2017

Engineering Designed Proteins For Light Capture, Energy Transfer, And Emissive Sensing In Vivo.

Joshua Andrew Mancini
University of Pennsylvania, joshuamancini07@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Biochemistry Commons, Biophysics Commons, and the Oil, Gas, and Energy Commons

Recommended Citation
https://repository.upenn.edu/edissertations/2458

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2458
For more information, please contact repository@pobox.upenn.edu.
Engineering Designed Proteins For Light Capture, Energy Transfer, And Emissive Sensing In Vivo.

Abstract
Proteins that are used for photosynthetic light harvesting and biological signaling are critical to life. These types of proteins act as scaffolds that hold small, sometimes metal-containing organic molecules in precise locations for light absorption and successive use. For signaling proteins, this energy can be used to induce a photoisomerization of the small molecule that can turn on or off a signaling cascade that controls the physiology of an organism. Alternatively, photosynthetic light-harvesting proteins funnel this energy in a directional manner towards a charge separating catalytic component that can change this light energy into chemical energy. The protein environment also serves to tune the photophysical properties of the small molecules. This is seen extensively with the linear tetrapyrroles that are used in both photosynthetic and signaling proteins.

Many efforts have been made to harness these natural proteins for societal use, including improving photophysical properties and interfacing capabilities with manmade catalytic components. Several methods of achieving improvement have entailed structurally guided mutation and directed evolution. However, these methods all have their limitations due to the inherent complexity and fragility of the natural proteins. This work presents an alternative more robust method to natural proteins.

My thesis states: that man-made proteins, known as maquettes, employing basic rules of protein folding, can be designed to become light harvesting and signaling proteins that can be assembled fully in vivo providing an alternative, robust, and versatile platform for meeting the diverse array of societal “green chemistry” and biomedical needs. This in vivo assembly is carried out by interacting with cyanobacterial protein and pigment machinery, both as stand-alone units and as protein fusions with natural antenna complexes. Additionally, this work offers insight for fast and tight binding of circular and linear tetrapyrroles to the maquettes both in vitro and in vivo. Design principles are also established for increasing the amount of linear tetrapyrrole attachment to the maquette as well as modulating their photophysical properties. Fast and tight binding of cofactors, high cofactor attachment yields, and control of cofactor photophysical properties are all prerequisites for the maquettes to be successful in vivo photosynthetic light harvesting and signaling proteins.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biochemistry & Molecular Biophysics

First Advisor
Peter L. Dutton

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2458
Keywords
excitation energy transfer, fusion protein, light-harvesting antenna, maquette, phycobiliprotein, synthetic protein

Subject Categories
Biochemistry | Biophysics | Oil, Gas, and Energy

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2458
ENGINEERING DESIGNED PROTEINS FOR LIGHT CAPTURE, ENERGY TRANSFER, AND EMISSIVE SENSING IN VIVO.

Joshua A. Mancini

A DISSERTATION

in Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

____________________________________
P. Leslie Dutton Ph.D., FRS
Eldridge Reeves Johnson Professor of Biochemistry and Biophysics and Director of the Johnson Foundation for Molecular Biophysics

Graduate Group Chairperson

____________________________________
Kim A. Sharp Ph.D.
Associate Professor of Biochemistry and Biophysics

Dissertation Committee:
Sergei A. Vinogradov Ph.D., Associate Professor of Biochemistry and Biophysics
Kim A. Sharp Ph.D., Associate Professor of Biochemistry and Biophysics
Brian Y. Chow Ph.D., Assistant Professor Bioengineering
Bohdana M. Discher Ph.D., Research Associate Professor of Biochemistry and Biophysics
Michael H. Hecht, Ph.D., Professor of Chemistry
Dedication

To my family who always encouraged my imagination, creativity, and curiosity.
ACKNOWLEDGEMENT

I have a lot of people to thank for making this work possible. I will start off with my advisor Les Dutton for giving me the chance to pursue a project that utilized the creativity and imagination that my parents instilled in me. Next is Goutham Kodali who was a post doc with Les when I joined the lab. He help me get my projects off the ground and was always there when I had questions or needed any type of help. Chris Moser for teaching me so much I could probably write a whole thesis on. From technical things in lab, to the organizing of ideas for writing, to the analyzing of data, Chris has always been there to help and guide me. I have become a much better scientist because of him.

I definitely need to thank my committee. Sergei Vinogradov much like Chris was always there to help me interpret and better understand my photophysical data. Kim Sharp’s feed back on my binding data revolutionized how the whole lab analyzed our non-heme binding data. Brian Chow’s understanding of the in vivo side of my project and believeing in my work provided inspiration and direction when the work was in it’s infancy. Bohdana Discher for helping me organize my scientific ideas and obtain much of the necessary equipment, funds, and other resources for making my work possible. I want to especially thank Michael Hecht for taking the time out of his busy schedule to travel from Princeton to be my external examiner. I’m not sure there is a more qualified person to assess the quality of this work.

iv
I have to thank my labmates for just generally being awesome. We had such a great time over the years with the old and new. Zhenyu Zhao and Chris Bialas joined lab the same time as me so we got to spend most of our graduate careers learning and laughing together. Nate Ennist and Bryan Fry were a few years ahead of me but I learned so much from both of them. Nate taught me so much about protein design and introduced me to the in and outs of setting crystal trays. Bryan Fry is a computational guru. Whenever I was having problems using software or personal code he was always there to guide me. We also had a blast together fixing the vacuum pumps for the lyophilizer. Our lab coats always looked like we had been in a war zone after a day with those pumps.

Martin Iwanicki and Dirk Auman joined towards the end of my tenure in the Dutton lab but we still had a great time. They will both do well in the lab. I need to especially thank Martin for being an awesome bay mate and helping me so much in the editing process for this thesis.

I would also like to thank my collaborators. Without the bilin biosynthetic machinery provided by Donald Bryant from Penn State this work would not have been possible. The expertise he provided along the way for characterizing the bilins we were attaching to the maquettes was also extremely helpful. Jonathan Lindsey and his post docs Jianbing Jiang, Kanumuri Ramesh Reddy, Olga Mass, and Aravindu Kunche from North Carolina State University provided the synthetic chlorins and bacteriochlorins that made work in chapter 2 and 4 possible.
Graduate students Nicholas Roach and Christopher Hobbs from David Officer’s lab at the University of Wollongong in Australia synthesized many of the tetrapyrroles in chapter 2. Those not synthesized by the latter two labs were synthesized by the talented postdoc Tatiana Esipova from Sergei Vinogradov’s lab. I also need to thank post doc, Molly Sheehan and graduate students Michael S. Magaraci, Ivan A. Kuznetsov, and Pimkhuan Hannanta-Anan of Brian Chow’s lab for finding an important direct application for bilimaquettes (chapter 3) in mammalian cells. Not only did they find a direct application but they also furthered our understanding of in vivo bilimaquette design.

This thesis would not have been possible without the collaborative nature of Penn and the Photosynthetic Antenna Research Center (PARC). Both of these organizations encourage collaboration across a vast array of disciplines which is essential for new and life changing discoveries to be made in this day and age. No one person or lab for that matter can possess all the knowledge, skills, and resources to do this. I am thankful that I was able to be a part of such wonderful collaborative environment. I learned more than I would have ever on my own.
ABSTRACT

Joshua A. Mancini
P. Leslie Dutton

ENGINEERING DESIGNED PROTEINS FOR LIGHT CAPTURE, ENERGY TRANSFER, AND EMISSIVE SENSING IN VIVO.

Proteins that are used for photosynthetic light harvesting and biological signaling are critical to life. These types of proteins act as scaffolds that hold small, sometimes metal-containing organic molecules in precise locations for light absorption and successive use. For signaling proteins, this energy can be used to induce a photoisomerization of the small molecule that can turn on or off a signaling cascade that controls the physiology of an organism. Alternatively, photosynthetic light-harvesting proteins funnel this energy in a directional manner towards a charge separating catalytic component that can change this light energy into chemical energy. The protein environment also serves to tune the photophysical properties of the small molecules. This is seen extensively with the linear tetrapyrroles that are used in both photosynthetic and signaling proteins.

Many efforts have been made to harness these natural proteins for societal use, including improving photophysical properties and interfacing capabilities with manmade catalytic components. Several methods of achieving improvement have entailed structurally guided mutation and directed evolution. However, these methods all have their limitations due to the inherent complexity
and fragility of the natural proteins. This work presents an alternative more robust method to natural proteins.

My thesis states: that man-made proteins, known as maquettes, employing basic rules of protein folding, can be designed to become light harvesting and signaling proteins that can be assembled fully in vivo providing an alternative, robust, and versatile platform for meeting the diverse array of societal “green chemistry” and biomedical needs. This in vivo assembly is carried out by interacting with cyanobacterial protein and pigment machinery, both as stand-alone units and as protein fusions with natural antenna complexes. Additionally, this work offers insight for fast and tight binding of circular and linear tetrapyrroles to the maquettes both in vitro and in vivo. Design principles are also established for increasing the amount of linear tetrapyrole attachment to the maquette as well as modulating their photophysical properties. Fast and tight binding of cofactors, high cofactor attachment yields, and control of cofactor photophysical properties are all prerequisites for the maquettes to be successful in vivo photosynthetic light harvesting and signaling proteins.
# TABLE OF CONTENTS

Dedication .................................................................................................................. iii

Acknowledgements....................................................................................................... iv

Abstract ....................................................................................................................... vii

List of Tables ............................................................................................................... xiii

List of Figures .............................................................................................................. xiv

Chapter 1: Introduction ............................................................................................... 1

1.1: Light harvesting in nature ...................................................................................... 1
    1.1.1: Chlorins ........................................................................................................... 3
    1.1.2: Bilins introduction .......................................................................................... 3
    1.1.3: Phytochromes ................................................................................................ 5
    1.1.4: Phycobiliproteins .......................................................................................... 5
    1.1.5: Types of bilins used in this work ...................................................................... 7
    1.1.6: Effect of local protein environment on bilin absorbance ............................... 9
    1.1.7: Using protein environment to modulate bilin conformation .......................... 11

1.2: Approaches to building nanoscale light-harvesting machinery ......................... 13

1.3: Designed proteins as an alternative light-harvesting antenna ............................ 14
    1.3.1: Different types of protein design methods .................................................... 14
    1.3.2: Maquettes as light-harvesting antenna ......................................................... 16

1.4: Goals of this thesis .............................................................................................. 23

1.5: References .......................................................................................................... 25

Chapter 2: Design and engineering of water-soluble light-harvesting maquettes using circular tetrapyrroles .......................... 31
## 2.1: Introduction

---

## 2.2: Determining rules for tetrapyrrole binding

- 2.2.1: Porphyrins
- 2.2.2: Chlorins

---

## 2.3: Intramaquette excitation energy transfer demonstration

---

## 2.4: Discussion and conclusions

---

## 2.5: References

---

### Chapter 3: Design and *in vivo* assembly of energy transfer competent synthetic bilin proteins

#### 3.1: Introduction

---

## 3.2: Overview of bilimaquette designs

---

## 3.3: *In vitro* bilin binding to maquette

---

## 3.4: *In vivo* bilin binding to maquette

---

## 3.5: Comparing *in vitro* and *in vivo* binding of PCB to maquettes

---

## 3.6: Palindromic lyase recognition site and further PCB spectral manipulation

---

## 3.7: Bilimaquette fluorescence

---

## 3.8: Comparing the developed bilimaquettes to natural phycobiliproteins

---

## 3.9: Intramaquette excitation energy transfer

- and *in vivo* near infrared imaging

---

## 3.10: Conclusion

---

## 3.11: References

---

### Chapter 4: Engineering an excitation energy transfer relay in a maquette/natural protein fusion

---
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.3: Chapter 4</td>
<td>129</td>
</tr>
<tr>
<td>A1.3.1: Gene construction</td>
<td>129</td>
</tr>
<tr>
<td>A1.3.2: Protein purification</td>
<td>130</td>
</tr>
<tr>
<td>A1.3.3: Confirmation of covalent bilin binding</td>
<td>131</td>
</tr>
<tr>
<td>A1.3.4: Bacteriochlorin attachment</td>
<td>132</td>
</tr>
<tr>
<td>A1.3.5: Folding and concentration determination</td>
<td>132</td>
</tr>
<tr>
<td>A1.3.6: Chlorin binding</td>
<td>133</td>
</tr>
<tr>
<td>A1.3.7: Spectroscopy</td>
<td>134</td>
</tr>
<tr>
<td>A1.4: References</td>
<td>134</td>
</tr>
</tbody>
</table>

Appendix 2: Supporting data for chapters ................. 137

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2.1: Chapter 2</td>
<td>137</td>
</tr>
<tr>
<td>A2.2: Chapter 3</td>
<td>138</td>
</tr>
<tr>
<td>A2.3: Chapter 4</td>
<td>158</td>
</tr>
<tr>
<td>A2.4: Small molecules for all chapters</td>
<td>166</td>
</tr>
<tr>
<td>A2.5: References</td>
<td>167</td>
</tr>
</tbody>
</table>
List of Tables

Chapter 3
Table 3.1: Bilin binding and quantum yields for Figure 3.4 ................. 61
Table 3.2: Bilin binding and quantum yields for Figure 3.5 ................. 64
Table 3.3: Bilin binding and quantum yields for Figure 3.7 ................. 70
Table 3.4: Bilin binding and quantum yields for Figure 3.8 ................. 74

Appendix 2
Table A2.2.1: Maquette amino acid sequences used in Ch. 3 ............. 140
Table A2.2.2: Parent DNA sequences for mutagenesis in Ch. 3 ........... 141
Table A2.2.3: Primers used for maquette mutagenesis in Ch. 3 .......... 143
Table A2.2.4: Modified CpcA sequence for PEB extraction ............... 144
Table A2.3.1: Primers used to make mutants in Ch. 4 ...................... 164
List of Figures

Chapter 1
Figure 1.1: Photosynthetic infrastructure .............................................. 1
Figure 1.2: Circular tetrapyrrole electronic excited states ................... 2
Figure 1.3: Formation of bilin types and biological functions .......... 4
Figure 1.4: Building blocks of the phycobilisome ......................... 6
Figure 1.5: Absorbance of bilin types used in thesis .................. 8
Figure 1.6: Degree of bilin coplanarity affects absorbance ............. 10
Figure 1.7: Bilin/amino acid contacts ............................................. 12
Figure 1.8: Maquette functional segregation .................................. 17
Figure 1.9: Basic maquette design ............................................... 18
Figure 1.10: Combinatorial design aspects of maquettes .......... 20
Figure 1.11: *In vivo* heme attachment to maquette .................. 23
Figure 1.12: Summary of thesis .................................................. 25

Chapter 2
Figure 2.1: Maquette and tetrapyrrole design for Ch.2 ................. 33
Figure 2.2: Hydrophobic partitioning drives tetrapyrrole binding ..... 34
Figure 2.3: Porphyrin titrations into maquette ........................... 36
Figure 2.4: Porphyrin K_d as function of amphiphilicity ............... 38
Figure 2.5: Effect of short and extended polar groups on binding ...... 40
Figure 2.6: Chlorins bind with high affinity ................................ 42
Figure 2.7: Tetrapyrrole binding increases thermal stability ........... 44
Figure 2.8: Demonstration of intramaquette EET ......................... 45

Chapter 3
Figure 3.1: Bilin binding to protein enhances photophysics .......... 55
Figure 3.2: Bilimaquette designs ................................................. 58
Figure 3.3: Phycobiliprotein consensus sequence ....................... 59
Figure 3.4: *In vitro* bilin binding ............................................ 60
Figure 3.5: Scan of *in vivo* PCB binding bilimaquette designs .... 63
Figure 3.6: Lyase effect on PCB binding ..................................... 66
Figure 3.7: PCB spectral manipulation ....................................... 69
Figure 3.8: Fluorescence properties of bilimaquettes .................. 73
Figure 3.9: PCB conformation in bilimaquettes ......................... 76
Figure 3.10: Extended PCB confirmation stabilizing residues ....... 77
Figure 3.11: Tripyrrole formation ............................................. 79
Figure 3.12: Absorbance spectra of maquette EET pigments .......... 80
Figure 3.13: EET demonstrations .............................................. 82
Figure 3.14: Maquette binding biliverdin in mammalian cells ....... 83
Chapter 4

Figure 4.1: Model of maquette/biliprotein fusion ........................................ 92
Figure 4.2: Bilin binding slot A of fusion ...................................................... 96
Figure 4.3: Covalent attachment of bilin to biliprotein .................................. 97
Figure 4.4: Titration of chlorins into slot B of maquette ................................. 98
Figure 4.5: Increased fluorescence upon chlorin binding ............................... 100
Figure 4.6: Covalent attachment of bacteriochlorin ..................................... 101
Figure 4.7: Absorbance of fully assembled fusion .......................................... 102
Figure 4.8: EET monitored by fluorescence emission .................................... 104
Figure 4.9: Geometry of circular tetrapyroles in maquette ............................ 109
Figure 4.10: Demonstration of a B slot energy transfer relay .......................... 111

Appendix A2: Supporting Data for chapters

Figure A2.1.1: Mass spec of maquette in Ch. 2. ........................................ 137
Figure A2.1.2: Absorbance spectra for ZnChlde construction .......................... 137
Figure A2.1.3: Increase in porphyrin (9) ε upon His coordination ................. 138
Figure A2.2.1: in vivo non-covalent BV affinity for Maquette ...................... 138
Figure A2.2.2: Lyase affect on bilin attachment .......................................... 145
Figure A2.2.3: Analytical HPLC of Maquettes in Figure 3.7 ......................... 146
Figure A2.2.4: Acidic urea PCB spectra for Figure 3.4 ................................ 147
Figure A2.2.5: Acidic urea PCB spectra for Figure 3.7 ................................ 148
Figure A2.2.6: Acidic urea PCB spectra for Figure 3.8-B ............................ 149
Figure A2.2.7: HPLC purification and MALDI-MS of tripyrrole .................. 150
Figure A2.2.8: Dual pigment maquette ....................................................... 151

absolute and difference absorbance spectra
Figure A2.2.9: SDS-PAGE of maquettes in chapter 3 ................................. 152
Figure A2.2.10: HPLC purification of PCB after methanolysis ..................... 153
Figure A2.2.11: MALDI-MS for purification of PCB ................................ 154
Figure A2.2.12: MALDI-MS for purification of PEB ................................ 154
Figure A2.2.13: HPLC of PEB after chloroform extraction ......................... 155
Figure A2.2.14: MALDI-MS verification of PCB ....................................... 156

in vitro attachment to Maquette
Figure A2.2.15: MALDI-MS verification of PEB ....................................... 157

in vitro attachment to Maquette
Figure A2.3.1: Fusion protein sequences .................................................... 158
Figure A2.3.2: HPLC estimates of PEB ..................................................... 159

attachment yields to fusion
Figure A2.3.3: Verification of BC1 attachment to maquette ......................... 160
Figure A2.3.4: Verification of BC1 attachment to fusion ............................... 161
Figure A2.3.5: Absorbance of H6 and H66 maquettes with BC1 ................. 162
Figure A2.3.6: Fusion plasmid construction .............................................. 163
Figure A2.3.7: Bilin biosynthetic machinery plasmids ................................ 165
Figure A2.4.1: All chemical structures used in thesis ................................. 166
Chapter 1: Introduction

1.1: Light Harvesting in Nature

**Figure 1.1: Photosynthetic infrastructure** (A) Schematic of light harvesting, energy transfer, and electron transfer components of photosynthesis and (B) the supramolecular assemblies organisms use to carry out these processes. Energy transfer components and electron transfer components are bracketed for both A and B. Crystal structures 2AXT and 5AQD were used in B. (C) Bilins, chlorins, and bacteriochlorins are the most common chromophores used for light harvesting in photosynthetic organisms. Examples of these chromophores’ absorbances are overlayed on earth’s solar irradiance spectrum.

Life on earth depends on light harvesting and charge separation performed in the photosynthetic processes of photoautotrophs.
1.1-A). To perform these tasks, organisms have built proteins and membrane infrastructure that are quite large and complex reaching sizes of 50 x 30 x 10 nm (Figure 1.1-B). Most of the mass of this infrastructure is from the protein scaffold needed to hold the chromophores in the correct position. This ensures that harvested light energy can be funneled to the reaction center where charge separation can occur. These chromophores, an important component of all protein based natural light-harvesting systems, consist of a conjugated electron pi system that are most commonly circular or linear tetrapyrroles. These

![Circular tetrapyrrole electronic excited states](image)

**Figure 1.2: Circular tetrapyrrole electronic excited states.** The absorbance spectra of porphyrin (black), chlorin (green), and bacteriochlorin (purple) in THF show visible to near-IR Q bands and near-UV to blue B bands. Band positions shift with changes in pi-conjugation, highlighted in yellow. 22, 20, and 18 pi electrons are found in the porphyrin, chlorin, and bacteriochlorin rings.
tetrapyrroles are called chlorins and bilins, respectively (Figure 1.1-C).

1.1.1: Chlorins

Three of the most common naturally occurring circular tetrapyrroles are porphyrins, chlorins, and bacteriochlorins. Photosynthetic organisms use the latter two for light-harvesting chromophores in their protein scaffolds. These circular tetrapyrroles have two main absorption bands in the blue and red regions of the visible spectra Figure 1.2. These two absorption bands, referred to as the Q and B (Soret) bands are electronic transitions to the first and second electronic excited singlet states. There are many differences between the spectra of these three types of tetrapyrroles. Yet, it is the increased red shifting in the Q bands and the increase in Q band extinction coefficient from porphyrin to chlorin to bacteriochlorin that is of the most importance for this work. This thesis work will exploit the spectral changes found in the chlorin and bacteriochlorin to create a light-harvesting scaffold that is much simpler and more robust than what is found in nature.

1.1.2: Bilin Introduction

In biological systems heme, a circular tetrapyrrole, is the precursor to forming linear tetrapyrroles called bilins. Heme is broken down to biliverdin (BV) by the enzyme heme oxygenase. BV can then be reduced by many different reductases to form the particular bilin an organism needs. To form the more blue shifted phycocyanobilin (PCB), four electrons need to be added to
BV, which is performed by the reductase PcyA. To form phycoerythrobilin (PEB) six electrons need to be added to BV. This can be carried out by either a reductase called PebS or by two reductases found in cyanobacteria and red algae\(^{12-14}\). Bilins gain functionality by insertion into a protein infrastructure. From here, the organism uses the protein environment and redox state of the bilin to tune the absorption properties of the bilin to the wavelengths necessary for life.

Two of the most common uses are (1) powering a bilin isomerization that drives a signaling cascade in an organism\(^{15}\) and (2) energy harvesting for photosynthesis.
The two most common bilin containing proteins are phytochromes and phycobiliproteins (Figure 1.3-B).

