. A. *110*, E1092–E1101.

Rollauer, S.E., Tarry, M.J., Graham, J.E., Jääskeläinen, M., Jäger, F., Johnson, S., Krehenbrink, M., Liu, S.M., Lukey, M.J., Marcoux, J., et al. (2012). Structure of the TatC core of the twin-arginine protein transport system. Nature *492*, 210–214.

Roszak, A.W., Howard, T.D., Southall, J., Gardiner, A.T., Law, C.J., Isaacs, N.W., and Cogdell, R.J. (2003). Crystal structure of the RC-LH1 core complex from Rhodopseudomonas palustris. Science *302*, 1969–1972.

Ruban, A. V. (2016). Nonphotochemical chlorophyll fluorescence quenching: mechanism and effectiveness in protecting plants from photodamage. Plant Physiol. *170*, 1903–1916.

Salares, V.R., Young, N.M., Bernstein, H.J., and Carey, P.R. (1977). Resonance raman spectra of lobster shell carotenoproteins and a model astaxanthin aggregate. A possible photobiological function for the yellow protein. Biochemistry *16*, 4751–4756.

Sambrook, J., Friysch, E.F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual (New York, USA: Cold Spring Harbor Laboratory Press)

Sancar, A. (1994). Structure and function of dna photolyase. Biochemistry *33*, 2–9.

Sargent, F., Bogsch, E.G., Stanley, N.R., Wexler, M., Robinson, C., Berks, B.C., and Palmer, T. (1998). Overlapping functions of components of a bacterial Sec-independent protein export pathway. EMBO J. *17*, 3640–3650.

Sasi Kumar, R., Padmanathan, N., and Alagar, M. (2015). Design of hydrophobic polydimethylsiloxane and polybenzoxazine hybrids for interlayer low k dielectrics. New J. Chem. *39*, 3995–4008.

Scheer, H. (1991). Chlorophylls. (Florida, USA: CRC Press, Inc).

Schiedt, K., and Liaane-Jensen, S. (1995). Isolation and analysis. In Carotenoids, Britton, G., Liaaen-Jensen, S., Pfander, H., eds. (Boston, USA: Birkhauser) pp. 81-107.

Schnappinger, T., Kölle, P., Marazzi, M., Monari, A., González, L., and De Vivie-Riedle, R. (2017). Ab initio molecular dynamics of thiophene: The interplay of internal conversion and intersystem crossing. Phys. Chem. Chem. Phys. *19*, 25662–25670.

Schott, S., McNellis, E.R., Nielsen, C.B., Chen, H.Y., Watanabe, S., Tanaka, H., McCulloch, I., Takimiya, K., Sinova, J., and Sirringhaus, H. (2017). Tuning the effective spin-orbit coupling in molecular semiconductors. Nat. Commun. *8*, 1–10.

Schulten, K., and Karplus, K. (1972). On the origin of a low-lying forbidden transition in polyenes and related molecules. Chem. Phys. Lett. *14*, 305–309.

Schweiger, A. and Jeschke, G. (2001). Principles of pulse electron paramagnetic resonance. (New York, USA: Oxford University Press, Inc)

Settles, A.M., Yonetani, A., Baron, A., Bush, D.R., Cline, K., Settles, A.M., Yonetani, A., Baron, A., Bush, D.R., Cline, K., et al. (1997). Sec-Independent Protein Translocation by the Maize Hcf106 Protein. *278*, 1467–1470.

Sharp, R.E., Moser, C.C., Rabanal, F., and Dutton, P.L. (1998). Design, synthesis, and characterization of a photoactivatable flavocytochrome molecular maquette. Proc. Natl. Acad. Sci. U. S. A. *95*, 10465–10470.

Shifman, J.M., Gibney, B.R., Sharp, R.E., and Dutton, P.L. (2000). Heme redox potential control in de novo designed four-α-helix bundle proteins. Biochemistry *39*, 14813–14821.

