

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Synthesis and Development of Viral Capsid Templated Light Harvesting Systems

Permalink

<https://escholarship.org/uc/item/5924t1kf>

Author

Finley, Daniel

Publication Date

2014

Peer reviewed|Thesis/dissertation

Synthesis and Development of Viral Capsid Templated Light Harvesting Systems

By

Daniel Thomas Finley

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Chemistry
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Matthew B. Francis, Chair
Professor Naomi S. Ginsberg
Professor Robert M. Glaeser

Spring 2014

Synthesis and Development of Viral Capsid Templated
Light Harvesting Systems

Copyright © 2014

By: Daniel Thomas Finley

Abstract

Synthesis and Development of Viral Capsid Templated Light Harvesting Systems

by

Daniel Thomas Finley

Doctor of Philosophy in Chemistry
University of California, Berkeley
Professor Matthew B. Francis, Chair

The self-assembling protein shells of viruses have provided convenient scaffolds for the construction of many new materials with well-defined nanoscale architectures. Of these numerous examples, one such application is the use of viral capsids to template synthetic light harvesting structures. Through the combination of modern molecular and chemical biology techniques, the symmetry and self-assembling nature of the viral proteins can be exploited to form regular arrays of chromophores on the protein surface in a site-specific manner. Tobacco Mosaic Virus coat protein (TMVcp) was used to template organic chromophores, forming light harvesting structures reminiscent of biological systems. While many of the processes in photosynthesis are well understood, there is still much mystery surrounding the physical basis for the efficiency seen in natural systems. We set about constructing protein-templated biomimetic light harvesting systems with tunable parameters such as protein attachment chemistry, inter-chromophore distances, and inter-chromophore geometries. Through genetic manipulation of the TMVcp, protein mutants were produced with various sites of reactivity on the monomeric protein surface, manifesting as regularly spaced arrays within the assembled capsid. The spatial and geometric relationships of the reactive sites were investigated, along with the resulting protein assembly state. Next, chromophores were introduced at various locations and investigated structurally and spectroscopically. The spectral signatures observed with a range of chromophores, geometries, and pigment-protein attachments were compiled for numerous TMVcp-templated systems, and the effects of protein-templating discussed. The overall goal of this work is to gain insight into the origin of the spectral and energetic effects behind the efficiencies observed in natural systems through the use of highly tunable synthetic biomimetic systems. For a truly biomimetic system to be produced, however, there must be a charge-separation event involved in the light harvesting process. We then turned our efforts to the synthesis of water-soluble phthalocyanines capable of protein attachment. Phthalocyanines are a near infrared absorbing class of chromophores, important for their sensitization properties, and of particular interest to us for their use as singlet oxygen generation materials in photodynamic cancer therapy. The synthesis of both symmetric and asymmetric phthalocyanines was accomplished, followed by characterization. Finally, an asymmetric phthalocyanine was integrated into the interior of the bacteriophage MS2 viral capsid, and its ability to generate singlet oxygen was investigated.

Dedicated to my family with love.

To my parents, whom I owe all of my past accomplishments.
To Carly, Jordan, and Ayla, my inspiration for future success.

Table of Contents

Chapter 1: Viral capsids as self-assembling templates for new materials	1
1.1 Abstract.....	1
1.2 Introduction.....	2
1.3 Covalent modification of viral capsids for applications in medical imaging and drug delivery.....	5
1.3.1 Introduction.....	5
1.3.2 Cowpea chlorotic mottle virus (CCMV).....	5
1.3.3 Cowpea mosaic virus (CPMV) and bacteriophage Q β	9
1.3.4 Canine parvovirus (CPV).....	11
1.3.5 Bacteriophage MS2.....	11
1.3.6 Summary.....	14
1.4 Viral capsids as templates for the construction of inorganic materials.....	14
1.4.1 Introduction.....	14
1.4.2 Cowpea chlorotic mottle virus (CCMV).....	15
1.4.3 Chilo iridescent virus (CIV).....	15
1.4.4 Cowpea mosaic virus (CPMV).....	16
1.4.5 Brome mosaic virus (BMV).....	16
1.4.6 Red clover necrotic mosaic virus (RCNMV).....	17
1.4.7 Bacteriophage M13.....	18
1.4.8 Tobacco mosaic virus (TMV).....	19
1.4.9 Summary.....	20
1.5 Capsid-based materials for optical and catalytic applications.....	20
1.5.1 Introduction.....	20
1.5.2 Cowpea mosaic virus (CPMV) and hepatitis B virus (HBV).....	21
1.5.3 The tobacco mosaic virus (TMV).....	22
1.5.4 Photocatalytic systems based on bacteriophage MS2.....	24
1.5.5 Bacteriophage M13.....	25
1.5.6 Summary.....	26
1.6 Conclusion and dissertation outline.....	26
1.7 Literature Cited.....	29
Chapter 2: The Tobacco Mosaic Virus as a versatile nanoscale scaffold	35
2.1 Abstract.....	35
2.2 The structure and assembly of TMV.....	36
2.2.1 History of TMV research.....	36
2.2.2 The 20S species.....	36
2.2.3 Assembly of the TMV coat protein.....	38
2.2.4 A detailed look at the TMVcp structure.....	39
2.3 TMV for conjugation of organic chromophore arrays.....	40
2.3.1 Background and introduction.....	40
2.3.2 Building a versatile new TMV assembly for bio-orthogonal functionalization...	42

2.3.3	The circular permutant of TMV.....	44
2.4	Testing the solution assembly of RR-TMVcp.....	45
2.5	Conclusion.....	48
2.6	Materials and methods.....	49
2.7	Literature cited.....	52
Chapter 3: Building an artificial photosynthetic analog.....		54
3.1	Abstract.....	54
3.2	Natural light harvesting systems – structure and design principles.....	55
3.3	TMV templated light harvesting systems – structure and design principles.....	58
3.4	Construction and spectral analysis of RR-TMVcp templated LH systems.....	60
3.4.1	Choice of chromophore.....	60
3.4.2	Modification at the C123 position.....	61
3.4.3	Modification at the K104 position.....	61
3.4.4	Modification with pyrene maleimide.....	62
3.4.5	Modification with coumarin 343.....	63
3.4.6	Modification of RR-TMVcp with xanthene dyes – OG488 and AF594.....	65
3.4.7	Fluorescence spectra of the templated arrays.....	66
3.4.8	Fluorescence quenching within the RR-TMVcp templated arrays.....	67
3.4.9	Fluorescence lifetime analysis.....	68
3.4.10	Fluorescence anisotropy analysis of RR-TMVcp systems – linker effects.....	69
3.5	Dual color systems.....	71
3.5.1	Construction and design of donor:acceptor systems.....	71
3.5.2	Absorption spectra of donor:acceptor systems.....	72
3.5.3	Fluorescence analysis of donor:acceptor systems.....	73
3.6	Discussion and interpretation of spectral data.....	74
3.6.1	Physical nature of RR-TMVcp templated chromophore arrays.....	75
3.6.2	Comparison of chromophores.....	77
3.6.3	Donor:acceptor systems.....	78
3.6.4	Conclusion and perspectives.....	78
3.7	Materials and methods.....	80
3.8	Literature cited.....	85
Chapter 4: Synthesis of a water-soluble phthalocyanine for bioconjugation.....		87
4.1.	Abstract.....	87
4.2	Introduction.....	88
4.3	Synthesis of phthalocyanines.....	90
4.3.1	Synthesis of tetra-carboxylic acid functionalized Pcs.....	90
4.3.2	Synthesis of asymmetric sulfonated Pcs.....	92
4.3.3	Installation of maleimide ring onto trisulfonated Pc.....	93
4.3.4	Characterization of trisulfonated maleimide Pc.....	94
4.4	Generation of singlet oxygen.....	94
4.5	Conclusion and perspectives.....	96
4.6	Materials and methods.....	97
4.7	Literature cited.....	101

Acknowledgements

There is a long list of people that I would like to thank that have in some way, either personally, professionally, or both, made my time at Berkeley truly wonderful and unforgettable. None of this work could have been accomplished without the guidance, support, and friendship of an amazing group of people.

First, I need to thank Professor Matt Francis. Your unwavering support and positivity as an advisor has made my tutelage enormously rewarding. From both a personal and professional standpoint, you have helped me to not just get through graduate school, but to enjoy my short time here and make the most of it. I cannot over emphasize just how much I have learned from you regarding science in general. Since joining your lab, I have rekindled my love of organic chemistry, as well as being introduced to countless new tools and techniques on the frontier of chemical biology. I always leave our research discussions with new perspectives, ready to improve on experiments, as well as thirsty for new knowledge. Whenever things weren't going so well in the lab (a not uncommon occurrence, in my experience), your contagious optimism consistently offered new solutions and outlooks to problems.

Next, I need to thank Professor Steven Emory, my research advisor throughout my undergraduate and M.S. work. He introduced me to research in a laboratory setting, and was extremely supportive throughout my undergraduate and also graduate work. He was instrumental in letting me experience science firsthand and finding a true passion of mine. It was during the time that I was in his lab that I ultimately decided to pursue science as a lifelong career, and am grateful that I was given this rare opportunity.

I have had the fortune of collaborating with an incredibly talented group of researchers while at Berkeley, from which many enlightening discussions, experiments, and projects have emerged. Naomi Ginsberg has been extremely helpful with discussions on spectroscopy, and a fruitful collaboration has developed between our labs. Her postdoctoral researcher, Rodrigo Noriega, has been vital in taking the spectroscopic analysis of my systems to the next level. Also at Berkeley, I have spent much time with the Whaley group conversing over quantum mechanics and the large scale implications of the emerging field of quantum biology. In particular, Mohan Sarovar, an amazingly brilliant and talented theorist, has provided copious amounts of knowledge and wisdom for my understanding of quantum processes. Donghyun Lee worked closely with me on modeling our synthetic light harvesting systems, and has been instrumental in elucidating some interesting research questions. Aside from agreeing to participate on both my qualifying exam committee, as well as my dissertation committee, I must thank Professor Robert Glaeser for his helpful guidance in an early electron microscopy collaboration with his student, Ross Walton. Other people that must be mentioned who have contributed insight and discussions about research are Loren Greenman, Aleksey Kocherzhenko, Samia Hamed, Florian Altvater, and Sahar Sharifzadeh.

As far as colleagues and lab mates here at Berkeley go, there is a huge number of people that I have worked with who have in some way contributed to my evolution as a scientist. Within the Francis group, I have overlapped with many classes of graduate students, all of which have left lasting impressions on me. Keeping with tradition in the Francis lab, each fellow lab mate that I have had the pleasure of working alongside in the lab will be acknowledged here.

When I arrived in the Francis lab, I was taken under the wing of recently-graduated Patrick Holder. Pat taught me a vast amount in just a few short months, and was amazingly

helpful with assisting me in assimilating both into the town of Berkeley, as well as into the chemistry department. When he left our lab, Pat also bequeathed to me the phthalocyanine project (for better or worse), eventually becoming the bane of many of my evenings. Thank you for that Pat.

The next class of graduate students was solely composed of Nick Stephanopoulos. However, this is only fitting. With Nick, we really didn't need anymore. Nick was my partner in crime on and off the clock, being my primary source of scientific knowledge, along with providing the majority of (inappropriate) jokes and keeping lab time productive, enjoyable, and hilarious all at the same time. Unfortunately, all things must pass, and Nick left us for Chicago, where he's enjoying all the success he deserves. Respect.

Next, we had the class of Zac, Wesley, Sonny, Kanna, Gary, and Michel. Zac was a great guy to work with, and knew more than I would ever want to know about molecular biology. Sonny and Wesley, you were always around, but were working so hard I unfortunately never got to know you very well. Kanna, I want to thank you for still coming to our group meetings when your other group let you, and I always enjoyed our times hanging out within or outside of lab. Ski trips or holiday parties were always more enjoyable when you were there too. Gary - you've been a great friend over the last few years, and I always have a blast when we get together. Always with a joke, while at the same time being a tremendously competent scientist when you feel like it. I'm glad that I still get to see you regularly around Berkeley. Michel - I owe all of my progress with the Tobacco Mosaic Virus to your molecular biology expertise, and none of this work could have been accomplished without the groundwork you laid. Additionally, you always had a bike pump on hand.

The next class in our group was Leah and Chris. I had the pleasure of spending lots of time with Leah, who always seemed to improve the mood of anyone around. She never ceased to impress me with her kind and giving attitude, her humor, her scientific knowledge and drive, her uncanny innate snow skiing aptitude, as well as the ability to slap me senseless whether I deserved it or not. Chris, I'm glad that we got to overlap so long and spend so much time together. Thank you for opening up your parent's ski chalet to us (especially Farkas), in addition to being an awesome infielder and lead-off batter for both of our softball teams. I'm happy that your postdoc and job hunting has gone expectantly well for you, and wish you and Nikki all the best in the bay area with your new jobs.

My fellow classmates come chronologically next in this saga of thanks. I've known Anna for a long time, and coming from WWU together, it was great having a friendly face in Berkeley, as well as in my lab. Congratulations on all your success, and I look forward to seeing Tyler grow up to be just like his Mom. Mike, you were a great addition to the lab, and were truly "one of us" Francis group folk, while still maintaining your allegiances to the Groves lab. Kristin, I was always impressed with your ability to get things done in the lab so quickly, and though we never got to know each other on a personal level very well, I always had a great time when we interacted outside of lab. Allie, you're next. You were ok at science, and kind of nice. Plus we got to exchange pictures of our newborn nieces, so that was good. I've shared more lunches with Troy than I ever have with anyone, and don't know if I could have gotten through graduate school without his constant wit and humor. You knitted me an awesome hat, you have infinite useless grammatical knowledge, and you were always there to discuss things related to science, graduate school, and life in general. Finally, there isn't enough to say about the personality that is trapped inside the physique of Amy Twite. She's been my lab mate,

roommate, and drinking buddy in various capacities during my time in Berkeley, and there is nothing I can express to do justice to the thanks I owe her for helping me survive graduate school.

The class that followed ours was comprised of an equally competent and awesome group of people. Jeff often said more with less words, but always had something insightful to include in a conversation. Jelly taught me about the black box that is cell biology, and was patient enough to accompany me in bona-fide biology experiments. It only took me four years, but I finally got to eat some of Jelly's pad kee mao. It was truly an authentic "drunken noodle" experience. Kareem was always coming up with new ideas, and it was great to see some of his projects come to fruition. Katherine, or "KMack", aka "Blue ribbon baker" was our unofficial lab purveyor of confectionaries. She was a polymer-protein extraordinaire, and we always managed to find a laugh – whether it be in lab, at a dinner party, or at some sort of extra-curricular gathering. Maybe most vital to both our graduate school careers, Katherine was always ready to tip her glass to mine when the lab work was less than motivational. Finally, the last person to mention in this class is Stacy. Stacy worked about 15 feet away from my desk, and not enough can be said about the infinite patience, grace, and understanding that she must have in order to endure being my sounding board for 5 years. Her tireless work ethic, along with her ability to push things to completion, are just two of the many things that make her a great researcher, in addition to her stellar personality. She had some very large shoes to fill following Nick as a lab mate, but I would say that she outdid herself.

Abby was our sole graduate student the following year, and is always full of positivity and smiles. She had thankfully taken over the undesirable group job of managing the computers, and was kind enough to help me with some analytical instrumentation that I was unfamiliar with.

The next year brought another large group of students. Richard, who has helped me with some coding and computer questions, also brews a mean ginger beer. Jake, my fellow Washingtonian, sat next to me for his time here, and was a great lab mate to have. I can hopefully transfer the light harvesting knowledge on, and I am certain Jake will be more than successful in his scientific endeavors. Jim has been amazingly productive in his early time here, and aside from all of the interesting science you do, it's nice to have someone to talk football with in a chemistry lab. Ioana, you work way too hard. It might be the Romanian in you, or the Cal Tech, but you are a fearless scientist who will accomplish anything you feel like. Just take a break every now and then. Kanwal, although we never really collaborated scientifically much, it has been great getting to know you in the lab. Whenever I need a pick-me-up, all I need to do is stop by your desk.

Jenna was a rotation student with me, and I'm glad that she made her way to the Francis group. She's worked incredibly hard since she's been here, and is undoubtedly a great addition to the lab. As far as the youngest students in our group, I never really got to know any of you very well, but you all seem to be starting out amazingly productively. Jess, Joel, Matt, and Matt, I wish you all the best of luck in the Francis group.

We have had a few wonderful postdocs during my time at Berkeley. Praveena, maybe the sweetest and most well-meaning person I've ever met, is both a talented chemist and a wonderful lab mate, always ready to lend a positive perspective to a conversation. Farkas, you epitomize the "work hard, play hard" ethos. Whether it was writing an extended grant, jetting back and forth across the bay with time sensitive radionucleii, or gearing up to go snowboarding, you were always in it 100%. Adel and Meera came during my last year in the Francis group, and

have definitely provided a huge sum of knowledge to the lab. Whenever I have a random question regarding anything remotely related to synthetic organic chemistry, one of these two will have the answer.

I also was fortunate to work with an undergraduate, Gary Suseno. Gary worked under me for a year or so, before going on to bigger things at UIUC. Although Gary didn't verbalize much, it was clear that his passion for science didn't need many words.

Finally, in addition to all of my colleagues I've had the pleasure of working with, I must thank a few people on a personal note. My friends and family have been incredibly supportive of me over the last few years, and I owe all my gratitude to them for putting up with my lack of correspondence and distance while I've been writing this thesis. Mom, Dad, Carly, Jordan, and Ayla, you've kept me going through this long process, and without your constant support, I never would have been able to finish in one piece.

Lastly, along with the personal interactions that have helped guide my graduate schooling, I must thank Bicycle Coffee Co. for helping me keep both consciousness and focus, and the Mead Kitchen for helping me retain a semblance of sanity throughout this process.

Chapter 1

Viral capsids as self-assembling templates for new materials

1.1 Abstract

The self-assembling protein shells of viruses have provided convenient scaffolds for the construction of many new materials with well-defined nanoscale architectures. In some cases, the native amino acid functional groups have served as nucleation sites for the deposition of metals and semiconductors, leading to organic/inorganic composites with interesting electronic, magnetic, optical, and catalytic properties. Other approaches have involved the covalent modification of the protein monomers, typically with the goal of generating targeting delivery vehicles for drug and imaging cargo. Covalently modified capsid proteins have also been used to generate periodic arrays of chromophores for use in light harvesting and photocatalytic applications. All of these research areas have taken advantage of the low polydispersity, high chemical stability, and intrinsically multivalent properties that are uniquely offered by these biological building blocks.

Portions of this work have appeared in published form:
Prog. Mol. Biol. Trans. Sci. **2011**, *103*, 353-392.

1.2 Introduction

Many of the most exciting ideas and opportunities in materials science require the construction of complex structures with nanoscale features. The realization of these designs presents a substantial synthetic challenge, however, as this size scale lies between those which are traditionally addressed using organic synthesis and lithographic patterning techniques. Although much progress is being made on both sides of this "spatial divide", a growing number of laboratories have turned to self-assembly as an effective means to generate integrated multicomponent materials with nanoscale architectures.¹ This approach can generate both discrete and extended structures with high efficiency as individual components come together to satisfy intermolecular recognition forces. Despite the many recent advances in the *de novo* creation of self-assembling materials, the most complex and functional examples of this concept are still found in biological systems. Through the assembly of proteins, nucleic acids, lipids, and carbohydrates that have evolved to have just the right shapes, flexibilities, solvation, and electrostatic properties, Nature can construct elaborate nanoscale "devices" that are capable of high level functions. Compelling examples include structures for energy transduction, complex molecule synthesis and selective degradation, information processing, and even self-replication. Inspired by these examples, a growing number of research laboratories have taken up the task of using biomolecular self-assembly to make artificial analogs that begin to mimic some of these functions.

As one possibility, many examples have shown that new materials can be generated using modified peptides, capitalizing on the ease of functional group introduction and the rich assembly behavior exhibited even by simple sequences.²⁻⁵ Nucleic acids have also been used to generate elaborate nanoscale structures that are difficult to create using other methods.^{6,7} A particular advantage of this approach is the ability to design completely new structures using computational methods. However, the bulk of the functional materials in nature involve self-assembling proteins, which typically involve defined three-dimensional shapes and complex interfaces that come together through large surface interactions. As a result of these properties, protein-based materials can possess a rigidity over length scales ranging from nanometers to microns. This property is often challenging to achieve using other building blocks, and is important for the generation of materials that can propagate nanoscale features into bulk materials with much larger dimensions. Proteins also offer the opportunity for synthetic tailoring, as site-directed mutagenesis can be used to introduce new amino acids chosen from the natural pool or the rapidly growing list of artificial groups that can be incorporated.^{8,9} This opens the door to the use of often sophisticated organic synthesis methods that can introduce new functional groups, from polymers to catalysts to nanoparticles, on the surfaces of the proteins either before or after the assembly occurs. As an additional advantage, many proteins can be produced on large scale through recombinant expression in bacteria or propagation in plants.

The protein shells of viruses, or "capsids", provide the most well-studied class of self-assembling proteins that has been used for materials science.¹⁰⁻¹⁹ These structures serve to protect the genomic nucleic acids from degradation and are often active participants in the cell-specific delivery of the genes required for propagation.²⁰ In the general sense, viral capsids possess a high degree of symmetry and are assembled from many copies of a small number of distinct protein monomers. These features have resulted in the atomic-level characterization of many capsids using X-ray crystallography and a growing number of detailed morphological

analyses using cryo-electron microscopy. Many viral capsids exist as hollow spherical structures with icosahedral symmetry, providing a high volume-to-surface area ratio for efficient genome encapsulation. Capsids with this geometry exhibit a characteristic " T " number, which is related to the total number of protein monomers ($T \times 60$) that comprise their structure.²⁰ As viral capsids with varying T numbers are known, it is possible to choose structures with different sizes and complexities to meet the synthetic challenge at hand. A number of viruses also exist as high aspect ratio rods, housing their nucleic acids in extended helices or circular loops. In some cases, purified viral capsid proteins can be used to generate more than one self-assembled state: tobacco mosaic virus (TMV) monomers can assemble into both disks or rods,^{21,22} the brome mosaic virus (BMV) can form spherical capsids with either $T=1$ or $T=3$ symmetry,²³ and the length of bacteriophage M13 can be altered by altering its genome size.²⁴ A structural comparison of several viral capsids used for materials science applications appears in **Figure 1**, and **Figure 2** provides a table listing many of their key properties.

Taken together, viral capsids offer many different structural starting points for materials construction. A number of the approaches detailed in the following sections use the intact particles without removing the nucleic acids inside, while some have shown that the genome can be removed to provide new spaces for chemical modification. In many examples, wild-type capsids can serve as scaffolds for the nucleation and growth of inorganic materials, while capsids bearing point mutations in the coat protein monomers are often used in biomedical and catalytic applications. The latter case allows the capsids to be modified using the rapidly growing set of chemoselective bioconjugation reactions,^{25,26} providing multivalent platforms for enhanced target recognition *in vivo* or "amplified" imaging agents that can be detected with very highly sensitivity. A summary of the chemical strategies that are commonly used for viral capsid modification appears in **Figure 3**.