1.1.3: Phytochromes

Phytochromes are found in bacteria, fungi, and plants. These biliproteins attach phycocyanobilin (PCB) and phytochromobilin (PΦB) via thioether linkages\textsuperscript{16}. Bacteriophytochromes and cyanobacteriophytochromes are subclasses of phytochromes that use BV as a chromophore\textsuperscript{17}. The protein structure of phytochromes consists of 4 domains. They are called the PAS, GAF, PHY, and Effector domains. BV is first bound to the GAF domain before a cysteine from the PAS domain forms a thioether bond with the vinyl group on the A ring of BV (Figure 1.3-A). Once bound, BV in the Z configuration (Figure 1.3-B) can be photo-isomerized to the E configuration causing a structural rearrangement that perturbs the Effector domain allowing for a signal to ensue (Figure 1.3-B).

1.1.4: Phycobiliproteins

Phycobiliproteins come together to form the giant 30 x 50 x 10 nm supramolecular structure called the phycobilisome (PBS) to serve as the light harvesting complex for photosynthesis\textsuperscript{20}. Individual phycobiliproteins are made up of an α and β subunit (Figure 1.4-A)\textsuperscript{21}. The number of bilins each subunit can bind varies; however, the β subunit generally binds more bilins than the α subunit. Phycocyanin is a phycobiliprotein that makes up the intermediate rods
in PBS (PC in Figure 1.3-B). Phycocyanin was shown in Figure 1.3-A, because it will be used extensively in this work.

To form the PBS, the biliproteins must oligomerize in a rod shape\textsuperscript{20,22}. These rods can then stack onto one another creating an energy funnel to the photosystem (Figure 1.3-B). Phycoerythrin (PE) rods are the most peripheral with absorption maxima around 570nm made possible by the bound phycoerythrobilin (PEB). Phycocyanin (PC) with absorption maxima around 630nm is the next most common biliprotein to accept the photon energy from PE. The core of PBS is made up of a biliprotein called Allophycocyanin (APC)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Building blocks of the phycobilisome (PBS) (A) The individual biliproteins are made up of $\alpha$ and $\beta$ subunits. To form an individual rod the biliproteins form oligomers of different sizes. (B, C, and D) Different views of PBS rods. (B) Phycocyanobilin bound. (C and D) Phycoerythrobilin bound. These rods stack on each other to form a structure that can funnel light energy from the green part of the solar spectrum to drive the water splitting photosystem. Structures in (A) are phycocyanin (PC) from\textsuperscript{18} PDB ID: 4F0T and phycoerythrin (PE) from\textsuperscript{19} PDB ID: 5AQD.}
\end{figure}
which has absorbance maxima closer to 650nm\textsuperscript{23}. There is a special linker protein between the APC rods and the photosystem called Lcm or ApcE\textsuperscript{24}. This linker protein is a unique biliprotein for two reasons. First, it contains a PCB molecule that is extremely red shifted to 665nm, and second, it is the only phycobiliprotein that can undergo autocatalytic attachment of PCB\textsuperscript{25}. The result of all these biliproteins working together is an efficient use of the solar spectrum in photosynthesis upon energy transfer from the PBS to the chlorophyll containing reaction center proteins\textsuperscript{6,26,27}.

1.1.5: Types of bilins used in this work

Three bilin types will be used in this work (Figure 1.5). BV is the farthest red shifted, bilin, absorbing from approximately 500 to 800nm, because the conjugation of the pi bonding system is the longest at 11 double bonds\textsuperscript{28}. PCB is the next farthest red shifted in absorbance due to the reduction at the C18\textsuperscript{1} bond of BV, bringing the conjugation length down to nine double bonds. This blue shifts the absorbance range about 50nm to about 450 to 750nm. The farthest blue shifted bilin used in this work is PEB. Breaking the conjugation at the C15 bound causes the length to drop down to seven. This blue shifts the absorbance range another 50nm to approximately 400 to 700nm. These bilins will be confined to a designed conformational binding pocket and covalently attached to a reactive cysteine on a protein to anchor and increase their fluorescence quantum yields, so that they can be used for efficient energy transfer. The
change in absorbance highlighted in Figure 1.5 illustrates the change that takes place upon binding and attachment.

**Figure 1.5: Absorbance of bilin types used in thesis.** Three common bilins used in nature for light harvesting and signal transduction. (Left panel) Structures of a linearized form of the 3 bilins. The conjugation is highlighted by a colored pi bond. Free bilins are shown on top and protein bound bilin via thioether linkage are shown on the bottom. (Conjugation length) shows the length of the pi bond conjugation that is responsible for the (Right Panel) range of absorbance of each bilin in the UV and visible regions of the radiation spectra. (Absorbance - top panel) shows the absorption spectra of bilins free in phosphate buffered saline (PBS) pH 7. (Absorbance - bottom panel) shows the absorption spectra of bilins bound to their natural protein counterpart in native form. PEB and PCB are bound to the α subunit of Phycocyanin\(^{29}\) and BV is bound to the bacteriophytochrome from *Deinococcus radiodurans*\(^{30}\).
1.1.6: Effect of local protein environment on bilin absorbance

Natural biliproteins can contain multiple bilins of the same type, but the protein’s local environment around the bilin affects its absorbance. This is illustrated in Phycocyanin (PC) from cyanobacterium *Synechococcus sp. PCC7002*, which contains three different PCBs, each having a unique structure and absorbance. This effect of the protein environment on PCB absorbance is further shown by an extreme red shift in the biliprotein AP-B, one of the terminal acceptors of harvested light, which contains an ApcD subunit that in the trimeric form has absorbance red shifted to 669nm. Upon dissociation into a monomer of ApcD and ApcB, the absorbance of the PCB bound to ApcD blue shifts to 621nm, much like what is seen in PC. It is this interaction between ApcB found in another monomer and more bulky residues that forces PCB in ApcD to be close to perfectly coplanar (Figure 1.6). This coplanarity extends the pi conjugation system and red shifts the absorbance almost 50nm.

Figure 1.6 takes a closer look into these four PCBs, which have been overlayed based on their B and C ring PDB coordinates. This shows that PCB from ApcD is more planar than the other three PCB molecules in Phycocyanin (PC). This increased coplanarity of the ApcD-PCB causes a red shifting of close to 70nm from the farther blue shifted CpcB-155-PCB. Red shifting of visible bands as a function of degree of coplanarity can also be seen for the three PCBs in PC. CpcB-155-PCB shows the greatest deviation from coplanarity when
angles of pyrrole rings C-D and B-A are summed. This translates to CpcB-155-PCB producing the largest absorbance blue shift. CpcA and B-84-PCB show similar sums of coplanarity to one another that is observably greater than CpcB-155-PCB, hence showing increased red shifting. A qualitative description for the amplitude of red shift of the visible bands for all four PCB molecules as correlated with degree of A and D ring coplanarity is seen in Figure 1.6-C. Chapter 3 of this thesis will utilize these types of correlations between known bilin structures and

Figure 1.6: Degree of bilin coplanarity affects absorbance (A) Pyrrole ring deviation from coplanarity. All four PCB molecules were superimposed in pymol using B and C pyrrole coordinates in pymol. (B) Perpendicular rotation of (A). (C) Qualitative description of A and D pyrroles conformation and how this affects amplitude of visible absorption band red shift. PCB are from Phycocyanin (Cpc) Structure 4F0T and Allophycocyanin (Apc) structure 4POS. A closer look at 3 PCBs in Phycocyanin can be seen in Reference 32.
absorbance to probe bilin structure providing a metric for engineering desired bilin photophysical characteristics.

1.1.7: Using protein environment to modulate bilin conformation

A logical engineering question to ask is how are single bilin types stabilized by protein environment to maintain conformationally rigid and extended versions that can span a range of 100nm$^{33}$ of absorbance maxima. These extended versions are higher energy conformations, and thus, must be kept from rotating back to the cyclical form. It has been generally stated that these shifts in absorbance and fluorescence peaks are brought about by noncovalent interateractions with the bilin in its apoprotein$^5$.

Hydrogen bonding and constraint by bulky hydrophobic residues are some of the most common noncovalent interactions responsible for bilin extension in a protein. Hydrogen bonding with the bilin commonly occurs between basic residues and the bilin deprotonated carboxyl groups. In Figure 1.7, this is seen to occur with a lysine (Lys) and arginine (Arg). The other side of the bilin is then balanced out via hydrogen bonds between the pyrrole nitrogens and the acidic aspartate (Asp). The planar parts of the bilin are then sandwiched between bulky hydrophobic residues phenylalanine (Phe), leucine (Leu), and tryptophans (Trp) (Figure 1.7-B, C). Figure 1.7-D shows the importance the bulky Phe plays in maintaining the rigidity and extension of the bilin. As this residue is mutated to a more polar tyrosine (Tyr), two indicators of the degree of bilin extension decrease, Qvis/UV ratio and the quantum yield of fluorescence. This is
not an isolated phenomenon; a similar hydrophobic replacement allowed for the formation of a fluorescent near infrared phytochrome containing biliverdin\textsuperscript{34}. These residue combinations are by no means an exhaustive way to control a bilin conformation in a protein, but they do illustrate the logic nature uses when it comes to utilizing noncovalent interactions for tuning of bilin photophysical properties.

Looking at the natural biliproteins such as phycobiliproteins and phytochromes allowed for me to see what nature has done to manipulate the photophysical properties of bilins to serve its purposes. This provided the framework for understanding what is possible when bringing the bilins into a

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Bilin/amino acid contacts (A-C) Closer look at noncovalent interactions between PCB and the biliprotein ApcE\textsuperscript{35}. (D) Shows absorbance (solid) and fluorescence (dotted) spectra for wild type (red) and F165Y (blue) biliproteins. Figure adapted from\textsuperscript{35}.}
\end{figure}

Looking at the natural biliproteins such as phycobiliproteins and phytochromes allowed for me to see what nature has done to manipulate the photophysical properties of bilins to serve its purposes. This provided the framework for understanding what is possible when bringing the bilins into a
manmade protein framework. For example, looking at tables of absorbance and fluorescence maxima and Qvis/UV ratios$^{36}$ for PCB and then comparing to the corresponding structures determined via X-ray chrystallography provides an idea for what the PCB looks like in the manmade framework. MD calculations and structure/absorbance comparisons verified via NMR can further augment this information by providing Qvis/UV Absorbance ratios for cyclical and intermediate bilin conformation$^{37-41}$.

1.2: Approaches to building nanoscale light-harvesting machinery

One can see from Figure1-B the large size and complexity of the natural light-harvesting scaffolding. This makes these structures difficult to redesign for synthetic purposes. A solution to this problem has been to create antennas from supramolecular assemblies made of natural and synthetic components$^{42}$. This method for assembly of spatially organized pigments, much like that seen in natural light-harvesting antennas, facilitates excitation energy transfer that can then be directed towards a variety of applications$^{42}$. Nanometer scale self-assembly of these supramolecular systems circumvents the considerable effort that direct organic synthesis of multi-nanometer scale structures frequently entails$^{43,44}$. While antenna designers have exploited micelles, gels, and DNA for pigment assembly$^{45-50}$, bacterial expression of both artificial and modified natural proteins provides a relatively inexpensive means to create adaptable, monodispersed molecular frameworks to spatially organize pigments on the one to few
nanometer scale at which excitation energy transfer EET typically takes place. Furthermore, an insulating protein coat can prevent unproductive excited state quenching from pigment contact.

Several different natural protein scaffolds have been modified for maleimide linkage of energy transfer pigments to cysteine residues. Pigment coupled natural tobacco mosaic virus protein subunits\(^{51}\) self-assemble into disk-like structures. Mixing these disks with different coupled pigments leads to self-assembly into stacks; however, energy transfer is complicated with this assembly method by the statistical organization of pigments in space, leading to a range of EET distances. A similar phenomenon is seen with amyloid based scaffolds\(^{52}\). In another scaffold rhodamine was used to fill the bacteriochlorophyll “green gap” by maleimide coupling to detergent solublized bacterial photosynthetic light harvesting complex II (LHII)\(^{53}\). Additionally, chlorin tetapyrroles (including BC1 used in chapter 4) were maleimide coupled in detergent solubilized truncated mostly α-helical subunits of natural photosynthetic bacterial light harvesting complexes LH\(_{i}\)\(^{54,55}\). Oligomeric self-assembly of subunits in detergent offered relatively fine control of spatial organization and energy transfer between coupled pigment and bacteriochlorophyll ligated by a histidine.

1.3: Designed proteins as an alternative light-harvesting antenna

1.3.1: Different types of protein design methods

In this thesis I offer an alternative protein scaffolding for spatial location of cofactors that is much less complex and more versatile than natural light-
harvesting proteins is offered. There are many names for these proteins, such as *de novo* designed, artificial, or synthetic. In this work these 4-α-helical-bundle proteins (Figure 1.9), designed from first principles of protein folding and cofactor binding will be called “maquettes”. Maquette design is just one of a handful of ways to make designer proteins. Computational methods are used by groups such as David Baker. They take advantage of computational power and knowledge that can be drawn from proteins with determined structures and apply this knowledge to their algorithms to be able to extrapolate structure and function in a sequence with unknown identity. This is illustrated in the protein design tool Rosetta. This takes an unknown protein sequence, divides it into smaller sections and compares them with sequences of known structures in the PDB\textsuperscript{56}. This has allowed for the production of many types of designed proteins with new enzymatic functions\textsuperscript{57,58}.

Groups such as Frances Arnold and even the publicly traded Codexis redesign natural proteins for new and robust enzymatic functions using the power of directed evolution. Exploring a large mutational landscape they select for mutants that begin to show the desired phenotype. This type of work is aided by the ability to make what I will term “smart mutations”. An example of a smart mutation is if there is an area of the protein that needs to be hydrophobic or the entire protein will unfold rendering the protein useless for further selection. That area will be mutationally avoided or will only be mutated with other types of hydrophobic residues. Another way these smart mutations can be carried
out is through semi-random mutations focused on certain spots. For example, if there is an active site in an enzyme, and the desire is to make it perform another function, the mutations can be focused at certain places in that active site. Even the types of amino acids (polar, nonpolar, acidic, basic) can be selected for randomization. It is through this method of protein design that they have been able to create enzymes with new function and over 4 million fold increases in enzymatic activity⁵⁹,⁶⁰.

A design method that more closely resembles how the maquette method will be used in this thesis was pioneered by Michael Hecht’s lab. They have used a technique called binary patterning⁶¹ to produce libraries of maquette like proteins. These 4-α-helical-bundle proteins are very diverse in sequence but for the most part they possess a hydrophilic exterior and hydrophobic interior to create a bundle Figure 1.9. This method, along with directed evolution techniques, has allowed for development of enzymatic activity⁶² and rescue of function in strains of E. coli that had lethal gene knockouts⁶³. This binary patterning design method shows the versatility of the 4-α-helical-bundle proteins and the ability of de novo designed proteins to function in vivo.

1.3.2: Maquettes as light-harvesting antenna
Maquettes offer a framework that is not juxtaposed with the numerous functions and evolutionary baggage that natural proteins have, illustrated in Figure 1.8. Natural proteins functions have many interrelated amino acids making those functions possible. This interdependency is thought to have
arisen from the principle described as multiple utility by Darwin\textsuperscript{64}. If the amino acids responsible for one function are mutated and there is another function in the protein that relies on those particular amino acids, that non-targeted function can be lost. For example, if mutations were made to adjust a cofactor redox midpoint potential, it is very likely that the thermal stability of the protein will be affected or lost. If the same changes were made to the maquette, changing amino acids to modulate redox potential of a cofactor would not have profuse affects on thermal stability.

**Figure 1.8-B** uses Muller’s Ratchet\textsuperscript{65} to explain this irreversible complexity that makes understanding and redesign of natural proteins very difficult. Natural proteins are illustrated in the left panel where two separate functions, (crossing the river and obtaining the fruit) can be accomplished utilizing the same infrastructure (yellow and blue components of arch). Now if the yellow part of the infrastructure is removed the blue part can no longer serve its purpose and neither obtaining the fruit nor crossing the river can be achieved. An analogous picture is provided for the maquette and independent amino acids responsible for separate functions. The right panel shows that the building blocks for obtaining the fruit and crossing the river can be changed independently without deleterious affects to the other function. Analogously changes can be made independently to the amino acids responsible for each function in the maquette providing excellent scaffolding for building in a light harvesting, charge separation, and eventually, a catalytic component.
The maquette method takes advantage of the α-helical secondary structure of proteins to form a bundle of 4 α-helices, commonly 26 amino acids long (Figure 1.9). The helices for this thesis contain a polar exterior and non-polar interior to make a water-soluble bundle that can sponge up hydrophobic cofactors into the core. This arrangement of polar and non-polar amino acids is called binary patterning. Seven amino acids make up approximately two turns of the α helix (Figure 1.9-B). These positions can be lettered a to g with positions a, d, and e consisting of non-polar residues and the other positions...
consisting of polar residues. The hydrophobic effect positions the bundles so that nonpolar purple residues pack together to exclude water from the bundle core providing the basic scaffolding for inclusion of hydrophobic cofactors. Early versions of the maquette were single α helices that interacted as a tetramer to form a 4-helix bundle\(^{66}\) (Figure 1.9-C). Newer versions connect helices into a monomer using flexible glycine rich loops\(^{67}\) shown in green (Figure 1.9-D). This single chain version has allowed for much functionality to be built into the infrastructure.

The maquette functions as a chromophore scaffold (Figure 1.10) formed via basic folding principles\(^{68}\), which can support electron transfer and energy transfer, functions that do not rely on the protein for sophisticated movement or catalysis. The main effecters of energy transfer are reliant on the chromophores and the distance and orientation they are relative to each other. Thus, all that is needed to create a protein-based light-harvesting antenna is to hold
chromophores in prescribed locations and orientations to elicit efficient energy transfer. The maquettes offer an engineering platform that is ideal for light-harvesting scaffolding. The hydrophobic core partitions chromophores readily from the surrounding solution, and specifically placed histidines and cysteines can anchor these chromophores at tailored locations in the core of the maquette.
These anchoring residues can be positionally tuned to increase or decrease energy transfer efficiencies.

There is a combinatorial aspect to all parts of the maquette design from the makeup of the helices and loops to the different types of cofactors that can be added to the maquette scaffold (Figure 1.10). The most popular scaffolds used currently consist of four α-helices, but as was shown in Figure 1.9, maquettes can form 4-α-helical bundles via tetramers\textsuperscript{66,69} as well as dimers\textsuperscript{70}. Currently the most used maquettes are single chain monomers that can be water-soluble\textsuperscript{67} or amphiphilic\textsuperscript{71}. The number of helices is not the only combinatorial scaffold modification that can be made to maquettes. Helices have been lengthened from the common 26 residues to 43 residues to allow for integration of more cofactors for light harvesting and charge separation\textsuperscript{72}.

Surface charge can also be modified to affect redox potential of cofactors as well as promote or inhibit interaction with other proteins (Figure 1.10-B)\textsuperscript{67,73}. Charges ranging from +11 to -16 with a -2 or 0 charge splitting the difference have been designed and successfully expressed in multiple species of both prokaryotic and mammalian cell types\textsuperscript{67,74}. This charge can be spread throughout a helix or it can be localized on either the N or C terminus halves of the helix to create a maquette with a dipole. For further reading on surface charge see\textsuperscript{73,75}. Closely related to this changing of surface charge is the modification of helical exteriors to be more hydrophobic so that bundles can be made to span
membranes as well (Figure 1.10-A)\textsuperscript{70}.

Changing charge and length is not limited to the helices; this can be performed in the loops as well (Figure 1.10-B). Longer loops of nine residues have had lysines added to make a net positive loop. Loops that have been shortened to five and four residues are also commonly used in the maquette toolbox\textsuperscript{76}. One of the main reasons for making these loop truncations is the increase seen in the thermal melt temperature of the 4-\(\alpha\)-helical bundle, thereby providing thermostable maquettes for use in higher temperature applications.

Tailored placement of cysteine (Cys) and histidine (His) residues has made the placement of many different types of cofactors into the maquette core possible. Both residues have been located at almost every position on the helix, and cysteines have also been placed in the loops. Cysteines have the added benefit of allowing for interior, exterior, or loop region placement of cofactors due to the covalent attachment commonly formed via thioether linkage with tetrapyrroles, such as bilins, porphyrins, and flavins\textsuperscript{67,72,76-78} (Figure 1.10-C). His residues are commonly used for coordination of metals, which generally is undertaken in the core of the maquette. The overall take away from Figure 1.10 is that the combinatorial nature of the maquette scaffolding and cofactors that can be placed in that scaffolding provide a vast array of tools for protein engineers and the like to address a wide scope of societies chemical needs.

The versatility of this platform has opened up avenues of \textit{in vivo} work in multiple different scaffolds. This is one of the most promising avenues for
building light-harvesting antennas using maquettes. Having the antenna produced and assembled with chromophores \textit{in vivo} eliminates the continual upkeep of the antenna as well as opens the door for emissive sensing and production of novel photochemical systems assembled fully \textit{in vivo}.

Initial \textit{in vivo} progress was made expressing maquettes that had a periplasmic export tag allowing the maquettes to be translocated to periplasm where the maquette could interface with cytochrome c maturation machinery (Figure 1.11)\textsuperscript{79}. This machinery would covalently attach heme c to a CXXCH motif in the maquette. This discovery laid the framework for the goals of this thesis.

![Figure 1.11: In vivo heme attachment to maquette](image)

(A) Expression of maquette with CIACH motif lead to covalent attachment of heme by interfacing with Ccm proteins in periplasm (B) Cell pellets were brick red after induction and purified protein was red due to c-type heme attachment.

1.4: Goals of this thesis

The ultimate goal of this work is to build a maquette light-harvesting antenna and emissive probe that can be fully assembled \textit{in vivo} (Figure 23)
1.12). This thesis work will build up incrementally how this can be accomplished. **Chapter 2** establishes the prerequisites for binding circular Zn tetapyrroles to histidines placed in the maquette scaffolding by changing the polar and non-polar substituents attached to tetapyrroles and performing binding titrations to the maquettes. Analyzing data to determine dissociation constants will allow for quantitation of binding. The binding prerequisites learned in chapter 2 will be applied to light-harvesting maquette antennas in **Chapters 3 and 4**. In **Chapter 3** a maquette antenna will be created by interfacing with bilin biosynthetic machinery much like what seen in **Figure 1.11** for covalent attachment of heme to the maquette. Rules for improving bilin to maquette binding yields and photophysical characteristics will be established. Excitation energy transfer (EET) will be performed by binding Zn Chlorophyllide from **Chapter 2** to this construct. It will also be shown here how these bilin-binding maquettes can be used as a near infrared imaging tool in mammalian cells.