Shockley, W., and Queisser, H.J. (1961). Detailed balance limit of efficiency of p-n junction solar cells. J. Appl. Phys. *32*, 510–519.

Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., Clair, J.L.S., Gallaher, J.L., Hilvert, D., Gelb, M.H., Stoddard, B.L., et al. (2010). Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. Science *329*, 309–313.

Simone, D., Bay, D.C., Leach, T., and Turner, R.J. (2013). Diversity and evolution of bacterial twin arginine translocase protein, TatC, reveals a protein secretion system that is evolving to fit its environmental niche. PLoS One *8*.

Singh, S., Jones, W.J., Siebrand, W., Stoicheff, B.P., and Schneider, W.G. (1965). Laser generation of excitons and fluorescence in anthracene crystals. J. Chem. Phys. *42*, 330–342.

Skalicky, J.J., Gibney, B.R., Rabanal, F., Bieber Urbauer, R.J., Dutton, P.L., and Wand, A.J. (1999). Solution structure of a designed four-??-helix bundle maquette scaffold. J. Am. Chem. Soc. *121*, 4941–4951.

Smith, M.B., and Michl, J. (2010). Singlet fission. Chem. Rev. *110*, 6891–6936.

Smith, M.B., and Michl, J. (2013). Recent advances in singlet fission. Annu. Rev. Phys. Chem. *64*, 361–386.

Solomon, L.A., Kodali, G., Moser, C.C., and Dutton, P.L. (2014). Engineering the assembly of heme cofactors in man-made proteins. J. Am. Chem. Soc. *136*, 3192–3199.

Spano, F.C., and Silva, C. (2014). H- and J-aggregate behavior in polymeric semiconductors. Annu. Rev. Phys. Chem. *65*, 477–500.

Stanley, N.R., Palmer, T., and Berks, B.C. (2000). The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in Escherichia coli. J. Biol. Chem. *275*, 11591–11596.

Stern, H.L., Musser, A.J., Gelinas, S., Parkinson, P., Herz, L.M., Bruzek, M.J., Anthony, J., Friend, R.H., and Walker, B.J. (2015). Identification of a triplet pair intermediate in singlet exciton fission in solution. Proc. Natl. Acad. Sci. *112*, 7656–7661.

Stern, H.L., Cheminal, A., Yost, S.R., Broch, K., Bayliss, S.L., Chen, K., Tabachnyk, M., Thorley, K., Greenham, N., Hodgkiss, J.M., et al. (2017). Vibronically coherent ultrafast triplet-pair formation and subsequent thermally activated dissociation control efficient endothermic singlet fission. Nat. Chem. *9*, 1205–1212.

Stevens, J.M., Daltrop, O., Allen, J.W.A., and Ferguson, S.J. (2004). C-type cytochrome formation: Chemical and biological enigmas. Acc. Chem. Res. *37*, 999–1007.

Sutherland, G.A., Grayson, K.J., Adams, N.B.P., Mermans, D.M.J., Jones, A.S., Robertson, A.J., Auman, D.B., Brindley, A.A., Sterpone, F., Tuffery, P., et al. (2018). Probing the quality control mechanism of the *Escherichia coli* twin-arginine translocase with folding variants of a *de novo* -designed heme protein. J. Biol. Chem. *293*, 6672-6681.

Tabachnyk, M., Ehrler, B., Bayliss, S., Friend, R.H., and Greenham, N.C. (2013). Triplet diffusion in singlet exciton fission sensitized pentacene solar cells. Appl. Phys. Lett. *103*.

Tarry, M.J., Schäfer, E., Chen, S., Buchanan, G., Greene, N.P., Lea, S.M., Palmer, T., Saibil, H.R., and Berks, B.C. (2009). Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. Proc. Natl. Acad. Sci. U. S. A. *106*, 13284–13289.