Based on these considerations, this chapter reviews the progress that has been made in the use of self-assembling viral capsids to make new materials. The examples have been grouped by the application area, including sections on carriers for biomedical imaging and drug delivery, templates for nanoparticle synthesis and the creation of inorganic structures, and positional scaffolds for complex materials with optical and catalytic behavior. Within each section, the material has been organized by the viral particle that has been modified, as most research groups currently publish a series of studies using one or a limited number of different capsids. Although the chapter does not include related work on the use of ferritin^{10,27-29} and heat shock protein cages,^{10,30-33} the reader is encouraged to refer to the referenced works for examples of the use of these proteins for many of the same applications.

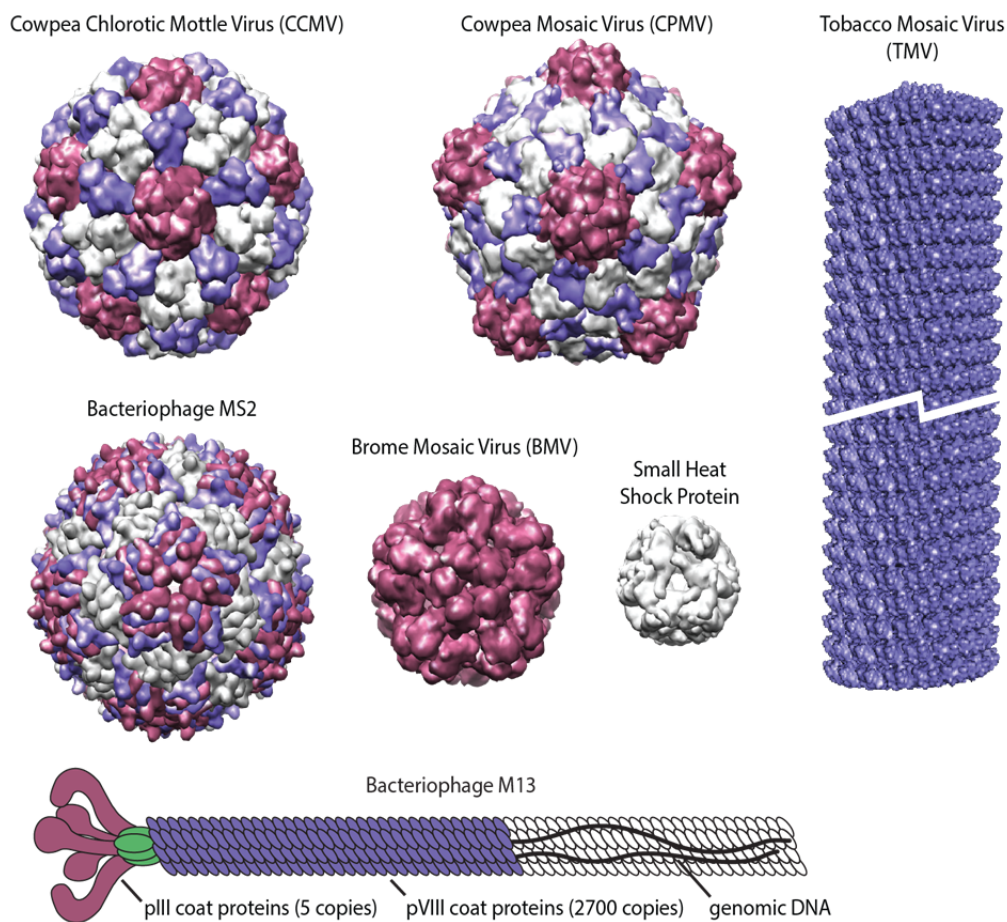


Figure 1. Structural comparison of viral capsids used for materials applications. Additional structural details are provided in the table in Figure 2.

virus	size	symmetry	monomers/ capsid	expression system (ref)	genome	pdb ID code
cowpea chlorotic mottle virus	28 nm	T=3	180	plant, <i>P. pastoris</i> (128)	ssRNA: 3171, 3100, 2173 bases	1CWP
cowpea mosaic virus	30 nm	pseudo T=3	120	plant (129), <i>S. cerevisiae</i>	ssRNA: 6.6, 3.8 kb	1NY7
bacteriophage MS2	27 nm	T=3	180	<i>E. coli</i> (infectious (66) and recombinant (75))	ssRNA: 3569 bases	2MS2
bacteriophage Q β	28 nm	T=3	180	<i>E. coli</i> (recombinant) (130)	ssRNA: 4217 bases	1QBE
canine parvovirus	26 nm	T=1	60	insect and mammalian cells (infectious) (131)	ssDNA: 5 kb	1P5Y
brome mosaic virus	28 nm	T=3	180	plant (132)	ssRNA: 3.2, 2.8, 2.1 kb	1JS9
brome mosaic virus	19 nm	T=1	60	plant (132)	—	1YC6
turnip yellow mosaic virus	30 nm	T=3	180	plant (133)	ssRNA: 6318 bases	1AUY
red clover necrotic mosaic virus	36 nm	T=3	180	plant (87)	ssRNA: 3889, 1448 bases	none
small heat shock protein*	12 nm	octahedral	24	<i>E. coli</i> (recombinant) (30)	—	1SHS
tobacco mosaic virus	17x300 nm	helical rod	2130	plant, (111) <i>E. coli</i> (recombinant) (114)	ssRNA: 6390 bases	2TMV
chilo iridescent virus	140 nm	icosahedral	thousands	whole insects, (134) cultured insect cells (135)	single linear dsDNA: 212,482 bp	none
bacteriophage M13	6.6x880 nm	helical rod	2700 (pVIII)	<i>E. coli</i> (infectious)	ssDNA: 6.4 kb	none

* from *Methanocaldococcus jannaschii*

Figure 2. Table of key properties for viral capsids commonly used to build new materials

1.3 Covalent modification of viral capsids for applications in medical imaging and drug delivery

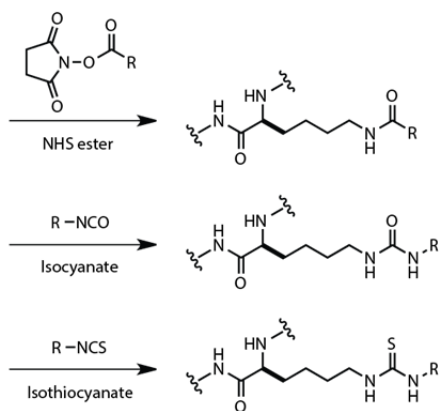
1.3.1 Introduction

Perhaps the most intuitive applications for virus based materials involve the delivery of therapeutic or imaggable cargo to specific locations in living organisms. As carriers, viral capsids offer the advantages of small particle size (10-100 nm, which is ideal for tumor accumulation),³⁴ low overall toxicity, and the ability to encapsulate and protect large numbers of cargo molecules within their interior volumes. They also provide a readily functionalizable external surface that can display multiple copies of a tissue targeting group. These features have inspired several research groups to develop a host of chemical methods that can add desired functionality to specific amino acid side chains introduced into the capsids before assembly. The strategies used to access these materials, the applications for which they are intended, and the rapidly advancing *in vivo* studies of viral capsid biodistribution and immunogenicity serve as the focus of this section. In addition to the applications that are discussed below, it should also be noted that filamentous viruses are commonly used to evolve peptides and small proteins with desired binding capabilities. As these applications lie outside of the scope of this chapter, the reader is instead directed to other reviews on this topic.^{35,36} The use of viruses to deliver genetic cargo in the context of gene therapy^{37,38} and the application of viruses as adjuvants for vaccines^{39,40} has also been reviewed elsewhere, and will not be covered here.

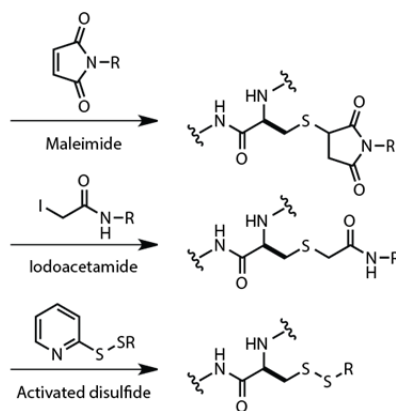
1.3.2 Cowpea chlorotic mottle virus (CCMV)

The first report of using viral capsids for materials applications appeared in 1998, when Trevor Douglas and Mark Young showed that polymeric and metal ion cargo could be contained inside the protein shell of CCMV.⁴¹ This early report relied on the ability of this capsid to "swell" when placed under specific pH conditions, allowing the access of exogenous molecules to the interior. They quickly followed up on this work with a series of publications involving the use of CCMV for the deposition of inorganic materials (described in the next section of this chapter), as well as the use of CCMV in a number of therapeutic applications. Regarding the latter, they demonstrated in their early work that the capsids could be chemically modified through the targeting of native lysine and carboxylic acid residues.⁴² This led to as many as 560 modifications on each capsid. They also introduced cysteine residues to allow the more controlled introduction of new functionality, such as 24 copies of a 24 amino acid peptide with antitumorogenesis properties. Although these high levels of modification caused the capsids to swell appreciably, they did not otherwise disrupt the morphology or lead to disassembly. In one study, they determined the capsid biodistribution by labeling them with fluorophores or radioactive iodine that could be tracked *in vivo*.⁴³ Their results showed that the viral capsids distributed throughout the animal, produced an immune response, and were significantly cleared after 24 hours.

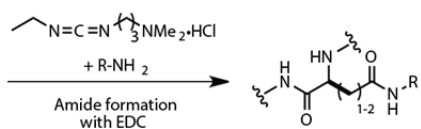
Lysine modification



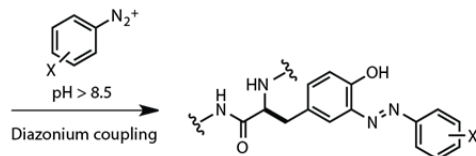
Cysteine modification



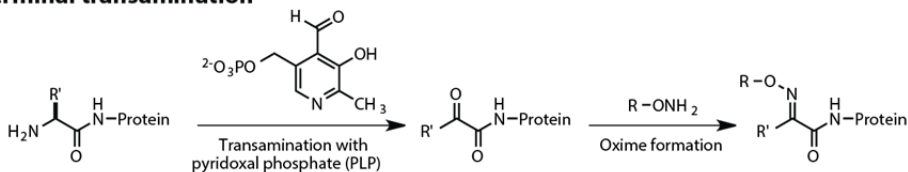
Aspartate and glutamate modification



Tyrosine modification



N-terminal transamination



Bioorthogonal modification strategies

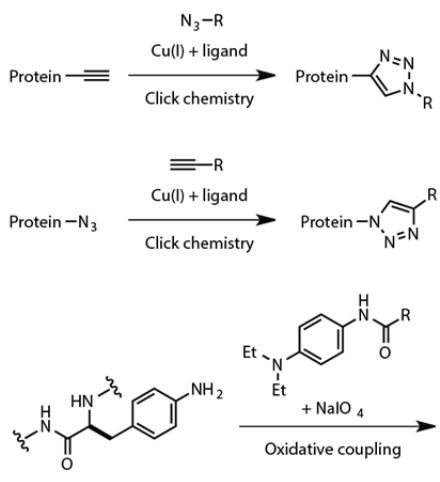


Figure 3. General chemical methods for the modification of specific amino acid side chains on viral capsid surfaces

By capitalizing on the ability of wild-type CCMV capsids to bind metal cations, the Douglas and Young groups were also the first to report a virus-based MRI contrast agent.⁴⁴ By exchanging the Ca^{2+} ions required for capsid assembly *in vivo*, they were able to bind roughly 140 Gd^{3+} ions to each structure. These capsids were also determined to have both the highest total relaxivity per particle and the highest relaxivity per Gd^{3+} ion reported to date. However, the relatively weak affinity for the toxic Gd^{3+} ions ($K_d = 31 \mu\text{M}$) prevented the use of these capsids in a clinical setting. Not to be discouraged, these labs began to develop capsid delivery systems with significantly tighter binding for the metal ions. To do this, Douglas and Young turned to the introduction of the chelating agent DOTA, which has a K_d of 10^{-20} M for Gd^{3+} .⁴⁵ They attached it to lysines on the surface of CCMV capsids using NHS-ester chemistry and then metallated it using GdCl_3 . The resulting capsids exhibited a 10-fold increase in relaxivity for each capsid-bound Gd(III)-DOTA complex compared to unbound analogs, which they attributed to the dramatically slower rotation of the capsids.⁴⁶ The relaxivity of the second-generation system was less than that with the Gd^{3+} bound directly to the capsids, which was explained by the lack of rigidity in the linker between the DOTA ligand and the protein backbone (thus undermining the slower rotation offered by the capsid). Alternately, the direct capsid binding motif could permit multiple water molecules to exchange at the same time, in contrast to the single water molecule bound by Gd(III)-DOTA complexes. Nonetheless, this compromise still yielded capsids with very high relaxivity values and a clinically useful Gd(III) affinity. The structures of the Gd(III) chelating ligands used in this study and in others are shown in **Figure 4a**.

In later studies, the Douglas and Young labs continued to pioneer the biomedical applications of viral capsids when they adapted CCMV for use in photodynamic therapy (PDT). To generate highly cytotoxic singlet oxygen in response to light, they attached a ruthenium bipyridine complex bound to an iodoacetamide modified phenanthroline ligand (**Figure 5a**) to cysteine residues introduced on the capsid surface.³³ Because the ruthenium bipyridine required blue light (which does not penetrate tissue well), the authors suggested that their PDT system could be most appropriate for treating biofilm infections of the oral cavity. In particular, they sought to target multi-drug resistant bacteria, such as *Staphylococcus aureus*, since resistance to singlet oxygen would be difficult if not impossible to develop.

Because singlet oxygen is indiscriminate but very short lived (the diffusion length is on the order of 50-100 nm⁴⁷), the photosensitizer must be tightly and selectively associated with the target. To explore this possibility, the authors compared two systems for targeting their CCMV-photosensitizer conjugates. The first approach was to coat the negatively charged capsids with poly-L-lysine, a cationic polymer, and then mix them with negatively charged bacteria. The second approach involved appending biotin to the surface lysines of the capsid via the corresponding NHS-ester, and then exposing the capsids to biotinylated antibodies coupled to streptavidin. The virtually irreversible association of biotin and streptavidin thus connected the antibodies to the photosensitizer-containing capsid. In this case, the antibody targeted *S. aureus*, but the broad utility of the system lies in the fact that *any* antibody raised to *any* biomolecular

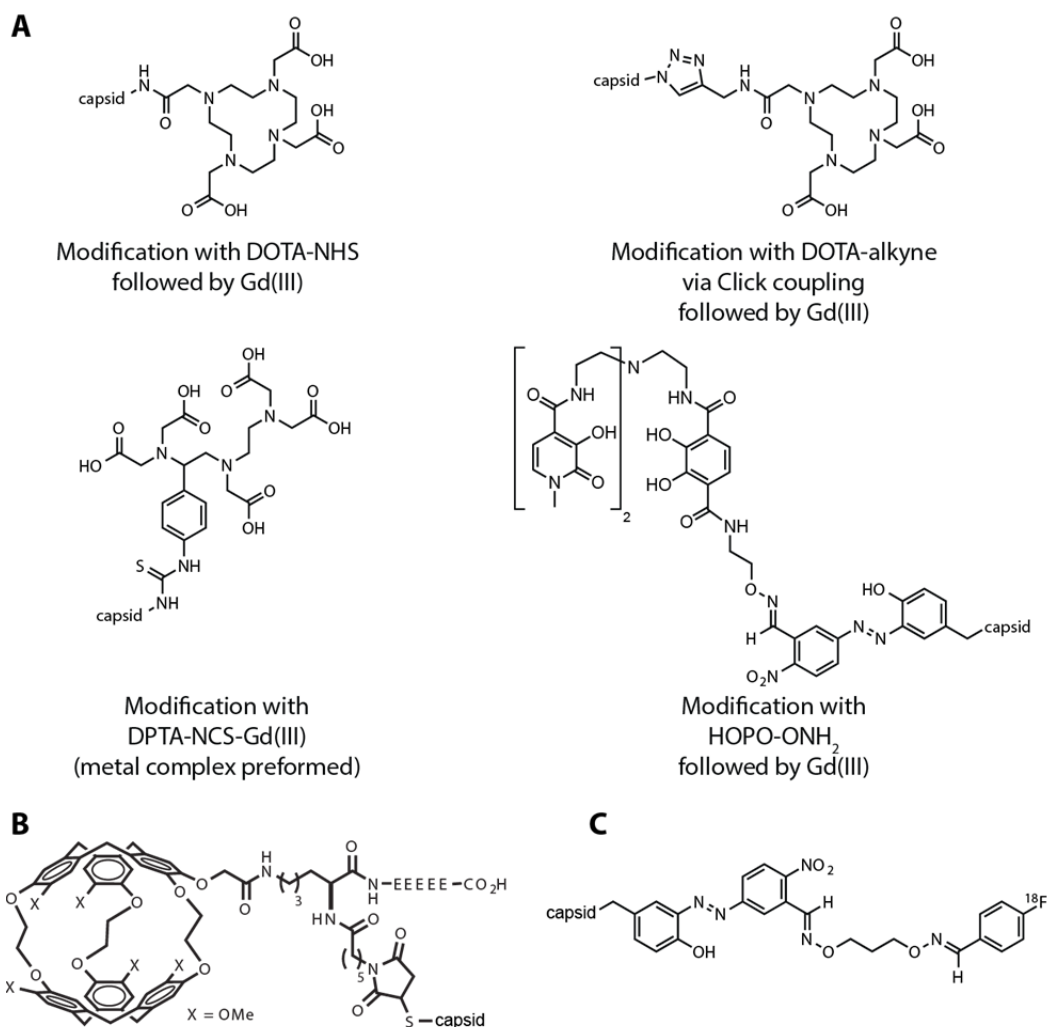


Figure 4. Summary of chemical functional groups attached to viral capsids for use in diagnostic imaging experiments. Applications of these conjugates include (A) proton MRI contrast enhancement, (B) ^{129}Xe -based MRI, and (C) positron emission tomography.

target could in principle be coupled using the same strategy. This second set of capsids led to significantly greater bacterial cell death upon illumination. These results were further substantiated by field emission scanning electron microscopy (FESEM) images, which showed sparse viral attachment with the electrostatic approach, and very heavy binding mediated by the antibodies. This report highlighted the importance of using capsids to combine multiple copies of different functional groups, especially when receptors are non-abundant, and therefore a maximum of cargo must be delivered to each.

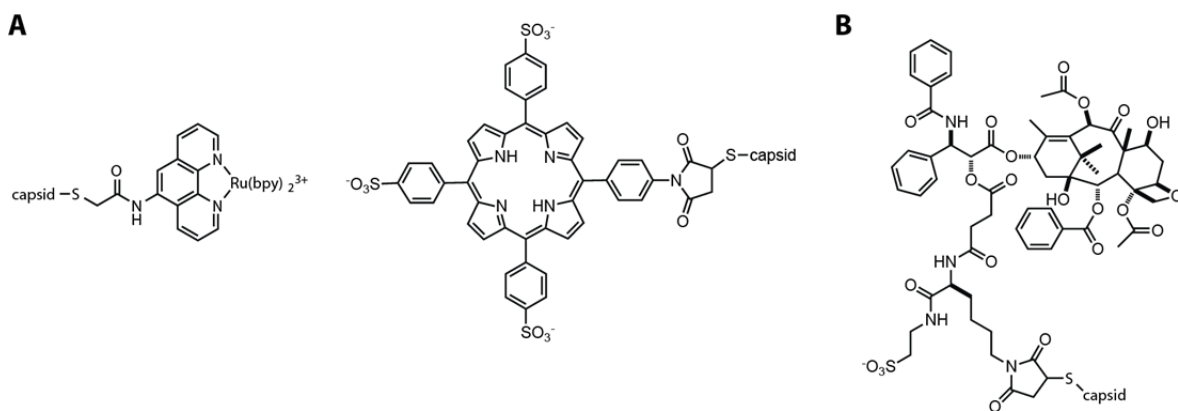


Figure 5. Therapeutic cargo attached to viral capsids for the purpose of killing cancer cells and *S. aureus*. The compounds in (A) are potent generators of singlet oxygen for photodynamic therapy, while Taxol (B) disrupts the cell cycle through inhibition of tubulin depolymerization.

1.3.3 Cowpea mosaic virus (CPMV) and bacteriophage Q β

In studies that were contemporary with much of the CCMV work, the Finn and Johnson groups also provided many early demonstrations of the addition of new synthetic functionality to viral capsid surfaces.⁴⁸⁻⁵⁰ In their initial reports, they developed a series of chemical methods that could target native lysine and introduced cysteine residues on the surface of the cowpea mosaic virus (CPMV). This gave them the ability to control both the positions and the overall coverage of the introduced groups, which they demonstrated using chromophores and other small molecules. In terms of adding biological targeting groups, they found in a later report that sugars derivitized with isothiocyanates could be attached to lysines on CPMV with varying surface density.⁵¹ To test if the protein-bound carbohydrates remained solution accessible, they mixed the CPMV conjugates with lectins, which are known to agglutinate natural glycoproteins. The observed aggregation confirmed that the sugars were displayed in a biologically recognizable manner.

Although carbohydrates can bind to some receptors *in vivo* and could lead to increased cellular uptake, the specific recognition of many cancer cell types is likely to require the addition of peptides and full-length proteins to the capsid carriers. Two of the first demonstrations of this capability were reported in a single publication by the Johnson lab in 2004.⁵² As the longest genetic sequence that can be inserted directly into a CPMV coat protein is about 30 amino acids, they opted instead to express intact targeting proteins independently and then couple them to the fully assembled capsids. In one example of this approach, they first prepared a mutant of CPMV's large subunit to introduce 60 thiol groups to the capsid surface. They then expressed LRR-InlB, a 22 kD protein shown to promote internalization of attached cargo, with its own free cysteine residue. To prepare the conjugate, a homobifunctional maleimide crosslinker was added to the CPMV capsids, followed by the LRR-InlB targeting group. The coupling of the proteins was confirmed using gel electrophoresis and electron microscopy, and could be carried out to the extent that western blots indicated all of the large subunits were linked to LRR-InlB.

To make a second type of targeted capsids the Johnson group also coupled them to the Int8 module of Herstatin, which has been shown to promote uptake by cancer cells. Rather than introducing a cysteine into Int8, they converted the N-terminal amino group to a thiolate before

using the homobifunctional crosslinking strategy described above. *In vitro* tests indeed showed the specific binding of the viral conjugate to cells expressing the HER2 receptor. Finally, a heterobifunctional NHS-ester-maleimide crosslinker was used to couple a cysteine mutant of T4 lysozyme to the lysines of wild-type CPMV, taking advantage of up to five native lysines per subunit for maximal loading. Importantly, they demonstrated the conjugated enzymes were still capable of hydrolyzing bacterial cell walls even after the coupling reaction.