**Chapter 4** will take a different approach to building a light-harvesting antenna. A maquette will be genetically fused to a natural biliprotein subunit. This biliprotein will be able to attach two different types of bilins during expression in *E. coli* while *in vitro* the fused maquette will be able to attach two types of Zn chlorins non-covalently and synthetic bacteriochlorin covalently. Using this construct full coverage of the visible solar spectra will be achieved. This will be joined by demonstrations of directional 3 step energy transfer from the biliprotein bound bilin to the maquette bound Zn chlorin and then bacteriochlorin. This
work will establish the precedence for plugging the maquettes into the much larger phycobilisome in cyanobacteria syphoning solar energy away from the light harvesting apparatus to the maquette based novel photochemical systems.

**Figure 1.12: Summary of thesis.** The goal of Chapter 2 is to establish circular tetrapyrrole binding principles to maquette. The goal of Chapter 3 is to establish in vivo bilin binding principles to maquette. The goal of Chapter 4 is to build a biliprotein/maquette light-harvesting antenna hybrid for demonstration of 3-step excitation energy transfer.

1.5: References

22. Glazer, A. N. in Cyanobacteria 167, 291–303 (Elsevier,
37. Göller, A. H., Strehlow, D. & Hermann, G. Conformational


51. Miller, R. A., Presley, A. D. & Francis, M. B. Self-assembling light-


64. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* (1859).


Chapter 2: Design and engineering of water-soluble light-harvesting maquettes using circular tetrapyrroles

2.1: Introduction

Natural photosynthetic systems embed nanometer scale light-harvesting protein complexes with tetrapyrrole and carotenoid cofactors to capture and direct light ranging from ultraviolet to near infrared. The particular wavelengths absorbed and emitted by these tetrapyrroles are tuned as they are chemically modified by peripheral substituents and as they interact with each other and the protein environment through metal ligation, hydrophobic partitioning, hydrogen bonding and aromatic pi-stacking. This chapter seeks to show how maquettes can be used as a scaffold for circular tetrapyrroles that have desirable photophysical characteristics for use in a light-harvesting antenna. Using the maquette scaffolding will allow us to understand the basic physical chemistry of protein and cofactors that drive nanometer scale self-assembly.

Binding principles will be established by systematically exploring the binding of Zn-tetrapyrroles with different patterns of polar and non-polar substituents to maquettes engineered with either two widely spaced histidines (positions 6 and 111) for independent Zn metal ligation or without ligating histidines altogether (sequences 2.1 and 2.2 respectively, Figure 2.1, mass verification Figure A2.1.1). We use phenyl rings with and without sulphonato or carboxy functional groups to vary tetrapyrrole polarity, obtaining dissociation constants (K_d values) through singular value decomposition (SVD) analysis of
visible spectra during binding titrations. We find that a balance of polar and non-polar substituents on porphyrins is key to fast and efficient binding for both porphyrins and chlorins. Amphiphilic tetapyrroles, with one non-polar end and one polar end, allow for efficient hydrophobic partitioning into the interior of the protein while the hydrophilic part remains exposed to the aqueous portion of the protein stabilizing the complex by polar interactions. Partitioning facilitates histidine ligation to a central Zn, which in turn thermally stabilizes the protein-tetapyrrole complex by ~3.5 kcal/mol.
2.2: Determining rules for tetrapyrrole binding

2.2.1: Porphyrins

The structural requirements for tetrapyrrole self-assembly in aqueous solution were first tested with Zn tetraphenyl porphyrins. The solubility in polar and non-polar environments was adjusted through combinations of hydrophobic and hydrophilic groups at the meso positions. Zn-tetrapyrrole pigments
discussed in this text are abbreviated with bold numbers, with full chemical names given in A1.1.2 and structures illustrated in the figures and Figure A2.4.1. When all four positions are substituted with sulfonatophenyl groups, as in Zn(II)

![Figure 2.2: Hydrophobic partitioning drives tetrapyrrole binding](image)

(A) and (B) Absorption spectra demonstrating that hydrophobic partitioning is key for maquette binding. Blue, green and orange spectra correspond to porphyrin in buffer, His-free maquette, or His-containing maquette, respectively. (A) and (B) show spectra for 4 μM porphyrins (1) or (2) respectively with 2 μM maquette, when present.
5,10,15,20-tetra(4-sulphonatophenyl)-porphyrin (1), the Zn tetapyrrole becomes highly water-soluble. Figure 2.2-A demonstrates that the spectrum remains unchanged when either maquette with histidines or without histidines is added, indicating that the tetapyrrole does not associate with either maquette. However, when one sulfonatophenyl group is replaced with the non-polar phenyl group in Zn(II) 5,10,15-tri[4-sulphonatophenyl]-5-phenyl-porphyrin (2) (Figure 2.2-B), the His-free control maquette shows broadening and noticeable red-shifting of the Soret absorbance compared to the tetapyrrole in buffer, indicating partial association of the tetapyrrole with the maquette, presumably through hydrophobic partitioning and partial burial of the porphyrin in the maquette interior. The His containing maquette generates a larger, 10 nm red shift (432 nm vs. 422 nm) and maintains the narrow Soret bandwidth, indicative of histidine ligation. These findings show that one non-polar edge of a Zn-porphyrin is sufficient for binding to the maquette.
Figure 2.3: Porphyrin titrations into maquette. The absorption spectra and binding curves for amphiphilic Zn porphyrins with various polar carboxyphenyl substituents (Orange) showing binding to maquettes. Gray curves: successive absorbance spectra of titration. Orange and green curves: bound and unbound tetrapyrrole spectra from SVD with extinction coefficients on right scale. Inset shows difference in absorbance for isobestic unbound porphyrin wavelengths as a function of total porphyrin.
To determine the acceptable range and pattern of hydrophobic vs. hydrophilic perimeters for Zn porphyrins, Nicholas Roach and Christopher Hobbs from the laboratory of David Officer at the University of Wollongong in Australia synthesized all permutations of phenyl and carboxyphenyl tetra meso substituents (Figures 2.3 and 2.4). The visible spectral changes that occurred as these porphyrins were titrated into a His maquette solution were fitted to a simple binding model with a single dissociation constant ($K_d$) for each His site using singular value decomposition$^6$. Porphyrins with either all-hydrophobic substituents, Zn(II) 5,10,15,20-tetraphenylporphyrin (3), or all-hydrophilic substituents, Zn(II) 5,10,15,20-tetra(4-carboxyphenyl)porphyrin (8), showed no detectable spectral shift, signifying that binding was weaker than 100 μM. Clear spectral shifts were observed for the mono-phenylcarboxylic acid (4), the syn-diphenylcarboxylic acid (5), the anti-diphenylcarboxylic acid (6) and the tri-phenylcarboxylic acid (7) with fitted $K_d$ values of $0.14 \pm 2$ nM, $18 \pm 3.0$ nM, $14 \pm 4$ nM and $8 \pm 3$ nM, respectively. At least one hydrophobic and one hydrophilic group substitution is necessary for porphyrin ligating to the maquettes with nanomolar affinities. Mono-carboxyphenylporphyrin (4) affinity is estimated at about two orders of magnitude tighter than both di-carboxyphenyl porphyrins (5) and (6) and the tricarboxyphenol porphyrin (7).

To determine if the pattern of enhanced binding affinity with amphiphilic character of 5,10,15,20 tetrasubstituted porphyrins also applies to simpler 5,15-disubstituted porphyrins Tatiana Esipova from Sergei Vinogradov’s lab
synthesized porphyrins 9 and 10. We characterized the binding of the relatively simple carboxyphenyl construct (9) (Figure 2.5-A). Although this amphiphilic Zn porphyrin apparently displays desirable < 100 nM affinity, stopped flow mixing performed by Lee Solomon of Leslie Dutton’s lab shows that binding equilibrium occurs slowly and reaches completion in 20 minutes (Figure 2.5-C Orange Trace, see also Figure A2.1.3). Figure 2.5-A also shows that the

**Figure 2.4: Porphyrin K_d as function of amphiphilicity.** Comparative binding affinity of amphiphilic Zn porphyrins to maquettes. Polar groups are highlighted in red. Binding of (3) and (8) are too weak to measure accurately.
unbound species has a red shifted Soret typical of multimerized tetrapyrrole, which complicates accurate determination of the binding affinity. Strengthening the polar character of the 15 substituent by replacing the single carboxylic acid group with a first generation Newkome dendrimer made up of three carboxylic acid groups (10), substantially increased the water solubility of the porphyrin (Figure 2.5B), eliminated pigment multimerization, and permitted binding on a few msec timescale (Figure 2.5-C Purple Trace).
Figure 2.5: Effect of short and extended polar groups on binding. (A-B) Maquette binding of amphiphilic 5,15-substituted Zn porphyrins with relatively short (9) and extended (10) polar groups. (C) Stopped flow binding data for (9) and (10). Color of trace corresponds to bound spectra in (A) and (B).
2.2.2: Chlorins

Chlorin absorption bands generally have significant red shifts compared to porphyrins providing a complementary spectral region of light-harvesting compared to porphyrins. To determine if the pattern of enhanced binding affinity of Zn porphyrins with amphiphilic character also applies to chlorins, Olga Mass and Aravindu Kunche from North Carolina State University designed and synthesized chlorins with the analogous amphiphilic profile of 5,15-substituted tetrapyrroles. In (11), a p-tolyl substituent at the 5-position provides a hydrophobic group and a carboxy phenyl group at the opposite, 15-position provides a polar group for water solubility. These chlorins also included a dimethyl group on the pyrrole ring to stabilize the macrocycle against adventitious dehydrogenation. This chlorin is a close analogue of porphyrin (9). A shift of the chlorin Soret band from 411 to 418 nm and increases in extinction coefficients in both the Soret and Q bands correlate with histidine ligation upon binding to the maquette (Figure 2.6-A). SVD analysis of the spectral cofactor titration resolves bound and unbound cofactor spectra. This allows us to select a wavelength absorption difference pair of 419 and 389 nm that is isosbestic for unbound cofactor, allowing a simpler view of the extent and stoichiometry of binding in the graphical insert, for a \( K_d \) of 80±10 nM.
Figure 2.6: Chlorins bind with high affinity. Chlorin induced bandshifts of (11), (12), and (13) upon binding to maquettes were used to estimate binding stoichiometry and $K_d$ values. Orange traces show the chlorin bound to maquette absorbance spectra while green traces show the chlorin unbound absorbance spectra resolved using SVD analysis. Isobestic points, from which the maximal difference between bound and unbound chlorin absorbance could be obtained, were determined from this SVD and plotted as insets.
Chlorin (11) binding also stabilizes the helical structure of the maquette as seen by circular dichroism spectroscopy (Figure 2.7). Thermal midpoint transition of $\alpha$-helical structure ($T_m$) increases from 46 to 61°C upon chlorin binding with a sigmoidal behavior suggesting cooperative melting. This $T_m$ shift corresponds to 6.8 kcal/mole of structure stabilization. One advantage of the chlorins over the porphyrins was seen in a stopped-flow mixing experiment where chlorin (11) was observed to bind much faster than the similarly substituted porphyrin (9).

We exploited the ability to manipulate tetrapyrrole substituents to shift light harvesting $Q_y$ absorption bands to the red, while maintaining a hydrophobic/hydrophilic balance, by replacing the 5-p-tolyl group of (11) with a 3-phenylethynyl group (12). SVD analysis clarified binding induced bandshifts, shown together with a graphic representation of the bound pigment titration at the 423 minus 413 nm isosbestic for unbound pigment in (Figure 2.6-B). While changing the phenyl group at 5-position to the bulkier phenylethynyl group at the 3-position lowers the affinity 4.5 fold ($K_d$ of $360 \pm 18$ nM), binding is still strong enough to drive the self-assembly of the protein tetrapyrrole complex.

Pheophorbide $a$ is a popular, comparatively soluble chlorin derivative of chlorophyll in which the hydrophobic phytol tail is cleaved to leave a propionic acid. The extra oxophorbine ring (compared to the synthetic chlorins) introduces an asymmetry that makes them less prone to aggregation. We replaced the
central Mg with Zn to make Zn-pheophorbide a also referred to as ZnChlide (13)

Figure 2.7: Tetrapyrrole binding increases maquette thermal stability. Thermal stability of the maquette increased upon tetrapyrrole binding. Circular dichroism at 222 nm monitors α-helical folding for 2 μM maquette without (black) or with (orange) 2 equivalents of (11) as a function of temperature. The data was normalized for fraction folded and modeled using a Boltzmann fit, \( y = A_2 + (A_1-A_2)/(1 + \exp((x-x_0)/dx)) \), where \( x_0 \) represents the melting temperature.

throughout this thesis (13) (See Figure A2.1.2). Like the synthetic chlorins (11) and (12), (13) has a charged group on one edge of the tetrapyrrole, but also includes some uncharged but polar groups on this same edge. SVD binding analysis and 435 minus 368 nm difference in absorption as a factor of total (13) concentration to track binding is shown in (Figure 2.6-C), with a tight \( K_d \) of 4 ± 3 nM.
2.3: Intramaquette excitation energy transfer demonstration

Figure 2.8: Demonstration of intramaquette EET

(A) Absorbance of maquette binding two different Zn-tetapyrroles, (13) and (14), independently (green or blue) and together (purple). Fluorescence emission profiles of the separate bound pigments are shown as green or blue dotted lines upon excitation at 590 or 720 nm. (B) Excitation spectrum for 780 nm emission from a maquette with both tetrapyrroles (purple) shows an excited state at 670 nm not present when only (14) is bound (blue) indicating energy transfer from (13) to (14).

When two different Zn-tetapyrroles are bound in a single maquette,
energy transfer over the 2 nm center-to-center separations is generally evident, as long as there is a reasonable overlap of the emission of one chromophore with the absorbance of the other, such as the combination of tetrapyrroles (11) and (12). Figure 2.8 illustrates another combination in which tetrapyrrole (13) with emission near 660 nm (green dashed lines, Figure 2.8-A) overlaps moderately well with the absorbance of (14), blue solid line, Figure 2.8-A). (14) binding alone in a similar maquette has been reported. The maquette with both chromophores shows a compound absorbance spectrum (purple). Monitoring the emission from (14) (Figure 2.8-B) at 780 nm, where there is very little emission from (13), shows a clear peak due to the absorbance of (13), revealing energy transfer between the tetrapyrroles.

2.4: Discussion and Conclusions

Both photosynthetic light energy and electron transfer between tetrapyrrole pigments are highly distance dependent. Thus, successful operation of both natural and artificial light harvesting and light-activated redox proteins depends on secure anchoring of the appropriate tetrapyrroles in the proper location in the protein frames. Just how natural Mg and Zn chlorophyll (Chl) and bacteriochlorophyll (BChl) cofactors are loaded in vivo into specific sites in the natural light-harvesting systems has not been resolved. It is clear that a molecular chaperone is needed to enable light-harvesting complex (LHC) proteins to insert properly into thylakoid membranes and that the protein folding
requires the presence of the Chl cofactors. It is not clear how the Chl cofactor is managed before protein binding and how the potential threat of free Chl photo-oxidative reactions is avoided. Chl can be reconstituted into natural LHC proteins outside the cell, but this requires the assistance of detergent. In these systems, long tailed esterifying alcohols on the BCHls and the carboxyl group on the mesoporphyrin ring are important structural requirements for binding\textsuperscript{13}. Many studies with artificial proteins also rely on detergent to solubilize light-active tetrapterrole cofactors and/or the proteins themselves\textsuperscript{14-17}. These systems are clarifying what factors determine the binding specificity and different protein-metal coordination preferences of detergent solubilized Chl a, b and c\textsuperscript{18}.

Artificial light-harvesting proteins offer a wider engineering freedom of design compared to natural proteins. Maquettes described here, which have no significant sequence similarity to natural proteins, have relatively simple sequences and structures that allow the roles of individual amino acids and cofactor substituents to be more easily isolated to resolve general engineering principles. The conformationally rigid protein scaffolds of maquettes have interior binding sites that are sufficiently malleable to accommodate structurally diverse tetrapterroles through a variety of protein-cofactor interactions both inside and outside the cell\textsuperscript{8,19-21}. Multi-cofactor binding maquettes demonstrate light absorption, energy transfer, and light-induced charge separation\textsuperscript{22-24}. Furthermore, because these proteins are designed from first principles of protein folding, they can be extremely stable, tolerating boiling temperatures in some
cases; as seen in Figure 2.7, cofactor binding generally increases the thermal stability of maquettes. These properties make the maquette platform attractive for engineering light activated energy converting systems outside of the cell as a biomaterial, and inside the cell, when integrated into natural bioenergetic pathways. Additionally, our ability to create water-soluble light-harvesting and charge-separating maquettes should allow us to design energy harvesting systems that work directly in the cytoplasm, free of cell membranes.

Previous work with water-soluble artificial proteins$^{8,25}$ recognized that the water insolubility of natural Chls and BChls and other light-active tetrapyrroles impedes binding. Often, water-soluble derivatives of Chl and BChl created by removing the hydrophobic tail were used instead$^8$. While increasing cofactor water solubility reduces aggregation in water, cofactor hydrophobicity is integral for tight cofactor binding. The hydrophobic effect for binding in the non-polar protein interior is estimated at 2.4 kcal/nm$^2$.$^{25}$ In this chapter we experimentally distinguish the energetic effect of hydrophobic portioning of a cofactor into a maquette interior from the effect of axial tetrapyrrole metal ligation by comparing maquette designs that are histidine-free from those with independent histidines orientated towards the bundle core Figure 2.2-B. Maquette circular tetrapyrrole binding occurs via a step-wise partitioning of the cofactor into the hydrophobic bundle interior, followed by histidine ligation.$^{26}$ Figure 2.2 demonstrates for successful incorporation of light-active tetrapyrrole cofactors into an artificial water-soluble protein frame, a balance must be struck between the
hydrophobic and hydrophilic character of tetrapyrrole ring substituents. The work reported in this chapter was initiated to systematically define what makes a successful balance in order to provide a guide for cofactor construction and customization for a range of artificial light-harvesting protein functions.

This worked parsed out a few key-binding principles for circular Zn tetrapyrroles to maquettes which are discussed in the next two paragraphs. Meso substitutions that are either all charged/polar (8) or all non-polar (3) represent extremes of the hydrophobic/hydrophilic spectrum that fall prey to far too weak partitioning into the hydrophobic protein interior or aggregation in aqueous solution (Figure 2.4). Iron tetrapyrrole self-assembly to bis-histidine sites displays a similar pattern. A charged/polar meso-substituent on one edge of the tetrapyrrole appears to offer the best compromise (4); this may be because hydrophobic burial of the other three edges may maximize hydrophobic effect forces. Adding charged groups to other faces ((5), (6), (7)) is most likely forcing energetically unfavorable protein adjustments to expose these charges to the aqueous phase, or the energetic penalty of protonation/deprotonation of the charged group to neutrality.

Porphyrians with a single charged and non-polar substituent on opposite sides of the tetrapyrrole perimeter without any other peripheral groups recapitulate the balance seen in the best tetra-substituted porphyrians, although binding may be slowed because of porphyrin stacking issues that can be present
in aqueous solution (e.g. (9)). Increasing the number of charge groups on one edge (e.g. (10)) can clearly speed binding (Figure 2.5-C). Note that Zn protoporphyrin IX, an early successful cofactor choice for light activated water-soluble maquettes\(^{27}\), falls into this category, with two charged propionates on one edge and non-polar methyl and vinyl groups on the other edges. A similar, even balance between a charged group and non-polar groups on opposite tetrapyrrole edges applies to successful chlorin self-assembly as well (Figure 2.6). The pattern that emerges for rapid and tight self association of Zn-tetrapyrroles to buried histidines in water-soluble helical bundle proteins is to place one or more charge groups on one edge of the tetrapyrrole perimeter, hydrophobic groups up to ~8 Å long on the opposite edge, with other non polar groups tolerated on the other two edges. Groups that disrupt pi stacking of tetrapyrroles in aqueous solution are also helpful.

This amphiphilic-cofactor binding model allows us to screen and choose among the variety of synthetic chlorins and bacteriochlorins with diverse spectroscopic and redox properties that have been synthesized so far\(^{28}\). It also provides a crucial focus for synthesizing new chlorins to construct light-harvesting maquettes for absorbing customized wavelengths of solar radiation. Because maquettes can be constructed with histidine ligating sites of different binding affinities within the same protein, we have been able to sequentially load different cofactors to different sites within the same maquette. Thus we should be able to exploit differences in the binding affinities of porphyrins, chlorins, Chls,
and BCHls to engineer maquette systems accommodating multiple cofactors with diverse spectroscopic properties analogous to natural protein systems of photosynthesis.

Maquette modularity, self-assembly and robust adaptability to changes in the pattern of exterior amino acids is now being explored in the construction of nano and meso scale architectures for solar light capture\textsuperscript{29}. The scope of this work is not limited to \textit{in vitro} applications. The greatest utility will come from intra cellular production and upkeep of designed light-harvesting proteins. Understanding circular tetrapyrrole binding \textit{in vitro} will aid our understanding of what is required to accomplish this \textit{in vivo}. In fact, preliminary work in cyanobacteria has already shown low yields of chlorophyllide, similar to (13), binding to maquettes \textit{in vivo}. The following chapter examines another \textit{in vivo} assembly of maquettes with natural bilin tetrapyrroles inside cells. All of these \textit{in vivo} cofactor-binding capabilities bring us closer to being able to tap into and divert energy flow of cellular bioenergetic systems towards a new class of light driven fuel production.

2.5: References


Chapter 3: Design and in vivo assembly of energy transfer competent synthetic bilin proteins

3.1: Introduction

Bilins, protein-bound pigments with wide-ranging, environmentally sensitive spectral absorbance and fluorescence emission bands (Figure 3.1), are the focus of attention in both natural\textsuperscript{1-5} and synthetic\textsuperscript{6} light-harvesting systems that capture sunlight for charge separation and catalysis. They are also of increasing interest as sensors, such as natural light-sensing phytochromes\textsuperscript{7-9} and engineered opto-genetic infrared reporters for deep-tissue and low-background fluorescence imaging\textsuperscript{10-14}.

Bilins are biosynthesized by oxidative cleavage of the heme cyclic tetrapyrrole by heme oxygenase to form the open-chain tetrapyrrole biliverdin (BV), which has absorption bands in the red and violet regions of the visible spectrum\textsuperscript{15}. Regio-specific reduction of BV generates phycocyanobilin (PCB) or phycoerythribilin (PEB)\textsuperscript{16-18}. This reductive shortening of the extended pi-conjugated chain of BV blue-shifts absorption bands and makes the resulting bilins useful phytochrome sensors for chlorophyll-containing photosynthetic organisms (Figure 3.1-a)\textsuperscript{17,19}. Similarly, scission of a pyrrole ring to form a tripyrrole (TPB) also blue-shifts the absorption\textsuperscript{20}.