Thomas, J.D., Daniel, R.A., Errington, J., and Robinson, C. (2001). Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli. Mol. Microbiol. *39*, 47–53.

Thompson, N.J., Wilson, M.W.B., Congreve, D.N., Brown, P.R., Scherer, J.M., Bischof, T.S., Wu, M., Geva, N., Welborn, M., Voorhis, T. Van, et al. (2014). Energy harvesting of non-emissive triplet excitons in tetracene by emissive PbS nanocrystals. Nat. Mater. *13*, 1039–1043.

Trinh, M.T., Zhong, Y., Chen, Q., Schiros, T., Jockusch, S., Sfeir, M.Y., Steigerwald, M., Nuckolls, C., and Zhu, X. (2015). Intra- to intermolecular singlet fission. J. Phys. Chem. C *119*, 1312–1319.

van Dongen, W., Hagen, Wilfred, H., van den Berg, W., and Cees, V. (1988). Evidence for an unusual mechanism of membrane translocation of the periplasmic hydrogenase of Desulfovibrio vulgaris (Hildenborough), as derived from expression in Escherichia coli. FEMS Microbiol. Lett. *50*, 9–5.

Vasilev, C., Johnson, M.P., Gonzales, E., Wang, L., Ruban, A. V., Montano, G., Cadby, A.J., and Hunter, C.N. (2014). Reversible switching between nonquenched and quenched states in nanoscale linear arrays of plant light-harvesting antenna complexes. Langmuir *30*, 8481–8490.

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. Nat. Biotechnol. *32*, 992–1000.

Walther, T.H., Gottselig, C., Grage, S.L., Wolf, M., Vargiu, A. V., Klein, M.J., Vollmer, S., Prock, S., Hartmann, M., Afonin, S., et al. (2013). Folding and self-assembly of the TatA translocation pore based on a charge zipper mechanism. Cell *152*, 316–326.

Wang, C., and Tauber, M.J. (2010). High yield singlet fission in a zeaxanthin aggregate observed by picosecond time-resolved resonance raman spectroscopy. J. Am. Chem. Soc. *132*, 13988–13991.

Watkins, D.W., Jenkins, J.M.X., Grayson, K.J., Wood, N., Steventon, J.W., Le Vay, K.K., Goodwin, M.I., Mullen, A.S., Bailey, H.J., Crump, M.P., et al. (2017). Construction and in vivo assembly of a catalytically proficient and hyperthermostable de novo enzyme. Nat. Commun. *8*, 1–9.

Watkins, D.W., Armstrong, C.T., and Anderson, J.L.R. (2014). De novo protein components for oxidoreductase assembly and biological integration. Curr. Opin. Chem. Biol. *19*, 90–98.

Weedon, B.C.L., and Moss, G.P. (1995). Structure and nomenclature. In Carotenoids, Britton, G., Liaaen-Jensen, S., Pfander, H., eds. (Boston, USA: Birkhauser) pp. 27-44.

Whitaker, N., Bageshwar, U.K., and Musser, S.M. (2012). Kinetics of precursor interactions with the bacterial Tat translocase detected by real-time FRET. J. Biol. Chem. *287*, 11252–11260.

Wichlacz, P.L., Unz, R.F., and Langworthy, T.A. (1986). Acidiphilium angustum sp. nov., Acidiphilium facilis sp. nov., and Acidiphilium rubrum sp. nov.: Acidophilic Heterotrophic Bacteria Isolated from Acidic Coal Mine Drainage. Int. J. Syst. Evol. Microbiol. *36*, 197–201.

Williams, R.M., and Verhoeven, J.W. (1994). On the applicability of the kasha-vavilov rule to C60. Spectrochim. Acta *50*, 251–254.

Wilson, M.W.B., Rao, A., Johnson, K., Gélinas, S., Di Pietro, R., Clark, J., and Friend, R.H. (2013). Temperature-independent singlet exciton fission in tetracene. J. Am. Chem. Soc. *135*, 16680–16688.