To allow the attachment of an even wider set of synthetic groups, the viral capsid of CPMV has been explored extensively as a substrate for copper-catalyzed “Click” chemistry by the Finn group.⁵³ Based on the [3+2] Huisgen cycloaddition of azides to alkynes, this reaction provides a highly effective way to couple diverse sets of molecules with virtually complete functional group tolerance. The reaction is particularly well suited for bioconjugation because it is accelerated in aqueous media and can be used in conjunction with many other protein modification reactions.⁵⁴ To introduce the reactive groups onto CPMV, they modified native lysine and cysteine residues with either alkyne-substituted NHS esters or azides attached to bromoacetamides, respectively. Treatment of the modified capsids with copper(I) and a small molecule bearing the complementary functional group provided very highly levels of modification. With the later availability of improved ligands for the copper(I) ions, the Finn and Manchester groups demonstrated the installation of highly hindered molecules, including the large glycoprotein transferrin.⁵⁵ This protein was chosen because it is known to mediate cancer cell targeting and uptake. In a subsequent paper, the Finn group used a further optimized Click reaction to attach transferrin to bacteriophage Q β .⁵⁶ They showed that the conjugates were specifically internalized by cells expressing transferrin receptors. Although the rate of cell uptake was significantly improved with increasing numbers of transferrin ligands, separate binding experiments showed that this polyvalency did not improve affinity or avidity. Despite these surprising results, their report shows that Q β has great potential as a drug delivery vehicle.

To begin the *in vivo* characterization of the rapidly growing collection of tissue-targeting CPMV conjugates, the Manchester and Finn groups sought to determine the fate of the virus during the first few days after its introduction.⁵⁷ By tracking both the viral RNA (which was packaged inside, serving as a model drug cargo) and externally attached fluorophores, they confirmed that both components were distributed to a wide range of tissues after either oral or intravenous administration. This suggested that the capsids remained intact and could indeed be used as successful delivery vehicles. To examine the biological uptake more closely, the Manchester and Stuhlmann labs published a subsequent paper describing the use of an AlexaFluor labeled capsid to image chick embryos to a depth of up to 500 μm .⁵⁸ Using microscopy of intact and microtomed tissues, they found uptake of the capsids into the perinuclear compartments of vascular endothelial cells. The attachment of PEG alongside the dye prevented this uptake, and also increased the circulation time. The superior resolution and brightness offered by the viral capsid carriers allowed the researchers to capture real time videos that reported vascular changes associated with tumor formation.

To image deeper into tissue than can be achieved using microscopy, the Finn and Manchester groups have also functionalized CPMV and bacteriophage Q β with MRI contrast enhancement agents.⁵⁹ In a two-step procedure, they first introduced azide groups on the exterior surface of the capsids using lysine/NHS-ester chemistry, after which an alkyne-modified DOTA ligand (**Figure 4a**) was installed using Cu(I) Click chemistry. The DOTA moiety was then complexed to Gd³⁺, which is the most common metal ion used in MRI contrast agents. When

compared to commercially-available Magenvist, the viral conjugates demonstrated a two- to three-fold increase in relaxivity for each Gd(III) complex. Despite the large size of the capsids, which should cause an increase in relaxivity by slowing rotation as noted above, they did not observe an increase over free Gd-DOTA complexes. This was attributed to a trade-off between the slower tumbling rate afforded by the capsid and a decreased access to water molecules. In a continuation of this collaboration, these groups demonstrated the use of the capsid conjugates for MRI contrast enhancement in mice, observing no *in vivo* toxicity.⁶⁰ For further reference, the Manchester group has published a review of the use of virus-based nanoparticles for diagnostic imaging.⁶¹

As a different type of cargo, the Manchester and Finn labs have reported the attachment of fullerenes, or “Buckyballs”, to the exterior of CPMV and bacteriophage Q β .⁶² The material properties of fullerenes have created a significant interest in their use in a number of areas, including the generation of photovoltaic devices and as photosensitizers for photodynamic cancer therapy, but they are ordinarily very difficult to solubilize in water. Using Click chemistry, they were able to achieve high levels of fullerene attachment to bacteriophage Q β capsids, compared to the levels that could be achieved using EDC/NHS coupling. These conjugates were observed to be taken up by cancer cells, suggesting promise for future applications in photoactivated tumor therapy.

A key consideration for the use of viral capsids for *in vivo* applications is the level to which they are recognized by the immune system. While the extent of this concern will vary according to the amounts of the capsid carriers that are introduced and the duration of the treatment, many applications will require the shielding of the capsid surfaces from antibody recognition at the very least. With a view toward this goal, the Finn, Johnson, and Manchester groups collaborated to attach poly(ethylene glycol) (PEG) chains to the surface of CPMV capsids using NHS-ester chemistry.⁶³ They found that this strategy was indeed successful in preventing the binding of monoclonal antibodies to the capsids.

1.3.4 Canine parvovirus (CPV)

The popularity of many of the viruses discussed herein is due in part to the fact that they infect plants or bacteria, rather than something resembling the scientists who are working with them. The downside of this is that the innate targeting capability of such viruses is irrelevant when placed inside a mammal, forcing researchers to go to great lengths to introduce their own targeting groups. Canine parvovirus, by contrast, has a natural affinity for transferrin receptors in dogs as well as in humans, which it can use to enter cells of either species. Certain tumors overexpress transferrin receptors, making CPV a promising scaffold for drug delivery to these sites. In the first report on its use as a potential drug delivery vehicle, the Manchester group modified up to 100 of the surface lysines with NHS-ester dyes, which served as a model drug cargo.⁶⁴ This loading was not found to interfere with capsid uptake, which still co-localized with transferrin. Furthermore, the uptake remained dependent on expression of the transferrin receptor. This report thus demonstrated the possibility of generating a new class of cancer targeting drug delivery vehicles from mammalian virus precursors.

1.3.5 Bacteriophage MS2

Another icosahedral virus that has been used extensively as a bioconjugation target is bacteriophage MS2, or simply MS2. Like other $T=3$ viruses, its capsid consists of 180 sequence-

identical monomers that form a 27 nm hollow sphere. As a somewhat unusual feature, the assembled MS2 capsid has 32 pores, each of which is about 2 nm in diameter. This allows small molecules to diffuse more readily into and out of the capsid. With a view towards the development of vaccines, early work by the Stockley group has shown that peptides can be inserted into the coat protein sequence for efficient display to the immune system.⁶⁵

In terms of chemical modification, the Francis lab has focused extensively on the development of new bioconjugation strategies for both the interior and the exterior surfaces of the MS2 capsid. The first of these reports was in 2004, when they showed that it was possible to remove the ssRNA genome of the virus under alkaline conditions to produce stable, "empty" capsids.⁶⁶ These assemblies exhibited high stability, tolerating temperatures up to 60 °C and pH conditions ranging from 3-11. This treatment exposed a set of native tyrosine residues on the interior capsid surface, which could be modified chemically using diazonium salts at pH 8.5 or above. This resulted in the introduction of up to 180 copies of a new functional group in locations that would presumably not interfere with capsid biodistribution or be subject to enzymatic degradation *in vivo*. Subsequent reactions showed that the azo groups introduced using this strategy could be reduced using sodium dithionite, and that the resulting aminophenol could be used in further functionalization reactions.

More recent work used the diazonium coupling method to install fluorescent dyes on the inside of MS2 capsids while independently modifying external lysine groups with NHS-ester PEG polymers. The PEG coating was shown to decrease the binding of polyclonal antibodies to the capsids by 90% using ELISA.⁶⁷ This study highlights the utility of these capsids for the creation of complex materials bearing different modifications on each surface. These structures thus also provided the first model of a virus-based delivery vector that could display targeting moieties on the external surface and over 100 synthetic cargo molecules within.

For use in imaging applications, the MS2 capsid has also been modified to display Gd(III)-based MRI contrast enhancement agents (**Figure 4a**). As was the case for CCMV and CPMV (described above), these studies highlight the advantage of decreasing the tumbling rate of the metal complexes while simultaneously amplifying their signal by attaching multiple copies to the same delivery vehicle. The first MS2-based example of this concept came from the Kirschenbaum group,⁶⁸ which used the external lysines of MS2 to attach isothiocyanate functionalized Gd-DTPA ligands. This installed over 500 Gd³⁺ atoms on each capsid and resulted in an overall relaxivity of 7,200 mM⁻¹s⁻¹ (over three orders of magnitude greater than that of a single Gd³⁺ ion). The Francis group also explored the use of MS2 as a scaffold to bind Gd³⁺,⁶⁹ but expanded the study to investigate the effects of internal vs. external attachment. They installed aldehyde functional groups either on the interior or the exterior of the capsid by targeting tyrosine residues with diazonium salts and targeting lysines with NHS-esters, respectively. A hydroxypyridonate(HOPO) ligand⁷⁰ functionalized with an aminoxy group was then attached to the aldehyde groups through oxime formation. They found that Gd-HOPO confined to the interior of the capsid exhibited higher overall relaxivities (up to 3,900 mM⁻¹s⁻¹) than the externally bound ligands, as well as much better solubility for the modified capsids. A follow up report⁷¹ showed that both the rigid linker attaching the Gd-HOPO complex to the capsid and the fast diffusion of water in and out of the protein shell were both important to obtain the high relaxivities that were observed.

As an alternative to proton based MRI, which directly observes the relaxation of water molecules occurring throughout the body, MRI imaging based on hyperpolarized xenon probes

shows much promise for the highly sensitive detection of nonabundant biomarkers. This is due to the large chemical shift window of ^{129}Xe (which gives it the ability to report its specific chemical environment) and the virtually nonexistent level of background signal. To develop viral-capsid based materials that could be detected using this imaging method, the Pines, Wemmer, and Francis groups attached 120 cryptophane cages (**Figure 4b**) to cysteine residues introduced inside genome-free MS2 capsids.⁷² The cryptophane cages bound to xenon atoms in aqueous media, giving them a specific chemical shift that could be distinguished from unbound xenon atoms. In combination with chemical exchange saturation transfer (HYPER-CEST),⁷³ solutions of capsids could be detected with concentrations as low as 700 fM. The high sensitivity of this technique, along with the non-toxicity of xenon, makes this a very attractive future candidate for *in vivo* MRI imaging.

MS2 capsids have also been used to house radioactive species, such as ^{18}F , for detection using positron emission tomography (PET). This technique allows the imaging of positron emitting radioisotopes with exceptionally high signal-to-noise ratios *in vivo*, and is widely used for cancer detection using [^{18}F]-fluorodeoxyglucose. To generate capsids that could be similarly tracked *in vivo* using this technique, aldehyde groups were introduced inside MS2 capsids using the diazonium coupling strategy described above.⁷⁴ A *bis*(aminoxy) linker was then used to attach [^{18}F]-fluorobenzaldehyde to these locations site selectively (**Figure 4c**). By following the biodistribution of the capsids in rats using dynamic PET, the authors observed that the capsids exhibited somewhat prolonged circulation times relative to free [^{18}F]-fluorobenzaldehyde. They also showed that additional small molecules attached to the inner surface of the capsid did not effect the biodistribution behavior.

In the previous examples of MS2 bioconjugation, the capsid exterior was only modified in a nonspecific fashion through the acylation of any of the three solvent-exposed lysine residues. This introduces many limitations for the installation of more complex targeting groups, as most of the cell binding proteins and peptides that one would want to attach will also contain lysine residues. Desiring a more elegant means of site-specifically modifying the external surfaces of MS2 capsids while retaining the ability to target cysteine residues on the interior, the Francis lab used the Schultz *in vivo* technique⁸ to introduce a uniquely reactive amino acid side chain into the coat protein monomers.⁷⁵ They chose to introduce *p*-aminophenylalanine (*paF*), producing 180 anilines on the external surface of each capsid. In the presence of sodium periodate (NaIO_4), the *paF* residues were shown to participate in a chemoselective oxidative coupling reaction with *N,N*-dialkyl phenylenediamines that had been reported previously by their group.⁷⁶ This reaction was shown to be fast, high-yielding, and selective for the *paF* residues over all of the other natural functional groups on the protein.

With this new construct in hand, the Francis group went on to elaborate MS2 as a targeted drug delivery vehicle. In their first report, the external *paF* residues were modified with *N,N*-dialkyl phenylene diamine substituted DNA aptamers.⁷⁸ The specific nucleic acid sequences used in this study were identified by the Tan group⁷⁷ as having a high affinity for protein tyrosine kinase 7 (PTK7), which is over expressed on many cancer cell types. The inside of the capsids were labeled with fluorescent dyes through the alkylation of introduced cysteine residues using maleimide chemistry. It was shown that these aptamer-functionalized capsids could effectively target Jurkat leukemia T cells with high specificity over other cell types, while capsids displaying an aptamer with a random sequence had no targeting capabilities.

A second report took this observation one step further by combining the capsids with

external PTK7 targeting aptamers with porphyrin molecules, which act as photodynamic sensitizers upon illumination.⁷⁹ To introduce them, water-soluble porphyrins bearing maleimide groups were constructed and used to alkylate the interior cysteine residue. Using this strategy, up to 180 copies of the porphyrins (**Figure 5a**) could be encapsulated within each capsid, and 20 copies of the DNA aptamers were displayed on the outside. As before, the dual-modified capsids were shown to bind to Jurkat cells effectively and showed low affinity for other cell lines. Once the modified viral particles had bound the receptors, illumination for 20 minutes caused the porphyrins to destroy the cancer cells through the generation of up to 300,000 molecules of single oxygen per capsid. Control experiments with cells lacking the PTK7 receptor displayed little to no capsid binding and subsequent cell death. Additionally, the authors mixed solutions of dual-modified Jurkat targeting MS2 capsids, Jurkat cells, and red blood cells. Jurkat cell death was observed, but the red blood cells were minimally perturbed. This illustrates the utility of photodynamic therapy in selective destruction of targets in complex mixtures. The high spacial resolution and selectivity of this technique stems from the exceptionally high reactivity of the active species, singlet oxygen. Singlet oxygen will undergo side reactions and become quenched before it can diffuse very far from where it is formed.

One final example of bacteriophage MS2 being used as a drug delivery platform also comes from the Francis lab. In this report, 120 copies of the potent but very hydrophobic anti-cancer drug Taxol (**Figure 5b**) were incorporated into the interior of the capsids.⁸⁰ This was accomplished by producing a water-soluble linker that provided each Taxol molecule with a maleimide group for cysteine alkylation. These untargeted constructs displayed the ability to destroy MCF-7 breast cancer cells, presumably initiated by uptake of the capsids into the cells through pinocytosis. Endosomal and lysosomal compartments then disassembled the capsid, exposing the drugs and allowing hydrolysis of the linker, and releasing free Taxol.

1.3.6 Summary

Viral capsids have the potential to overcome a number of the challenges for the targeted delivery of therapeutic and diagnostic cargo. They offer monodisperse non-toxic structures that can solubilize hydrophobic drugs introduced through encapsulation or bioconjugation, facilitate chemotherapy by targeting the payload to cancer cells and minimizing exposure of healthy tissues, increase uptake into tumors through the EPR effect, and extend the half-life by preventing premature filtration by the kidneys. All of these roles require the addition of functionality to the capsid, performed using a rapidly growing set of bioconjugation reactions.

1.4 Viral capsids as templates for the construction of inorganic materials

1.4.1 Introduction

Many of the first examples of using viral capsids to make new materials involved their use as templates for the controlled deposition of metal ions and the precise positioning of inorganic nanoparticles. In general, these studies have been motivated by the monodisperse nanoscale sizes of viral capsids, as well as the powerful ability to introduce metal-binding amino acid side chains in specific locations on their surfaces. Often, the assembled capsid structures are subjected to combinations of metal ion salts and reducing agents, which causes metal to be deposited selectively on the protein surface. This can produce inorganic nanostructures with discrete

shapes and sizes that are difficult to realize using other methods. Hollow spherical viral capsids can sequester nanoparticles within their cavities or be used as “nano-reactors” to grow nanoparticles of well-defined sizes. Filamentous and rod-shaped viruses can be used to template the growth of one-dimensional nanomaterials, and can produce extended structures through liquid crystal assembly properties.

1.4.2 Cowpea chlorotic mottle virus (CCMV)

The earliest report of any viral capsid-based material came from the labs of Trevor Douglas and Mark Young in 1998.⁴¹ In this work, they used 28 nm CCMV capsids that were devoid of genetic material as empty shells for the encapsulation of small molecule guests. Key to this work was the discovery that the CCMV capsid swelled at pH values above 6.5, leading to increased porosity and more facile diffusion of materials into the structure. This property, coupled with the natural positive charge lining the interior of the cavity stemming from 1,620 basic residues, allowed the controlled sequestration of anionic species within the capsid. Initially, soluble tungstate anions (WO_4^{2-}) were introduced into the capsids at a high pH. The pH was then decreased, causing oligomerization of the tungstate to form insoluble paratungstate ($\text{H}_2\text{W}_{12}\text{O}_{42}^{10-}$) while simultaneously decreasing the porosity of the capsid and trapping the polyoxometalate within the cavity (**Figure 6a**). This methodology was additionally used to mineralize decavanadate ($\text{V}_{10}\text{O}_{28}^{6-}$) inside CCMV. In this same report, the positive lining of the capsid was also used to capture the polyanionic polymer poly(anethol sulfonate) through electrostatic interactions. Expanding on this seminal work, Douglas, Young, and coworkers introduced genetic modifications into the viral coat protein in order to increase the range of materials that could be trapped inside.⁸¹ This was driven by the desire to form iron oxide nanoparticles with controlled dimensions. To accomplish this, the positively charged interior lining was altered by mutating nine basic residues on each monomer to glutamic acids. The modified capsids were shown to bind ferrous ions, which then reacted with oxygen to form iron oxide particles with a size that was constrained by the interior diameter of the capsid. In separate work, they were able to use similar approaches to produce platinum nanoparticles within heat shock proteins and demonstrate their use for photocatalytic hydrogen production.³¹

As an alternative to this procedure, Dragnea and coworkers (along with Douglas and Young) reported the use of fully formed inorganic nanocrystals as templates for the assembly of CCMV coat proteins (**Figure 7a**).⁸² The motivation for this work was to investigate the assembly process of CCMV and to determine if the nanoparticles could assist or even direct it. In this study, the authors removed a large portion (34 amino acids) of the protein N-terminus, which resulted in a broadening of the resulting capsid size distribution. When assembled in the presence of gold nanocrystals, however, the capsids that encapsulated the particles once again formed a monodisperse population that was dependent on the nanocrystal size.

1.4.3 Chilo iridescent virus (CIV)

The chilo iridescent virus (CIV) is not a simple icosohedral virus, instead exhibiting a layered structure composed of a dsDNA-protein core that is surrounded by a lipid bilayer. This is in turn encased in a second protein shell, giving the entire ensemble a diameter of 140 nm. With the goal of making materials with desirable optical properties, Radloff and Vaia reported the use of CIV

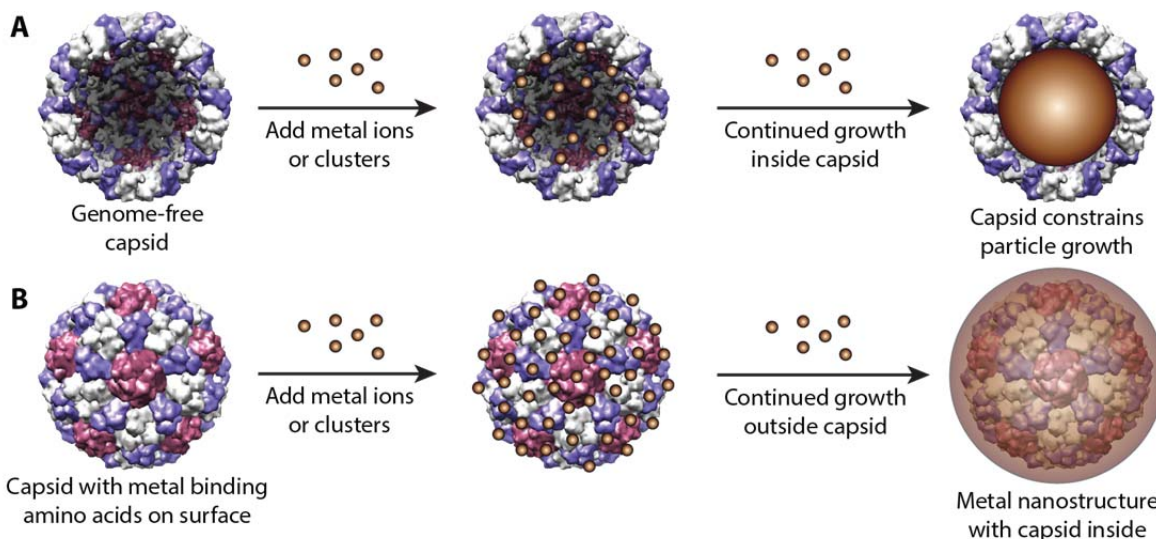


Figure 6. General strategies for the templated growth of metal nanoparticles using viral capsids.

as a scaffold to position gold nanocrystals and subsequently form continuous nanoshells using electroless deposition techniques (**Figure 6b**).⁸³ For the initial step, the authors relied on the native surface residues of CIV to bind negatively charged gold nanocrystals through electrostatic interactions. Then, reduction of gold(III) salts from solution afforded a thin shell of gold surrounding the entire virion. This approach of modifying the exterior of the capsid is complimentary to the examples of CCMV, where the inner surface of the capsid was exploited for inorganic particle nucleation.

1.4.4 Cowpea mosaic virus (CPMV)

In addition to its applications in medical imaging and drug delivery, CPMV has also found use as a template for inorganic materials. In an intriguing report by Colvin, Johnson and coworkers, body centered cubic crystals of CPMV were formed to provide empty cavities and channels comprising about 50% of the overall crystal volume.⁸⁴ These repeating void spaces were then used for platinum and palladium deposition through the addition and subsequent reduction of metal salts. By using this approach, extended metal structures were formed with periodic nanoscale features that were templated by the regular packing arrangement established by the CPMV capsids.