Free bilins in aqueous buffer have many degrees of rotational flexibility, which leads to relatively broad absorption bands (Figure 3.1-b). This same flexibility provides a means of non-radiative quenching of the light-excited
states, which results in relatively low quantum yields of fluorescence\textsuperscript{21}. Bilin vinyl or ethylidene groups may react with free cysteine (Cys) thiols to form a thioether bond that covalently anchors the pigment in a natural protein binding pocket usually, but not always, with the assistance of a natural bilin lyase\textsuperscript{22-24}. Attachment sharpens absorbance features and increases fluorescence quantum yields (Figure 3.1-c), presumably because of constrained pigment motion in a non-cyclical conformation supports a stronger transition dipole moment\textsuperscript{25}.

![Figure 3.1](image)

\textbf{Figure 3.1: Spectral diversity of bilins and increase in visible absorption band upon binding to protein.} (a) Structure of free bilins and picture in phosphate buffered saline (PBS), pH 7. (b) Absorbance of free bilins in (PBS), pH 7. (c) Absorbance of bilins bound to protein. TPB and BV are bound to maquette while PCB and PEB are bound to the CpcA subunit of the protein fusion discussed in chapter four\textsuperscript{27}.

The general rules for biliprotein design that would inform an intentionally engineered approach to controlling pigment optical properties are largely unknown. Most mutagenic work on natural light-harvesting
phycobiliproteins and natural phytochromes has focused on identifying the specific Cys residues that interact with bilins and native bilin lyases\textsuperscript{26,27}. Although there has been some structure-guided mutational studies aimed at increasing fluorescence yield\textsuperscript{28,29}, mutagenesis is typically random combined with screening for fluorescence properties. As unplanned products of natural selection natural proteins are often structurally complex, leading to physico-chemical interactions between amino acids within the proteins and between amino acids and pigment cofactors being obscure and difficult to resolve. Attempting to add together many mutational changes in progressive redesign of a natural protein frequently results in disrupted protein folding and poor stability and yields. The maquette method addresses these difficulties by using first principles of protein folding to intentionally build compact, structurally transparent protein frameworks that are resilient to extensive mutation. Not only will this allow for elucidation of bilin to protein attachment principles but this will also allow for a light-harvesting maquette component to be quickly tied into a charge-separation component creating novel photochemical systems that can be directed toward synthetic applications\textsuperscript{30-32}.

The first step in bilin bound to maquette (bilimaquette) design is to achieve a basic understanding of the requirements for efficient covalent binding of bilin pigments to specific cysteine (Cys) amino acids in a protein framework and to determine how to manipulate the protein environment around the ligation site to control the absorption and emission properties. This work was made
possible by the observation that when expressed in *Escherichia coli*, bilins may spontaneously associate with the hydrophobic interiors of bundles, even in the absence of a bilin-ligating Cys (Figure A2.2.1). Maquette-bound bilins undergo shifts in absorbance to the red and narrowing of band-width analogous to natural biliproteins (green trace Figure 3.1-c). This chapter describes design principles associated with efficient bilin binding, light absorption, fluorescence and energy transfer needed to support engineering for desirable photophysical properties and novel applications in light harvesting and sensing.

3.2: Overview of bilimaquette designs

Maquette frames are highly tolerant to changes in sequences that maintain the general binary patterning. We exploit this tolerance to move a common, four amino acid, cysteine-X-arginine-aspartate (CXRD) bilin-binding motif freely around the helical and loop regions of the maquette, change the net charge of individual helices, alter lengths of loops, and make changes in the hydrophobic core without compromising the robust folding of the protein. Figure 3.2, 3.3, and Tables 3.1-3.4 describe maquette sequences (see Appendix 2, Table A2.2.1) that combine these parameters, indicating the position in the sequence in which Cys residues have been inserted to provide a bilin-binding site and provide binding and spectroscopic quantitation of the resulting bilimaquettes.
3.3: In vitro bilin binding to maquette

Control of bound-bilin photophysics requires engineering specific Cys attachment sites. An initial survey of Cys locations using only single cysteine variants for in vitro ligation of three types of bilins, (BV, PEB and PCB) to maquettes revealed that loop-anchoring provided the best yield for attachment for all bilins (Figure 3.4, sequence 3.2). This suggests that cysteine exposure facilitates binding. However, burying the Cys in the core (sequence 1)
generally enhanced the quantum yield of fluorescence, indicating that the core restricts bilin mobility and reduces the quenching of bilin fluorescence.

While there are no universal amino acid consensus sequences for Cys bilin attachment, the CXRD motif is found at conserved sites in the alpha and beta subunits of phycocyanin, allophycocyanin, phycoerythrin, and phycoerythrocyanin\(^4\) (Figure 3.3). To test the hypothesis that CXRD motifs improve \textit{in vitro} bilin binding, we inserted a CLRD motif downstream from the buried helical core position at the beginning and end (sequences 3.3 and 3.4) of the

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.3.png}
\caption{Phycobiliprotein consensus sequence Bilin lyase recognizing consensus sequences developed from\(^34\).}
\end{figure}
Figure 3.4: *In vitro* bilin binding. Absorbance (left) and emission (right) properties of bilimaquettes ligated *in vitro* with BV, PEB or PCB. Isolated Cys for covalent bilin attachment in helix core (blue, 3.1) or loop (green, 3.2), and natural consensus sequence towards the beginning (yellow, 3.3) or end (red, 3.4) of a helix. Absorbance spectra are normalized for protein concentration: 100, 25, and 30 mM for BV, PEB and PCB, respectively. Emission spectra are normalized for absorbance at 600, 560 and 580 nm for BV, PEB and PCB, respectively. Excitation spectra are normalized for 600, 575 and 575 nm for BV, PEB and PCB, respectively. Maquettes were excited at 600, 560, and 580 nm for BV, PEB and PCB, respectively. Spectra are from HPLC purified samples. All BV work in this figure was performed by Molly Sheehan working in Brian Chow’s lab.
second helix. Neither CLRD position improved bilin binding or fluorescence quantum yields compared to isolated Cys sequences – in fact, BV binding was lowered. This contrasts with in vivo binding (described below), suggesting that the CLRD motif gains significance through modulating interactions with bilin lyases.

<table>
<thead>
<tr>
<th>ID</th>
<th>Helix charge</th>
<th>Loop type</th>
<th>Cys insertion</th>
<th>Cys location</th>
<th>Binding Yield in vitro</th>
<th>Quantum Yield (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3</td>
<td></td>
<td></td>
<td>BV</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>- - - -</td>
<td>N N N</td>
<td>C</td>
<td>2</td>
<td>6</td>
<td>70 (1.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BV</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>- + - +</td>
<td>P S P</td>
<td>CGR</td>
<td>2</td>
<td>3</td>
<td>100 (0.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BV</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>2</td>
<td>9.7 (0.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BV</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>25</td>
<td>7.1 (0.7)</td>
</tr>
</tbody>
</table>

Table 3.1: Bilin binding and quantum yields for Figure 3.4. Maquette designs in Figure 3.4, showing the pattern of helix charge (negative or positive), loop design (neutral (N), positive (P), or short (S)) Cys insertion sequence and position along helix or loop. In vitro bilin binding yield is relative to sequence 3.2. Fluorescence quantum yield was determined by comparison to a cresyl violet standard.

3.4: In vivo bilin binding to maquette

Maquette design utility is enhanced by integration into natural proteins in living cells and will be a critical part of maquettes functioning as light-harvesting components or as sensors. We exploited the expression system developed by Alvey et al., which imports bilin cellular maturation and attachment machinery for ligation with diverse phycobilins into E. coli. The plasmids code for the following: heme oxygenase-1 to catalyze BV synthesis from heme; a reductase to reduce BV to either PCB or PEB; and a bilin lyase. We replaced the plasmids encoding natural phycobiliproteins with our maquette plasmid. Although both PCB and PEB plasmids led to maquettes with covalently bound bilins,
the PCB plasmid was generally preferred because it led to greater yields of holo-
maquettes with bound chromophore. When maquettes are over-expressed for
ease of analysis, the interaction between the bilin S-type lyase, CpcS from
*Thermosynechococcus elongatus* is critical. Without PCB lyase, maquettes
were purified with minimal amounts of bound PCB along with small amounts of
bound BV or heme (Appendix Figure A2.2.2).

An initial scan of *in vivo* bilin binding (Figure 3.5) shows the bilin
absorbance of various fully folded maquette sequences at similar protein
concentrations to compare the relative yield of PCB attachment (Table 3.2). This
initial scan allowed for a few rules to be established for application in future
designs. Rule 1 illustrated in Figure 3.5-a shows that an isolated Cys in a loop
(sequence 5) effectively ligates bilin during expression *in vivo* as it did *in vitro*.
However, placing an isolated Cys in a helix was less efficient for *in vivo* ligation,
for both buried heptad core positions “a” or “d” (sequence 3.1 or 3.9) and surface-
exposed helix heptad positions “b” and “c” (sequence 3.7 or 3.8). Rule 2 is the
yield of PCB binding to a loop Cys and conformational rigidity is enhanced by
introducing a nearby arginine (Figure 3.5-b, sequence 3.2 and 3.6), as reflected
in the sharpening of the visible Q band (Qvis) absorbance around 600nm.

Rule 3 illustrated in Figure 3.5-c shows that unlike *in vitro* bilin ligation, *in
vivo* ligation yield generally increased when a CXRD motif was inserted into a
helix (sequences 3.3, 3.10, 3.11), most likely by improving interaction with the
bilin lyase. However, placing the Cys of the CXRD motif within four residues
of the end of the helix results in relatively poor bilin binding (sequences 4, 16, 17). Rule 4 illustrated in Figure 3.5-d demonstrates that placing the extended KAKCARD motif, found in phycobiliprotein CpcA, with ligating Cys in a core “a” position (sequence 3.13) or exposed “b”, “c”, or “g” heptad position (sequences 3.14, 3.15 or 3.12) in the second half of the first helix significantly increased the
yield of bound PCB. Bilin attachment to Cys positions 7 and 8 residues from the end of the helix (sequences 3.12 and 3.13) generated more structured absorption bands that imply greater bilin rigidity and/or a decrease in the degree of helicity $^{21,37,38}$.

<table>
<thead>
<tr>
<th>ID</th>
<th>Helix charge</th>
<th>Loop type</th>
<th>Cys insertion</th>
<th>Cys location</th>
<th>Binding Yield in vivo (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>1 2 3</td>
<td></td>
<td>Helix</td>
<td>Loop</td>
</tr>
<tr>
<td>1</td>
<td>- - - -</td>
<td>N N N</td>
<td>C</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>- + - +</td>
<td>P S P</td>
<td>CGR</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>- + - +</td>
<td>P S P</td>
<td>C</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>- + - +</td>
<td>P S P</td>
<td>CGRD</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>- + - +</td>
<td>P S P</td>
<td>C</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>- + - +</td>
<td>P S P</td>
<td>C</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>- + - +</td>
<td>P N P</td>
<td>C</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>- + - +</td>
<td>P N P</td>
<td>KAKCARD</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>13</td>
<td>- + - +</td>
<td>P N P</td>
<td>KAKCARD</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>14</td>
<td>- + - +</td>
<td>P N P</td>
<td>KAKCARD</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>15</td>
<td>- + - +</td>
<td>P N P</td>
<td>KAKCARD</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>16</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>17</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>23</td>
</tr>
</tbody>
</table>

Table 3.2: Bilin binding and quantum yields for Figure 3.5. Maquette designs in Figure 3.5 showing the helix and loop patterning, Cys insertion sequence and residue number. In vivo bilin binding yield is computed relative to sequence 10. Abbreviations are the same as used in table 3.1.

3.5: Comparing *in vitro* and *in vivo* binding of PCB to maquettes

Before moving on to the establishment of more in depth *in vivo* bilimaquette rules a brief comparison between *in vitro* and *in vivo* PCB attachment between sequences 1-4 will be given. The crucial change in moving from *in vitro* bilin to maquette attachment to *in vivo* attachment is the addition of
lyases. In natural systems, lyases not only improve the yield of bilin ligation, but they also modulate the stereochemistry assumed by the bound bilin, thus changing bilin absorption and emission properties. Autocatalytic ligation of PCB to natural phycobiliproteins in vitro tends to form species with red-shifted Qvis absorption maxima and smaller Qvis/UV absorbance ratios compared to the same phycobiliproteins exposed to a lyase either in vivo or in vitro after bilin ligation. For example, PCB binding to CpcB undergoes a 635 to 629 nm blue shift on exposure to S-type lyase\(^{39}\) while allophycocyanin APC-A2 blue shifts from 625 to 610 nm\(^{40}\). Autocatalytic ligation of PCB to natural apo-C-phycocyanin is also red-shifted compared to ligation in the presence of CpcE-CpcF lyase (647 nm compared to 622 nm). However, this relatively large blue shift comes from a 2-electron oxidation of PCB to mesobiliverdin (MBV) during autocatalytic binding. PCB is not oxidized during CpcE-CpcF lyase action in vivo\(^{26,41}\). The general belief is that lyases not only select for a specific stereochemistry, but they also act as a chaperone to aid in the formation of a higher energy conformation of PCB (more extended) in the natural phycobiliprotein binding pocket\(^{42}\).

Because maquettes have minimal sequence similarity to natural phycobiliproteins, they are not expected to form specific complexes with natural lyases. Nevertheless, besides boosting bilin binding in vivo (see Appendix Figure A2.2.2), the CpcS lyase confers different maquette bilin conformations than the ones seen in lyase-free autocatalysis. Autocatalytic PCB binding to maquettes produces spectral forms with broad absorption bands with maxima around
650 nm and 600 nm (Figure 3.6-a). Maquette designs with a Cys in the central loop do not show much of a band shift upon in vivo expression in the presence of lyase (dashed green line Figure 3.6-a). However, for maquettes with Cys
residues placed in helix positions, lyase action in vivo led to significant blue-shifts, analogous to the shifts seen in natural phycobiliproteins. These absorption maxima fall in the typical 610 to 650 nm range of the majority of natural biliproteins carrying bound PCB\textsuperscript{44,45}.

**Figure 3.6-b** shows that lyases exert control over the E-Z isomerization at the C15-C16 double bond of PCB in maquettes. When ligated to maquettes in the presence of lyase in vivo, PCB largely maintains the Z isomer which has an absorbance maximum at 661 nm under acidic denaturing conditions which remove the influence of the protein environment on the bilin absorbance\textsuperscript{46}. Corresponding denatured spectra of maquettes carrying PCB autocatalytically ligated in vitro confirmed that there had been no oxidation to mesobiliverdin (MBV). However, these spectra show that during autocatalytic ligation in the dark, the Z isomer of purified PCB is converted to predominantly E isomers (absorbance maximum 593 nm\textsuperscript{46}). Although photo-isomerization between Z and E configurations of PCB is a critical part of the signaling process in phytochromes\textsuperscript{7}, there is so far no evidence for analogous photo-isomerization in bilimaquettes.

In summary in vitro attachment of PCB produces broad absorption features that appear to be the E-PCB isomer when unfolded in acidic urea whereas in vivo attachment with the CpcS lyase produces a blue shifted PCB species that is confirmed as the Z isomer when unfolded in acidic urea.
3.6: Potential palindromic lyase recognition site and further PCB spectral manipulation

The versatility of the maquette chassis is further illustrated in the palindromic placement of lyase recognition sites in the maquette helices demonstrating that nearby aspartate and arginine residues may still be effective even when they are not in a natural consensus sequence. The inverted sequence DRXC inserted into the middle of the fourth helix is still capable of bilin ligation (Figure 3.7), comparable to that seen with a loop Cys (sequence 3.23 shown in gray). Also of interest is that the DRLC binding yield (sequences 3.21 and 3.22) is generally greater than DRAC (sequences 3.18 and 3.19). Protein blast searches of CLRD and CARD natural phycobiliprotein consensus sequences show that the S-type lyase we employed is strongly associated with the CLRD sequence. It appears from the improvement of PCB ligation to CLRD over CARD that even in a manmade protein the S-type lyase shows preference for the CLRD motif. This hypothesis needs to be further tested in sequences that do not also contain an LRK sequence upstream of the Cys in the 4th helix. Though testing of maquette sequence 9, which contains CLRK, demonstrates this sequence is not enough to improve lyase recognition in inter-helical positions on helix 2 (Figure 3.5-A). Overall, this palindromic sequence recognition suggests that it is coarse-scale biophysical chemistry rather than fine-scale pattern recognition that facilitates the PCB binding catalysis.
Figure 3.7: PCB spectral manipulation. PCB binding to an inverted DRXC motif is comparable to the forward CXRD consensus motif in either loops or helices. (a) Maquettes were His tag purified followed by two separate runs on a size exclusion column. Spectra taken in PBS, pH 7.4 normalized to 362μM total protein. Locations for Cysteines for each sequence can be seen in yellow on both the schematic and pymol structure (b) Samples were HPLC purified and run on a SDS-PAGE gel. Coomassie stained maquettes can be seen in visible panel and the fluorescence emission from PCB covalently attached to maquettes can be seen in UV panel. MALDI-MS spectra for HPLC samples are provided below.
Table 3.3: Bilin binding and quantum yields for Figure 3.7. Maquette designs in Figure 3.7 showing the helix and loop patterning, Cys insertion sequence and residue number. Units for extinction coefficient ($\varepsilon$) are in $\text{mM}^{-1}\text{cm}^{-1}$. Absolute bilin binding yield was determined by acidic urea and analytical HPLC. Higher ratio of Qvis to UV absorption bands indicate greater bilin rigidity. Helix charge is Helix 1 -, Helix 2 +, Helix 3 -, and Helix 4 + for all constructs in this experiment. Abbreviations are the same as used in table 3.1.

PCB spectral manipulation rules could be established for palindromic consensus sequences as well (Figure 3.7-a). Histidines (His) are seen to have opposing effects on bilin rigidity, depending on where they are placed. Replacing a core His distant from a CXRD motif with a non-polar Alanine (sequences 3.21 and 3.22) is expected to improve maquette core packing and exclude water molecules. This improvement of core packing leads to an increase in PCB rigidity, observed as an increase in the Qvis to UV absorbance ratio\textsuperscript{47}. Similar effects from exclusion of water molecules in the bilin binding pocket of phytochrome have shown that this loss of water molecules causes a restriction in the conformational flexibility of the bilin which leads to enhancement of fluorescence properties\textsuperscript{28}. The opposing affect of histidine placement on bilin rigidity was shown when adding a histidine near the CXRD motif.
(sequence 3.22) further improved bilin rigidity, presumably by providing a
hydrogen-bonding partner for the bilin.

In order to verify that all of these spectral changes were occurring on Cys-attached PCB, a variety of methods were used for assessment. First, overlap of Coomassie stained maquette bands with fluorescence emission bands in the SDS-PAGE gels (Figure 3.7-b) of HPLC purified bilimaquette samples (see Appendix 2 Figure A2.2.3) verified covalent attachment of PCB. Second, MALDI-mass spectrometry (MALDI-MS) also confirmed that PCB was covalently bound with the assistance of lyase (Figure 3.7-b). Mass spectra correspond to HPLC purified maquettes in Figure 3.7-a. The spectra color and corresponding maquette are as follows: Red- sequence 3.18; Orange- 3.19; Green- 3.20; Blue-3.21; Purple- 3.22; Gray- 3.23. The high laser power of MALDI tends to fragment bilin from maquette leading to lower mass values seen in the left peak. Since the absorbance properties of bilins are considerably modified by the protein environment and by pH, we verified the type of bilin bound and quantitated the yield of bilin binding spectroscopically with conditions developed for the analysis of natural biliproteins, namely unfolding in acidic 8M urea or 6M guanadinium\(^3\) (see Appendix 2, Figures A2.2.4, A2.2.5, and A2.2.6). Absolute yields of bound bilins in Tables 3.1 and 3.3 were calculated by using published bilin extinction coefficients\(^{48,49}\).
3.7: Bilimaquette fluorescence

To develop a successful bilimaquette antenna, it is important that the PCB bound maquette is fluorescent. **Figure 3.8-a** compares the absorbance and fluorescence emission spectra of three sequences described in **Figure 3.7**, exciting at 580 nm near the peak of the PCB absorbance and calibrating the integrated fluorescence yield using cresyl violet perchlorate as a reference fluorophore. The Qvis/UV absorbance ratio is also compared here, with a higher ratio being correlated with a more rigid and extended bilin conformation\textsuperscript{21}. Using these metrics, the DRLC motif not only binds more PCB than the DRAC motif, but this variant also has a higher fluorescence quantum yield. Just as in the experiment in **Figure 3.7**, introducing a His near the end of the helix adjacent to the DRLC motif further improves the Qvis/UV ratio and the fluorescence quantum yield.

**Figure 3.8-b** shows that changing the external charge patterning from near-neutral to negative modestly improves the fluorescence quantum yield from 2 to 3\% (Sequence 3.26 and 3.24 respectively). These yields are comparable to that observed for the PCB binding domain of a natural cyanobacteriochrome\textsuperscript{50}. Shortening the loops had only modest effects on the absorbance and fluorescence. Breaking the central loop (Sequence 3.25), so that the 4-helix bundle assembles from two separate subunits, each with Cys residues, resulted in more than a twofold increase in the yield of bilin binding per Cys, presumably because of the greater accessibility of the Cys residues for the lyase.
However, despite the higher Qvis/UV absorption ratio, the fluorescence quantum yield in the bundle dimer was relatively poor. This suggests that although held rigidly, the bound bilin had a higher susceptibility to nonradiative energy decay. 

Figure 3.8: Fluorescence properties of bilimaquettes (a) A DRLC motif (3.20, 3.22) binds more PCB and has higher fluorescence quantum yield than a DRAC motif (3.18). Absorbance spectra not normalized. Raw data from purification. (b) Changing external charge patterning and loop length (3.10, 3.24, 3.26) had minimal effects on bilin attachment and quantum yield. Breaking the 4-helix bundle into a dimer of two helices (3.25) increases bilin ligation efficiency but lowered the fluorescence quantum yield. Spectra were normalized to 50μM protein.
from increased solvent collisions that resulted from the loss of hydrophobic area from deletion of 2 helices. For the newly presented maquettes SDS-PAGE gels in Figure 3.8-c provide evidence of covalent attachment of bilin and is supplemented with acid urea spectra (Appendix 2 Figure A2.2.6).