Woolfson, D.N. (2005). The design of coiled-coil structures and assemblies. Adv. Protein Chem. *70*, 79–112.

Wright, W.W., Carlos Baez, J., and Vanderkooi, J.M. (2002). Mixed trehalose/sucrose glasses used for protein incorporation as studied by infrared and optical spectroscopy. Anal. Biochem. *307*, 167–172.

Xia, D., Yu, C., Kim, H., Xia, J., Kachurin, A.M., Zhang, L., Yu, L., and Deisenhofer, J. (1997). Crystal structure of the cytochrome bc 1 complex from bovine heart mitochondria. Nature *926*.

Yang, J.S., Seo, S.W., Jang, S., Jung, G.Y., and Kim, S. (2012). Rational engineering of enzyme allosteric regulation through sequence evolution analysis. PLoS Comput. Biol. *8*.

Yarkony, D.R. (1998). Conical intersections: Diabolical and often misunderstood. Acc. Chem. Res. *31*, 511–518.

Yoon, S., Kim, S., and Kim, J. (2010). Secretory production of recombinant proteins in Escherichia coli. Recent Pat. Biotechnol. *4*, 23–29.

Yu, J., Fu, L.M., Yu, L.J., Shi, Y., Wang, P., Wang-Otomo, Z.Y., and Zhang, J.P. (2017). Carotenoid Singlet Fission Reactions in Bacterial Light Harvesting Complexes As Revealed by Triplet Excitation Profiles. J. Am. Chem. Soc. *139*, 15984–15993.

Yuan, Y., Giri, G., Ayzner, A.L., Zoombelt, A.P., Mannsfeld, S.C.B., Chen, J., Nordlund, D., Toney, M.F., Huang, J., and Bao, Z. (2014). Ultra-high mobility transparent organic thin film transistors grown by an off-centre spin-coating method. Nat. Commun. *5*, 1–9.

Zajac, G., Machalska, E., Kaczor, A., Kessler, J., Bouř, P., and Baranska, M. (2018). Structure of supramolecular astaxanthin aggregates revealed by molecular dynamics and electronic circular dichroism spectroscopy. Phys. Chem. Chem. Phys. *20*, 18038–18046.

Zhang, Y., Wang, L., Hu, Y., and Jin, C. (2014). Solution structure of the TatB component of the twin-arginine translocation system. Biochim. Biophys. Acta - Biomembr. *1838*, 1881–1888.

Zimmerman, P.M., Bell, F., Casanova, D., and Head-Gordon, M. (2011). Mechanism for singlet fission in pentacene and tetracene: From single exciton to two triplets. J. Am. Chem. Soc. *133*, 19944–19952.

Appendix

Appendix section 1: Growth media

The quantity of each component corresponds to the mass required for making 1 L of each media in QH_2O

A1.1 Luria-Bertani (LB) agar

LB agar was provided as a pre-formulated powder (FORMEDIUM), with 40 g added to 1 L QH_2O and the solution sterilised by autoclaving (125 °C, 15 p.s.i., 20 min).

Compound	Mass (g)
Tryptone	10
Yeast extract	5
NaCl	10
Agar	15

Final pH 7.0 ± 0.2 at 25 °C.

A1.2 LB broth

LB broth was provided as a pre-formulated powder (FORMEDIUM), with 25 g added to 1 L QH_2O and the solution sterilised by autoclaving (125 °C, 15 p.s.i., 20 min).

Compound	Mass (g)
Tryptone	10
Yeast extract	5
NaCl	10

Final pH 7.0 ± 0.2 at 25 °C.

A1.3 Selenomethionine minimal medium

The following compounds were added to 1 L QH_2O. 480 mL was transferred to two 2 L conical flasks and 40 mL to a separate glass container, with the solutions sterilised by autoclaving (125 °C, 15 p.s.i., 20 min).