1.4.5 Brome mosaic virus (BMV)

The capsid of BMV is a simple icosahedral assembly with the ability to form both 28 nm ($T=3$, 180 monomers) and 19 nm ($T=1$, 60 monomers) particles. Similar to (and actually preceding) their work with CPMV, Dragnea and coworkers sought to assemble the BMV coat protein around preformed gold nanocrystals (**Figure 7a**).⁸⁵ A critical step in doing so was determining that a tetraethylene glycol (TEG) coating on the gold nanoparticles was required for their effective encapsulation. They examined the effects of nanoparticle size on protein assembly, leading to the finding that the capsid size could be tuned by altering the diameter of

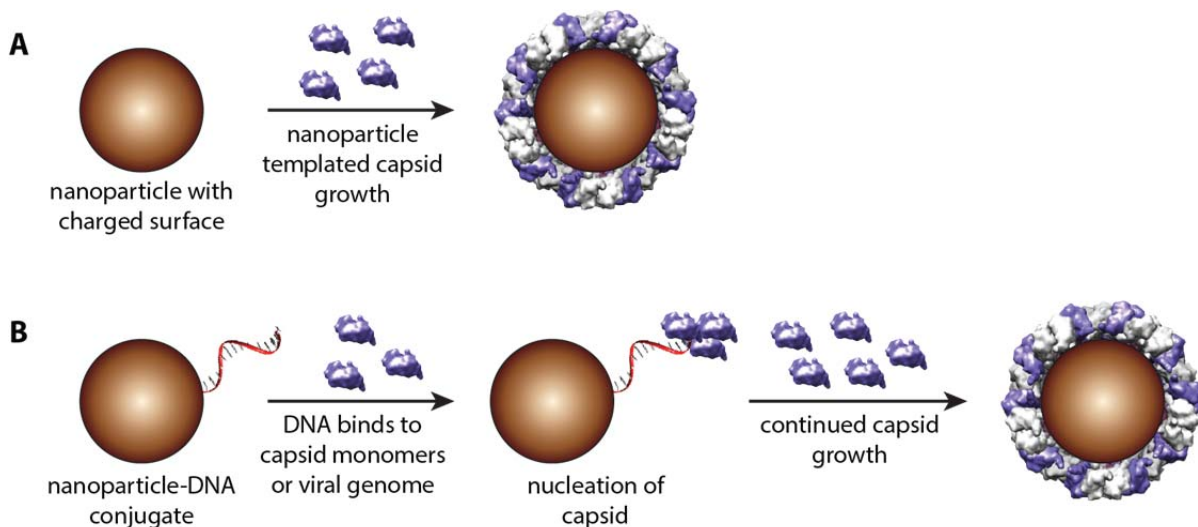


Figure 7. General strategies for the templated growth of viral capsids on the surfaces of metal nanoparticles.

the gold particle.²³ This result suggested that strong interactions took place between the encapsulated particle and the coat proteins during the assembly. A mechanism for this process was proposed, in which initial nonspecific protein adsorption first occurs on the nanocrystal surface. Once the nanoparticle has been covered with protein, the surface-bound capsomers can then fully assemble on the gold surface in response to a drop in pH. Through similar methodology, the Dragnea lab also showed that fluorescent quantum dots can also be encapsulated in BMV.⁸⁶

1.4.6 Red clover necrotic mosaic virus (RCNMV)

An additional example of an icosahedral virus that has been used for nanoparticle encapsulation is provided by the 35 nm capsid of RCNMV. This virus, much like the CCMV, swells at pH values above 6.5. By increasing the pH even further, the capsid can be completely disassembled. Through the careful lowering of the pH, it is then possible to reassemble the intact capsid without the genomic RNA inside. Interestingly, the RNA-free capsid retains its native icosahedral geometry, but the diameter is reduced slightly to 30 nm. Franzen and coworkers used this controllable assembly property to encase gold nanoparticles using the RCNMV proteins, but took a different approach than previous examples by instead exploiting the natural RNA sequence to guide the assembly process.⁸⁷ To do this, they used gold nanoparticles functionalized with ssDNA sequences that were complementary to the RNA of RCNMV. Isolated RCNMV RNA hybridized to the nucleic acids on the nanoparticles, and the subsequent introduction of coat protein allowed the directed assembly of capsids on the gold surfaces (**Figure 7b**). It was shown that multiple sizes of gold nanoparticles could be encapsulated within the RCNMV, but when the size of the particle exceeded the size of the virus cavity, no complex structure formed. The authors then showed that both CoFe_2O_4 nanoparticles and quantum dots could be incorporated into RCNMV using the same chemistry,⁸⁸ displaying the versatility and promise of this technique. Fundamentally different from the examples above, in which the inorganic material was used as the scaffold to assemble the capsid, this method is unique in its use of the

genomic RNA to template capsid formation around metal particles.

1.4.7 Bacteriophage M13

Unlike the spherical viruses described above, M13 is a filamentous phage with a large aspect ratio (880 nm in length, 6.6 nm in diameter) and a complex coat protein assembly. Its size and shape allows the virus to form liquid crystalline domains on surfaces due to the high degree of interparticle attraction along the rod axis. Along the length of this large virus, a ssDNA genome is protected by 2,700 copies of a pVIII major coat protein assembled in a helical pattern. One end of the virus displays five copies of a pIII minor coat protein, which is often fused to libraries of protein sequences for use in phage display techniques. Briefly, this process entails repeated cycles of mutation, selection of phage that display pIII proteins with desired binding capabilities, and amplification. In this way, phage can be identified with the ability to bind to virtually any desired biological or even inorganic target. In-depth discussions of the phage display technique can be found in other references.^{35,36}

The M13 phage display platform has been widely used by the Belcher lab to evolve peptides that can bind a variety of inorganic materials. In 2000, they showed that by generating large libraries of phage with differing peptides displayed on the pIII coat proteins, followed by extensive screening for binding against semiconductor crystals,⁸⁹ they were able to identify peptides that selectively bound to a specific semiconductor material. In this study, the single crystal semiconductors GaAs(100), InP(100) and Si(100) could be selectively recognized by the evolved peptides, and the researchers were even able to show selectivity towards binding different faces of the same material, with the peptides able to differentiate between GaAs(100) and GaAs(111).

Through further evolution of peptides on the pIII protein, the researchers were able to generate M13 phage bearing peptides capable of binding ZnS quantum dots.⁹⁰ They then used the liquid crystalline behavior of the M13 to form ordered domains of the quantum dots that extended to millimeters in length. The application of a magnetic field was used to align the liquid crystal domains, providing an elegant means of ordering nanoscale objects over long distances.

Though most of the early work on inorganic material-binding peptides was done using the pIII protein, the pVIII major coat protein has also shown promise for the attachment of inorganic particles along the length of the phage. By evolving peptides on the pVIII protein in a manner that was analogous to the pIII protein evolution described above, the Belcher group was able to install 2,700 sites on each M13 capsid for the binding of ZnS, CdS, CoPt, and FePt nanoparticles. This led to the construction of one-dimensional nanowires with high aspect ratios.⁹¹ Two years later, the Belcher lab used anionic tetraglutamate sequences displayed on the pVIII coat proteins to deposit continuous nanowires of cobalt oxide (Co₃O₄).⁹² These structures were shown to be attractive candidates for anodes in lithium ion batteries due to their high surface area, ability to produce thin 2D arrays, and mechanical flexibility afforded by their core-shell structure. By also introducing a gold binding peptide motif alongside the tetraglutamate peptides, gold was incorporated into the Co nanowires to increase the battery capacity further. As an alternative approach to creating flexible, high capacity electrodes for advanced lithium ion batteries, the Belcher lab has also used the pVIII protein to attach amorphous iron phosphate (α -FePO₄) while binding the pIII protein to carbon nanotubes.⁹³ Electrodes constructed with these nanowires surpassed the capacity and conductivity of bulk iron phosphate electrodes.

The Schaak laboratory used similar methodology for creating nanowires on M13, but

instead exploited the native charge of the pVIII proteins to bind metal cations.⁹⁴ In this fashion, nanowires of Rh, Pd, or Ru nanoparticles could be created in a much less time- and energy-consuming process than through the use of phage display and directed evolution.

1.4.8 Tobacco mosaic virus (TMV)

The tobacco mosaic virus (TMV) is another rod-shaped virus that has seen extensive exploration for the fabrication of new materials. The capsid of the native virus is formed from the assembly of 2,130 copies of a single protein monomer around a ssRNA genome. This results in low polydispersity helical rods that are 18 nm in diameter and 300 nm in length. These tube-like structures are hollow, with a 4 nm central pore that spans the entire length of the virus.

In an early report, the labs of Mann, Douglas, and Young used the naturally charged amino acids of TMV capsids to deposit a range of inorganic materials along their sides.⁹⁵ Various nanotube structures were created by passing H₂S gas through solutions containing TMV and CdCl₂ or Pb(NO₃)₂, leading to the nucleation of CdS or PbS on the viruses. Iron oxide mineralization and the sol-gel deposition of SiO₂ on the TMV exterior surfaces were also reported, displaying the versatility of TMV as the template.

Like the bacteriophage M13, TMV also will form liquid crystal domains due to its high aspect ratio. The Mann group was able to use the liquid crystal forming behavior of TMV to their advantage as a template for the sol-gel condensation of SiO₂. Through this process, the void spaces between viruses were filled with silica, producing a porous silica mesostructure.⁹⁶ In a different report, Velev and coworkers were able to gain control of the long-range alignment of TMV liquid crystals by dragging the meniscus of a TMV solution between two glass plates.⁹⁷ This created a shear gradient in the solution, aligning the liquid crystal domains over centimeter distances. The introduction of NHS-modified gold nanoparticles to these liquid crystals resulted in the positioning of nanocrystals along the capsids. These sites were then used to direct the subsequent electroless deposition of silver to afford ordered nanowires aligned over distances greater than a centimeter.

In the case of metal deposition, the preceding example uses the strategy of first attaching preformed particles to the capsids, and then increasing their size through the addition of metal salts and reductants in solution. An interesting alternative approach was taken by Demir and Stowell, who initially attached the dye Cascade Blue to the exterior surface of TMV through NHS-ester chemistry.⁹⁸ UV irradiation of the Cascade Blue molecules caused the photochemical reduction of Cu(I) ions in solution, leading to the formation of copper nanoparticles on the surface of TMV.

In addition to metals and metal oxides, conducting nanowires can also be templated by TMV using organic conducting polymers. Although the building blocks are different, the general concept is the same—the capsid serves as an initial nucleation site for the material, which grows to cover the protein surface. The Wang group displayed this methodology when they attached and polymerized pyrrole and aniline on the exterior surface of TMV to create conductive wires.⁹⁹ This serves as yet another example of the wide scope of materials that can be generated using capsid templates.

The central pore of TMV capsids has also served as a surface for controlled metal deposition. This was first demonstrated by the Culver and Mann laboratories in 2003, where they demonstrated the ability to control whether nucleation of metal nanoparticles occurred on the interior or the exterior surface by modifying the surface charge of the virus.¹⁰⁰ Later that year,

Bittner and coworkers demonstrated the successful construction of continuous nickel and cobalt nanowires that were only a few atoms thick within the TMV pore.¹⁰¹ Further work showed that this technique could also be applied to form copper nanowires in the same fashion.¹⁰²

To improve metal deposition on the virus surface, the Culver lab turned to genetic manipulation of the coat protein. Cysteine residues were introduced in positions 2 and 3, which positioned them on the exterior surface and 3'-face of the rods. This promoted metal nucleation along the length due to thiol-metal interaction and lead to a denser and more stable coverage of gold, silver, or palladium nanoparticles.¹⁰³ In a subsequent paper, a coat protein mutant was produced with a single cystine in position 3, which proved solvent accessible on the 3'-face of TMV rods, but less reactive along its length.¹⁰⁴ This design allowed the Culver lab to attach TMV viruses to a gold surface in an “end-on” configuration through the formation of gold-sulfur covalent linkages. Because only one end of the virus displayed the thiol groups, directionally oriented 2-dimensional arrays of TMV capsids were formed with perpendicular angles to the surface. Electroless plating of nickel or cobalt on these systems then produced metal nanotube arrays with a high surface area, displaying a two-fold increase in battery capacity when used as an electrode in a nickel-zinc battery system.

Further versatility of TMV as an inorganic scaffold is illustrated in the work of Ozkan, Yang, and coworkers, who created components for digital memory devices by aligning platinum nanoparticles along the sides of TMV.¹⁰⁵ One final example comes from a closely related virus—the tomato mosaic virus (ToMV). In this report, Yamashita and coworkers were able to align magnetic nanoparticles within the pore of the rod-shaped ToMV.¹⁰⁶

1.4.9 Summary

These examples highlight the utility of viral capsids for the construction of inorganic nanomaterials. When used as templates for metals and metal oxides, a number of spherical, core-shell, one-dimensional, and mesoscale architectures with nanometer-sized features can be created. Their well-defined sizes and shapes, coupled with the stability of these viral capsids to the often aggressive conditions that are required for metal ion deposition, have established them as promising building blocks for future electronic and optical materials.

1.5 Capsid-based materials for optical and catalytic applications

1.5.1 Introduction

In order to capture the energy of sunlight, photosynthetic organisms arrange an impressively diverse collection of chromophores and photoprotective molecules around centralized electron transfer groups (**Figure 8a**).^{107,108} This is achieved through the use of a protein scaffold that establishes the crucial spacings between and relative orientations of the components in order to maximize the efficiency of inter-chromophore Förster resonance energy transfer (FRET). This architecture has inspired a variety of successful artificial light harvesting systems based on polymers and dendrimers. However, these efforts are often hindered by the difficulties associated with the synthesis of organic frameworks containing multiple chromophores and the lack of available scaffolds that are rigid enough to maintain discrete inter-chromophore distances over large length scales. Viral capsids provide promising alternative scaffolds for the positioning of light harvesting molecules, as they readily assemble into regular, highly symmetric, rigid shapes

with an ideal periodicity for FRET to occur between attached components. The following examples demonstrate the use of the capsid monomers to integrate multiple chromophores and electron transfer groups, with the goal of creating complex functional assemblies with efficient energy collection and transduction capabilities.

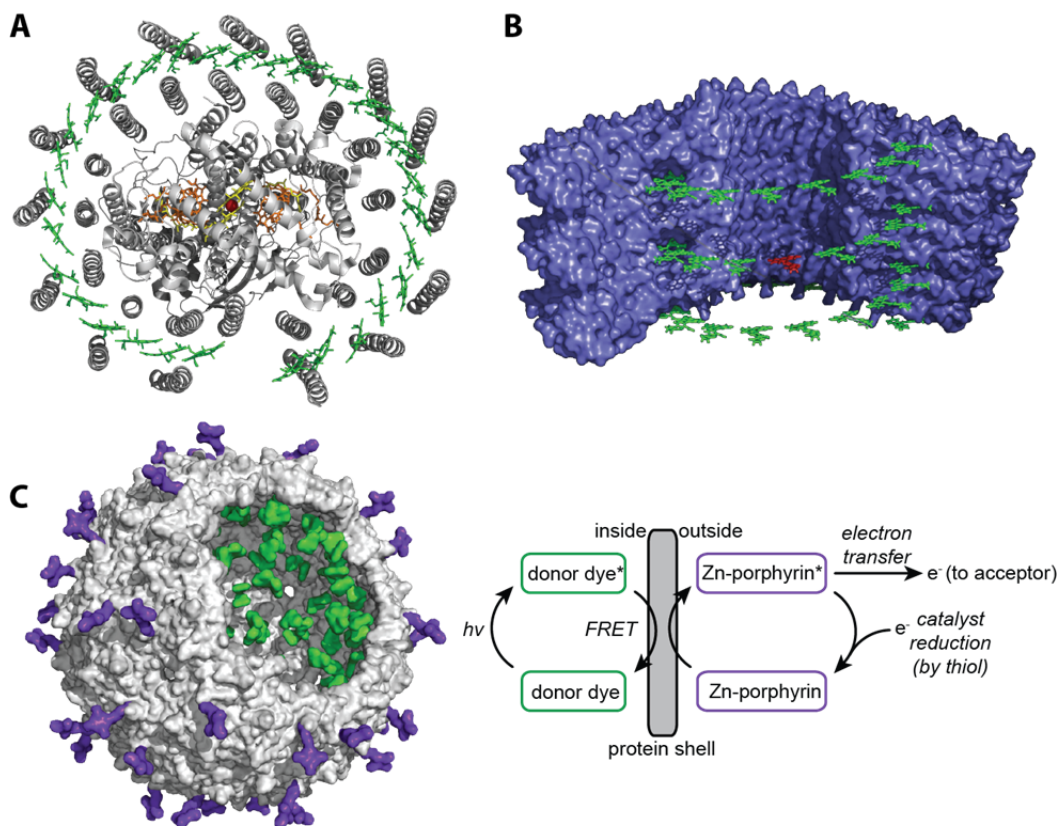


Figure 8. Light harvesting and energy conversion systems templated by self-assembling proteins. (A) A structure is provided for the LHI complex and the photosynthetic reaction center of *Rb. Palustrus* (from reference 136), showing the highly optimized arrangement of bacteriochlorophylls (green) for light collection. Viral capsid-templated chromophore assemblies that mimic some aspects of natural light harvesting systems have been prepared by using (B) TMV (ref. 114), (C) bacteriophage MS2 (ref. 123), and bacteriophage M13 (ref. 126, not shown).

1.5.2 Cowpea mosaic virus (CPMV) and hepatitis B virus (HBV)

As noted above, CPMV was one of the first viral capsids to be covalently modified to create nanoscale materials.⁴⁸ In 2002, the Finn and Johnson groups began their work with virus based materials by exploring the reactivity of the wild-type capsid with a maleimide fluorescein derivative. They observed that roughly 60 dyes could be attached to a capsid comprised of two different monomers. They then installed an additional, much more reactive cysteine in a loop on the exterior surface of one of the monomers, which they found could be alkylated using fewer equivalents and shorter reaction times. This difference in reactivity allowed them to perform sequential reactions, selecting for the introduced cysteine with moderate conditions, followed by more forcing conditions to modify the native cysteine. Using this differential reactivity, they first

introduced fluorescein, followed by tetramethylrhodamine, but did not analyze energy transfer. While these systems were not expressly built for optical or catalytic applications, they provided a proof of principle and inspiration for much of the work to follow.

In subsequent papers published the same year, the Finn and Johnson groups explored the reactivity of lysine residues and pushed the envelope for chromophore bioconjugation.⁴⁹ They showed that up to 240 chromophores could be installed using NHS-ester chemistry. Lysines could also be targeted to introduce a biotin tag, which caused the formation of extended networks upon addition of avidin.⁵⁰ They also demonstrated the sequential and independent labeling of cysteine and lysine residues.

Four years later, Evans and coworkers took advantage of the high reactivity of CPMV's lysines to introduce 240 copies of NHS-ester modified ferrocene groups on each capsid.¹⁰⁹ This represented a new record for the number of redox-active groups on a single scaffold, surpassing previously reported dendrimer-based systems by almost a factor of four. Despite the crowding expected from placing four metal centers on each monomer, the ferrocene units were observed to behave independently using cyclic voltammetry. The authors proposed that the modified CPMV capsids could be used as electron storage reservoirs with the capability for multi-electron transfer. Taking a different approach with another 30 nm icosahedral virus, the Finn group explored non-covalent catalyst attachment to Hepatitis B (HBV). To do this, they introduced a His₆ tag, commonly used for protein purification over immobilized nickel ions, and instead used it to bind iron protoporphyrin IX (or heme) groups from solution.¹¹⁰ Despite the bulk of both the His₆ tag and the heme, they were able to attach an average of 80 copies to each capsid. As with the ferrocene case, the attached hemes were observed to behave as independent redox entities despite their proximity.

1.5.3 The tobacco mosaic virus (TMV)

The coat protein of the Tobacco Mosaic Virus (TMV) has appeared in several reports as a means to arrange large arrays of synthetic chromophores. In contrast to the work done interfacing TMV with inorganic materials, light harvesting systems have been built from both the intact virus and the partially assembled coat proteins, either extracted from viruses or produced recombinantly. The self-assembly properties of TMV have been thoroughly studied, allowing researchers to access different metastable structures by changing the pH and ionic strength of the solution.²¹ The most appealing structures for synthetic manipulation have been double disks, consisting of two stacked rings of 17 monomers each, and helical rods, which contract slightly to a periodicity of 16.33 monomers/turn and can range from tens to thousands of nanometers in length. The subunits in both structures are tightly packed so that attaching a chromophore to the same position on each protein yields a well ordered array with interchromophore spacing well suited for energy transfer via FRET.

TMV has multiple distinct surfaces but few inherently reactive sites for covalent modification, which researchers have overcome in a number of creative ways. In the first example of using covalent chemistry to produce TMV-based materials, the Francis lab targeted native tyrosines on the exterior surface using a newly developed diazonium coupling reaction, and targeted native carboxylic acids in the pore of the rod to insert polymers, chromophores, and chelating groups.¹¹¹ Shortly following, two reports were published introducing cystine mutants to facilitate the modification of different surfaces of the TMV rod. The groups of Harris and Culver introduced a cystine on the exterior, which they were able to functionalize with chromophores.¹¹²

They also developed a clever surface immobilization strategy in which the end of the genomic RNA was revealed and hybridized to a complementary strand patterned on a gold surface. To modify the pore of TMV, the Majima group introduced a cysteine at position 99 or 100, which they were able to modify with a pyrene maleimide when the protein was disassembled into monomers.¹¹³ Rather than interfering with the assembly, the pyrene seemed to encourage it by promoting the formation of longer rods than were formed from unmodified protein. Fluorescence measurements of the pyrene-containing rods indicated the formation of excimers, suggesting the increased assembly may be due to favorable π -stacking in the pore.

The Francis group developed a system to collect light across a broad spectrum by creating a closely spaced array of mixed chromophores with spectral properties that allowed for FRET down an energy gradient, analogous to natural photosynthetic systems.¹¹⁴ Recombinantly derived TMV coat protein lacks genomic RNA, which usually winds through an internal channel in the assembled rods. Taking advantage of this newly available 'surface' for covalent modification, the Francis group introduced a cysteine into position 123 to allow the straightforward modification of monomers with a variety of commercially available maleimide dyes. Subsequent assembly into rods or disks positioned the dyes to within 1.5 nm, which is well within the Förster radius for efficient energy transfer between identical dyes as well as donor-acceptor neighbors (**Figure 8b**).

The modular nature of the system allowed for measuring the energy transfer with a variety of chromophore ratios and combinations, in both the disk and rod assemblies. Studying systems of two spectrally distinct donor and acceptor dyes provided easier deconvolution and characterization, but limited efficiency (up to 47%). Much higher efficiencies, exceeding 90%, were observed following the incorporation of a third dye to bridge the gap between the donor and acceptor of the simpler systems. This highly efficient architecture allows broad spectrum light to be collected and transferred over relatively large distances, permitting an acceptor to collect light from a maximal number of donors.

More detailed analysis of the energy transfer rates were performed using time resolved spectroscopic methods in a subsequent collaboration with the Fleming group.¹¹⁵ Additional studies performed with the Geissler group showed that energy transfer is possible not only to the nearest neighbor, but to neighbors in all directions, providing many paths for energy to travel from the light-absorbing chromophore to the final acceptor.¹¹⁶ This property is crucial to provide defect tolerance, for example, arising from the photobleaching of pigments.

In addition to the synthetic chromophores discussed above, the empty RNA channel has played host to molecules as large as porphyrins, as reported by the Majima group.¹¹⁷ By introducing a cysteine at position 127, they created a similar system in which their donor zinc porphyrins and acceptor freebase porphyrins were conjugated to TMV via a maleimide linker. This held them in close proximity to one another upon monomer assembly into short rods. One advantage of using porphyrins is their very high extinction coefficient, though the large gap between excitation and emission prohibited significant donor-to-donor FRET, and therefore long-distance energy transfer.

Having established the possibility of collecting light with TMV based system, the next challenges were converting the light into useful energy and collecting that energy on a large scale. To address the first challenge, the Francis group redesigned the TMV coat protein to introduce a new reactive site in the pore.¹¹⁸ Their ultimate goal was to place an electron transfer catalyst there to act as a final energy acceptor from the surrounding chromophores in the RNA

channel. To do this, they first cut and reassembled the coat protein gene to link the pre-existing termini on the exterior and introduce new ones in the pore—a process called circular permutation. The repositioned N-termini were then selectively converted into aldehydes using a PLP-mediated transamination reaction,¹¹⁹ providing an orthogonal reactive handle in the central channel. Donor chromophores were placed in the RNA channel as before to serve as an 'antenna', and an acceptor was coupled to the aldehydes just a few nanometers away, allowing it to collect light from the surrounding donors either in a rod or disk assembly. In the latter arrangement, the acceptor mimicked the reaction center of the LH1 complex, with the future goal of replacing it with an electron transfer catalyst. This permuted TMV further differed from known TMV coat protein by forming robust disk assemblies under a broader range of conditions, making them more promising candidates for integration into devices.