<table>
<thead>
<tr>
<th>ID</th>
<th>Helix charge</th>
<th>Loop type</th>
<th>Cys insert</th>
<th>Cys location</th>
<th>His Position on Helix</th>
<th>Binding Yield in vivo Abs.</th>
<th>Quantum Yield (%)</th>
<th>PCB</th>
<th>Qvis/UV</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>- - - -</td>
<td>N N N</td>
<td>CLRD</td>
<td>2</td>
<td>1</td>
<td>6 6 6 6</td>
<td>0.9 (2.9)</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>- + - +</td>
<td>S S S</td>
<td>DRAC</td>
<td>4</td>
<td>9</td>
<td>6 6 6 -</td>
<td>6.2 (0.9)</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>- + - +</td>
<td>S S S</td>
<td>DRLC</td>
<td>4</td>
<td>9</td>
<td>6 6 6 -</td>
<td>7.2 (1.3)</td>
<td>1.27</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>- + - +</td>
<td>S S S</td>
<td>DRLC</td>
<td>4</td>
<td>9</td>
<td>- 6 6, 26 -</td>
<td>9.5 (1.6)</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>- - - -</td>
<td>S4 S5 S5</td>
<td>CLRD</td>
<td>2</td>
<td>1</td>
<td>6 6 6 6</td>
<td>1.3 (3.0)</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>- +</td>
<td>S</td>
<td>CLRD</td>
<td>2</td>
<td>1</td>
<td>6 6</td>
<td>5.6 (0.9)</td>
<td>4.49</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>- + - +</td>
<td>S S S</td>
<td>CLRD</td>
<td>2</td>
<td>1</td>
<td>6 6 6 6</td>
<td>2.1 (2.3)</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>- + - +</td>
<td>S S S</td>
<td>CLRD</td>
<td>2</td>
<td>1</td>
<td>- 6 6 -</td>
<td>(11.6)</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4: Bilin binding and quantum yields for Figure 3.8. Maquette designs in Figure 3.8 showing the helix and loop patterning as in Table 3.1, Cys insertion sequence and residue number. Absolute bilin binding yield was determined by protein unfolding in acidic urea and analytical HPLC. Sequence 3.27 evolved from 3.10, 3.24, 3.25, and 3.26 and will be used for EET demonstrations. Abbreviations are the same as used in table 3.1.

3.8: Comparing the developed bilimaquettes to natural phycobiliproteins

The Q-band absorbance maxima and the Qvis/UV absorbance ratios for high-yield bilimaquettes lie within the range of natural phycobiliproteins ApcA, ApcD, ApcB, ApcE, and CpcB previously reconstituted in E. coli, 615/357, 672/370, 616/370, 662/370, and 592/354 respectively\(^5\). This suggests that PCB is being significantly constrained by the maquette and capable of assuming
at least partly extended conformations (Figure 3.9). Similar comparisons could be made with the phycobiliproteins, CpcB and PecB, from other E. coli reconstitution studies; however, the quantum yields of fluorescence in the range of 1-3% for bilimaquettes are smaller than the 12% and 32% of these two natural phycobiliproteins subunits\textsuperscript{52,53}, suggesting that maquette chromophore rigidity can be further improved. Nevertheless, these levels of fluorescence are adequate for adapting bilimaquette designs as sensors and energy-transfer agents, as shown in Figure 3.11.

A more detailed window into maquette bound PCB geometry can be estimated by comparing the absorption spectra with absorption spectra of known PCB geometries. Bilin chromophore geometry affects the pi-conjugation of the linear tetrapyrrole with significant effects on the bilin absorption spectra. Unrestrained bilins in solution tend to have absorption spectra that are broad in the visible region with a $Q_{vis}/UV$ ratio of around 0.5\textsuperscript{37,38,54}. $Q_{vis}/UV$ absorbance ratios of <1 are associated with mostly helical arrangements of the pyrrole rings. When the bilin conformation is constrained, the $Q_{vis}$ absorbance becomes more narrow and structured\textsuperscript{37}. $Q_{vis}/UV$ absorbance ratios > 3 are associated with the most extended conformations and values between 1 and 3 are associated with intermediate extension\textsuperscript{40,55,56}. 
Earlier versions of the maquettes (Sequence 3.10 from Figure 3.5-c) were exhibiting constrained conformations due to the tight core packing of the maquette; however, their Qvis/UV ratios of <1 indicated the bilin was in a mostly helical conformation. The latter sequences improved to have Qvis/UV ratios that were greater than 1 with maquette 3.25 achieving a ratio of near 4 indicating successful engineering of more extended conformations of PCB. These engineering improvements are highlighted in Figure 3.9, where the Qvis/ UV absorbance ratios of 10 maquettes are compared with theoretical and experimental absorbance ratios as a function of PCB conformer extensions. This comparison shows indications of both forced helical conformation and partly
extended conformations when PCB is constrained by binding to maquettes.

Insight gained from natural biliproteins\textsuperscript{11,61,62} has shown that in order to obtain the more extended conformation of PCB it is important to stabilize the propionate groups on the B and C rings as well as the whole D ring. Using maquette sequence 3.27 \textbf{Figure 3.10} shows a theoretical model for stabilizing an extended PCB in the maquette core. Cys 31 provides the covalent attachment point to the bilin while R35 and K93 provide hydrogen-bonding partners for stabilization of the propionate groups on rings B and C. Stabilization of the D ring is accomplished by a combination of bulky hydrophobic residues that force the D ring into a near planar orientation with rings B and C.

\textbf{Figure 3.10: Extended PCB confirmation-stabilizing residues.} Model of maquette sequence 27 PCB stabilization. Using a molecular dynamics simulation of the sequence 3.27 base structure (provided by Bryan Fry)\textsuperscript{63}, PCB was manually docked into a theoretical stabilization configuration using Pymol. Points of possible contacts between PCB and Maquette are highlighted in (a) and (b)
3.9: Intramaquette excitation energy transfer and *in vivo* near infrared imaging

Once bilimaquettes with adequate photophysical properties were designed, experiments were performed to demonstrate excitation energy transfer (EET) between chromophores bound to the maquette. Sequences 3.10, 3.24, 3.25, and 3.26 gave rise to sequence 3.27, which made it feasible to incorporate a core His opposite the bilin-binding site in the maquette, allowing for the binding of light-active Zn-tetapyrrole macrocycles\(^6^4\). These chromophores function as excitation-energy acceptors in a two-pigment system that includes a bilin excitation donor. We chose two types of bilins to act as energy-transfer donors: PCB and a blue-shifted tripyrrole TPB\(^6^5\) derived from PCB.

PCB-bound maquette 3.27 showed the typical 665 nm absorbance in acidic denaturing conditions. When the pH of a denaturing 6 M guanadinium solution is raised to pH 8, the PCB absorbance broadens and blue-shifts to 650 nm (*Figure 3.11-a*). Raising the pH to 8.5 results in a rapid and dramatic redshift of the absorption to 739 nm, presumably due to chromophore deprotonation\(^6^5\). Over the next 30 minutes, this band disappeared and was replaced by an absorbance feature at 586 nm. Upon lowering the pH to 2, the spectra further blue shifted to 568 nm (*Figure 3.11-b*), indicating an irreversible chemical change had
occurred. Mass spectrometry revealed a mass loss of 109 Da, which is consistent with the scission of one pyrrole ring to form a tripyrrole bilin.

Figure 3.11: Tripyrrole formation (a) Formation of TPB by alkaline scission of PCB bound to Maquette 3.27. Bound PCB in 6 M guanidinium pH 8 has a broad absorption around 650 nm (cyan). Adjusting the pH to 8.5 rapidly shifts to 739 nm (blue); then converts on a minutes time scale to the 586 nm absorbing species (purple) associated with a tripyrrole. (b) PCB at pH 2 converts to TPB upon pH change to 8.5. Lowering the pH back to 2 blues shifts the TPB to a single peak.
(TPB)\textsuperscript{21,66} (see Appendix 2, Figure A2.2.7). The detailed chemical structure of this reproducible tripyrrole derivative has not yet been determined, but the high quantum yield of fluorescence (26\%) makes it an excellent candidate for energy transfer.

With either PCB or TPB anchored to a Cys at the beginning of helix 2, Zn-chlorophyllide \textit{a} (ZnChlide) was ligated to a distal position His on the third helix. These residues are nearly 6 helical turns, or about 32 Å, apart. ZnChlide binding does not distort the bilin absorption spectrum, as confirmed by the absorbance spectrum of the dyad (see Appendix 2 Figure A2.2.8); although, the longer PCB chromophore may weaken the binding affinity for ZnChlide by physically extending towards the binding site for the chlorin. Figure 3.12 shows absorbance spectra for individual energy-transfer pigments. As ZnChlide is titrated into the His binding site, there is a stepwise decrease in bilin emission (Figure 3.13-b,d). Bilin emission levels off at 34\% for the PCB and 19\% for the TPB constructs. Absorbance spectra of the titrations in Figure 3.13 show that bilin spectra are not perturbed upon addition of ZnChlide to the core of the maquette. Care was also taken to avoid
absorbance above Optical Density 0.05 near bilin emission to avoid inner filter effects which would be a false positive for EET between bilins and ZnChlide\textsuperscript{67}.

The Förster resonance energy transfer (FRET) formalism\textsuperscript{68} allows for the energy transfer efficiencies and estimated distances to provide a window into the orientation of chromophores in the maquette core. The FRET distance for randomly orientated bilin-ZnChlide pigment combinations is greater than 50 Å\textsuperscript{69}, which is longer than the maquette itself. Bilin-ZnChlide energy-transfer efficiencies of 66\% and 81\% indicate immobilization of the pigments within the maquette framework such that the optical transition dipoles are in a sub-optimal orientation for energy transfer. Because the singly charged acidic constituent of the amphiphilic ZnChlide chromophore favors surface exposure, the Q\textsubscript{y} optical transition dipole is expected to be nearly perpendicular to the long axis of the helical-bundle. The transition dipoles of the bilins are commonly assumed to be along the direction of the B to D pyrroles\textsuperscript{70}. The relatively non-polar pyrrole rings of the bilins will favor burial in the bundle core to support a partly extended bilin conformation mostly parallel to the long helical axis. These nearly orthogonal transition-dipole geometries are relatively unfavorable for energy transfer, consistent with observed energy-transfer efficiencies. However, maquettes bind a wide range of Zn tetrapyrrroles with different orientations\textsuperscript{64}, allowing a range of dipole geometries. In addition, maquette designs now being characterized include a third cofactor to enable multistep energy transfer and charge separation (see chapter 4).
Another exciting application for intra cellular binding of bilins to maquettes was developed in collaboration with Molly Sheehan in Brian Chow’s lab. Derivatives of maquette sequence 3.2 were found to bind biliverdin in both rat hippocampal neurons and HEK293 human cells (Figure 3.14). Not only did they bind biliverdin but it was bound in such a way that the quantum yields of fluorescence were high enough for intra cellular near infrared imaging.
This finding laid the ground work for a whole host of further applications to be explored using the bilimaquettes in the *in vivo* non-invasive imaging field.

![Diagram](image)

**Figure 3.14: Maquette binding biliverdin in mammalian cells.** Work performed by Molly Sheehan of Brian Chow’s lab showed the bilimaquette’s ability to bind Biliverdin (BV) (a) in mammalian cells: (b) rat hippocampal neurons and (c,d) HEK cells. Upon excitation with approximately 630nm light maximum fluorescence emission around 665nm was observed.

### 3.10: Conclusion

By developing synthetic protein bilimaquettes, this work has explored minimal rules underlying ligation and conformational restriction responsible for the strong and variable absorbance and fluorescence evident in natural biliproteins. These rules when distilled down are as follows: 1) in the absence of
lyases, bilins bind in highest yields to maquettes in vitro and in vivo to Cys residues placed in loops; 2) in vivo PCB attachment yields are highest in the presence of lyases when consensus sequences are placed towards the amino end or in the middle of helices. Binding to Cys is hindered when placed near the carboxyl end of helices; 3) the CXRD motif can be transposed and still interact with the bilin lyases; 4) whether the bilin consensus sequence is transposed or not, interaction with the lyase leads to extension of the bilin conformation to a more linear form. Just like natural phyobiliproteins and phytochromes these rules have allowed us to construct bilimaquettes that can span the whole visible spectrum.

Through the process of establishing these principles, the blue-shifted tripyrrole (TPB) formed from maquette bound PCB was discovered. TPB displays a quantum yield of 26%, the highest recorded in a designed biliprotein. These two types of bilins support intramaquette excitation energy transfer to chlorophyllide acceptors, which are also cofactors that can be incorporated into the maquette in vivo. Gaining control of in vivo design and assembly of bilins in maquettes promises new routes to understanding natural engineering of biliproteins, as well as to novel synthetic light-harvesting photochemical systems and and far-red-light sensors. These bilimaquettes may also provide utility as biosensors due to the attached bilins ability to chelate various metals leading to a modulated output in both the absorbance and fluorescence emission spectra. The present levels of cofactor binding and emission quantum yields set the
stage for optimization of synthetic protein sequence selection through directed evolution.

3.11: References


46. Hirose, Y., Shimada, T., Narikawa, R., Katayama, M. & Ikeuchi,


Chapter 4: Engineering excitation energy transfer relay in a maquette/natural protein fusion

4.1: Introduction

Supramolecular assembly of spatially organized pigments into light-harvesting antenna complexes facilitates excitation energy transfer (EET) that can then be directed towards a variety of applications including photochemical reactions\(^1\). A lot of effort has been invested in creating these functionally adaptable and efficient antennas at low cost, because of the advances they will allow society to make in the development of green, sustainable chemistry\(^2\). Cellular expression of both artificial and modified natural proteins provides a relatively inexpensive means to create adaptable, mono-dispersed molecular frameworks to organize pigments spatially on the one-to-few nanometer scale at which EET typically takes place. Furthermore, an insulating protein coat can prevent undesirable side reactions, including excited-state quenching from pigment contact with other molecules.

Several research groups have enhanced natural photosynthetic and non-photosynthetic protein scaffolds through maleimide linkage of energy-transfer pigments to cysteine residues. An example of this are the use of pigment-coupled natural tobacco mosaic virus protein subunits\(^3\), which can self-assemble into disk-like structures. Mixing disks with different maleimide-coupled pigments leads to self-assembly into stacks; however, this assembly method results in a statistical organization of pigments in space, leading to a range of EET
distances. In another scaffold, rhodamine was used to fill the bacteriochlorophyll “green gap” by maleimide-coupling to detergent-solubilized bacterial photosynthetic light-harvesting complex II (LHII)\(^4\). Additionally, synthetic tetrapyrroles (including bacteriochlorin BC1 used in this chapter) were maleimide-coupled in detergent-solubilized, truncated, mostly alpha-helical b subunits of natural photosynthetic bacterial light-harvesting complexes, LHI\(^5,6\). These oligomeric self-assembly of subunits in detergent offered relatively fine control of spatial organization and energy transfer between coupled pigment and natural bacteriochlorophyll a ligated by a histidine. However they suffer from the complexity of working in a detergent based environment making the technology less adaptable for cellular production.

An alternative approach is to fuse maquettes that bind a light-active Zn-tetrapyrrole to a water-soluble natural light-harvesting phycobiliprotein subunit, creating a protein fusion (Figure 4.1) that can in principle be exploited for supramolecular assembly in vivo. In previous work, truncated natural biliprotein subunit ApcEDelta was fused to a two alpha-helix, two histidine maquette, which dimerizes to a 4-helix bundle\(^7\). Energy transfer between a bilin and Zn-tetrapyrrole pigment pair was successfully demonstrated, but dimeric assembly, and multiple His ligation sites complicated the task of site-specific control of pigment binding. The work reported in this chapter creates an energy-transfer cascade by fusing the natural bilin-binding protein Phycocyanin subunit, CpcA, with a single-chain maquette. This maquette is designed for high site-
selectivity in pigment binding to specific positions within the fusion protein frame to support a combinatorial approach to pigment triad selection and energy transfer relay design. It provides a test-of-concept to move towards the goal of melding natural antenna proteins with synthetic antenna systems to redirect energy flow inside a cellular environment. Unlike earlier work with a truncated biliprotein sequence\(^7\), the fused maquette designs introduced here have relatively high yields of \textit{in vivo} bilin attachment and do not compromise the phycobiliprotein spectral/structural integrity. These are traits that are needed for future development of novel maquette-based photochemical system assembly in the
cell as part of a larger energy-harvesting phycobilisome antenna complex.

4.2: Modular artificial antenna design

The structural robustness of a maquette designed to bind two macrocyclic tetrapyrrole pigments readily permits fusion to the initial helix of natural biliprotein CpcA, a subunit in the middle of the phycobilisome photon-funneling cascade. Panel A of Figure 1 shows a representative Pymol model using the crystal structure of CpcA in gray (PDB 4F0T) including bound bilin (purple) fused to a molecular dynamics structure of a maquette helical bundle (cyan) binding chlorin (gold) and bacteriochlorin (red) cofactors. Panel B details the maquette sequence of four helices joined by three short, glycine-rich loops. A single histidine (His) on either the first or third helix of the maquette is available for ligation of light-active Zn-tetrapyrroles. A cysteine (Cys) placed in the loop joining the first and second maquette helices enables ligation of a free base tetrapyrroles through a maleimide linkage (See Appendix 2 Figure 2.3.1 for full fusion sequence).

A variety of cofactors in different combinations covering a wide region of the visible spectrum can be inserted into the three cofactor-binding slots. For \textit{in vivo} binding of a pigment to slot A, we use phycocyanobilin (PCB) or the relatively blue-shifted phycoerythrobilin (PEB), which anchor to the ligating Cys of CpcA. \textit{In vivo} bilin binding exploits the biosynthetic and attachment machinery common to cyanobacteria and red algae imported into the \textit{E. coli} expression system. This includes heme oxygenase, to cleave the heme macrocycle.
oxidatively to form biliverdin, as well as a biliverdin reductase, either PcyA or PebS for the respective production of PCB or PEB\textsuperscript{11-13}. Although phycobilins can be non-specifically bound to apoproteins \textit{in vitro} without catalytic assistance\textsuperscript{14-16}, we achieve stereochemically correct bilin attachment \textit{in vivo} by expressing bilin lyase subunits CpcE and CpcF\textsuperscript{17}.

Work in \textbf{Chapter 2} demonstrated that amphiphilic Zn-tetrapyrroles with opposed polar and non-polar substituents balance aqueous solubility with tight binding in hydrophobic maquette cores\textsuperscript{18} for easy assembly. Here, we describe two examples: a Zn-chlorophyllide a (ZnChlide (13) in Ch. 2) or an amphiphilic synthetic Zn-chlorin (ZnC (12) in Ch. 2)\textsuperscript{19} binding to histidine in pigment slot B. In the final slot, we used a free-base bacteriochlorin as the most red-shifted pigment and final energy-transfer acceptor\textsuperscript{5}. Fully assembled, the soret and Q absorption bands of pigments in slots A, B, and C cover much of the visible spectrum for light harvesting. Furthermore, the nm-scale spacing between pigments in the fusion maquette is appropriate for Förster resonance energy transfer (FRET), while at the same time suppressing energy-consuming, excited-state electron tunneling between pigment cofactors\textsuperscript{20}.

4.3: \textbf{Relay pigment combinations}

By selecting pigments with absorption maxima in well-separated regions of the spectrum, we achieved both wide-spectrum, light capture and spectral clarity in the contribution of each pigment to energy transfer. The different
degrees of pi-conjugation found in the linear phycobilins and the macrocyclic chlorins and bacteriochlorins spread absorption bands across the visible spectrum while maintaining relatively high extinction coefficients. We control the phycobilin type that binds in vivo to slot A of natural biliprotein CpcA by altering the expression of natural bilin synthases and lyases\textsuperscript{10}.

Spontaneous, rapid binding by a metallochlorin to a histidine in the hydrophobic core of the water-soluble maquette bundle requires the tetrapyrrole to have an amphiphilic nature\textsuperscript{18}. Although this work exploits facile maquette expression in non-photosynthetic \textit{E. coli}, we are also developing fusion maquette expression in a photosynthetic organism to integrate with a natural light-harvesting antenna. When maquettes are expressed in the chlorophyll-producing cyanobacterium \textit{Synechocystis} sp. PCC 6803 rather than \textit{E. coli}, maquette protein spontaneously binds detectable amounts of a chromophore with absorbance peaks and a molecular weight similar to water soluble Mg-chlorophyllide. Thus, binding and energy-transfer assays of the widely available Zn-analogue, ZnChlde, should be relevant for engineering in vivo assembly in photosynthetic organisms. The 650 to 700 nm Q\textsubscript{y} absorption band of B-slot chlorin ZnChlde complements the 550 to 650 nm visible absorption and 620 to 700 nm emission of bilin PCB in slot A. The chlorin 400 to 450 nm B absorption band remains distinct from the other pigments.

The higher energy 510 to 570 nm absorption and 550 to 600 nm emission of bilin PEB in slot A is better paired with the more blue-shifted 575 to 640
nm $Q_y$ absorption of synthetic chlorin ZnC. Both chlorin emissions are compatible with the 700 to 730 nm $Q_y$ absorption band of bacteriochlorin BC1 for energy transfer.

4.4: *In vivo* anchoring of bilins to slot A of fusion proteins

**Figure 4.2: Bilin binding slot A of fusion.** Fusion protein (Sequence 4.2 and 4.4) co-expressed with PEB or PCB synthesis machinery yield deeply colored pink or blue cells. Normalized absorbance (solid) and fluorescence emission (dashed) spectra of the fusion protein with PEB (magenta) or PCB (blue) synthesis are similar to that seen in native CpcA with either bilin (lightly colored solid and dashed curves). Excitation wavelengths were 540 and 580 nm for PEB and PCB, respectively.

Figure 4.2 shows that expression of the maquette/CpcA fusion (Sequence 4.2) and PEB synthesis plasmids (described in the methods A1.3) in *E. coli* results in intensely colored cell pellets; the absorption maximum of the PEB fusion is 555 nm (solid pink trace), essentially identical to the absorption spectrum of the native CpcA-containing PEB (light pink)\(^{10}\). When excited at 540 nm, the fluorescence emission maximum is 565 nm (dashed magenta...
curve), again much like the native protein (light dashed curve), confirming that fusion to the maquette has not interfered with the native-like conformation of CpcA. Substituting PCB synthesis plasmids generated deep-blue cell pellets, and the absorption maximum of the purified fusion protein (Sequence 4.4) was 625 nm (solid blue trace), essentially identical to the absorption spectrum of the native CpcA-containing PCB (light blue). When excited at 580 nm, the fluorescence emission maximum was 640 nm (dashed blue curve), again much like the native protein (light dashed curve). Despite high expression yields, upwards of 50% attachment of bilin to fusion protein can be

![Figure 4.3: Covalent attachment of bilin to biliprotein.](image)

Figure 4.3: Covalent attachment of bilin to biliprotein. His-tagged fusion protein with either PCB (A) or PEB (B) (Sequence 4.2 and 4.3 respectively) was precipitated using 50% saturated ammonium sulfate, resolubilized in acidic 8 M Urea (100 mM HCl) and applied to an HPLC reversed-phase column at pH 2. (C) SDS-PAGE analysis of recombinant proteins. Proteins carrying bilins were identified by fluorescence excited by UV illumination of the electrophoretically separated proteins. His-tag and TEV-tag cleaved PEB and PCB fusion proteins are shown before and after reduction (R) with 14mM 2-mercaptoethanol. (D) MALDI-MS shows a major peak at 33987 Da for phycobilin-bound fusion protein. High MALDI-MS laser power appears to be correlated with bilin destruction and generates a 568 Da lower molecular mass fragment.