Compound	Mass (g)
K_2HPO_4	10.5
$(NH_4)_2SO_4$	1.0
KH_2PO_4	4.5
Tri-sodium citrate·$2H_2O$	0.5
Glycerol	5.0
Adenine	0.5
Guanosine	0.5
Thymine	0.5
Uracil	0.5

From the remaining sterilised 40 mL of SeM+ minimal medium 20 mL was used to resuspend the desired cell pellet and 20 mL was used to solubilise the following amino acid components. As the components shown below cannot be autoclaved or filtered, they were sterilised by exposure to ultraviolet light. Note that the following list uses mg L^{-1} rather than g L^{-1}.

Compound	Mass (mg)
$MgSO_4 \cdot 7H_2O$	1000
Thiamine	4
Seleno-L-methionine	40
Ampicillin	100
L-lysine	100
L-phenylalanine	100
L-threonine	100
L-leucine	50
L-valine	50
L-isoleucine	50

Final pH 8.0 at 25 °C.

10 mL of the cell resuspension and 10 mL of the amino acid solution was then added to the sterilised 480 mL culture medium.

Appendix section 2: *E. coli* strains and plasmid vectors

A2.1 *Escherichia coli* strains used in this study

Strain	Abbreviation	Properties	Source
JM109	-	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB*$^+$ *Δ(lac-proAB) e14-[F' TraD36 proAB*$^+$ *lacI*^{q} *lacZΔM15] hsdR17 (r*$_K^-$*m*$_K^-$*)*	Promega
BL21 (DE3)	-	Strain B. F$^-$ *ompT gal dcm lon hsdS*$_B$(*r*$_B^-$*m*$_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K-12}$(λ^S)	Studier and Moffat, (1986)
W3110 'TatExpress'	TatExpress	K-12 strain. F$^-$ λ$^-$ rph-1 INV(rrnD, rrnE) with a ptac promoter upstream of tatABCD	Browning et al., (2017)
XL1-Blue	-	*endA1 gyrA96(nal*^{R}*) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB*$^+$ *lacI*^{q} *Δ(lacZ)M15] hsdR17(r*$_K^-$ *m*$_K^+$*)*	Agilent

A2.2 Plasmid vectors used in this study

Plasmid	**Description‡**	**Source/reference**
pJexpress414	High copy number, IPTG inducible T7 promoter, Amp^R.	DNA2.0 (now ATUM)
pJexpress414::BT6	BT6 with C-terminal His6 tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 3
pJexpress414::BT64A	BT64A with C-terminal His6 tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 3
pJexpress414::BT64F	BT64F with C-terminal His6 tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 3
pJexpress414::BT64A6M	BT64A6M with C-terminal His6 tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 5
pJexpress414::BT64A_A47Y	BT64A_A47Y with C-terminal His6 tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 5
pJexpress414::BT64A_Q49Y	BT64A_Q49Y with C-terminal His6 tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 5
pJexpress414::BT6M1	Quikchange using pJexpress414::BT6 as template to generate BT6M1 variant.	Chap 6
pJexpress414::BT6M0	BT6M0 with C-terminal His6-tag cloned into *Nde*I and *Xho*I sites of pJexpress414.	Chap 6
pJexpress414::BT6_STOP	Quikchange on pJexpress414::BT6 template to introduce stop codon for untagged protein production.	Chap 6
pJexpress414::BT6M1_STOP	Quikchange on pJexpress414::BT6M1 template to introduce stop codon for untagged protein production.	Chap 6
pJexpress414::BT6M0_STOP	Quikchange on pJexpress414::BT6M0 template to introduce stop codon for untagged protein production.	Chap 6
pEXT22	Low copy number (R100 *ori*), IPTG inducible *tac* promoter, *lacI*^{q}, Kan^R.	Dykxhoorn et al., (1996)
pEXT22::BT6	BT6 with C-terminal His6-tag cloned into *Kpn*I and *Xba*I sites of pEXT22.	Chap 6
pEXT22::BT6M1	BT6M1 with C-terminal His6-tag cloned into *Kpn*I and *Xba*I sites of pEXT22.	Chap 6
pEXT22::BT6M0	BT6M0 with C-terminal His6-tag cloned into *Kpn*I and *Xba*I sites of pEXT22.	Chap 6
pEXT22::TorA-BT6	TorA-BT6 with C-terminal His6-tag cloned into *Kpn*I and *Xba*I sites of pEXT22.	Chap 6
pEXT22::TorA-KR-BT6	TorA signal peptide KR version generated by Quikchange using pEXT22::TorA-BT6 as template.	Chap 6