The second challenge involves interfacing TMV-based light harvesting systems with devices to collect the high energy electrons produced, either in the form of chemical bonds or as a photocurrent. For the latter system, an electrically conductive component might be required to connect light harvesting components to an electrode. Carbon nanotubes have a number of properties that would make them ideal in this role, such as high conductivity and a demonstrated ability to interface with a variety of photocatalytic materials.^{120,121} Since TMV and nanotubes do not associate on their own, the Francis group developed a method to pre-treat both TMV rods and carbon nanotubes such that they would bind together. In parallel, an exterior tyrosine of TMV was modified to display aldehydes using a diazonium coupling method,¹¹¹ while the carbon nanotubes were solubilized with a surfactant-like pyrene-PEG-aminoxy bifunctional linker, which bound the nanotubes with the hydrophobic pyrene.¹²² Upon mixing, the aminoxy and aldehydes reacted to form oximes, linking the TMV and nanotubes. In this proof of principle work, the PEG was chosen to solubilize the nanotubes despite its poor conductivity. In the future, it would need to be replaced with a more conductive polymer to transfer electrons from TMV to nanotubes.

1.5.4 Photocatalytic systems based on bacteriophage MS2

Photosynthesis uses the energy of absorbed light to form high-energy electrons, which are ultimately used to generate chemical bonds. As an alternative to TMV-based photovoltaic systems emphasized above, artificial light harvesting systems can also be designed to generate high energy electron carriers, for example reducing NADP⁺ to NADPH. The requirements for this system are essentially identical to those described above: the ability to array donor chromophores within FRET distance of both each other and a photocatalytic acceptor. The icosahedral protein shell of MS2 provides these capabilities in a distinctly different structure. In a report by the Francis group, an array of 180 donor chromophores was created by modifying a cysteine introduced to the interior of the empty capsid.¹²³ A zinc porphyrin was chosen as a suitable acceptor, and targeted to the unnatural amino acid *p*-aminophenylalanine (*paF*) installed on the exterior surface through the Schultz *in vivo* method.^{8,75} Because the protein shell is only about 1 nm thick, the separation between the interior donors and the exterior acceptors was well within FRET distance (**Figure 8c**). The large number of exterior reactive sites (180) allowed the researchers to screen a variety of catalyst loadings (20-120 zinc porphyrins) to study the energy transfer process.

The flexibility of the system also allowed for a comparison between donors designed to sensitize the porphyrin at either of two wavelengths (350 or 500 nm) where the porphyrin did not

strongly absorb. Upon illuminating the donor, the energy transfer to the zinc porphyrin could be measured by the reduction of methyl viologen, which produced a colorimetric response. The catalytic cycle was completed by addition of the sacrificial reductant, β -mercaptoethanol, to reduce the resulting porphyrin radical cations. When coupled to the chromophore array, the porphyrins were capable of photoreduction over a broader set of input wavelengths. Illumination at the peak donor absorption band produced an almost fourfold increase in the reduction of methyl viologen compared to a system with no donor.

1.5.5 Bacteriophage M13

Bacteriophage M13 is another helical virus that has been used to template arrays of dyes and photocatalysts. Its utility for light harvesting was first demonstrated by Scolaro and coworkers, who found that cationic porphyrins would spontaneously adsorb to its surface.¹²⁴ These porphyrins served as FRET acceptors for energy absorbed by a surface tryptophan introduced through mutation. However, the light harvesting efficiency was limited both by the low extinction coefficient of tryptophan and the poor overlap between donor emission and acceptor excitation.

The Belcher lab has created light harvesting catalytic systems with M13 as well, taking advantage of their extensive experience interfacing phage with inorganic particles to design complex structures. They began by attaching zinc porphyrins to the 2,700 pVIII coat proteins that run the length of the virus.¹²⁴⁵ These were introduced by targeting the N-terminus and a lysine on each monomer, providing distances of 1-2.4 nm between the reactive sites. This was close enough to allow for significant amounts of electronic coupling. Using transient pump-probe spectroscopy, exciton delocalization was indeed observed in the porphyrin system as quenched fluorescence emission and decreased lifetime. Building on this work in a subsequent report, Belcher and co-workers used phage display to develop a M13 mutant that could also bind iridium oxide (IrO_2) clusters after porphyrin attachment.¹²⁶ These clusters could then nucleate the encapsulation of the virus-porphyrin assembly with an iridium oxide shell under mild conditions. The porphyrins then served as photosensitizers, oxidizing the iridium shell, which in turn oxidized water. The integration of the porphyrins and the catalysts with the viral template was found to improve the quantum yield relative to systems with no porphyrin or no M13 template, highlighting the potential of using self-assembling protein structures to arrange nanoscale components for optimal solar energy collection and conversion.

In a final example, the Belcher group used a similar approach to develop catalysts for conversion of ethanol into hydrogen and carbon dioxide, displaying the versatility of the M13 phage for templating nanoscale catalysts.¹²⁷ Through a process of biopanning, or selected evolution, the lab identified a M13 mutant capable of binding metal ions along its length. This binding is based on extra glutamate residues in the pVIII protein as discussed in the preceding section, and can be used to nucleate the formation of CeO_2 nanowires doped with nickel and rhodium. When tested in reactions and compared to non-templated catalysts, these nanowires displayed greater stability, which was attributed to decreased phase separation, greater resistance to surface deactivation, and smaller catalyst pore size. These demonstrations of using M13 to template improved inorganic materials suggest this technique has great potential to enhance other catalytic systems.

1.5.6 Summary

The past 10 years of research has provided strong demonstrations of the utility of viral capsids as scaffolds for light harvesting and catalysis. The ability to position orthogonal reactive groups with precision allows the integration of different interacting device components, and the self-assembly of repeating units permits large arrays to be constructed with relative ease. Artificial systems mimicking various aspects of photosynthesis have been developed, and should increase in complexity as additional reactive handles are introduced to attach components such as photoprotection groups. In the future, these viral components may be incorporated into a range of devices.

1.6 Conclusion and dissertation outline

The examples discussed in this chapter clearly indicate that viral capsids (and by extension, many other self-assembling proteins) have much to offer for the construction of new materials. Although this success has been based on many different capsids that have been modified using a highly creative list of synthetic approaches, the common theme is the use of the well-defined protein structure to template nanoscale architectures that are very difficult to achieve on this size scale. These examples also highlight many of the future challenges that must be addressed before viral capsids reach their full synthetic potential.

First among these is the development of new chemical methods that can be used to attach synthetic functionality to their surfaces. Although many of the above examples have made effective use of cysteine and lysine modifications, the generation of more complex materials with increased number of functional components will require additional chemistries that can be combined with tried-and-true bioconjugation methods. Artificial amino acids provide one promising way to implement these new reactive strategies.

It should also be noted that most viral capsids exhibit a very high level of symmetry that is difficult to break in a controlled fashion. While this situation is generally satisfactory for the creation of core/shell materials for drug delivery, many future applications in energy collection and conversion will require asymmetric arrays of chromophores and photocatalysts that can direct the energy flow. Perhaps the best way to achieve this will be to design new protein monomers that can mimic the main features of viral capsid assembly, but through the use of a richer set of sequences that come together with defined positional relationships in the final structures. These approaches will certainly benefit from the availability of emerging theoretical tools that can model and predict complex self-assembly behavior.

For optical and electronic materials, completely new methods are required to interface biologically templated materials with device surfaces. Our current understanding of the ways in which proteins interact with surfaces is limited at best, making the generation of well-oriented protein arrays a very difficult task. This is coupled with the fact that proteins have inherent insulating properties, and thus will most likely require embedded metals or arrays of well-positioned metal complexes to achieve efficient electron transfer. In general, proteins have good photostability, and many can withstand a significantly wide range of temperature and pH conditions without losing their three-dimensional structure. However, even proteins extracted from thermophilic organisms will place limitations on the range of temperatures that can be used for processing and fabrication. While not insurmountable, these considerations will provide

many challenges as biomolecules are wed with inorganic materials for commercial applications.

In terms of biomedical applications, careful studies are needed to obtain a better picture of the fate of modified viral capsids that find their way into the bloodstream. While many of them are likely to elicit an immune response, it may be possible to attenuate this through shielding with polymer chains or altering the sequences of the protein monomers. In addition, there are still few if any examples of using viral capsids for the recognition of specific targets *in vivo*. While this concept is likely to be demonstrated in the coming years, many such studies will be required to understand the complex and currently non-intuitive influences of multivalency, overall charge, size, aspect ratio, and cargo loading on the biodistribution of these particles. Although drug delivery may be an end goal of these studies, *in vivo* imaging techniques are currently poised to provide much of this information.

This dissertation will detail strategies to address some of these current challenges in viral capsid based material design. It will focus heavily on the Tobacco Mosaic Virus, with a small section dedicated to the bacteriophage MS2 platform. The ultimate goal of this body of work is the production of protein-templated light harvesting systems where the structural aspects and symmetry of capsid proteins can be combined with modern chemical biology techniques to yield highly diverse and functional chromophore-protein complexes. Through both genetic and chemical methods, it is possible to integrate chromophores and sensitizers into proteins in a highly site-specific manner. Throughout this discussion, a variety of techniques will be described in the manipulation of the TMVcp scaffold for the construction of biomimetic light harvesting systems. Additionally, the development of electron transfer catalysts will be outlined with a brief description of their utility for sensitization and application towards new technologies.

Chapter 2 will describe our manipulation of the TMVcp structure. The overall experimental design will be outlined, as well as our motivation behind the strategies employed. Multiple site-specific mutations have been engineered into the TMVcp in order to generate uniquely reactive sites on the monomeric protein surface, and characterization of these protein capsids will be presented. Through computer modeling and rational design of the location of these reactive sites, capsids capable of templating chromophore arrays with range of geometries and pigment environments were produced. Additionally, the assembly properties of the TMV capsids will be assessed following mutagenesis and our observations compared to the current picture of TMV assembly states.

Chapter 3 will build on the protein systems described in detail in Chapter 2, and describe the integration of chromophores into the protein scaffold. Biomimetic light harvesting arrays were generated through the use of organic chromophores, and the structural and spectral characterization is presented. Specific aspects of the protein-pigment systems will be addressed, such as: the choice of chromophore, inter-chromophore distances/geometries, and protein-pigment interactions. The impact of these results, and the relationship to current questions in both natural and synthetic light harvesting will be addressed.

Chapter 4 no longer focuses on TMV, but rather on the synthesis of phthalocyanines, a class of macrocycles that can act as sensitizers upon photoexcitation in the red region of the visible spectrum. Efforts toward synthesis of a water-soluble and protein-reactive

phthalocyanine are described, followed by experimental observations upon conjugation of the phthalocyanine to a viral capsid. The ability of the synthesized phthalocyanines for application in singlet oxygen generation and photodynamic therapy will be emphasized, and initial observations of the singlet oxygen sensitization capabilities of the synthesized phthalocyanines will be presented.

1.7 Literature cited

1. G. M. Whitesides, J. P. Mathias, C.T. Seto. *Science* **1991**, *29*, 1312-1319.
2. H. Cui, M. J. Webber, S. I. Stupp. *Peptide Science* **2010**, *94*, 1-18.
3. K. Rajagopal, J. P. Schneider. **2004**, *14*, 480-486.
4. S. Zhang. *Nat. Biotechnol.* **2003**, *21*, 1171-1178.
5. X. Zhao, F. Pan, H. Xu, M. Yaseen, H. Shan, C. A. Hauser, S. Zhang, J. R. Lu. *Chem. Soc. Rev.* **2010**, *39*, 3480-3498.
6. N. C. Seeman. *Annu. Rev. Biochem.* **2010**, *79*, 65-87.
7. J. Nangreave, D. Han, Y. Liu, H. Yan. *Curr. Opin. Chem. Biol.* **2010**, *14*, 608-615.
8. J. Xie, P. G. Schultz. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775-782.
9. A. J. Link, M. L. Mock, D. A. Tirrell. *Curr. Opin. Biotechnol.* **2003**, *14*, 603-609.
10. M. L. Flenniken, M. Uchida, L. O. Liepold, S. Kang, M. J. Young, T. Douglas. *Curr. Top. Microbiol. Immunol.* **2009**, *327*, 71-93.
11. T. Douglas, M. Young. *Science* **2006**, *312*, 873-875.
12. D. J. Evans. *J. Mater. Chem.* **2008**, *18*, 3746-3754.
13. M. Fischlechner, E. Donath. *Angew. Chem. Int. Ed.* **2007**, *46*, 3184-3193.
14. T. Lin. *J. Mater. Chem.* **2006**, *16*, 3673-3681.
15. C. M. Niemeyer. *Angew. Chem. Int. Ed.* **2001**, *40*, 4128-4158.
16. P. Singh, M. J. Gonzalez, M. Manchester. *Drug Dev. Res.* **2006**, *67*, 23-41.
17. N. F. Steinmetz, D. J. Evans. *Org. Biomol. Chem.* **2007**, *5*, 2891-2902.
18. M. Uchida, M. Klem, M. Allen, P. Suci, M. Flenniken, E. Gillitzer, Z. Varpness, L. Liepold, M. Young, T. Douglas. *Adv. Mater.* **2007**, *19*, 1025-1042.
19. M. Young, W. Debbie, M. Uchida, T. Douglas. *Annu. Rev. Phytopathol.* **2008**, *46*, 361-384.
20. S. C. Harrison. "Principles of Virus Structure" in *Virology*, 27-44; ed. B.N. Fields *et al.*; Raven Press, New York; 1985.
21. A. Klug. *Phil. Trans. R. Soc. London, Ser. B* **1999**, *354*, 531-535.
22. P. J. G. Butler. *Phil. Trans. R. Soc. London, Ser. B* **1999**, *354*, 537-550.
23. J. Sun, C. DuFort, M. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, C. C. Kao, B. Dragnea. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 1354-1359.
24. L. Specthrie, E. Bullitt, K. Horiuchi, P. Model, M. Russel, L. Makowski. *J. Mol. Biol.* **1992**, *228*, 720-724.
25. Hermanson, G. T. *Bioconjugate Techniques*, 2nd Ed. Academic Press, San Diego, 2008.
26. Tilley, S. D., Joshi, N. S., Francis, M. B. *The Chemistry And Chemical Reactivity Of Proteins* in *The Wiley Encyclopedia of Chemical Biology*, Eds. Begley, T. Wiley-VCH: Weinheim, 2008.
27. M. Allen, D. Willits, J. Mosolf, M. Young, T. Douglas. *Adv. Mater.* **2002**, *14*, 1562-1565.
28. M. Allen, D. Willits, M. Young, T. Douglas. *Inorg. Chem.* **2003**, *42*, 6300-6305.
29. D. Ensign, M. Young, T. Douglas. *Inorg. Chem.* **2004**, *43*, 3441-3446.
30. M. L. Flenniken, D. A. Willits, S. Brumfield, M. Young, T. Douglas. *Nano Lett.* **2003**, *3*, 1573-1576.
31. Z. Varpness, J. W. Peters, M. Young, T. Douglas. *Nano Lett.* **2005**, *5*, 2306-2309.
32. M. J. Abedin, L. Liepold, P. Suci, M. Young, T. Douglas. *J. Am. Chem. Soc.* **2009**, *131*, 4346-4354.

33. P. A. Suci, Z. Varpness, E. Gillitzer, T. Douglas, M. Young. *Langmuir* **2007**, *23*, 12280-12286.
34. H. Maeda, J. Wu, T. Sawa, Y. Matsumura, K. Hori. *J. Control. Release* **2000**, *65*, 271-284.
35. G. P. Smith, V. A. Petrenko. *Chem. Rev.* **1997**, *97*, 391-410.
36. J. W. Kehoe, B. K. Kay. *Chem. Rev.* **2005**, *105*, 4056-4072.
37. I. M. Verma, M. D. Weitzman. *Annu. Rev. Biochem.* **2005**, *74*, 711-738.
38. D. V. Schaffer, J. T. Koerber, K. Lim. *Annu. Rev. Biomed. Eng.* **2008**, *10*, 169-194.
39. R. L. Garcea, L. Gissmann. *Curr. Opin. Biotechnol.* **2004**, *15*, 513-517.
40. M. J. Gonzalez, E. M. Plummer, C. S. Rae, M. Manchester. *PLoS ONE* **2009**, *4*, e7981.
41. T. Douglas, M. Young. *Nature* **1998**, *393*, 152-155.
42. E. Gillitzer, D. Willits, M. Young, T. Douglas. *Chem. Commun.* **2002**, 2390-2391.
43. C. R. Kaiser, M. L. Flenniken, E. Gillitzer, A. L. Harmsen, A. G. Harmsen, M. A. Jutila, T. Douglas, M. Young. *Int. J. Nanomed.* **2007**, *2*, 715-733.
44. M. Allen, J. W. M. Bulte, L. Liepold, G. Basu, H. A. Zywicke, J. A. Frank, M. Young, T. Douglas. *Magn. Reson. Med.* **2005**, *54*, 807-812.
45. L. Liepold, S. Anderson, D. Willits, L. Oltrogge, J. A. Frank, T. Douglas, M. Young. *Magn. Reson. Med.* **2007**, *58*, 871-879.
46. P. Caravan, J. J. Ellison, T. J. McMurry, R. B. Lauffer. *Chem. Rev.* **1999**, *99*, 2293-2352.
47. S. Hatz, L. Poulsen, P. R. Ogilby. *Photochem. Photobiol.* **2008**, *84*, 1284-1290.
48. Q. Wang, T. Lin, L. Tang, J. E. Johnson, M. G. Finn. *Angew. Chem. Int. Ed.* **2002**, *41*, 459-462.
49. Q. Wang, E. Kaltgrad, T. Lin, J. E. Johnson, M. G. Finn. *Chem. Biol.* **2002**, *9*, 805-811.
50. Q. Wang, T. Lin, J. E. Johnson, M. G. Finn. *Chem. Biol.* **2002**, *9*, 813-819.
51. K. S. Raja, Q. Wang, M. G. Finn. *ChemBioChem* **2003**, *4*, 1348-1351.
52. A. Chatterji, W. Ochoa, L. Shamieh, S. P. Salakian, S. M. Wong, G. Clinton, P. Ghosh, T. Lin, J. E. Johnson. *Bioconjugate Chem.* **2004**, *15*, 807-813.
53. Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn. *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193.
54. J. E. Moses, A. D. Moorhouse. *Chem. Soc. Rev.* **2007**, *36*, 1249-1262.
55. S. Sen Gupta, J. Kuzelka, P. Singh, W. G. Lewis, M. Manchester, M. G. Finn. *Bioconjugate Chem.* **2005**, *16*, 1572-1579.
56. D. Banerjee, A. P. Liu, N. R. Voss, S. L. Schmid, M. G. Finn. *ChemBioChem* **2010**, *11*, 1273-1279.
57. C. S. Rae, I. Wei Khor, Q. Wang, G. Destito, M. J. Gonzalez, P. Singh, D. M. Thomas, M. N. Estrada, E. Powell, M. G. Finn, M. Manchester. *Virology* **2005**, *343*, 224-235.
58. J. D. Lewis, G. Destito, A. Zijlstra, M. J. Gonzalez, J. P. Quigley, M. Manchester, H. Stuhlmann. *Nat. Med.* **2006**, *12*, 354-360.
59. D. E. Prasuhn, Jr., R. M. Yeh, A. Obenaus, M. Manchester, M. G. Finn. *Chem. Commun.* **2007**, 1269.
60. P. Singh, D. Prasuhn, R. M. Yeh, G. Destito, C. S. Rae, K. Osborn, M. Finn, M. Manchester. *J. Control. Release* **2007**, *120*, 41-50.
61. M. Manchester, P. Singh. *Adv. Drug Deliv. Rev.* **2006**, *58*, 1505-1522.
62. N. F. Steinmetz, V. Hong, E. D. Spoerke, P. Lu, K. Breitenkamp, M. G. Finn, M. Manchester. *J. Am. Chem. Soc.* **2009**, *131*, 17093-17095.
63. K. S. Raja, Q. Wang, M. J. Gonzalez, M. Manchester, J. E. Johnson, M. G. Finn.

- Biomacromolecules* **2003**, *4*, 472-476.
64. P. Singh, G. Destito, A. Schneemann, M. Manchester. *J. Nanobiotechnol.* **2006**, *4*, 2.
 65. R. A. Mastico, S. J. Talbot, P. G. Stockley. *J. Gen. Virol.* **1993**, *74*, 541-548.
 66. J. M. Hooker, E. W. Kovacs, M. B. Francis. *J. Am. Chem. Soc.* **2004**, *126*, 3718-3719.
 67. E. W. Kovacs, J. M. Hooker, D. W. Romanini, P. G. Holder, K. E. Berry, M. B. Francis. *Bioconjugate Chem.* **2007**, *18*, 1140-1147.
 68. E. A. Anderson, S. Isaacman, D. S. Peabody, E. Y. Wang, J. W. Canary, K. Kirshenbaum. *Nano Lett.* **2006**, *6*, 1160-1164.
 69. J. M. Hooker, A. Datta, M. Botta, K. N. Raymond, M. B. Francis. *Nano Lett.* **2007**, *7*, 2207-2210.
 70. J. Xu, S. J. Franklin, D. W. Whisenhunt, K. N. Raymond. *J. Am. Chem. Soc.* **1995**, *117*, 7245-7246.
 71. A. Datta, J. M. Hooker, M. Botta, M. B. Francis, S. Aime, K. N. Raymond. *J. Am. Chem. Soc.* **2008**, *130*, 2546-2552.
 72. T. Meldrum, K. L. Seim, V. S. Bajaj, K. Palaniappan, W. Wu, M. B. Francis, D. E. Wemmer, A. Pines. *J. Am. Chem. Soc.* **2010**, *132*, 5936-5937.
 73. L. Schröder, T. Lowery, C. Hilty, D. Wemmer, A. Pines. *Science* **2006**, *314*, 446-449.
 74. J. Hooker, J. O'Neil, D. Romanini, S. Taylor, M. B. Francis. *Mol. Imaging Biol.* **2008**, *10*, 182-191.
 75. Z. M. Carrico, D. W. Romanini, R. A. Mehl, M. B. Francis. *Chem. Commun.* **2008**, 1205-1207.
 76. J. M. Hooker, A. P. Esser-Kahn, M. B. Francis. *J. Am. Chem. Soc.* **2006**, *128*, 15558-15559.
 77. D. Shangguan, Y. Li, Z. Tang, Z. C. Cao, H. W. Chen, P. Mallikaratchy, K. Sefah, C. J. Yang, W. Tan. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11838-11843.
 78. G. J. Tong, S. C. Hsiao, Z. M. Carrico, M. B. Francis. *J. Am. Chem. Soc.* **2009**, *131*, 11174-11178.
 79. N. Stephanopoulos, G. J. Tong, S. C. Hsiao, M. B. Francis. *ACS Nano* **2010**, *4*, 6014-6020.
 80. W. Wu, S. Hsiao, Z. Carrico, M. Francis. *Angew. Chem. Int. Ed.* **2009**, *48*, 9493-9497.
 81. T. Douglas, E. Strable, D. Willits, A. Aitouchen, M. Libera, M. Young. *Adv. Mater.* **2002**, *14*, 415-418.
 82. S. E. Aniygyei, C. J. Kennedy, B. Stein, D. A. Willits, T. Douglas, M. Young, M. De, V. M. Rotello, D. Srisathiyarayanan, C. C. Kao, B. Dragnea. *Nano Lett.* **2009**, *9*, 393-398.
 83. C. Radloff, R. A. Vaia, J. Brunton, G. T. Bouwer, V. K. Ward. *Nano Lett.* **2005**, *5*, 1187-1191.
 84. J. C. Falkner, M. E. Turner, J. K. Bosworth, T. J. Trentler, J. E. Johnson, T. Lin, V. L. Colvin. *J. Am. Chem. Soc.* **2005**, *127*, 5274-5275.
 85. C. Chen, M. Daniel, Z. T. Quinkert, M. De, B. Stein, V. D. Bowman, P. R. Chipman, V. M. Rotello, C. C. Kao, B. Dragnea. *Nano Lett.* **2006**, *6*, 611-615.
 86. S. K. Dixit, N. L. Goicochea, M. Daniel, A. Murali, L. Bronstein, M. De, B. Stein, V. M. Rotello, C. C. Kao, B. Dragnea. *Nano Lett.* **2006**, *6*, 1993-1999.
 87. L. Loo, R. H. Guenther, V. R. Basnayake, S. A. Lommel, S. Franzen. *J. Am. Chem. Soc.* **2006**, *128*, 4502-4503.
 88. L. Loo, R. H. Guenther, S. A. Lommel, S. Franzen. *J. Am. Chem. Soc.* **2007**, *129*, 11111-11117.
 89. S. R. Whaley, D. S. English, E. L. Hu, P. F. Barbara, A. M. Belcher. *Nature* **2000**, *405*, 665-