97
achieved, as determined by HPLC (see Appendix 2 Figure A2.3.2). Covalent attachment of the bilin pigments was confirmed by UV illumination of SDS gel and by MALDI mass spectrometry (Figure 4.3).

4.5: High affinity chlorin binding to slot B enhances energy transfer properties

Figure 4.4: Titration of chlorins into slot B of maquette. Binding of ZnChlide and ZnC to slot-B. (A) A red shift in absorbance occurs for ZnChlide when it goes from free in phosphate buffer (black) to bound in either the fused (Sequence 4.4, dashed) or unfused (Sequence 4.3, solid) maquette. Approximate orientation of Q_y transition dipole shown as double arrow. (B) Titration of ZnChlide to non-fused maquette with His at position 66. SVD analysis extracts dissociation constants and bound (green) and unbound (black) spectra. Inset: same titration at the 434 minus 377 nm wavelength pair isosbestic point for unbound ZnChlide. (C) Analogous titration for the fusion protein with absorbance of the chlorin-free PCB-bound fusion protein shown in blue. (D) A red shift in absorbance occurs for ZnC when it goes from free in phosphate buffer (black) to bound in either the fused (Sequence 4.2, dashed) or unfused (Sequence 4.1, solid) maquette. Approximate orientation of Q_y transition dipole shown as double arrow. (E) Titration of ZnC to non-fused maquette with His at position 6. SVD analysis extracts dissociation constants and bound (gold) and unbound (black) spectra. Inset: same titration at the 426 minus 412 nm wavelength pair isosbestic point for unbound ZnC. (F) Analogous titration for the fusion protein with absorbance of the chlorin-free PEB-bound fusion protein shown in purple.
We create a second slot for pigment binding in the fusion protein by placing a single His deep enough into the maquette bundle core to allow burial of the hydrophobic regions of Zn-chlorins but still close to the point of maquette-phycobiliprotein fusion. Both position 6 (Sequence 4.2 and 4.4) in the first helix and position 66 in the third helix meet these criteria. Narrowing and red-shifting of absorption bands indicate pigment binding to His. Figure 4.4 shows that ZnChlide binding to His 66 shifts the absorbance from 411 to 425 nm and from 661 nm to 669 nm. Singular Value Decomposition (SVD) analysis of binding titrations to unfused four-helix bundle (Figure 4.4-B) and to the phycobilin-bound fusion protein (Figure 4.4-C) gives ~100 nM dissociation constants (74 ± 11 nM and 245 ± 35 nM). Absorption profiles for the bound and unbound species match the reference spectra in Figure 4.4-A. It was discovered that His 66 provides a stronger binding site for ZnChlide than His 6 (data not shown) therefore further photophysical characterization with ZnChlide in the B slot was performed with the H66 maquette variant.

The blue-shifted 555 nm absorption and 565 nm fluorescence emission maxima of bound PEB are more distant from the absorption bands of ZnChlide, hindering efficient energy transfer. However, an amphiphilic synthetic chlorin (ZnC) with a hydrophobic phenylethynyl group and a polar carboxyphenyl group located on opposite sides of the macrocycle (Figure 4.4-E, 21) has Q-band absorptions nearer the PEB emission and readily binds to maquettes. Figure 4.4-E and Figure 4.4-F show binding-induced bandshifts from 421 to 424 nm.
and from 634 nm to 636 nm for the unfused maquette with a His at position 6 (Sequence 4.1), and to 426 and 638 nm for corresponding fusion protein (Sequence 4.2). Analogous shifts are seen in other maquette designs. SVD analysis of binding titrations to the unfused four-helix bundle (Figure 4.4-E) and to the bilin-bound fusion protein (Figure 4.4-F) provide near μM dissociation constants (1.6 ± 0.2 and 1.0 ± 0.1 μM). This augmentation of ZnC photophysical characteristics is further seen by the increase in ZnC fluorescence by ~69% upon His ligation to the maquette (Figure 4.5). In both the PCB/ZnChlide and PEB/ZnC pigment combinations, pigment binding to the fusion proteins is equimolar with no distortion of the independent pigment absorption profiles.

**4.6: Maleimide-bacteriochlorin binding in slot C**

The energy-transfer pigment triad is completed by placing a Cys in the loop connecting helices 1 and 2, such that a synthetic maleimide-functionalized
bacteriochlorin BC1\textsuperscript{5,19} can bury the bacteriochlorin in the hydrophobic maquette core when the maleimide linker attached to the Cys is fully extended (Figure 4.1). Covalent attachment was confirmed HPLC purification (Appendix Figures A2.3.3 and A2.3.4) followed by mass spectrometry (see Figure 4.6). Figure 4.7 shows absorbance band shifts of the free pigment upon hydrophobic burial, with a sharp increase in the extinction coefficient of the Q band at 713 nm. A similar spectrum is seen when this pigment is bound to the detergent-solubilized beta polypeptide of natural antenna complex LH1 of \textit{Rhodobacter sphaeroides}\textsuperscript{22}. Maquette-bound BC1 spectra are the same with His

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure46.jpg}
\caption{	extbf{Figure 4.6: Covalent attachment of bacteriochlorin.} MALDI mass spectra of (A) Maquette (Sequences 4.1 and 4.3) and (B) Fusion (Sequences 4.2 and 4.4) without and with BC1 covalently attached.}
\end{figure}
Absorbance spectra for both PCB and PEB fusions with BC1 attached can be seen in the insets in Figure 4.7-B and C. These spectra were taken from HPLC fractions, further demonstrating covalent attachment of bilins and BC1 to the same fusion construct. The fusion proteins were then folded and reconstituted with B slot cofactors. Upon adding an excess of ZnChl to the B slot of the PCB:BC1 (Sequence 4.4) fusion protein, followed by PD-10 size exclusion column removal of

Figure 4.7: Absorbance of fully assembled fusion. BC1 slot C binding. A: Free maleimide-bacteriochlorin BC1 (black) narrows absorption features on reacting with maquette (Sequence 4.3) loop Cys and burial in the maquette core (red). B: pigment absorption peaks are resolved across the visible spectrum in both the PCB:Zn-Chl:BC1 (Sequence 4.4) (B: blue:green:red) PEB:ZnC:BC1 (Sequence 4.2) pigment fusion-protein triads (C: purple:gold:red). Insets for B and C shows HPLC photodiode array spectra of samples during purification.
unbound pigment to create a pigment triad, moderately well-separated absorption peaks of the individual pigments can be seen across the visible spectrum (Figure 4.7-B). Similarly, upon adding an excess of ZnC to the B slot of the PEB:BC1 fusion protein (Sequence 4.2), followed by PD-10 size exclusion column removal of unbound pigment to create a pigment triad, clearly resolved absorption peaks of the individual pigments can be seen across the visible spectrum (Figure 4.7-C).

4.7: Energy transfer in two and three pigment assemblies infers compact fusion geometry

The fluorescence emission spectra of Figure 4.8 show that pigment dyads and triads display EET between pigments in all three slots with a range of efficiencies. Figure 4.8-A shows that in unfused maquette (Sequence 4.3) with only slots B and C occupied, ZnChlide emission is quenched upon binding of BC1, and that BC1 emission is greatly amplified when ZnChlide is bound. This emission quenching will be used to determine energy transfer efficiencies ($\eta$) between chromophores by using equation 4.1\(^2\) where $F_{DA}$ is the integrated fluorescence emission with acceptor present and $F_D$ is the integrated fluorescence emission without acceptor.

$$\eta = 1 - \frac{F_{DA}}{F_D} \quad (4.1)$$

Using equation 4.1 to analyze the data in Figure 4.8-A the direct EET is calculated to be 75% in this pigment dyad. In the PCB fusion maquette of Figure
4.8-B (Sequence 4.4), PCB fluorescence emission is quenched upon addition of either ZnChlide in slot B or BC1 in slot C, with the greatest quenching occurring when all three pigments are bound. About 33% of the energy is transferred from PCB to ZnChlide in this pigment dyad, ~55% in the PCB:BC1 dyad, and ~82% to ZnChlide and BC1 in the pigment triad.

Figure 4.8: Excitation energy transfer monitored by fluorescence emission in the various pigment combinations of maquettes. A: Emission from unfused maquette excited at 655 nm with ZnChlide alone (1), BC1 alone (2), or in combination (3). B: Emission from fused maquette excited at 580 nm with PCB alone (1), PCB and ZnChlide (2), PCB and BC1 (3), and PCB, ZnChlide, and BC1 (4). C: Emission from unfused maquette excited at 625 nm with ZnC alone (1), BC1 alone (2), or in combination (3). D: Emission from fused maquette excited at 555 nm with PEB alone (1), PEB and ZnC (2), PEB and BC1 (3), and PEB, ZnC, and BC1 (4).

Insets: Fluorescence quantum yields (Φ) of individual pigments and EET efficiencies (η) and Förster radii (R₀) of pigment pairs in the two maquettes. Pigment to pigment distances (r) calculated from R₀ and η using FRET formalism. A possible fusion protein geometry rotated about the point of fusion consistent with isotropic Förster distances (R₀) is shown. This is the same geometry as figure 1, viewed from a different perspective. Double arrows: possible transition dipole orientations consistent with bound cofactor amphiphilic character.
Efficiencies are estimated from the integrated wavelength range shaded in the figure. **Figure 4.8-C** shows the corresponding energy transfer behavior when pigments ZnC and BC1 fill slots B and C of the unfused maquette (Sequence 4.1). At an equimolar equivalent of ZnC, BC1 emission upon excitation at 625 nm increases by 546%. **Figure 4.8-D** shows nearly equal PEB emission quenching in the PEB fusion maquette (Sequence 4.2) upon the addition of either ZnC in slot B or BC1 in slot C, with the greatest quenching occurring when all three slots are filled. Direct EET is 85% in this pigment dyad. Approximately 48% of excitation energy is transferred from PEB to ZnC in this pigment dyad, 46% in the PEB:BC1 dyad, and ~73% to ZnC and BC1 in the pigment triad.

**Figure 4.8** contains fusion structures that summarize the quantitation of the efficiency of energy transfer (η) between pigment pairs, based on the change in emission upon the addition of the second pigment. This figures also includes the quantum yield of fluorescence (Φf) of phycobilin and chlorin pigments in the absence of other pigments, as calibrated by comparison with reference fluorophores. The figure also shows the isotropic Förster energy transfer radii (R0) between bilins and chlorins and between bilins and bacteriochlorin.23 Pigment to pigment distances (r) are calculated using (η) and (R0) and the FRET formalism illustrated in equation (4.2). These isotropic Förster distances are readily achieved because the fusion protein is likely to be highly flexible at the fusion linkage. In fact simple rotation at this link shows geometries
consistent with the isotropic Förster distances.

The EET efficiencies of 33% and 48% between slots A and B for the PCB-ZnChlide and PEB-ZnC donor-acceptor pairs suggest that in the absence of a unique orientation of pigment transition dipoles, the distance between pigments is similar, $48 \pm 4$ Å. The analogous EET efficiencies of 55% and 46% for the PCB-BC1 and PEB-BC1 donor-acceptor pairs suggest distances between pigments of about $34.5 \pm 2.5$ Å. The difference in the bilin-bacteriochlorophyll isotropic Förster distances suggests that there may be some constraint on the fusion protein orientation that modestly enhances PEB to BC1 energy transfer. These bilin-to-bacteriochlorin EET values are comparable to those for a previous phycobiliprotein-maquette fusion. The flexibility expected from the linking amino acids at the site of fusion can accommodate these Förster distances with a relatively compact structure for the fusion protein (Figure 4.8). Removing this flexible region by fusing a maquette to a truncated CpcA, in which 22 or 32 amino acids are removed from the N terminus, leads to less efficient pairwise bilin energy transfer, presumably by making it more difficult to achieve the compact geometry that brings the phycobilin close to the other pigments (Data not shown).

The 85% and 75% EET efficiencies for the ZnC-BC1 and ZnChlide-BC1 donor-acceptor pairs are comparable to those reported in modified natural LHI antennas and correspond to isotropic Förster distances of 33 Å and 21 Å, respectively. However, the donor cofactor is significantly constrained by the Zn-histidine ligation on binding, placing the tetapyrrole plane between the
long axes of the alpha helices, and by the orientation of the hydrophilic
tetrapyrrole substituents towards the aqueous exterior, as seen in X-ray crystal
structures of Zn-tetrapyrrole binding maquettes\textsuperscript{24}. The $Q_y$ transition dipole may
adopt one of two perpendicular orientations depending on which face of the Zn-
tetrapyrrole receives histidine ligation. For ZnC, orienting the polar carboxylic
acid towards the aqueous surface leads to the two $Q_y$ transition orientations
making an angle of nearly $45^\circ$ with respect to the long bundle axis (Figure 4.9-
A). For ZnChlde orienting the polar carboxylic acid towards the aqueous surface,
angles of nearly $25^\circ$ and $65^\circ$ (Figure 4.9-B) result for the two $Q_y$ orientations.
The BC1 acceptor orientation is constrained by hydrophobic forces that drive the
bacteriochlorin and the linker to the loop cysteine into the helical bundle core
between alpha helices. The center-to-center distances between either chlorin and
the bacteriochlorin are about 16 Å in the likely geometry with a mostly extended
linker that places the $Q_y$ transition dipole of BC1 at a $45^\circ$ to the long axis of the
bundle (Figure 4.9-C and D).

This known 16 Å distance ($r$) and calculated energy transfer efficiencies
($\eta$) from Equation 4.1 allow Equation 4.2 to be solved for $R_0$ independent of $\kappa^2$
Equation 4.3.

$$\eta = \frac{R_0^6}{R_0^6 + r^6} \quad (4.2)$$

$$R_0 = \left(\frac{r^6 \times \eta}{\eta - 1}\right)^{1/6} \quad (4.3)$$

Plugging in the sixth root value for $R_0$ that is real into equation (4.4) and solving
for $\kappa^2$ gives a value to plug into Equation 4.5 and determine angles for dipoles of cofactors.

$$R_0 = 9.78 \times 10^3 \left( \kappa^2 n^{-4} Q_D J(\lambda) \right)^{1/6} \quad (4.4)$$

$$\kappa^2 = (\cos \theta_T - 3 \cos \theta_D \cos \theta_A)^2 \quad (4.5)$$

For a 16 Å distance between donor and acceptor, the 85% energy-transfer efficiency between ZnC and BC1 gives a Förster kappa squared value of 0.045, indicating that the transition dipoles are nearly orthogonal to each other. Near orthogonality is only possible if the tetapyrrole planes of the donor and acceptor are nearly parallel to one another, as would be expected if the widening of the inter-helix spacing to accommodate His ligation and binding of ZnC also facilitated insertion of the bacteriochlorin ring (Figure 4.9). Furthermore, the requirement of near orthogonality indicates that bound ZnC has one strongly preferred face for histidine ligation, and that BC1 also has a preferred $Q_y$ orientation within the bundle, likely reflecting a preferred core amino acid packing around a bent bacteriochlorin macrocycle.

Extending the polar carbonyl on the fifth ring of ZnChlide towards the aqueous domain is likely to tilt the $Q_y$ transition of this chlorin by about 20° relative to the orientation of ZnC. If the BC1 assumes a similar binding orientation as with ZnC, then the ZnChlide-BC1 donor-acceptor transition dipoles will be considerably less orthogonal than for ZnC-BC1 pair. This is borne out in the EET value of 75% and the corresponding kappa squared value of 0.33, which
indicates a roughly 65° angle between the transition dipoles Figure4.9-B. This is consistent with the ZnChlide macrocycle, also assuming an orientation approximately parallel to the BC1 with a strong preference for His ligation to one face of the chlorin.

![Figure 4.9: Geometry of circular tetrapyrroles in maquette.](image)

Relationship between $K^2$ and donor emission and acceptor dipoles angles for (A) ZnC and BC1 and (B) ZnChlide and BC1 bound to maquette.

The energy transfer efficiency from the bilin to the other two pigments in the fusion protein pigment triads ($\eta_{ABC}$) can be estimated independently from the pairwise EET values ($\eta_{AB}$ and $\eta_{AC}$) as follows: $\eta_{ABC} = (\eta_{AB} + \eta_{AC} - 2 \eta_{AB} \eta_{AC})/(1 - \eta_{AB} \eta_{AC})$. The calculated values of 63% and 64% for the PCB-ZnChlide-BC1 and PEB-ZnC-BC1 triads compare with observed values of 82% and 73%, respectively. The observed values of 82% and 73% are notably larger than the calculated values of 63% and 64% for the PCB-ZnChlide-BC1 and PEB-ZnC-BC1 triads. An increase in energy transfer efficiency is expected if the
fusion protein assumes a small range of compact geometries, where geometries that place the bilin further from the chlorin, and thus decrease this energy transfer, are compensated by placing the bilin closer to the BC1, and increase this energy transfer.

4.8: Three-pigment energy transfer relay

In addition to parallel energy transfer from phycobilin to chlorin and from phycobilin to bacteriochlorin, serial energy transfer from phycobilin to chlorin to bacteriochlorin is evident from the enhanced PEB absorbance cross section for BC1 emission when ZnC is inserted as an intermediate in an energy-transfer cascade (Figure 4.10). Although similar relay behavior is expected for the PCB-ZnChlide-BC1 triad, smaller quantum yields of ZnChlide emission lower the relay effect. Excitation spectra monitoring the lowest energy emission from BC1 provide clear evidence of the supporting role of the slot B pigment in energy transfer from bilins in slot A to BC1 in slot C. When loaded with all three phycobilin, chlorin, and bacteriochlorin pigments, the fusion maquettes gain substantial absorption across much of the visible spectrum. Nevertheless, the spectral absorption peaks of individual pigments are relatively well resolved (Figure 4.10-A). By monitoring the BC1 emission at 720 nm as a function of the excitation wavelength, the contribution of each pigment to EET and the final emission by BC1 are readily determined (Figure 4.10-B). The PEB:BC1 pigment dyad (brown) reveals the contribution of the PEB absorption peak near 555 nm to
BC1 emission. Adding the chlorin, ZnC, to slot B (blue) adds to BC1 emission absorption by the chlorin near 640 nm.

Figure 4.10: Demonstration of a B slot energy transfer relay ZnC in slot B assists energy transfer from PEB in slot A to BC1 in slot C. A: Absorption spectrum (solid lines) of fusion PEB:BC1 dyad (brown) and PEB:ZnC:BC1 triad (blue) (Sequences 4.2). B: Corresponding excitation spectra (dashed lines) collected at 720 nm emission. Emission from fusion protein with just PEB (black) or PEB and ZnC is minimal at 720 nm. Overlay of excitation spectra and absorbance spectra normalized to BC1 absorbance at 714 nm (C) PEB:BC1 and (D) PEB:ZnC:BC1.

Critical evidence of the relay effect of EET from PEB to ZnC to BC1 is seen at the 555 nm absorption band of PEB (Figure 4.10-B) where ZnC has
virtually no absorbance (Figure 4.5). The cross section for PEB light absorption resulting in BC1 emission nearly doubles when ZnC is bound. If there were no relay, this cross section would drop instead, because EET to ZnC would rob energy otherwise flowing to BC1. The doubling shows that indirect relay EET via ZnC is just as effective as direct EET from PEB to BC1. In a construct in which PEB was held distant from BC1, we expect this relay enhancement will be greater than a factor of two, although overall energy transfer efficiency would suffer. Analogous EET relay behavior takes place when chlorin, ZnChlide, is added to slot B of a PCB:BC1 fusion protein; however, the broad spectral emission of PCB itself overlaps with BC1 emission and makes separation of component contributions to EET difficult.

4.9: Conclusion

Successful engineering of solar-driven catalytic systems must address stability, ease of synthesis, and regeneration of degraded components. The approach of fusing natural and artificial proteins for multi-step energy transfer described here directly addresses these needs. The maquette chassis provides a protective, high thermal-stability environment for varied light absorbing pigments inside or outside the cell. Cellular expression of these fusion proteins provides a facile and adaptable means of synthesizing the architecture to spatially orient various energy-transfer pigments. With the help of synthases and lyases, phycobilins insert during expression much as in the natural phycobiliproteins,
while the maquette His ligation sites scavenge metallotetrapyrroles. Insertion of different colored bilins into maquette-phycobiliprotein biohybrid constructs is effectively controlled by plasmid selection. Nanometer energy-transfer distances are effectively selected by positioning ligating amino acids in the maquette framework. Unlike chemically synthesized arrays, regeneration of degraded components is continuous in cellular expression systems.

The challenge now is to adjust the amino acid environments of the maquette tetrapyrrole binding sites to favor site-specific binding of naturally available or genetically introduced tetrapyrrole (porphyrin, chlorin and bacteriochlorin) cofactors. In vivo bis-His ligation of hemes with natural electron-transfer activity in maquettes is already established\textsuperscript{27,28}. These hemes are used as partners in maquette-based light-induced electron-transfer systems that produce oxidants and reactants, as will be needed to support solar-powered catalysis. Moving the maquette fusion protein expression system into a natural phycobiliprotein-producing photosynthetic organism, such as \textit{Synechocystis} sp. PCC 6803, will provide both light-active pigments for self-assembly and a means to engineer an association of the fusion maquette with the phycobilisome via formation of a phycocyanin protein that is made of CpcB and CpcA/Maquette subunits. Such assembly is a plausible means to create energy-diverting artificial protein modules to power photo-catalysis.

\textbf{4.10: References}


Chapter 5: Conclusions

The work performed in this thesis moves the maquette platform further into the realm of complete in vivo assembly. This is an area where protein design arguably offers the most potential, because no longer does the laborious task of protein purification and assembly with cofactors have to be performed in vitro to gain usefulness from these designed proteins. Hypothetically, high-energy products produced by maquettes in vivo can be syphoned away from the cells without needing to kill the cells. Each of the three chapters in this thesis offers a key area of progress towards in vivo holo-assembly of functionally useful maquettes.

Chapter 2 establishes the principles for binding of circular tetrapyrroles to the maquettes. This is important because these cofactors are crucial for the functionality of natural photochemical systems. If we hope to design novel photochemical systems using the maquettes, we need to understand how to utilize the naturally available tetrapyrroles. The understanding that it is the amphiphillic nature of the tetrapyrrole that drives it into the core of the maquette\textsuperscript{1,2} provides justification for moving in vivo work forward in photosynthetic microorganisms. Chlorophyll and many of the derivatives that are naturally available possess this amphiphillic nature. Preliminary work has shown that it is
in fact possible to bind chlorophyll derivatives to maquettes in cyanobacteria, further providing evidence that the amphiphillic nature of biologically available tetrapyrroles is a viable mode of exploitation for holomaquette assembly \textit{in vivo}. The future task of exploring a much larger mutational landscape to elevate levels of chlorophyll binding in maquettes \textit{in vivo} is one of the most promising and exciting new avenues inspired by the work in chapter 2.