pEXT22::TorA-KK-BT6	TorA signal peptide KK version generated by Quikchange using pEXT22::TorA-KR-BT6 as template.	Chap 6
pEXT22::TorA-BT6M1	TorA-BT6M1 with C-terminal His6-tag cloned into *Kpn*I and *Xba*I sites of pEXT22.	Chap 6
pEXT22::TorA-BT6M0	TorA-BT6M0 with C-terminal His6-tag cloned into *Kpn*I and *Xba*I sites of pEXT22.	Chap 6
pET-21a(+)	Optional N-terminal T7-tag and C-terminal His-tag, Amp^{R} (*bla* gene) for immunoblot loading control.	Novagen
pET-22b(+)	N-terminal PelB signal sequence, optional C-terminal His6-tag, IPTG inducible T7 promoter. Amp^{R}.	Novagen
pET-22b(+)::BT6	BT6 cloned into *Nco*I and *Xho*I sites of pET-22b(+) so has N-terminal PelB signal peptide and C-terminal His6-tag.	Chap 6
pET-22b(+)::BT6M1	BT6M1 cloned into *Nco*I and *Xho*I sites of pET-22b(+) so has N-terminal PelB signal peptide and C-terminal His6-tag.	Chap 6
pET-22b(+)::BT6M0	BT6M0 cloned into *Nco*I and *Xho*I sites of pET-22b(+) so has N-terminal PelB signal peptide and C-terminal His6-tag.	Chap 6