- 668.
90. S. Lee, C. Mao, C. E. Flynn, A. M. Belcher. *Science* **2002**, *296*, 892-895.
 91. C. Mao, D. J. Solis, B. D. Reiss, S. T. Kottmann, R. Y. Sweeney, A. Hayhurst, G. Georgiou, B. Iverson, A. M. Belcher. *Science* **2004**, *303*, 213-217.
 92. K. T. Nam, D. Kim, P. J. Yoo, C. Chiang, N. Meethong, P. T. Hammond, Y. Chiang, A. M. Belcher. *Science* **2006**, *312*, 885-888.
 93. Y. J. Lee, H. Yi, W. Kim, K. Kang, D. S. Yun, M. S. Strano, G. Ceder, A. M. Belcher. *Science* **2009**, *324*, 1051-1055.
 94. K. N. Avery, J. E. Schaak, R. E. Schaak. *Chem. Mater.* **2009**, *21*, 2176-2178.
 95. W. Shenton, T. Douglas, M. Young, G. Stubbs, S. Mann. *Adv. Mater.* **1999**, *11*, 253-256.
 96. C. E. Fowler, W. Shenton, G. Stubbs, S. Mann. *Adv. Mater.* **2001**, *13*, 1266-1269.
 97. D. Kuncicky, R. Naik, O. Velez. *Small* **2006**, *2*, 1462-1466.
 98. M. Demir, M. H. B. Stowell. *Nanotechnol.* **2002**, *13*, 541-544.
 99. Z. Niu, J. Liu, L. A. Lee, M. A. Bruckman, D. Zhao, G. Koley, Q. Wang. *Nano Lett.* **2007**, *7*, 3729-3733.
 100. E. Dujardin, C. Peet, G. Stubbs, J. N. Culver, S. Mann. *Nano Lett.* **2003**, *3*, 413-417.
 101. M. Knez, A. M. Bittner, F. Boes, C. Wege, H. Jeske, E. Maiß, K. Kern. *Nano Lett.* **2003**, *3*, 1079-1082.
 102. S. Balci, A. Bittner, K. Hahn, C. Scheu, M. Knez, A. Kadri, C. Wege, H. Jeske, K. Kern. *Electrochim. Acta.* **2006**, *51*, 6251-6257.
 103. S. Lee, E. Royston, J. N. Culver, M. T. Harris. *Nanotechnol.* **2005**, *16*, S435-S441.
 104. E. Royston, A. Ghosh, P. Kofinas, M. T. Harris, J. N. Culver. *Langmuir* **2008**, *24*, 906-912.
 105. R. J. Tseng, C. Tsai, L. Ma, J. Ouyang, C. S. Ozkan, Y. Yang. *Nat. Nanotech.* **2006**, *1*, 72-77.
 106. M. Kobayashi, M. Seki, H. Tabata, Y. Watanabe, I. Yamashita. *Nano Lett.* **2010**, *10*, 773-776.
 107. N. Nelson, A. Ben-Shem. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 971.
 108. A. Freer, S. Prince, K. Sauer, M. Papiz, A. H. Lawless, G. McDermott, R. Cogdell, N. W. Isaacs. *Structure* **1996**, *4*, 449-462.
 109. N. Steinmetz, G. Lomonosoff, D. Evans. *Small* **2006**, *2*, 530-533.
 110. D. Prashun Jr., J. Kuzelka, E. Strable, A. K. Udit, S. Cho, G. C. Lander, J. D. Quispe, J. R. Diers, D. F. Bocian, C. Potter, B. Carragher, M. Finn. *Chem. Biol.* **2008**, *15*, 513-519.
 111. T. L. Schlick, Z. Ding, M. B. Francis. *J. Am. Chem. Soc.* **2005**, *127*, 3718-3723.
 112. H. Yi, S. Nisar, S. Y. Lee, M. A. Powers, W. E. Bentley, G. F. Payne, R. Ghodssi, G. W. Rubloff, M. T. Harris, J. N. Culver. *Nano Lett.* **2005**, *5*, 1931-1936.
 113. M. Endo, H. Wang, M. Fujitsuka, T. Majima. *Chem. Eur. J.* **2006**, *12*, 3735-3740.
 114. R. A. Miller, A. D. Presley, M. B. Francis. *J. Am. Chem. Soc.* **2007**, *129*, 3104-3109.
 115. Y. Ma, R. A. Miller, G. R. Fleming, M. B. Francis. *J. Phys. Chem. B.* **2008**, *112*, 6887-6892.
 116. R. A. Miller, N. Stephanopoulos, J. M. McFarland, A. S. Rosko, P. L. Geissler, M. B. Francis. *J. Am. Chem. Soc.* **2010**, *132*, 6068-6074.
 117. M. Endo, M. Fujitsuka, T. Majima. *Chem. Eur. J.* **2007**, *13*, 8660-8666.
 118. M. T. Dedeo, K. E. Duderstadt, J. M. Berger, M. B. Francis. *Nano Lett.* **2010**, *10*, 181-186.
 119. J. Gilmore, R. Scheck, A. Esser-Kahn, N. Joshi, M. Francis. *Angew. Chem. Int. Ed.* **2008**, *47*, 7788-7788.

120. D. M. Guldi, G. M. A. Rahman, F. Zerbetto, M. Prato. *Acc. Chem. Res.* **2005**, *38*, 871-878.
121. D. Tasis, N. Tagmatarchis, A. Bianco, M. Prato. *Chem. Rev.* **2006**, *106*, 1105-1136.
122. P. Holder, M. Francis. *Angew. Chem. Int. Ed.* **2007**, *46*, 4370-4373.
123. N. Stephanopoulos, Z. M. Carrico, M. B. Francis. *Angew. Chem. Int. Ed.* **2009**, *48*, 9498-9502.
124. L. M. Scolaro, M. A. Castriciano, A. Romeo, N. Micali, N. Angelini, C. Lo Passo, F. Felici. *J. Am. Chem. Soc.* **2006**, *128*, 7446-7447.
125. Y. S. Nam, T. Shin, H. Park, A. P. Magyar, K. Choi, G. Fantner, K. A. Nelson, A. M. Belcher. *J. Am. Chem. Soc.* **2010**, *132*, 1462-1463.
126. Y. S. Nam, A. P. Magyar, D. Lee, J. Kim, D. S. Yun, H. Park, T. S. Pollom, D. A. Weitz, A. M. Belcher. *Nat. Nanotech.* **2010**, *5*, 340-344.
127. B. Neltner, B. Peddie, A. Xu, W. Doenlen, K. Durand, D. S. Yun, S. Speakman, A. Peterson, A. Belcher. *ACS Nano* **2010**, *4*, 3227-3235.
128. S. Brumfield, D. Willits, L. Tang, J. E. Johnson, T. Douglas and M. Young. *J. Gen. Virol.* **2004**, *85*, 1049-1053.
129. D. J. Siler, J. Babcock, G. Bruening. *Virology* **1976**, *71*, 560-567.
130. T. M. Kozlovska, I. Cielens, D. Dreilina, A. Dislers, V. Baumanis, V. Ose and P. Pumpens. *Gene* **1993**, *137*, 133-137.
131. W. Yuan, C. R. Parrish. *Virology* **2001**, *279*, 546-557.
132. J. H. Sun, S. Adkins, G. Faurote, C. C. Kao. *Virology* **1996**, *225*, 1-12.
133. H. N. Barnhill, R. Reuther, P. L. Ferguson, T. Dreher, Q. Wang. *Bioconj. Chem.* **2007**, *18*, 852-859.
134. R. J. Webby, J. Kalkmakoff. *Virus Res.* **1999**, *59*, 179-89.
135. S. Barray, G. Devauchelle. *Arch. virol.* **1985**, *86*, 315-326.
136. A. W. Roszak, T. D. Howard, J. Southall, A. T. Gardiner, C. J. Law, N. W. Isaacs, R. J. Cogdell. *Science* **2003**, *302*, 1969.

Chapter 2

The Tobacco Mosaic Virus as a versatile nanoscale scaffold

2.1 Abstract

This chapter describes in detail our genetic manipulation of the self-assembling Tobacco Mosaic Virus coat protein (TMVcp) and the resultant protein capsids. Several changes were made to the primary sequence of TMVcp in order to create capsids containing arrangements of uniquely and/or orthogonally reactive sites. These reactive sites can then be modified with small molecule probes to produce arrays on the protein surface whose dimensions are determined both by the location within the protein framework as well as the symmetry of the self-assembled protein monomers. In this way, it is possible to create arrays with varied spatial and geometrical properties by carefully choosing the site of interest on monomeric protein. However, prior to discussing in detail our protein manipulations, it is relevant to discuss the structure of the wild-type TMV. The rich history of research into the chemical and physical nature of TMV has been ongoing since the 19th century, and the impact that TMV has had on such fields as virology, structural biology, protein self-assembly, and x-ray crystallography cannot be overstated.

2.2 The structure and assembly of TMV

2.2.1 History of TMV research

TMV is a helical rod-shaped virus composed of approximately 2130 protein subunits. These 17.5 kDa protein monomers self-assemble around a single RNA strand into a helical arrangement with $16\frac{1}{3}$ subunits per helical turn and a pitch of 2.3 nm, resulting in a capsid of 300 nm in length and 18 nm in diameter.^{1,2} The RNA is contained within an empty groove on the interior of the protein forming close contacts with the protein,³ and a 4 nm diameter pore extends along the long axis in the center of the rod. Its structure has been investigated for over a century, and has played a pivotal role in the development of analytical tools such as fiber diffraction, x-ray crystallography, and electron microscopy.^{4,5,6,7} In fact, TMV was the first virus to be imaged in the then-new transmission electron microscope.⁸ TMV was also the first virus to be crystallized (technically into paracrystals), achieved by Stanley in 1935, who also demonstrated that the crystallized material was still infectious after resuspension.⁹ This finding revolutionized the field of virology by producing the first hard evidence that virus particles behaved chemically similarly to proteins, and could be crystallized while retaining their infectivity. Subsequent groups went on to study these TMV ‘crystals’, and by the mid 1950’s, the structure of TMV was determined to be helical.^{10,11} It was around this time that the current model for generic virus structure and infectivity was developed, in part due to elegant experiments carried out by Fraenkel-Conrat and Williams¹² involving TMV that showed the origin of infectivity was indeed nucleic acid, with protein acting as a protective barrier to isolate the labile RNA from cellular degradation.

By examining the infectivity of isolated TMV RNA, TMVcp, and the intact virus, researchers saw that the RNA alone showed very little infectivity due to its lability. The isolated protein carried no infectivity. However, when reconstituted, solutions of the RNA mixed with protein regained infectivity.¹³ This observation, along with other breakthroughs at the time began to paint a picture of the general virus design we know today. Currently, there exists a high resolution (2.9 Å) fiber diffraction structure for the intact virus containing RNA,¹⁴ as well as a 2.4 Å model for a 17-fold symmetric disk structure that crystallizes from isolated TMPcp.¹⁵ The relationship between these two protein isomorphs will be discussed further when addressing the assembly process of TMV.

2.2.2 The 20S species

Early on in TMV research, an isomorphous assembly of the TMVcp lacking helical symmetry was observed. This structure appeared as small disk species under an electron microscope (EM). Analytical ultracentrifugation showed this species to sediment as 20S, corresponding to about two turns of the protein.¹⁶ The 20S aggregate possessed 17-fold symmetry rather than the $16\frac{1}{3}$ subunits per turn geometry of the TMV helix, implying that it is composed of two stacked and closed rings of 17 protein subunits. Close investigation of electron micrographs show species with clearly discrete rings of protein, lacking helical character as in **Figure 1**. These disks are also able to form elongated stacks that can be visualized with EM when subjected to high salt and alkaline conditions.¹⁷ The simplest explanation for the quaternary structure of the 20S particle would posit that it is half of the crystallized 4 disk structure (A-B ring pair from **Figure 1F**) as a polar double disk of 34 subunits. Observations by Klug and Butler suggested that under virus reassembly conditions (100 mM phosphate buffer,

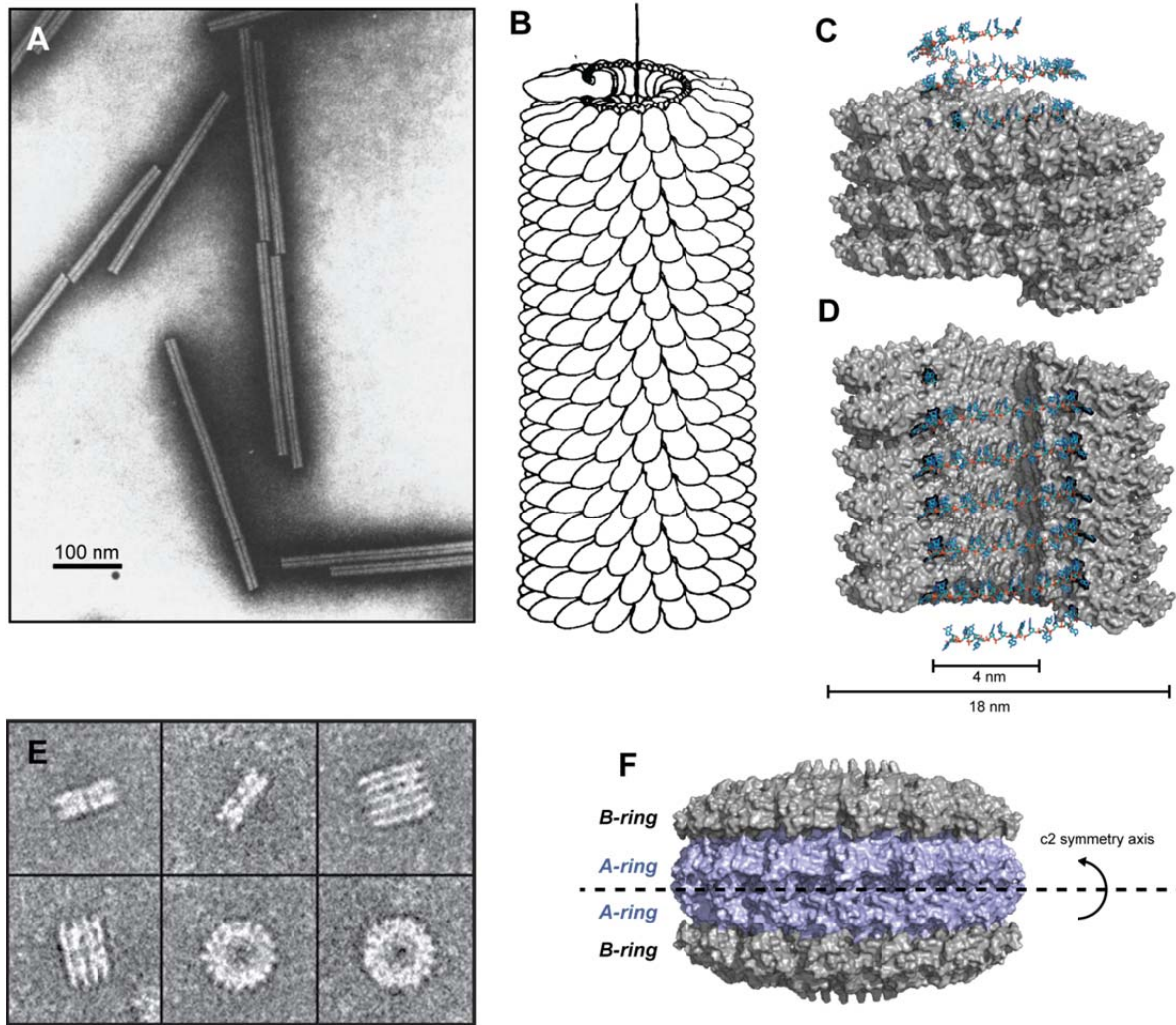


Figure 1. Structure of the Tobacco Mosaic Virus (TMV) and TMV coat protein (TMVcp). (A) Electron micrograph of the intact virus. (B) and early model for the protein assembly around RNA (Klug, 1999). High resolution crystal structures are shown (pdb ID: 2tmv) with RNA present and with (C) and without (D) the pore visible. (E) Electron micrograph images of the disk assembly of recombinantly expressed TMVcp. (F) Crystal structure of TMVcp (pdb ID: 1ei7) as a quad disk.

pH 7.3) the 20S (disk) form of the protein played a vital role in assembly. Through monitoring of the assembly kinetics, they observed that the disks both served to bind the RNA in the initial steps of assembly, as well as to act in subsequent elongation.¹⁸ The exact nature of the 20S species in solution, and its relation to the lock washer and the crystallized quadruple-layer disk, is still unclear,^{19,20} as will be discussed later.

2.2.3 Assembly of the Tobacco Mosaic Virus coat protein

Along with structural investigations, there was also the question of how the virus assembled. From the symmetry gleaned from early fiber diffraction data, it was not a far stretch to imagine that the protein subunits self-assembled around the nucleic acid. The mechanism of this assembly, however, was not initially clear. When dealing with isolated TMVcp, polymorphic aggregates of the protein are visible, as depicted in **Figure 2**. From the diagram shown, the equilibrium assembly state of TMVcp can be altered through control of the ionic strength and pH of the solution. Although there is still a slight uncertainty in the direct role of each species during the assembly process, there are a few undisputed observations that warrant stating. The TMVcp can be isolated from viral RNA, and if allowed to incubate will self-assemble without addition of any genomic material.²¹ Depending on solution conditions, the protein will adopt an equilibrium of assembly states, primarily denoted as helical, stacked disk, lock washer (20S), double disk (20S), or A protein (4S).¹

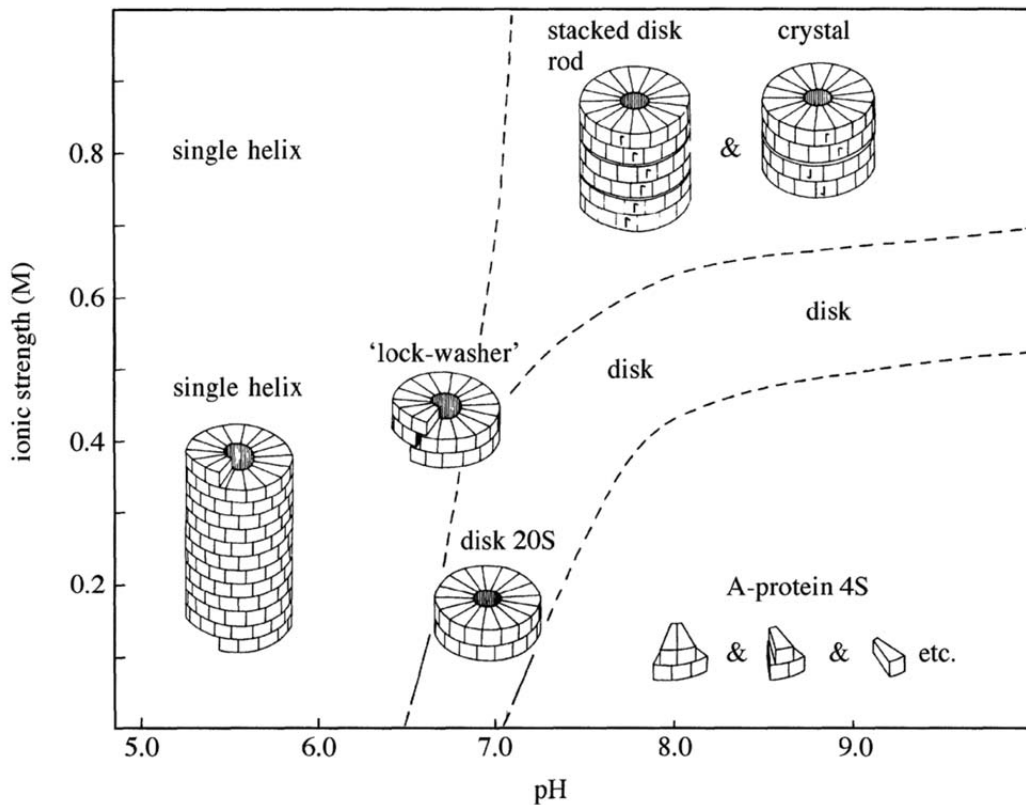


Figure 2. Diagram of TMV assembly and the dependence on pH and ionic strength. Major species are shown in the equilibrium conditions where they predominantly exist.

Without RNA to template the length of the helices, they are formed with various lengths, typically less than the 300 nm native virus. These observations came from the use of electron microscopy and analytical ultracentrifugation in the direct observation of TMV particles.²¹

Finally, under crystallization conditions (alkaline NH_4SO_4), the protein arranges into 17-fold symmetric quadruple disk structures composed of a dimer of polar double disks.¹⁵ Each 2-disk segment of this crystallized assembly is presumed to be identical to the 20S species observed in solution.