Chapter 3 provides another avenue for \textit{in vivo} maturation of tetrapyrrole containing maquettes. This work shows that the biologically available linear tetrapyrroles called bilins, also an amphiphillic molecule, have a propensity for partitioning into the core of the maquette. There are many advantages to adding bilins to the \textit{in vivo} cofactor-binding repertoire. Bilins offer greater spectral tunability without modification to the chemical structure, have high quantum yields, and allow the maquette antennas to absorb in the green region of the visible spectrum. Another important finding shown here is the maquette’s ability to directly interface with bilin lyases to not only increase bilin attachment yields, but also, to affect the geometry of bilin insertion into the maquette interior. This shows that the earlier interaction of the maquettes with c-type heme maturation machinery is not an isolated incident\textsuperscript{3}. Without any evolutionary mechanism taking place, the maquettes are again able to interface with natural proteins. Improvement on these processes should easily be taken care of by applying an evolutionary selection mechanism mixed with structurally guided information. The work in this chapter not only provides the means to assemble light-
harvesting antennas *in vivo*, but it also strengthens the argument for a continued search for other biological means of maquette interfacing. It seems the possibilities are endless.

The work in chapter 3 was not only a proof of principle, but also showed a direct application to the biomedical field. Work performed in collaboration with the Chow lab showed that maquettes could also assemble with biliverdin in mammalian cells. This provided a tool for non-invasive near-infrared imaging with a protein that is half the size or smaller than green fluorescent protein (GFP). This is only the beginning; once these bilimaquettes are located to the membrane they can be further utilized as voltage indicators because of the effect electric fields have on bilin fluorescence.

Chapter 4 provides yet another path for maquette *in vivo* assembly. The fully functional protein fusion of the maquette with a biliprotein subunit provides a potential avenue for maquettes to connect to the supramolecular phycobilisome complex in cyanobacteria. Redirection of harvested light energy to the maquette anchored to the phycobilisome offers a starting place for a spatially oriented photochemical system. Fusing maquette constructs that have been designed for chlorophyll binding from work inspired in chapter 2 could make *in vivo* charge separation a reality. A fusion has already been made with the charge-separating maquette designed by Nathan Ennist. His x-ray crystal structure information\(^4\) for this maquette should greatly speed up the whole design process.

Combining the work performed in this thesis with larger guided
mutational landscape searches offers the starting point for an endless amount of possible photochemical and catalytic designs. The potential for this work is supported by discoveries made by Michael Hecht who has shown the versatility of these 4 alpha-helical bundles in redesigning the proteomes of *E. coli* that were lacking essential proteins for life.

As a subset of protein design, this type of work has yet to blossom, but it offers an approach that is incremental and compartmentalized. As we begin to build more individual functional components in the maquettes, we will be able to tie them together to create the types of function that make the maquette and similar approaches no longer a just a subset, but an industry standard for protein design. A key part of this blossoming is the integration of the maquette technology into living systems. This thesis brings complete functional integration one-step closer.

### 5.1: References

Appendix 1: Materials and Methods

A1.1: Chapter 2

A1.1.1: Maquette expression

Codon optimized synthetic genes were obtained from DNA2.0 in PJ414 vector. The protein was expressed with a histidine tag in *E. coli* BL21 (DE3) cells for 5 hours at 37° C, after induction with isopropyl-thiogalactopyranoside (IPTG) (0.5 mM). The cells were harvested by centrifugation, resuspended in KH₂PO₄ buffer with octylthioglucoside (1%), and lysed by sonication with a micro-tip attachment. Lysate was centrifuged at 25,000 g for 25 minutes, with supernatant applied to a Ni nitritetriacetic acid superflow resin (Qiagen) on an Akta FPLC. The His-tag was cleaved by recombinant tobacco etch virus N1a protease overnight, and final purification was via Waters reverse-phase HPLC. Molecular weight was assayed by MALDI mass spectrometry.

A1.1.2: Porphyrins

Zn(II) 5,10,15,20-tetra(4-sulphonatophenyl)-porphyrin (1), Zn(II) 5,10,15-tri(4-sulphonatophenyl)-5-phenyl-porphyrin (2) and Zn(II) 5,10,15,20-tetraphenylporphyrin (3) were purchased from Frontier Scientific, all other Zn porphyrins with different substituent groups were synthesized and are referred to with the following numbers in Chapter 2, Zn(II) 5-(4-(carboxyphenyl)-10,15,20-triphenylporphyrin (4), Zn(II) 5,15-di(4-carboxyphenyl)-10,20-diphenylporphyrin (5), Zn(II) 5,10-di(4-carboxyphenyl)- 15,20-diphenylporphyrin (6), Zn(II)
5,10,15-tri(4-carboxyphenyl)-5-phenylporphyrin (7), Zn(II) 5,10,15,20-tetra(4-carboxyphenyl)porphyrin (8), Zn(II) 5-phenyl-15-(p-carboxyphenyl)porphyrin (9) and Zn(II) 5-phenyl-15-(p-carboxyphenyl)porphyrin with first generation Newkome dendrimer (10).

A1.1.3: Chlorins

Amphiphilic chlorins 15-(4-carboxyphenyl)-17,18-dihydro-18,18-dimethyl-5-p-tolylporphyrin (11) and Zn(II) 15-(4-carboxyphenyl)-17,18-dihydro-18,18-dimethyl-3-(phenyl–ethynyl) porphyrin (12) were synthesized with a de novo method that enables introduction of substituents at desired sites about the perimeter of the macrocycle. Acid-promoted condensation of the Eastern and Western halves is followed by metal-mediated oxidative cyclization. The selective introduction of substituents relies on (i) the use of substituted precursors (Eastern and Western halves); or (ii) bromination of the chlorin macrocycle followed by Pd-mediated coupling reaction. Both strategies were used in the synthesis of target chlorins. Pheophorbide a is purchased from Frontier scientific and Zn(OAc)$_2$ is inserted by refluxing 5 molar equivalents of ZnCl$_2$ in methanol as previously described. Bacteriochlorophyll was extracted from _Rb. sphaeroides_, then demetalated, phytol chain cleaved before Zn insertion to form Zn bacteriochlorophyllide. More detailed information on synthesis and characterization of porphyrins can be found in the supplementary data of reference.
A1.1.4: UV/Visible and circular dichroism (CD) spectroscopy

Protein solutions were prepared in CHES buffer (20 mM, 150 mM KCl, pH 9.0). Binding was monitored by UV/Vis Soret band absorbance on a Varian Cary-50 spectrophotometer at room temperature in a 1 cm path quartz cuvette. Secondary structure was monitored by CD spectroscopy (Aviv Model 410) at 25°C with a 1 mm path quartz cuvette. Thermal denaturation was followed by monitoring the ellipticity at 222 nm every 5°C after 15 minutes of equilibration. Melting temperatures were calculated using a Boltzmann equation with one term for each observed transition.

A1.1.5: Cofactor binding affinity

Pigment stock solutions were weighed out and solubilized in dimethylsulfoxide to give concentrations of 1 mM Zn porphyrins, 1 mM chlorin (11), 500 μM chlorin (12), and 2 mM chlorin (13). Binding of Zn porphyrins shift the Soret peak from 422 nm to 432 nm. Binding of chlorins shift the Soret band from 415 nm to 423 nm for (12), 411 nm to 425 nm for (13), and 409 nm to 417 nm for (11). Dissociation constants (K_d values) were determined as follows. To 1 ml of protein solution at 25°C 0.2 (porphyrin) or 0.1 (chlorin) equivalent aliquots were added successively to obtain a spectrum from 300 to 700 nm. Titrations typically ended at 4 equivalents of cofactor. The matrix of spectral absorbance
from 300 to 700 nm for each of the additions of cofactor was subjected to singular value decomposition\textsuperscript{6} using a customized Mathematica program, revealing two dominant singular values, as expected for a system with two spectral species, one bound and one unbound. The amplitudes of the first two principle components as a function of total pigment added fit well to a mathematical model of a single dissociation constant for each protein binding site to generate fitted spectral extinction coefficients at each wavelength for the unbound and bound pigment spectra.

Dissociation constants for chapters 2 and 4 were determined using a derivative of the equation used in\textsuperscript{2}. The measured absorbance ($\text{Abs}$) of the titration is the sum of the extinction coefficients of bound ($\varepsilon_{\text{bound}}$) and unbound ($\varepsilon_{\text{unbound}}$) tetrapyrrole multiplied by the concentrations of bound $[\text{Tet}_{\text{bound}}]$ and unbound tetrapyrrole $[\text{Tet}_{\text{unbound}}]$ respectively (Eq. 5.1).

$$\text{Abs} = (\varepsilon_{\text{bound}} \times [\text{Tet}_{\text{bound}}]) + (\varepsilon_{\text{unbound}} \times [\text{Tet}_{\text{unbound}}])$$ \hspace{1cm} \text{Eq. 5.1}

Taking absorbance differences of the $\text{Tet}_{\text{unbound}}$ isobestic wavelengths determined from the SVD analysis allow us to negate any absorbance contribution from $\varepsilon_{\Delta m\_unbound}$ because at these wavelengths $\varepsilon_{\Delta m\_unbound} \equiv 0$ (Eq. 5.2).

$$\Delta \text{Abs} = \varepsilon_{\Delta m\_bound} \times [\text{Tet}_{\text{bound}}]$$ \hspace{1cm} \text{Eq. 5.2}

Using three equations we can solve $[\text{Tet}_{\text{bound}}]$ for terms that all can be easily empirically determined.

$$K_d = \frac{[\text{Tet}_{\text{unbound}}] \times [\text{Mac}_{\text{unbound}}]}{[\text{Tet}_{\text{bound}}]}$$ \hspace{1cm} \text{Eq. 5.3}
\[ [\text{Tet}_{\text{bound}}] + [\text{Maq}_{\text{unbound}}] = [\text{Maq}_{\text{total}}] \quad \text{Eq. 5.4} \]

\[ [\text{Tet}_{\text{bound}}] + [\text{Tet}_{\text{unbound}}] = [\text{Tet}_{\text{total}}] \quad \text{Eq. 5.5} \]

With these 4 equations we can now write an equation with all variables known besides \( K_d \). I have left out concentration brackets for ease of reading.

\[
\Delta \text{Abs} = \epsilon_{\Delta nm, \text{bound}} \times \frac{(K_d + \text{Maq}_{\text{tot}} + \text{Tet}_{\text{tot}} - \sqrt{(K_d + \text{Maq}_{\text{tot}} + \text{Tet}_{\text{tot}})^2 - 4\text{Maq}_{\text{tot}}\text{Tet}_{\text{tot}}})}{2}
\]

Eq. 5.6

The change in absorbance data was then plotted as a function of concentration and Mathematica was used to find best-fit values for \( K_d \).

**A1.1.6: Stopped-Flow spectroscopy**

Tetrapyrrole and protein were added to separate syringes of an OLIS RSM 1000 stopped flow spectrophotometer, which takes a full visible spectrum every millisecond through a 2 cm flow cell after rapid (~2 ms deadtime) mixing. Temperature was controlled with a Fischer-Scientific IsoTemp 3031 water bath. Individual wavelengths were selected for further kinetic analysis. All experiments were performed in triplicate, and the data were averaged together for further analysis.

**A1.2: Chapter 3**

**A1.2.1: Gene constructions**

Codon-optimized synthetic genes were obtained from DNA2.0 in a PJ414 expression vector. Mutations in the parent vector were made with primers.
synthesized by IDT or Invitrogen (see Appendix, Table A2.3.1, 2, and 3).
Sequences were verified using UPenn DNA sequencing core.

**A1.2.2: Maquette expression**

A plasmid encoding the CpcS lyase of *Thermosynechococcus elongatus* was coexpressed with a plasmid encoding *pcyA* and *ho1* (PCB biosynthetic enzymes). His-tagged maquettes and bilin biosynthetic and CpcS lyase-attachment enzymes were expressed in *E. coli* BL21 (DE3). Strains initially containing bilin biosynthetic plasmids were made competent to take up the maquette plasmid using the Hanahan method. Cells were grown in Terrific Broth to an OD value between 0.8 and 1.0 at 37°C with shaking at 205 rpm; cells were then induced by addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and cultures were further incubated for 21 h at 20°C with shaking at 260 rpm in beveled flasks. (Size of cultures used for Figures 3.5-100mL, Figure 3.6-100mL, and Figure 3.7-500mL.)

**A1.2.3: HPLC methods**

Solvents used for pigment and protein purification were a mixture of acetonitrile (ACN) and 0.1% trifluoroacetic acid (TFA) in water. A Waters HPLC system was used with VydaC C18 analytical and preparative reversed-phase columns. Before HPLC injection, protein and pigment solutions matched the starting gradient TFA and ACN concentrations.

**A1.2.4: Protein purification**

Pelleted cells were resuspended and homogenized in 300
mM NaCl, 20 mM imidazole, 50 mM NaH$_2$PO$_4$ pH 8 buffer prior to lysis by sonication. The lysate was centrifuged at 25,000 x $g$ for 25 minutes, and the resulting supernatant was applied to column containing Ni-NTA superfloow resin, and the protein was purified by a gravity–flow column method (Qiagen https://tools.thermofisher.com/content/sfs/manuals/MAN0011700_HisPur_NiNTA_Resin_UG.pdfPage 3). PCB-biliprotein 27, expressed without a His-tag was purified by solvent extraction; cell pellets were washed with 0.1% TFA in water and then sonicated with 50% acetonitrile, 0.1% TFA in water. Initial extractions lacked blue color and were discarded; after ~60 mL, subsequent blue extractions were pooled and evaporated under vacuum until dry.

Unless otherwise noted, both tagged and untagged proteins were purified by HPLC using a 30-mL gradient from 30 to 42% ACN; in-line absorption spectra were collected during the gradient, and the elution profiles were used to quantify the efficiency of PCB attachment to maquette variants (Figure 3.6; also see Appendix, Figure A2.2.3). Purified bilimaquettes were collected, lyophilized, and stored at -20° C in the dark until used. The molecular weight was determined by both SDS-PAGE (Invitrogen Novex NuPAGE electrophoresis system, 4–12% bis-Tris gels, MES running buffer) and mass spectrometry (MALDI)$_2$ (in figure and see Appendix, Figure A2.2.9).

**A1.2.5: Pigment purification**

PCB was extracted using a modified procedure from$^{11}$. The thioether bond of PCB was cleaved by refluxing a fusion protein (50 mg) between CpcA and
maquette\textsuperscript{12} overnight with methanol. The resulting mixture was evaporated to dryness under vacuum and extracted with 1 mL of ethanol. The crude PCB was purified by reversed-phase HPLC using a 20 to 70\% ACN gradient (500 ml) on a C-18 column (see Appendix, Figure A2.2.10). Care was taken to avoid fractions with spectra indicative of bilin metal chelation. MALDI-MS verified the mass of PCB (see Appendix, Figure A2.2.11).

PEB was obtained from a modified CpcA sequence missing the ligating Cys (Appendix. Table A2.2.4) that binds PEB noncovalently. After purification by immobilized metal affinity chromatography, an aliquot of protein (25 mg) was precipitated by adding methanol to 50 followed by a half volume of CHCl\textsubscript{3}. The pigmented CHCl\textsubscript{3} phase was dried by rotary evaporation at 40° C. MALDI mass spectrometry was performed using a 2,5-dihydroxybenzoic acid matrix (see Appendix, Figure A2.2.12). Analytical reversed-phase HPLC verified the purity (see Appendix, Figure A2.2.13). Pigment concentrations were calculated by using extinction coefficients (\(\varepsilon\)) for free PEB of 25.2 mM\textsuperscript{-1} cm\textsuperscript{-1} at 591 nm\textsuperscript{13} and for free PCB \(\varepsilon = 37.9\) mM\textsuperscript{-1} cm\textsuperscript{-1} at 690 nm in 5\% (v/v) HCl/methanol.

**A1.2.6: \textit{In vitro} bilin attachment**

Maquettes were reduced for at least one hour with 5 mM DTT, followed by removal of DTT by PD-10 size-exclusion chromatography before addition of the bilin. Reaction with BV was performed in PBS buffer at pH 7.4 for 4 h with 100 mM protein and 500 mM BV. Free BV was separated from the BV-maquette...
by chromatography on a size-exclusion column that had been equilibrated with PBS pH 7.4. *In vitro* treatment with PEB was performed overnight in 50 mM Tris-HCl buffer, pH 8.0, while PCB attachment was performed in 50 mM NaPO₄ buffer, pH 7.0 as described in¹⁴. Protein concentrations and bilin concentrations, 60 mM and 30 mM, respectively, were used. The products of *in vitro* PEB/maquette attachment reactions were purified using a gradient (700 mL) from 30% to 60% ACN for sequences 1, 3, and 4 and a 500-mL gradient from 20% to 70% ACN for sequence 2. *In vitro* maquette attachment reactions with PCB were purified using a 500-mL gradient from 20% to 70% ACN for maquette sequences 1 through 4. Attachment of PEB and PCB was verified via MALDI-MS (see Appendix, Figures A2.2.14 and A2.2.15). To form the tripyrrole, sequence 27 with bound PCB was solubilized in 6 M guanidinium-HCl, pH 8 and raised to pH 8.5 with NaOH. After 30 min, the desired protein was purified by reversed-phase HPLC using a 500-mL gradient from 20% to 70% ACN.

**A1.2.7: Protein, PCB concentration, and quantum yield determination**

UV/Vis absorbance spectra were measured with a Varian Cary-50 spectrophotometer at room temperature in a 1-cm pathlength quartz cuvette. Fluorescence spectra for determining quantum yields and fluorescence emission spectra were measured with a Horiba Fluorolog 2 fluorimeter at 20° C in a 1-cm pathlength quartz cuvette. The PCB concentration was calculated by denaturing the recombinant protein in 8 M urea at pH 2 and using the molar extinction
coefficient ($\varepsilon$) at 660 nm = 35.4 mM$^{-1}$ cm$^{-1}$.$^{15}$ Spectra of HPLC purified material showed that PCB absorbance at 280 nm is minor in comparison to tryptophan (Trp) absorbance so protein concentrations were determined using 280 nm absorbance. Relative bilin content was assayed by normalizing the 660 nm absorbance in acidic urea to the 280 nm Trp absorbance in PBS pH 7.5 after purification.

To determine the fluorescence quantum yield of PEB and PCB bound to maquettes using the method in$^{16}$, residual imidazole was removed by PD-10 size exclusion chromatography. Cresyl violet perchlorate, with a quantum yield of 0.54 in methanol$^{17}$, was used as a standard. To determine the quantum yield of BV bound to maquettes, four dilutions of each BV-maquette and Cy5 standard in PBS were made. Absorbance at 600 nm was plotted against integrated emission from 635 nm to 830 nm. Based on the previously published quantum yield of Cy5 of 0.27, the quantum yield was determined from the slopes of the lines determined for the dilutions relative to that for Cy5$^{18}$.

**A1.3: Chapter 4**

**A1.3.1: Gene construction**

The maquette sequence with an amino terminal octa-His tag (see Appendix A2.3.1 and A2.3.6) was contained on a PJET expression plasmid (DNA 2.0). This plasmid was linearized by PCR at the 3’ end of the maquette gene. The gene for CpcA from Synechocystis sp. PCC 6803 was
amplified from plasmid BS414v\textsuperscript{19,20} by PCR (see Appendix Figure A2.3.6). Primers used for amplification of \textit{cpcA} (see Appendix Table A2.3.1) had homologous sequences to the 3’ end of the maquette gene added to the 5’ end of the primers. The \textit{cpcA} gene and linearized maquette vector were assembled using Gibson Assembly\textsuperscript{21} with components provided by New England Biolabs, to create the maquette/\textit{cpcA} gene fusion. Unneeded histidine residues in the helices were removed using a point mutation protocol (Stratagene Quick Change) to leave a single His Zn-Chlorin ligation site on either helix 1 or 3. Sequences were verified at the DNA sequencing core facility at the University of Pennsylvania. Fusion constructs were transformed into \textit{E. coli} BL21DE3 strains containing plasmids with heme oxygenase genes (for biliverdin production), bilin lyase subunits \textit{CpcE} and \textit{CpcF}, and phycoerythrobilin synthase (\textit{pebS}) for production of phycoerythrobilin\textsuperscript{19}. Alternatively, for the production of phycocyanobilin, \textit{pebS} was inactivated by a three amino acid deletion near the bilin-binding Cys, and a second plasmid containing the 3Z-phycocyanobilin:ferredoxin oxidoreductase gene (\textit{pcyA}) was co-expressed (See Appendix, Figure A2.3.7).

\textbf{A1.3.2: Protein purification}

Cultures (2 L) were grown with shaking at 205 rpm in a New Brunswick Innova 4230 Refrigerated Benchtop Incubator Shaker at 37° C in Terrific Broth\textsuperscript{22} to an OD\textsubscript{600} between 0.8 and 1, then induced with 1 mM isopropyl-
thiogalactopyranoside (IPTG) for 20 h at 20° C. Cell pellets were suspended in 300 mM NaCl, 20 mM imidazole, 50 mM NaH$_2$PO$_4$ pH 8 buffer, homogenized and lysed by sonication using a microtip probe. Cell debris was then pelleted by centrifugation for 20 min at 25,000 × g. The supernatant was added to NiNTA superflow resin (Qiagen) on an Akta FPLC, washed with 5 volumes of suspension buffer and eluted with 500 mM imidazole. Phycobilin-bound and bilin-free fusion proteins were separated before His-tag removal using a C4 reversed phase column. A 100 ml wash of 20% acetonitrile (ACN) 80% water, 0.1% trifluoroacetic acid (TFA), pH 2 was followed by a 500 ml gradient from 47% ACN to 55% ACN. The His-tag was removed following buffer exchange with a PD-10 desalting column (GE Health Care) into 100 mM NaCl, 50 mM Tris pH 8 by overnight TEV protease cleavage including 14 mM β-mercaptoethanol (BME). A second Ni-NTA purification step removed cleavage fragments and undigested protein. Final purification used a 500 ml HPLC gradient from 20% to 80% ACN on a C18 column. Chromatogram and MALDI-MS verification of mass can be seen in Figure 4.3.

**A1.3.3: Confirmation of covalent bilin binding**

Covalent attachment of the bilin was verified by fluorescence after denaturing SDS-PAGE using a UVP TMW-20 gel transilluminator on the UV setting prior to Coomassie-blue staining using ThermoFisher SimplyBlue™ SafeStain to visualize all proteins. Protein were dissolved in 3 parts
MilliQ water to 1 part NuPAGE LDS buffer, denatured for 10 min at 70° C, and applied onto a SDS-PAGE gel (Invitrogen Novex NuPAGE electrophoresis system, 4–12% bis-Tris gels, MES running buffer). More precise masses before and after bilin ligation were obtained using MALDI-MS. Lyophilized samples were solubilized in a 60%ACN/40%water pH 2 with saturated sinapinic acid matrix and applied to a MALDI-MS plate for analysis.