Appendix section 3: Additional absorption spectra

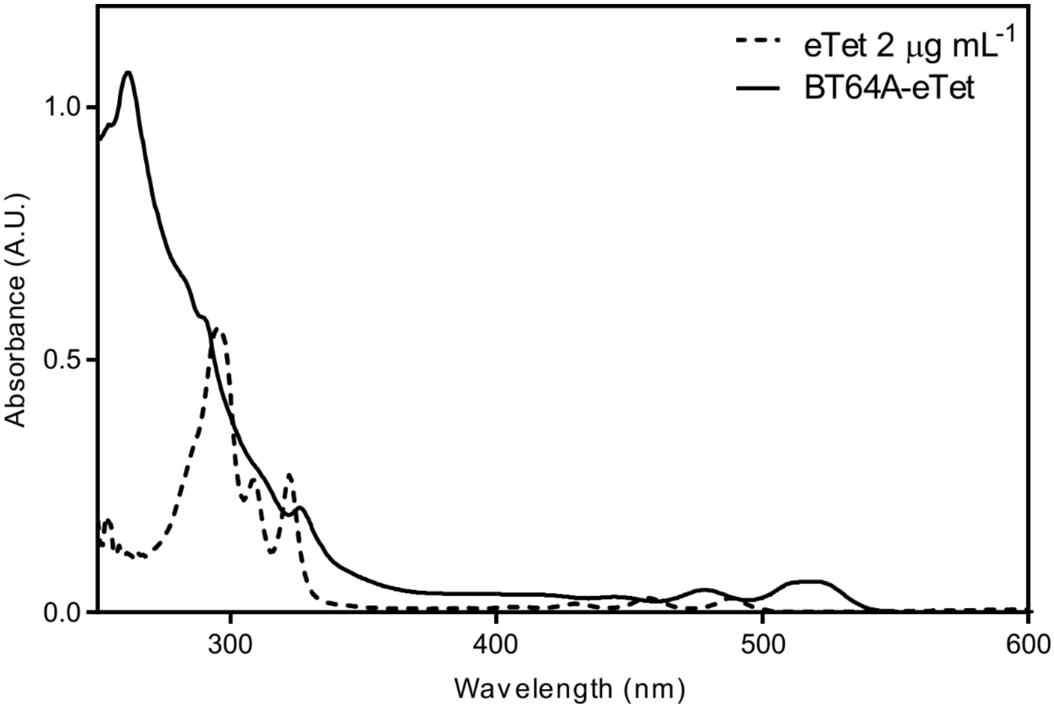


Figure A1. Absorption spectra of eTet in BT64A (solid line) and 100 % DMSO (dashed line)

Appendix section 4: Additional NMR data

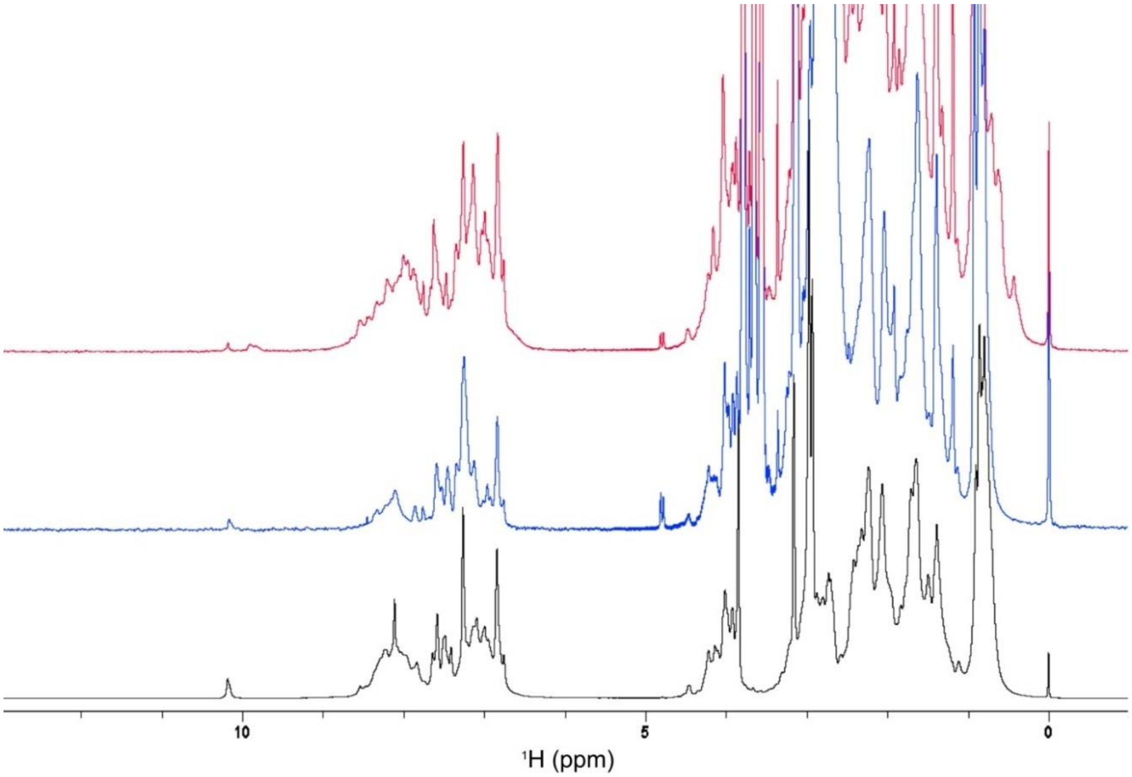


Figure A2. Full ^{1}H NMR of maquette constructs. BT6 (black), BT6M1 (blue) and BT6M0 (red).