2.2.4 A detailed look at the TMVcp structure

The TMVcp monomer is composed of 158 amino acids and has a molecular weight of 17,492 Da. X-ray crystallographic structures have revealed that the protein is primarily α -helical with very small β -sheet character toward the perimeter. The protein extending into the inner pore region is a mobile loop not visualized by crystallography, and both the N- and C-termini extend outwards from the capsid. Aromatic amino acids are concentrated towards the periphery of the capsid, and form a type of ‘hydrophobic girdle’ in the high radius portion of the protein. In terms of chemically reactive residues, there are multiple Asp and Glu residues, two Lys residues, and one Cys. The cysteine is buried and only solvent accessible in conditions that compromise the secondary structure of the protein. The two lysine residues, however, are available for chemical modification and must be taken into consideration when labeling with amine-reactive probes. The high frequency and solvent accessibility of carboxylic acid residues allows the introduction of amine probes through amide formation, which has been accomplished in the past.²² Finally, we should not forget to address the one uniquely reactive site that *every* protein contains: the N-terminus. While small molecules such as PLP (pyridoxyl, 5'-phosphate) can transaminate the terminus to yield bio-orthogonal ketone or aldehyde groups,²³ it was found that although this reaction proceeds to acceptable yields, it only does so under prolonged (overnight) conditions and at elevated temperatures of 37 °C. This procedure is also somewhat sequence dependent, and the acylated N-terminus of the wild-type virus is not amenable to transamination. Alternatively, by exploiting the native N-terminal residue of serine, it is possible to oxidize the N-terminus from the primary amino alcohol to the corresponding glyoxaldehyde in a rapid fashion under mild conditions with NaIO_4 . This glyoxaldehyde moiety is unique on the protein surface, and can be selectively modified through the use of alkoxyamine reagents.²⁴

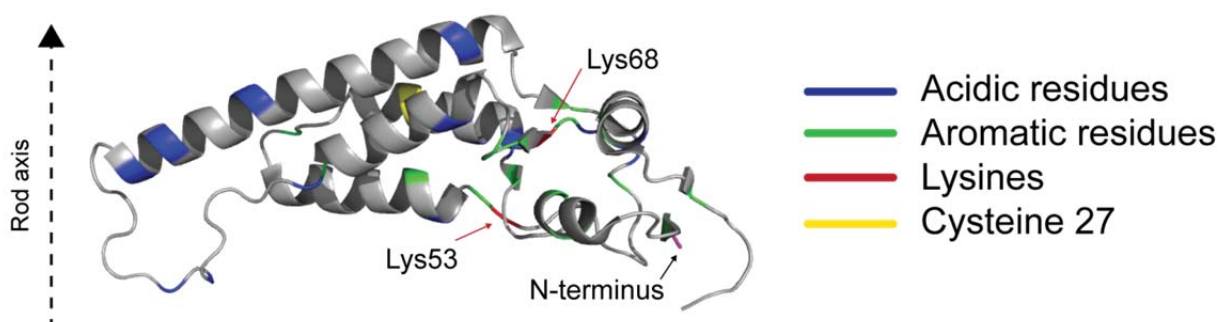


Figure 3. Structure of the TMV monomer. Viewed in the plane of the disk, from the side. Relevant native amino acids are highlighted.

2.3 TMV for conjugation of organic chromophoric arrays

2.3.1 Background and introduction

We feel that the TMV protein is an ideal candidate to act as a scaffold on which organic chromophores can be assembled into arrays due to the symmetry of TMV capsids, which arises from the self-assembling nature of the protein subunits. Through control of the TMVcp primary sequence, uniquely reactive residues can be engineered into various locations on the monomeric protein surface. This allows site-specific attachment of molecules, creating arrays with well-defined geometries and spacings based on the TMVcp quaternary symmetry and location of attachment to the protein.

As described in Chapter 1, the modification of TMVcp has been accomplished by many groups *via* a variety of strategies. While there are multiple native amino acids to target, current work in our group has focused on the construction of TMVcp with uniquely reactive non-native residues. Our strategy involved the installation of non-native lysine and cysteine residues to facilitate site-specific modification with readily available amine- and thiol-reactive probes, respectively. The first of these attempts was accomplished by the Francis group in 2007. They showed that the introduction of a cysteine at position S123 of the TMVcp allowed a solvent accessible, reactive thiol to be housed within the empty RNA groove of the recombinant protein.²⁵ Commercially available maleimide chromophores were then introduced in order to form arrays taking on the geometry of the protein assembly. The authors were interested in examining the energy transfer among chromophores in this system, so they created donor-acceptor arrays using two and three color systems capable of energy transfer. To generate these assemblies, recombinant TMVcp was initially disassembled into the 4S aggregates by taking advantage of TMV's environmental dependence on assembly state. Under low salt alkaline conditions, the protein is primarily in the 4S and 20S states, and the proteins were modified with chromophores in this state. After purifying the protein-pigment conjugates, dialysis into buffers promoting either rod or disk formation was carried out. If a single chromophore was used, then single-color arrays were produced with the desired assembly state. However, if donor-labeled protein was mixed with acceptor-labeled protein prior to dialysis and assembly, the resulting complexes contained donor:acceptor (D:A) arrays with statistically determined donor/acceptor ratios. While the overall ratio is determined by the experimenters through control of D:A stoichiometry, the precise order and location of the dyes is random. In this study, the energy transfer was monitored by analysis of fluorescence excitation spectra collected at the acceptor emission. By varying D:A ratios as well as by comparing disk assemblies to helical rods, a simple Förster model of energy transfer was produced, and efficiencies of energy transfer were calculated. Follow up work on this system investigated the faster timescales of excited state dynamics by using picosecond time-resolved fluorescence spectroscopy.²⁶ Additionally, a classical model of energy hopping within the templated arrays was created to investigate at the possible pathways and directionality of energy flow. Within the helical assembly, there can be horizontal, vertical, and diagonal transfer pathways each with different rate constants for energy transfer. By incorporating "bleached" chromophores within photo-active arrays it was possible to disrupt these pathways. A ballistic model of energy hopping between chromophores was built, and theoretical and experimental data was combined to compare the efficiency of energy transfer in various directions, as well as determine the tolerance of energy transfer to bleached sites.²⁷

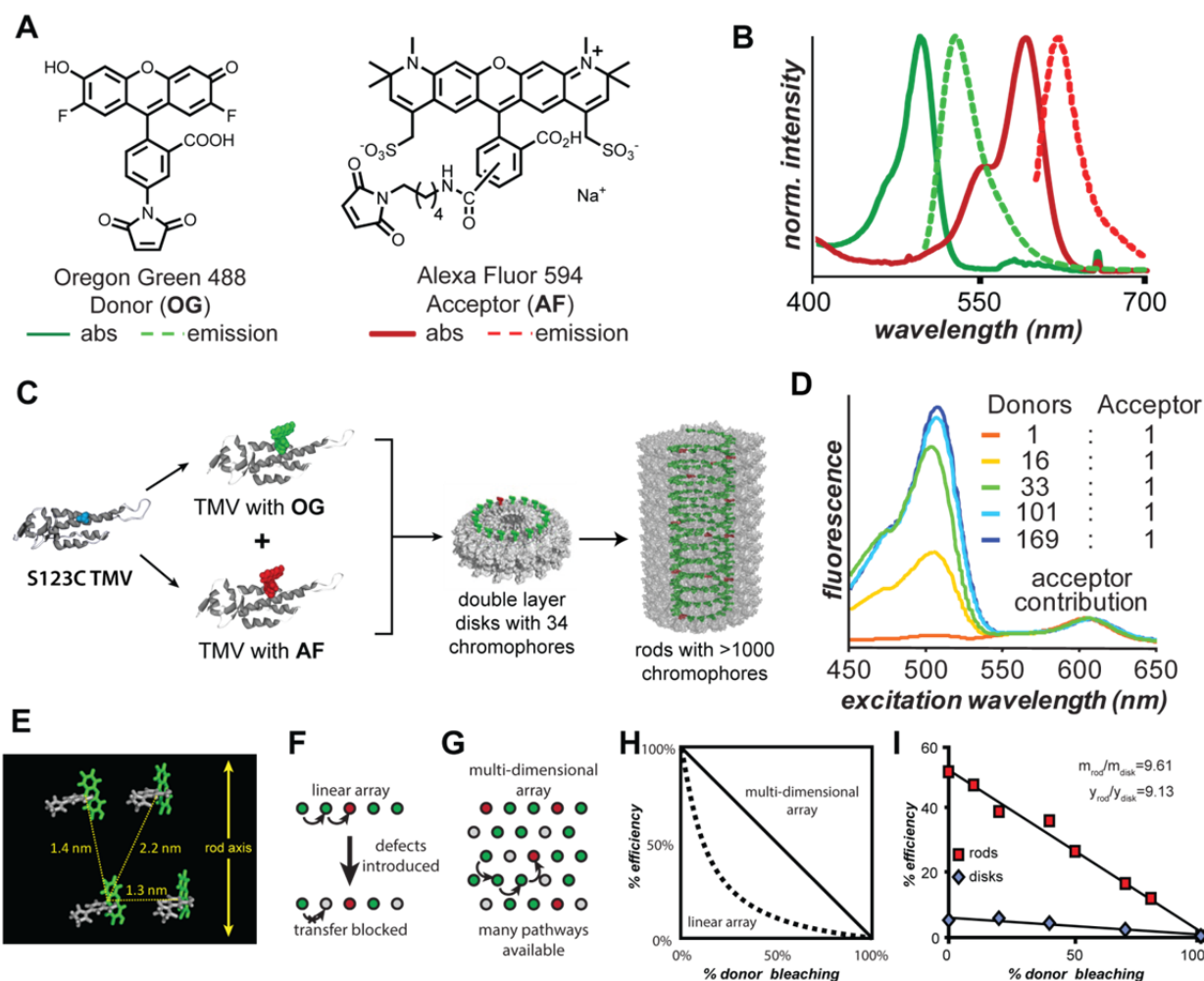


Figure 4. Self-assembling light harvesting systems based on the coat protein of the tobacco mosaic virus (TMV). (A) Chemical structures of typical donor (OG) and acceptor (AF) chromophores used in these systems. (B) Normalized absorption and emission spectra for these chromophores, showing the degree of spectral overlap. (C) S123C TMV monomers covalently labeled with either OG or AF were combined and self-assembled under disk and rod-forming conditions. (D) Excitation spectra for rod assemblies indicate light harvesting behavior through the increase in sensitivity at 500 nm with increasing donor fraction. The emission is monitored at 650 nm, and the spectra are normalized to the acceptor excitation maximum (597 nm). (E) The calculated horizontal and vertical distances are shown for the chromophores in the TMV lattice. (F) In a 1D array or a ring, defects can block transfer pathways. (G) Multidimensional systems allow redundant energy transfer pathways, and thus exhibit improved defect tolerance. (H) Mean-field models predict a linear decrease in performance for multidimensional systems, relative to 1D arrays. (I) This linear decrease was observed for TMV rods containing controlled numbers of defects, confirming that multiple energy transfer pathways are operational. (A-D) are adapted from ref. 24; and (E-F) are adapted from ref. 26.

2.3.2 Building a versatile new TMV assembly for bio-orthogonal functionalization

To construct the most versatile and highly functionalizable protein scaffold, Michel Dedeo in the Francis group constructed a series of TMVcp mutants that could be modified up to three times on each protein monomer in bio-orthogonal and chemically distinct steps. First, retention or omission of the previously described cysteine at position 123 allowed possible reactivity with thiol reactive probes. Next, N-terminal oxidation/transamination was achievable, and could be facilitated by extension of the N-terminus, presumably to enhance accessibility. Third, by converting the two native lysines to arginine residues, a single lysine was inserted at alternate locations to obtain constructs with a uniquely reactive amine within the protein. In this way, protein could be produced that is amenable to thiol modification, amine modification, and modification at the oxidized N-terminus. Depending on the location of reactive residues within the protein, arrays can be templated with varying inter-site spacings. Ultimately, three sites of reactivity were selected for the experiments in this dissertation, and will be described below.

Extending the work of Miller *et al.* described above, the first site of modification that was used frequently was the S123C site. This leaves a solvent-accessible reactive thiol within the empty RNA groove of the helical assembly, or on the top surface of the single layer disk at a radius of 5 nm from the center of the pore. These show very high reactivity, and batches of protein can be produced that have 100% modification at the S123C location. The isosteric nature of the serine to cysteine mutation imparts very minimal change to the protein structure, and does not alter the assembly state at all. Therefore, after modification, the protein can be equilibrated between the 4S, disk, and helical form.

The installation of a uniquely reactive lysine was also accomplished, but turned out to be more difficult. In order to install a single lysine residue in the TMVcp sequence, the native lysines at positions 53 and 68 must be removed through mutation. This was done, and both native lysines were changed to arginines. Although the protein expressed in high yield and remained soluble, the assembly was perturbed from that of the native lysine-containing TMVcp. This turned out to be an interesting finding. The double mutant K53R K68R TMVcp (which will be denoted as RR-TMVcp) showed an extremely strong preference for the 20S assembly state. In fact, after expression and purification, it was not possible to disassemble it into the 4S species under typical conditions, and the RR-TMVcp also would not adopt the helical conformation. The disk species persisted, and could be induced to form stacked-disk structures, but no equilibration to the helical rod was observed. Through electron microscope imaging, light scattering, and size exclusion chromatography, the RR-TMVcp disks were indistinguishable from disks formed from normal TMVcp. We interpreted these observations as the RR-TMVcp forming 20S disks typically observed at moderate pH and low ionic strength, but with increased stability that did not readily interconvert between isoforms. The origin of the alteration in assembly character can be rationalized by viewing the crystal structure of the crystallized disk. Both lysines are located near each other, and are at the direct interfaces of the A-A and A-B ring pairs. These axial interfaces have obvious importance in the self-assembly of the protein. It has been shown²⁴ that the vertical contacts between subunits are primarily polar in nature, being moderated by salt bridges, polar contacts, and tightly held water molecules bridging the interface. The introduction of the guanidinium groups of arginine therefore, not surprisingly, changes the polar contacts and affects the vertical packing, albeit to a small degree. There are

acidic sidechains very close to the ϵ -amino groups of these two lysines that could form strong interactions in the native sequence.

Regardless of the specific origin, double arginine mutant TMVcp (RR-TMVcp) forms extremely stable disks that apparently do not disassemble to the 4S species within pH 6-9. Neutral conditions of 25 mM buffer, pH 7.0 form populations of monodisperse disks. The solution properties were determined by light scattering to be 18.8 nm in accordance with the expected disk diameter. Gel filtration of the protein solutions show a single elution corresponding with the disk form, as compared with previous work in our lab. Electron micrographs confirm these measurements by showing a population of disks along with short stacks of disks. No larger species are observed in any solution-phase sizing experiments (i.e. size-exclusion chromatography or dynamic light scattering), so it is likely that these stacks of disks are either transiently formed in solution in low concentration, or are an artifact of sample preparation prior to electron microscopy. Additionally, no 4S protein is detected through these methods, opposed to native TMVcp, which contains about 20% 4S protein at moderate pH and low ionic strength.¹⁹ This assembly property of the RR-TMVcp is particularly appealing to us in that it allows discrete, monodisperse populations of seemingly hyper-stable disks to be formed with no apparent equilibrium to the other isoforms.

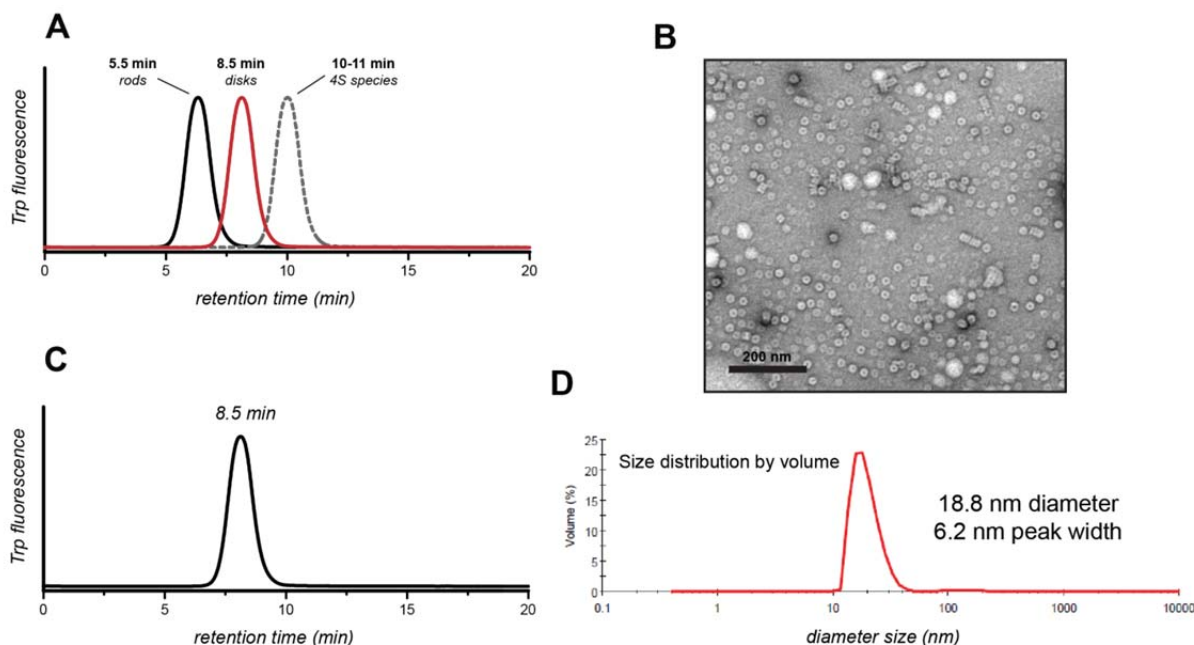


Figure 5. Assembly state of RR-TMVcp. (A) Size exclusion gel filtration chromatography (SEC) elution profile of TMVcp as a function of assembly state. (B) TEM image of RR-TMVcp after incubation at 20 mM TEA, pH 7.2 showing its preference for the disk form. (C) SEC trace of RR-TMVcp after incubation at 20 mM TEA, pH 7.2 showing no presence of rod or 4S species. (D) Confirmation of particle size by dynamic light scattering (DLS).

2.3.2 The circular permutant of TMV

Another major structural change to the TMVcp was made by Dedeo when he created a circular permutant of the protein, moving the N- and C-termini from the perimeter of the protein into the center of the pore.²⁸ By cleaving the sequence between residues 99 and 100, and connecting the native termini by inserting a spacer, a variant of the TMVcp was created that retained the overall secondary structure of the native protein. The protein crystallized under similar conditions as the native protein, and a 2.45 Å structure was solved by molecular replacement using the “A” ring of the quadruple disk structure. Interestingly, the crystals of cpTMV were formed of only two-layer c2 symmetric disks virtually identical to the A-A ring pair (see **Figure 1F**). This contrasts the crystallized native TMVcp, which is composed of a c2 symmetric dimer of the A-B ring double layer disk in **Figure 1F**. The permuted protein monomer shows minimal change in folding of the monomer, and is virtually indistinguishable from TMVcp when overlaid. This is not surprising, considering the structural changes imparted are located at peripheral positions and within the pore, where the protein is known to be quite flexible. The well-defined secondary structure of the protein at moderate radii is where most inter subunit contacts are located.

Circular permuted TMV (cpTMV) also displays a strong preference for the disk assembly. Through screening buffer conditions, the author was able to disassemble the cpTMV in a 20 mM TEA buffer, pH 8. Electron micrographs of the cpTMV disks were identical to those of native TMVcp, and analysis by gel filtration showed that they elute simultaneously with native TMVcp disks. Rod forming behavior of cpTMV mimicked that of native TMVcp, and at low pH was observed to assemble into rods 100-200 nm in length. The similarity of assembly with native TMVcp was exemplified when it was observed that mixtures of chromophore labeled (donor) native TMVcp and chromophore (acceptor) labeled cpTMV co-assembled into rods containing both protein variants when subjected to low pH. This was confirmed through FRET studies, where the rods assembled from donor and acceptor labeled proteins displayed energy transfer.

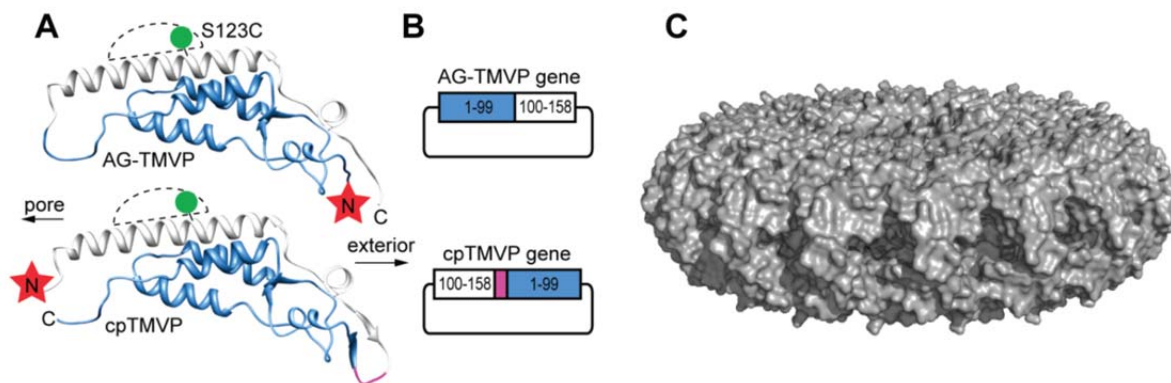


Figure 6. Structure of the circular permutant TMV (cpTMV). (A) Comparison of native TMV and cpTMV where it is clear that the change in position of the termini does not significantly alter the secondary structure. (B) Schematic of the gene construction, where the blue and white regions are switched and a pink spacer region is introduced. (C) Crystal structure of the cpTMV shows strong similarity to the apolar A-A ring pair of TMVcp.

2.4 Testing the solution assembly of RR-TMVcp

Both the circular permutant, as well as the RR-TMVcp disks described above, display a strong preference for the 20S disk structure at moderate to high pH. This property was extremely attractive to us because we were interested in examining nanoscale assemblies templated by the lower-dimensional disks as a compliment to rod-templated materials. The increased disk stability allows for monodisperse populations of protein disks at a larger range of pH values, and does not reveal equilibrium into other aggregates at moderate pH and ionic strength. This allows both modification and analysis of a stable, homogeneous population of TMV disks. However, before further elaborating on templated materials, the solution phase nature of this 20S species must be addressed. There is still, to our knowledge, a debate in the literature of the exact nature of the 20S aggregate. The last publication on the structure of the TMVcp disk was in 2000, and two opposing views emerge from the literature. What is clear and well agreed upon is that the 20S species corresponds to two layers of protein that form 17-fold symmetric disks as detailed in Section 2.2.3. Regardless of its role in assembly, this TMVcp isoform is composed of non-helical, discrete rings of protein and crystallizes as the c2 symmetric dimer of polar A-B ring dimers shown in **Figure 1**. The most logical identity of the two-layered 20S species is therefore the A-B ring pair, which exists as a polar two-layer structure in solution that dimerizes under crystallization conditions into the 4-layer structure crystallized structure. This is primarily the interpretation put forth by Klug and coworkers based on electron microscopy and reassembly experiments.¹ Alternatively, there is a conflicting view posited by Casper and coworkers suggesting that the 20S species is actually composed of the A-A ring pair. Through EM reconstruction and analysis of TMVcp in stacked-disk aggregates, they concluded that the 34-mer double disk is in fact the A-A ring pair.^{29,30,31} Further evidence by Dore and coworkers was also produced by binding antibodies specific to the 5' face of TMVcp to the disks and stacked disks and analyzing electron micrographs.³² They found that the 20S structure was apolar, and contained the 5' face on both sides of the disk structure. All of the data on both sides of the argument are indirect at best, and there is reasonable evidence for both views. Ultimately, this does not have much effect in the general shape and structure of the capsid, assembly being another story altogether, but has profound importance when using the protein as a surface to template nanoscale materials. As visualized in **Figure 1**, polar A-B pairs do not show axial symmetry, and when the protein is modified, produce sites with different microenvironments on each layer. Additionally, due to the c2 symmetry of the A-A ring pair, any modification of the protein monomers will result in two symmetric arrays with identical environments.