**A1.3.4: Bacteriochlorin attachment**

The synthesis of (BC1) has been described\(^\text{23}\). For pigment coupling, 0.1 \(\mu\)moles of lyophilized protein was incubated in 1 mL of 6 M guanidinium-HCl, 20 mM Tris pH 7.3, 14 mM BME for 1 hour at room temperature followed by PD-10 buffer exchange into phosphate buffered saline (PBS, 10 mM \(\text{Na}_2\text{HPO}_4\), 1.8 mM \(\text{KH}_2\text{PO}_4\), 2.7 mM KCl, and 137 mM NaCl, pH 7.4) and deoxygenated by purging with argon for 20 min. BC1 was added in 10-fold molar excess along with DMF to 33% for overnight reaction. The pH was adjusted to 2.0 with 10% trifluoroacetic acid with ACN added to 20% before HPLC purification. The PEB fusion protein used a 500 mL 40 to 80% ACN gradient on an C18 column, while both the PCB fusion and free maquette proteins used a 500 mL 30 to 70% ACN gradient.

**A1.3.5: Folding and concentration determination**

Refolding of the CpcA fusion proteins used successive dilution of urea\(^\text{24}\) followed by a buffer exchange using a PD-10 size exclusion column (GE
Healthcare). Non-fusion maquettes were solubilized without urea directly from lyophilized powder into PBS pH 7.4. Q band extinction coefficient (ε) values of Fusion-PEB and Fusion-PCB were calibrated by using established phycobilin extinction coefficients in denaturing acidic urea of 33.2 mM⁻¹ cm⁻¹ at 662 nm for PCB²⁵ and 53.7 mM⁻¹ cm⁻¹ at 550 nm for PEB²⁶. After folding the ε was 105 mM⁻¹ cm⁻¹ at 625 nm for PCB and 139 mM⁻¹ cm⁻¹ at 555 nm for PEB. Fusion proteins used for determining Q-band ε when folded were purified via HPLC to obtain 100% chromophorylated fusion.

A1.3.6: Chlorin binding

The synthesis of amphiphilic chlorin ZnC has been described²⁷. ZnChlide was prepared by Zn(OAc)₂ addition to pheophorbide a (Frontier Scientific) as described⁴. DMSO-solubilized ZnChlide stock concentrations were determined by dilution into methanol using an extinction coefficient of 64 mM⁻¹ cm⁻¹ at 656 nm²⁸. Binding affinities of ZnC and ZnChlide to the His site of slot B were measured by room temperature titrations, adding aliquots of 1 to 2 mM pigment stock solutions in DMSO to few μM protein solutions at molar equivalents from 0.1 to 4. Conspicuous visible region absorbance bandshifts for the bound and unbound pigments were analyzed by SVD to determine stoichiometry and dissociation constants, as described⁵. For fluorescence emission and excitation, chlorins were incubated at 2-fold molar excess for 30 minutes before PD-10 column removal of excess pigment.
A1.3.7: Spectroscopy

UV/Visible and fluorescence spectroscopy of proteins in PBS buffer were taken on a Varian Cary-50 spectrophotometer. Fluorescence emission and excitation spectra of solutions purged with argon were recorded on a Horiba Fluorolog 2 at 20° C. The quantum yield of bilin emission was determined by comparing integrated emission spectra with the 0.54 yield of cresyl violet perchlorate in methanol\textsuperscript{17}. The quantum yield of fluorescence of ZnChlide was determined by comparing the integrated emission spectra with the 0.016 yield of ZnChlide in 10\% water to ethanol (V/V)\textsuperscript{28}. The isotropic Förster radius estimates for EET between pigment pairs were calculated with a fluorescence resonance energy transfer (FRET) formalism\textsuperscript{16} using pigment extinction coefficients determined in this work; the isotropic Förster radius used a kappa squared value of 0.67. Calculations were performed in PhotoChemCAD\textsuperscript{29}.

A1.4: References


Appendix 2: Supporting data for chapters

A2.1: Chapter 2

Figure A2.1.1: Verification of mass of Maquette H6H111 sequence 2.1 used in Chapter 2 via MALDI-TOF MS.

Figure A2.1.2: Creation of ZnChlide for all of thesis: Before and after Zn insertion.
Figure A2.1.3: Absorbance spectra of porphyrin (9) in 2-His (red) and His-free (black) maquette. Increase in extinction coefficient upon His coordination.

A2.2: Chapter 3

Figure A2.2.1: In vivo non-covalent BV affinity for Maquette 3.28. *E. coli* cultures expressing maquette were doped with 500 μM Levulinic Acid (I) and the growth media (II) and purified protein (III) came out green (with BV bound) through the NiNta purification. Absorbance spectrum of purified protein shown in solid green trace. Dashed trace shows free BV in PBS pH 7.
Table A2.2.1: Maquette amino acid sequences used in chapter 3. Sequences can be read from left (N-terminus) to right (C-terminus). Mutations are highlighted in red.
Table A2.2.2: Parent DNA sequences for maquette mutagenesis. A-F Maquettes are the parent sequences used to obtain the maquettes shown in tables in chapter 3. Maquette helical sequence and residue number begins at EIWK following the amino poly-histidine tag.
<table>
<thead>
<tr>
<th>Parent Plasmid</th>
<th>Moquette ID</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>plasmid purchased</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>GAAACGAAAGAAGGTGGCTGGCTGGTCATGATGGAACGCAGCAGAGGA</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>CTCTCGTGTGTTTCCAGATACGAGCAGCGCAGGACACTTTTTTCTTC</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>GGTTCGGCTGAGCTGAGTCTGAGAGGTAGGATGCTGAGG</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>CGCGGACATGCTGGGATCAGGAGCGGACAGGACGAC</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>TCCAGGAGGAAATGGTACTGTCAGGACGAGCATAGG</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>TTTCCAGAAGAAGGGTAGGTAAGGTGAGGAAAGGAG</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>AAAGAGCTCAAGTAGGCGGCAGGCAATGTGCTGTTGAT</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>GAGTGACGACGACGAGAGGATGCAGGGATGGGGT</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>TCCAGGAGGAAATGGTACTGTCAGGACGAGCATAGG</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>TGGTGGTGGTGAGTCTGTCAGGAGGATGCTGAGG</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>CGAGATCGCTGAAATGGTACTGTCAGGACGAGCATAGG</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>TGGTGGTGGTGAGTCTGTCAGGAGGATGCTGAGG</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>GCTGTGGGATGCGGTGAGGAAAGGCTGAGGAGG</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>GCGCGGACATGCTGGGATCAGGAGCGGACAGGACGAC</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>TCCAGGAGGAAATGGTACTGTCAGGACGAGCATAGG</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>GCTGTGGGATGCGGTGAGGAAAGGCTGAGGAGG</td>
</tr>
</tbody>
</table>

**Note:** The table above represents the primer sequences for different Moquette IDs associated with specific plasmids. Each Moquette ID is linked to a specific primer sequence, which is crucial for genetic engineering and molecular biology applications. The sequences are designed to amplify specific regions of DNA, facilitating various laboratory processes such as PCR (Polymerase Chain Reaction).
Table A2.2.3: Primers used for maquettes mutagenesis. Column 1: parent plasmids associated with maquettes with ID shown in column 2. Column 3: forward and reverse primers sequences for minor modification of maquette parent plasmid sequence. Intermediate maquette sequences not described in the main text of chapter 3 are assigned a non-numerical ID based on amino acid changes. Maquette sequence 2 and 6 used mixed base primers “S” or “W” to generate a wider variety of sequences, then verified at the University of Pennsylvania Sequencing Core.
<table>
<thead>
<tr>
<th>Table A2.2.4: Modified CpcA sequence for PEB extraction.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>His\textsubscript{6}-tagged CpcA sequence</strong></td>
</tr>
<tr>
<td><strong>Modified His\textsubscript{6}-tagged CpcA sequence for PEB extraction</strong></td>
</tr>
</tbody>
</table>
Figure A2.2.2: Lyase significantly increases bilin attachment yield as well as blue shifts the Qvis spectra for maquette with bilin bound. Maquettes expressed with the PcyA and Heme Oxygenase but not the CpcS lyase show very little attachment of bilin seen by comparing the dotted traces with the solid traces.
Figure A2.2.3: Analytical HPLC of Maquettes in Figure 3.7. HPLC separates maquettes with no bilin bound (Trp absorbance only at 280 nm) from bilin bound maquettes (absorbance at both 280 and 640 nm).
Bilimaquettes purified by size-exclusion column chromatography in the presence of 8M urea adjusted to pH 2 using HCl. Maquette concentrations were calculated from 280 nm tryptophan absorbance for sequences 3.1, 3.2, 3.3, and 3.4 are 8.5, 13, 8.9, and 6.8 µM respectively.
Figure A2.2.5: Acidic urea PCB spectra for Figure 3.7 and Table 3.3. Maquettes from figure 5 and table 3 were diluted 10-fold into 8 M urea adjusted to pH 2 using HCl. Maquette concentrations calculated from 280-nm tryptophan absorbance for sequences 3.18, 3.19, 3.20, 3.21, 3.22 and 3.23 were 22, 35, 31, 36 and 20 µM, respectively.
Absorbance spectra of PCB bound to maquettes diluted 10-fold in 8 M urea (adjusted to pH 2 using HCl) matches PCB bound to other natural biliproteins when denatured in acidic urea. Some BV may be present in sequence 25 as seen in the red shifting of Qvis band\(^1\). Maquette concentrations calculated from 280-nm tryptophan absorbance for sequences 3.10, 3.24, 3.25, and 3.26 were 58, 56, 19 and 41 µM, respectively.
Figure A2.2.7: HPLC purification and MALDI-MS of TPB from PCB bilimaquettes. Maquette sequence 3.27 with PCB bound was subjected to treatment in 6 M guanidinium pH 8.5 and then purified via HPLC. (A) and (B) 1st and 2nd purification of TPB-maquette on a RP-318 BioRad Reverse Phase column. 300ul then 50ul injection volumes were used respectively. Samples were solubilized in 50% ACN and 50% H2O with 0.1% TFA before injection. Chromatograms are shown on the left and photodiode array spectra for TPB-bilimaquette peaks are shown on the right. (C) MALDI-MS spectra for TPB-bilimaquette peak is shown in pink and the PCB-maquette peak is shown in blue. Principle mass peak for TPB-bilimaquette is near the expected mass for loss of 1 pyrrole ring.
Figure A2.2.8: Dual pigment maquette absolute and difference spectra (A). Absorbance of PCB bilimaquette 3.27 without ZnChlide (red, solid line) and with added ZnChlide (green). Subtracting the absorbance of bound ZnChlide in the bilin-free maquette (black) yields the dashed red line, indicating the ZnChlide has little effect on PCB absorbance. However, PCB tends to red-shift the ZnChlide spectrum indicating weakening of ~70 nM ZnChlide binding in the presence of bilin. Red-shifting of ZnChlide is also evident when the PCB spectrum is subtracted (gray dashed line). Buffer conditions were 1M MgSO₄ and 10 mM NaPO₄ at pH 7.4. (B). Absorbance of TPB bilimaquette 3.27 without ZnChlide (red, solid line) and with added ZnChlide (green). Subtracting the absorbance of bound ZnChlide in the bilin-free maquette (black, from Figure 4.4-A) yields the dashed red line, indicating the ZnChlide has little effect on TPB absorbance; a slight red-shift of the ZnChlide spectrum indicates a slight weakening of ZnChlide binding affinity. A slight red shift of ZnChlide is also evident when the PCB spectrum is subtracted (gray dashed line). Buffer conditions were 1M NaCl, PBS pH 7.4.
Figure A2.2.9: SDS-PAGE of all maquettes found in chapter 3. Visible illumination of SDS gels after staining are shown in the left panels. UV illumination of the SDS gel prior to staining are shown on the right. Mass in kDa for protein standards are shown in white. White text describes purification method.
A2.2.10: HPLC purification of PCB after methanolysis. After cleaving the PCB via overnight methanol reflux and drying, PCB was solubilized with a solution containing 0.1% TFA; 20% ACN, and 80% water and then HPLC purified. PCB was purified by HPLC using a 500 mL gradient of 20% to 70% ACN with 0.1% TFA on a Vydac C-18 column. The 18-min peak was used for in vitro coupling. Spectra of peaks are shown at right. The side peak at 16 min is likely to be a PCB isomer. The peaks near 27 min appear to be isomers of meso-biliverdin (MBV).
A2.2.11: MALDI-MS for purification of PCB. Free protonated PCB has a mass of 587.

A2.2.12: MALDI-MS for purification of PEB. The 587 peak shows free protonated PEB.
A2.2.13: HPLC purification of PEB after chloroform extraction. PEB purification using a 500 mL gradient of 20% ACN to 70% ACN containing 0.1% TFA on a Vydac C-18 column. PEB was extracted with chloroform from CpcA mutant without ligating Cys (Table A2.2.4) and purified by HPLC.
A2.2.14: MALDI-MS verification of PCB in vitro attachment to Maquette. (A) Cys locations of maquette sequences 3.1 (blue), 3.2 (green), 3.3 (yellow) and 3.4 (red). (B) MALDI-MS mass increases on PCB binding to each sequence.
A2.2.15: MALDI-MS verification of PEB in vitro attachment to maquette. (A) Cys locations of maquettes 3.1 (blue), 3.2 (green), 3.3 (yellow) and 3.4 (red). (B) MALDI-MS mass increases on PEB binding to each sequence.
A2.3: Chapter 4

Fusion sequence with His at position 6: Sequence 4.1 and 4.2

\[ \text{4.1 and 4.2} \]

\[ \text{Sequence} \]

\[ \text{G}E\text{I} W\text{K} Q \text{H} \text{E} \text{D} \text{A} \text{L} Q \text{K} \text{F} \text{E} \text{E} \text{A} \text{L} N \text{Q} \text{F} \text{E} \text{D} \text{L} K \text{Q} \text{L} \text{G} \text{C} \text{G} \text{E} \text{I} \text{K} \text{Q} \text{R} \text{A} \text{E} \]  

\[ \text{D} \text{A} \text{L} R \text{K} Q \text{E} \text{E} \text{A} \text{L} K \text{R} \text{F} \text{E} \text{D} \text{L} K \text{Q} K \text{G} \text{G} \text{S} \text{G} \text{E} \text{I} \text{W} \text{K} Q \text{A} \text{E} \text{D} \text{A} \text{L} Q \text{K} \text{F} \text{E} \text{E} \]  

\[ ALNQFEDLKQLGGSEIKQRAGEDALKRFEDLKQKMKTPLTEAVSTADSGRFLSSTELQIAFGRLRQANAGLQAIAKALTQDNLAVNGAAQAVYNKFPYTQTNQGNNFAADQRGKDCKARDIGYLYLRIVTYCLVAGTGGPGLDEYLIAGIDEINRTFDSLPSWYWVEALKYIKANHGLSGDARDEANSYLDYAINALS \]

His at position 66: Sequence 4.3 and 4.4

\[ \text{4.3 and 4.4} \]

\[ \text{Sequence} \]

\[ \text{G}E\text{I} W\text{K} Q \text{A} \text{E} \text{D} \text{A} \text{L} Q \text{K} \text{F} \text{E} \text{E} \text{A} \text{L} N \text{Q} \text{F} \text{E} \text{D} \text{L} K \text{Q} \text{L} \text{G} \text{C} \text{G} \text{E} \text{I} \text{K} \text{Q} \text{R} \text{A} \text{E} \]  

\[ \text{D} \text{A} \text{L} R \text{K} Q \text{E} \text{E} \text{A} \text{L} K \text{R} \text{F} \text{E} \text{D} \text{L} K \text{Q} K \text{G} \text{G} \text{S} \text{G} \text{E} \text{I} \text{W} \text{K} Q \text{H} \text{E} \text{D} \text{A} \text{L} Q \text{K} \text{F} \text{E} \text{E} \]  

\[ ALNQFEDLKQLGGSEIKQRAGEDALKRFEDLKQKMKTPLTEAVSTADSGRFLSSTELQIAFGRLRQANAGLQAIAKALTQDNLAVNGAAQAVYNKFPYTQTNQGNNFAADQRGKDCKARDIGYLYLRIVTYCLVAGTGGPGLDEYLIAGIDEINRTFDSLPSWYWVEALKYIKANHGLSGDARDEANSYLDYAINALS \]

Figure A2.3.1: Fusion protein sequence used for attaching PEB or PCB. Histidines used for chlorin ligation and cysteines used for bilin attachment are shown in bold red letters. Numbering for His position starts at the beginning of the first helix after initial Glycine.
Figure A2.3.2: HPLC estimates of PEB attachment yields to fusion protein 4.2. A 500-ml gradient of 40% to 47% acetonitrile (ACN), pH 2, was used with a C4 prep column to purify HisTagged-fusion with PEB bound (2) from unbound (1) after 20 min reduction in 2-mercaptoethanol. Despite high expression yields, upwards of 50% attachment of bilin to fusion protein can be achieved, as determined by HPLC. Chromatograms are shown in A and photodiode array detector (PDA) spectra for each peak in the chromatogram are shown in B.
A2.3.3: HPLC and MALDI-MS verification of BC1 attachment to unfused maquette protein. (A) HPLC elution profile monitored at 280 nm and 712 nm for maquette H66 (Sequence 4.3) with maleimide-anchored BC1. (B). HPLC elution profile monitored by absorbance at 280 nm and 712 nm for maquette H6 (Sequence 4.1) with maleimide-anchored BC1. Absorbance spectra for maxima of peaks shown in (A) and (B) are shown in the panel labeled “photodiode array (PDA) spectrum.”
A2.3.4: HPLC and MALDI-MS verification of BC1 anchoring to fusion protein. (A) HPLC elution profile monitored by absorbance at 560 nm and 712 nm for Fusion-PEB (Sequence 4.2) with maleimide-anchored BC1. (B) HPLC elution profile monitored by absorbance at 637 nm and 712 nm for Fusion-PCB (Sequence 4.4) with BC1 attached. Photodiode array (PDA) spectrum: absorbance spectra for maximum of peaks shown in panels (A) and (B).
A2.3.5: Absorption spectra of unfused maquette with bound BC1 are similar for His at either position 6 or 66 (Sequence 4.1 and 4.3).
Figure A2.3.6: Plasmid construction Part A shows the plasmids for each component and the primers that were used to amplify the gene. Gibson Assembly was used to attach the cpcA gene and maquette vector. Part B shows the DNA sequence for each gene. Part C shows the translation of each DNA sequence.
Table A2.3.1: Primers used to make mutants in chapter 4.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Purpose</th>
<th>Sequence 5' to 3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linearize Maquette plasmid for CpcA gene fusion FWD</td>
<td>CCCCTAGCATAACCCCT</td>
</tr>
<tr>
<td>2</td>
<td>Linearize Maquette plasmid for CpcA gene fusion RVS</td>
<td>TTCTGCTCGAGTCCG</td>
</tr>
<tr>
<td>3</td>
<td>Overlap C terminus of Maquette and N terminus of CpcA FWD</td>
<td>AGGACTTGAAGCAGAAAATGAAACCCCTTTAAGAAG</td>
</tr>
<tr>
<td>4</td>
<td>Overlap C terminus of CpcA with T7 terminatorSeq</td>
<td>GGTATGCTAGGGGCTAGTACAGCATATTGAG</td>
</tr>
<tr>
<td>5</td>
<td>changing non ligating Cystiene to Alanine in CpcA FWD</td>
<td>CGCATCGTTACCTACGCCGTATTTGCTGGT</td>
</tr>
<tr>
<td>6</td>
<td>changing non ligating Cystiene to Alanine in CpcA RVS</td>
<td>ACCACCAGCAAACGTAGGTAAGCATGAG</td>
</tr>
<tr>
<td>7</td>
<td>deloMDLfromPEBsFWD</td>
<td>TGCTCTGTTTTGATATGGTTAGT</td>
</tr>
<tr>
<td>8</td>
<td>deloMDLfromPEBsRVS</td>
<td>ATCTAGAACTTCATACCAAACAAAGGAG</td>
</tr>
<tr>
<td>9</td>
<td>Maquette H6A FWD</td>
<td>GAGATCGAGAAAGCTGAGGACGCAGGAG</td>
</tr>
<tr>
<td>10</td>
<td>Maquette H6A RVS</td>
<td>CTGCAGAGCTCCTGAGCTCTGAGAAG</td>
</tr>
<tr>
<td>11</td>
<td>Maquette H36A FWD</td>
<td>GAGATTAAAGCAGGAGGCAAGATGGGAG</td>
</tr>
<tr>
<td>12</td>
<td>Maquette H36A RVS</td>
<td>CGCAGCGCATCTCGGCACCACCTGCTTTAACATC</td>
</tr>
<tr>
<td>13</td>
<td>Maquette H66A FWD</td>
<td>GAGATTGGAACAGCTGAGCAGATGCAAG</td>
</tr>
<tr>
<td>14</td>
<td>Maquette H66A RVS</td>
<td>TGCACTGACATTCTCACTCACTGAGGAG</td>
</tr>
<tr>
<td>15</td>
<td>Maquette H96A FWD</td>
<td>GAGATCAAGAGCGAGCGAGAGATGAGTCTCAGT</td>
</tr>
<tr>
<td>16</td>
<td>Maquette H96A RVS</td>
<td>ACGCAGAGCATCTTCTGCGCGTCTTGAATC</td>
</tr>
</tbody>
</table>


Figure A2.3.7: Bilin biosynthetic machinery plasmids. Panel A shows machinery for PEB production and attachment. Primers 7 and 8 were used to make PebS non-functional so plasmid in Panel B could be transformed alongside that shown in Panel A to allow for PCB attachment to the fusion protein. The pCOLAduet and pACYCH plasmids each contain T7 promoters. The PCOLAduet plasmid under the selection by kanamycin is used to express CpcE/CpcF, PebS, and Heme Oxygenase in E. coli. These are the proteins responsible for the production and attachment of PEB to the Maquette/CpcA gene fusion. The pACYCH plasmid is maintained under selection by chloramphenicol and produces PcyA for synthesis of PCB from biliverdin in E. coli.
**A2.4: Small molecules for all chapters**

Figure A2.4.1: Chemical structures with numerical IDs and full chemical names for circular tetrapyrroles and abbreviations for linear tetrapyrroles used in thesis.
A2.5: References

1. Fushimi, K. et al. Photoconversion and fluorescence properties of a red/green-type cyanobacteriochrome AM1_C0023g2 that binds not only phycocyanobilin but also biliverdin. *Front Microbiol* 7, 588 (2016).