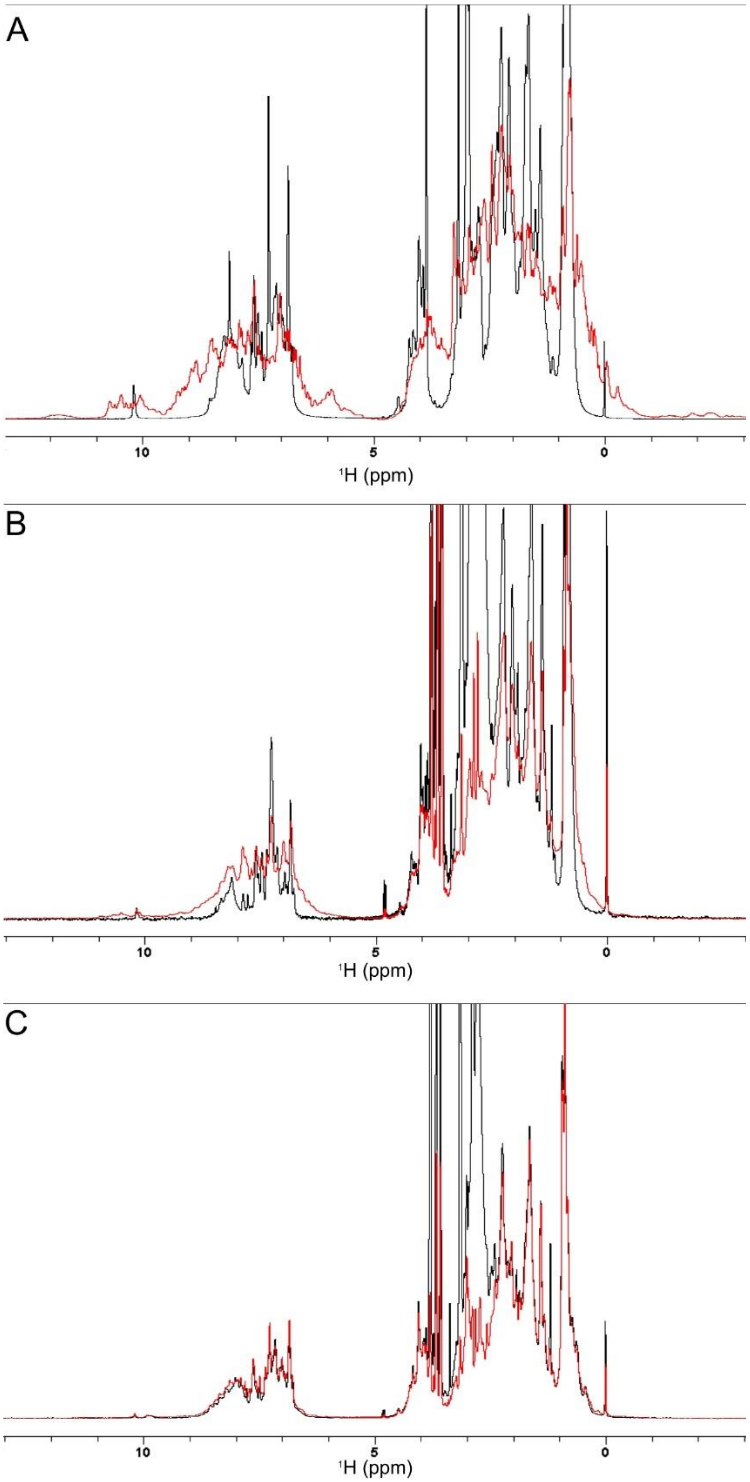


Figure A3. Full ^{1}H NMR of maquette constructs in the absence (black) and presence (red) of heme. (A) BT6, **(B)** BT6M1, **(C)** BT6M0.