We felt that our RR-TMVcp was an ideal candidate to further test this discrepancy in the identity of the 20S structure. By introducing a lysine at various radii in the RR-TMVcp, we could produce disks with different but well-defined lateral distances between reactive lysines. Additionally, by placing the lysine either on the top or the bottom face of the protein, different vertical distances between lysine residues could be attained. Then, by addition of a bis-succinimidyl ester small molecule, we could test the proximity of sites by monitoring the degree of crosslinking occurring at the lysine sites. After reaction with the cross-linker, analysis by using a denaturing SDS-PAGE gel detects protein dimers in mutants where the lysine residues are spaced closer than the size of the cross-linker. Through careful choice of the position of lysine insertion, we hoped to differentiate between the A-A and A-B ring pairs in solution. The two positions we eventually decided on were the R46K and T104K mutations. The cross-linker

di-succinimidyl glutarate (DSG) was chosen, a bifunctional molecule capable of covalently cross-linking two separate nucleophilic amines. The mutated sites along with DSG are shown in **Figure 8**, where the residues of interest are highlighted within the protein framework. Comparison of the A-A ring and the A-B ring pairs show a clear difference in the geometry and relative proximity of the lysine rings.

Based on distances extracted from the crystal structure and the size of DSG, the two mutants are expected to crosslink very differently based on the assembly state. R46K RR-TMVcp is expected to crosslink in the A-A form, but not in the A-B ring conformation. The T104K construct should crosslink and dimerize in both cases.

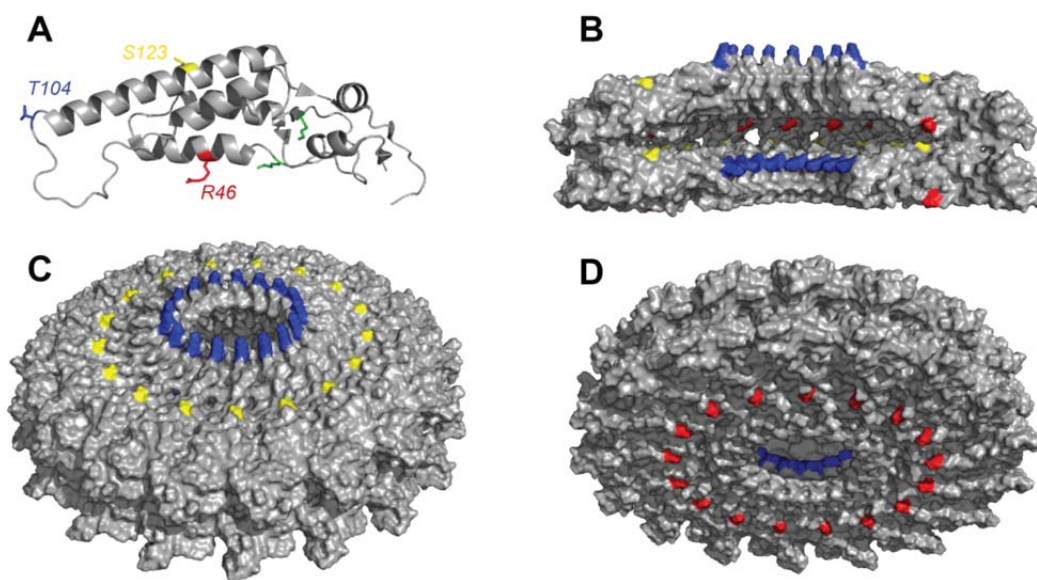


Figure 7. Structures of RR-TMVcp with reactive residues highlighted. T104(K) is colored blue; S123(C) is colored yellow; and R46(K) is colored red. (A) The protein monomer as viewed from the side. (B) The A-B ring pair with chains removed for visualization of the interior. (C) Top and (D) bottom views of the A-B ring pair with the reactive residues highlighted.

Solutions were prepared of each dilute (10 μ M) protein in phosphate buffer, pH 8.0. To each protein solution was added DSG in various amounts. The concentration of protein was kept low to minimize reactivity between different 20S species, and a range of sub-stoichiometric to excess DSG was tried. After the reaction was complete, the samples were immediately denatured and analyzed under standard SDS-PAGE conditions. Staining of the gel revealed both monomeric 17.5 kDa protein, as well as dimerized protein, in both cases, as displayed in **Figure 8**. The T104K mutants show a large degree of crosslinking, which is expected based on the close lateral proximity of the 104 positions. This site is highly reactive toward small molecular weight reagents in other cases (see chapter 3), in agreement with the current observation. No vertical cross-linking is possible at the 104 position due to the large vertical distance between the 104

positions in both A-A and A-B disks. The R46K mutant displayed a small, but relevant level of crosslinking with DSG. Lateral distances between the 46 positions on different subunits are separated by 19 Å, so no cross-linking is expected to be possible in that direction. This measurement was taken from the crystal structure, so it is possible to shorten the inter-lysine distance if two adjacent K46 sidechains are oriented toward each other, but this will still not bring them close enough for the 7.7 Å needed for DSG crosslinking. Also, in the A-B ring pair, the vertical distance between position 46 sites is much too large for DSG to span. However, in the A-A ring pair, the two 46K residues are in close proximity, and orient themselves into an ideal geometry for reaction with DSG and cross-linking. Based on these results, we can only come to the conclusion that an active solution species of our RR-TMVcp represents the A-A ring pair, and not the A-B ring pair. This is the only reasonable explanation for the cross-linking observed in both T104K RR-TMVcp as well as R46K RR-TMVcp.

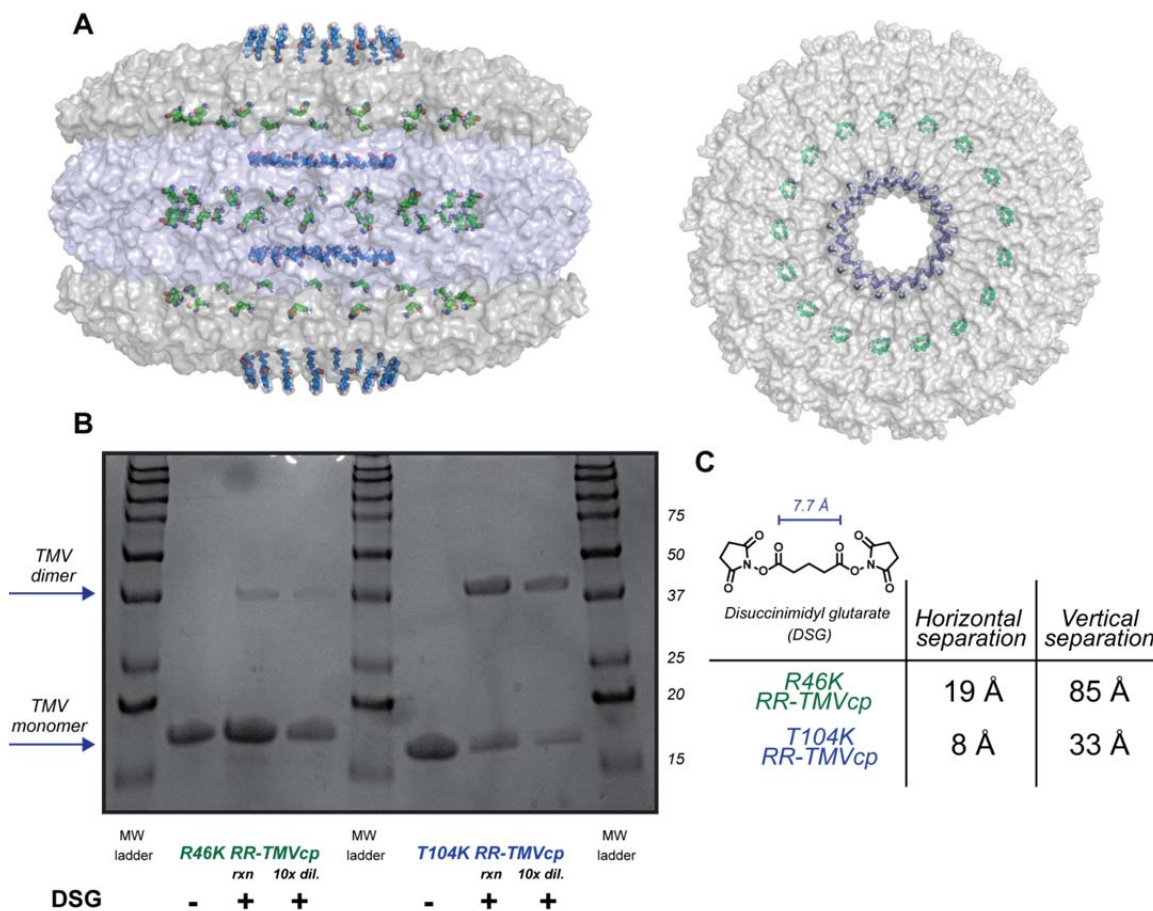


Figure 8. Overview of crosslinking experiment. (A) Side-on and top view of the crystallized quadruple TMVcp disk with R46K (green) and T104K (blue) highlighted. (B) SDS-PAGE gel of the results from crosslinking experiments. Dimer formation is observed at 35 kDa for both the R46K and T104K, suggesting a solution state of A-A disks. Each protein's three lanes contain non-crosslinked protein, the reaction mixture directly loaded, and a 10x dilution for viewing. (C) Structure and length of the crosslinker disuccinimidyl glutarate (DSG) and distances between adjacent lysine sidechains in the RR-TMVcp mutants.

It should be noted that it is possible that the double RR mutation affects the inter-subunit packing of the protein capsid, and the 20S species in RR-TMVcp is inherently different from that of native TMVcp. The introduction of the two Arg sidechains could stabilize interactions at the A-A interface and produce destabilizing interactions at the A-B ring interface, leading to the observed results. There is no way of determining this currently, but in either case, the RR-TMVcp represents a new and extremely useful protein scaffold. The c_2 symmetry of the RR disks dictates that each site on the monomeric protein will have identical environments to the corresponding sites on other monomers. This property will be heavily utilized in the next chapter when describing scaffolds for light harvesting systems.

2.5 Conclusion

This chapter discussed the detailed atomic structure and assembly of the TMVcp, and introduced novel changes to the protein structure that have been accomplished by previous work in our lab. The permutant form of TMV was described, and both its structure and assembly properties addressed. The RR-TMVcp constructs developed previously in our lab were also introduced, and their unique assembly properties and affinity for the disk form were rationalized. Building on the RR-TMVcp framework, different sites of modification were introduced through lysine residues, and the solution phase structure analyzed by observing small molecule cross-linking within the disk assemblies. These data corroborate the view of the A-A ring pair as the identity of the 20S species, correlating to a departure from the classic 20S picture put forth by Klug. In the next chapter, both the RR-TMVcp as well as the cpTMV will be described as scaffolds on which to assemble organic light harvesting systems. The c_2 symmetry of their structures will be exploited to yield highly tunable and well-defined light harvesting arrays templated by these protein assemblies.

2.6 Materials and methods

General methods

Unless otherwise noted, all chemicals and solvents were of analytical grade and used as received from commercial sources without further distillation or purification. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F₂₅₄ plates with visualization by ultraviolet (UV) irradiation at 254 nm and/or potassium permanganate stain. Purifications by flash chromatography were performed using EM silica gel 60 (230-400 mesh). The eluting system for each purification was determined by TLC analysis. All organic solvents were removed under reduced pressure using a rotary evaporator. Water (dd-H₂O) used as reaction solvent was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Spin concentration steps were performed using 100 kDa molecular weight cutoff spin concentrators from Millipore (Billerica, MA). Centrifugations were performed with an Eppendorf 5415 benchtop centrifuge, (Eppendorf, Hauppauge, NY).

Instrumentation and sample analysis

Spectroscopy. Protein concentration was routinely determined by UV/Vis analysis on a Nanodrop 1000 instrument (Nanodrop, USA) by monitoring absorbance at 280 nm.

Mass Spectrometry. Protein bioconjugates were analyzed using an Agilent 1200 series liquid chromatograph (Agilent Technologies, USA) that was connected in-line with an Agilent 6224 Time-of-Flight (TOF) LC/MS system equipped with a Turbospray ion source. Spectra were quantified by peak area integration using Origin 8.0 software (Originlab, USA).

High Performance Liquid Chromatography. HPLC was performed on Agilent 1100 Series HPLC Systems (Agilent, USA). Sample analysis for all HPLC experiments was achieved with an in-line diode array detector (DAD) and in-line fluorescence detector (FLD). Analytical reverse-phase HPLC of small molecules was accomplished using a C18 stationary phase and a H₂O/MeCN with 0.1% TFA gradient mobile phase. Size exclusion chromatography (SEC) was performed using a Phenomenex polysep GFC-P-5000 column (4.6 x 250 mm) (Phenomenex, USA) at 1.0 mL/min using a mobile phase of 10 mM phosphate buffer, pH 7.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus (Bio-Rad, Hercules, CA), using a 10-20% precast linear gradient polyacrylamide gel (Bio-Rad). The sample and electrode buffers were prepared according to Laemmli.³³ All protein electrophoresis samples were heated for 5-10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of disulfide bonds. Gels were run for 70 min at 120 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-

250 (Bio-Rad). Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA). ImageJ was used to determine the level of modification by optical densitometry.

Dynamic light scattering. Dynamic light scattering (DLS) measurements were obtained using a Malvern Instruments Zetasizer Nano ZS. Data plots and standard deviations are calculated from an average of three measurements, each of which consists of 10 runs of 45 seconds each. Measurement data are presented as a volume distributions.

Protein expression and modification

Construction of TMVcp expression plasmids. The starting point for the RR TMV mutants was a gene for the coat protein of the TMV U1 strain optimized for the codon usage of *E. coli* (Genscript, Piscataway, NJ). This sequence contained the mutation S123C. Site-directed mutagenesis was performed using QuikChange mutagenesis (Stratagene, Santa Clara, CA). Development and optimization for the mutations described here are attributed to Michel Dedeo, a former graduate student in the Francis group.

Buffer A: 20 mM TEA, pH 7.2

Buffer B: 10 mM phosphate, pH 7.0

Protein expression and purification. Tuner DE3pLysS competent cells (Novagen) were transformed with the vector described above, and cultured in Terrific Broth with 100 µg/L ampicillin. When cultures reached an optical density of 0.6 to 0.8, IPTG was added to a final concentration of 30 µM. Cultures were grown 24 h at 30 °C, harvested by centrifugation, and stored at -80 °C. Cells (from a 1 L expression batch) were thawed, resuspended in 20 mL of 20 mM TEA pH 7.2 containing 10 µg/mL DNase and RNase, and lysed by sonicating with a 2 s on, 6 s off cycle for a total of 20 minutes using a standard disruptor horn at 90% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared by ultracentrifugation for 45 min at 11,000 rpm using a SLA-600TC rotor in a Sorvall RC5C Plus centrifuge (Waltham, MA). The clarified lysate was decanted, warmed to rt, and stirred while adding a saturated solution of ammonium sulfate dropwise to a final concentration of 50% (v/v). After 5 min, the white ppt that formed was pelleted by ultracentrifugation (45 min at 11,000 rpm using a SLA-600TC rotor), washed with deionized water, and resuspended in 5-10 mL buffer A. The resulting protein solution was next loaded onto a DEAE column equilibrated with buffer A and eluted with a 0 – 300 mM NaCl gradient (in buffer A). Purity was confirmed by SDS-PAGE and HPLC/ESI-MS. Pure fractions were pooled, and fractions containing desired TMV in addition to impurities were re-run in identical DEAE conditions. This preparation provided pure RR-TMVcp in yields up to 100 mg/L culture. The resulting RR-TMVcp was only observed to be in the disk form under all laboratory conditions (pH 6-9). Aliquots of 10 mg/mL RR-TMVcp were flash frozen and stored at -20° C for further use.

Crosslinking of RR-TMVcp experimental procedure. 0.6 mL Eppendorf tubes were charged with either T104K RR-TMVcp or R46K RR-TMVcp (100 µL, 10 µM) in 10 mM phosphate, pH

8.0. To the protein solutions were added 1 μL of 100x solutions of DSG in DMF to obtain a final DSG concentration of 20 μM . The reactions were allowed to incubate at room temperature 1 h, at which time they were heated to 95 $^{\circ}\text{C}$ in the presence of DTT to facilitate protein denaturation, and run on an SDS-PAGE gel for analysis.

2.7 Literature cited

1. Klug, A. *Philos. Trans. R...* **1999**, 354, 531–535.
2. Butler, P. J. *Gen. Virol.* **1984**, 253–279.
3. Namba, K., Stubbs, G. *Science* **1986**, 231, 1401–1406.
4. Wyckoff, R. W. G., Corey, R. B. *J. Biol. Chem.* **1936**, 16, 51-55.
5. Holmes, K. C. *J. Struct. Biol.* **2010**, 170, 184–191.
6. Rossmann, M. *Q. Rev. Biophys.* **2013**.
7. Stanley, W. M., Anderson, Thomas F. *J. Biol. Chem.* **1942**, 146, 25-30.
8. Kausche, G.A., Ruska, H. *Biochemische Zeitschrift* **1939**, 303, 221-230.
9. Stanley, W. M. *Science* **1935**, 81, 644-645.
10. Watson, J. D. *Biochimica et biophysica acta* **1954**, 13, 10-19.
11. Franklin, Rosalind E. *Nature* **1955**, 175, 379-381.
12. Fraenkel-Conrat, H., Williams, R.C. *PNAS* **1955**, 41, 690-69.
13. Butler, P.J.G. Assembly of Regular Viruses. In *International Review of Biochemistry, Chemistry of Macromolecules IIB*; Offord, R.E., Ed.; University Park Press: Baltimore, 1977; Vol. 25; p 205-237.
14. Namba, K., Pattanayek, R., Stubbs, G. *J. Mol. Biol.* **1989**, 208, 307-325.
15. Bhyravbhatla, B., Watowich, S.J., Caspar, D.L. *Biophys. J.* **1998**, 74, 604-615.
16. Durham, A., Finch, J.T., Klug, A. *Nature* **1971**, 229, 37-42.
17. Díaz-Avalos, R., Caspar, D. L. *Biophys. J.* **1998**, 74, 595–603.
18. Butler, P. J. G. *Cold Spring Harb. Symp. Quant. Biol.* **1972**, 36, 461–468.
19. Butler, P. J. G., Bloomer, A.C., Finch, J. T. *J. Mol. Biol.* **1992**, 224, 381–394.
20. Schuster, T. M. *Biochemistry* **1985**, 3298.
21. Durham, A. C. H. *J. Mol. Biol.* **1972**, 67, 289-305.
22. Schlick, T. L., Ding, Z., Kovacs, E. W., Francis, M. B. *J. Am. Chem. Soc.* **2005**, 127, 3718–3723.
23. Gilmore, J.M., Scheck, R.A., Esser-Kahn, A.P., Joshi, N.S., Francis, M.B. *Angew. Chem., Int. Ed.* **2006**, 45, 5307-5311.
24. Stephanopoulos, N., Francis, M.B. *Nat. Chem. Biol.* **2011**, 7, 876-84.
25. Miller, R. A., Presley, A. D., Francis, M. B. *J. Am. Chem. Soc.* **2007**, 129, 3104–3109.
26. Ma, Y. Z., Miller, R. A., Fleming, G. R., Francis, M. B. *J. Phys. Chem. B* **2008**, 112, 6887–6892.
27. Miller, R. A., Stephanopoulos, N., McFarland, J. M., Rosko, A. S., Geissler, P. L., Francis, M. B. *J. Am. Chem. Soc.* **2010**, 132, 6068–6074.
28. Dedeo, M. T., Duderstadt, K. E., Berger, J. M., Francis, M. B. *Nano Lett.* **2010**, 10, 181–186.
29. Díaz-Avalos, R., Caspar, D. L. *Biophys. J.* **1998**, 74, 595–603.
30. Raghavendra, K., Salunke, D. M., Caspar, D. L., Schuster, T. M. *Biochemistry* **1986**, 25, 6276–6279.
31. Raghavendra, K., Kelly, J. A., Khairallah, L., Schuster, T. M. *Biochemistry* **1988**, 27, 7583–7588.
32. Dore, I., Ruhlmann, C., Oudet, P., Cahoon, M., Caspar, D. L., Van Regenmortel, M. H. *Virology* **1990**, 176, 25–29.
33. Laemmli, U. K. *Science* **1970**, 227, 680–685.

Chapter 3

Building an artificial photosynthetic analog

3.1 Abstract

This chapter will briefly review the structures of natural photosynthetic machinery and the general principles that allow efficient solar light collection. The discussion will then lead to how these principles and structural aspects can guide the rational design of synthetic LH materials. In particular, the focus will be given to artificial analogs of natural protein-templated chromophoric systems that can be created by using the tobacco mosaic virus coat protein (TMVcp) as a protein scaffold. The construction of artificial LH systems templated on the RR-TMVcp scaffold will be discussed, and the considerations of chromophore identity, relative chromophore geometry, inter-chromophore distances and couplings, and protein-dye linkages will be addressed. Spectral and physical characterization of the constructed TMVcp templated light harvesting structures will also be presented.

3.2 Natural light harvesting systems – structure and design principles

Over the course of roughly 3 billion years, life has evolved the capacity to harness the power of the sun in amazing ways, adapting to specific environments using a variety of strategies.¹ From bacteria and algae, to higher plants, the range of complex structures in photosynthetic machinery is simply astounding.^{2,3,4,5,6,7} Despite the enormous diversity in natural light harvesting (LH) systems, there are several common themes shared between all photosynthetic organisms. The vast majority of photosynthetic machinery is composed of rigidly held organic chromophores embedded within a protein matrix. All organisms contain antenna systems whose sole purpose is to absorb sunlight, and ultimately transfer this energy to a reaction center where a charge separation event occurs.⁸ This charge separation at the reaction center is then used to reduce NADP^+ to NADPH and the harvested solar energy stored as chemical energy. It is of interest both for academic purposes as well as with direct application to new technologies to elucidate and understand the physical processes and subtle elements of these natural light harvesting structures that allow such efficient collection of light in virtually all of Earth's environments.⁹

Nearly all of the photosynthetic machinery on Earth is composed of pigments embedded within the interior of proteins with well-defined geometries and massively complex energy collection and funneling pathways.^{10,11} Fundamentally, there are three parameters that dictate the overall LH capacity and efficiency of a system: The identity and spectral profile of the individual chromophores, how these chromophores are arranged spatially and geometrically relative to each other, and the local chromophoric microenvironments. These three parameters will be briefly discussed in the context of photosynthetic light collection and energy transfer. Additionally, strategies for controlling and testing these within synthetic systems will be introduced.

Chlorophylls play a critical role in all photosynthetic reaction centers, as well as being the primary light collectors in antennas. A wide variety of chlorophylls have evolved to be produced in different organisms over complex biosynthetic pathways.^{12,13} The other major light collection class of molecules, phycobilins, are linear tetrapyrroles that have much in common with chlorophylls.¹⁴ Slight changes in the structure of these pigments manifest in large spectral changes, energy transfer dynamics, and ultimately charge separation/electron transfer.¹⁵ To harvest solar energy with maximum efficiency, the majority of the solar spectrum is used by integrating different color pigments. Adaptation to particular environments drives the diversity in pigment colors, with the need to cover as much of the incident solar spectrum as possible. Also, due to different solar spectra seen in various niche environments, antennas that maximally absorb the available wavelengths of light are crucial. Examples of this is the photosynthetic algae that thrive over 1000 meters below the surface of the Red Sea. They have evolved a unique antenna system called the *chlorosome* that absorbs the few long wavelength photons that manage to make it to those depths.¹⁶

While the spectral profiles of the individual chromophores are clearly vital to the photosynthetic process, another equally important factor is the relative geometry and ordering of the individual pigments. Chromophores within photosynthetic machinery are held in fixed and controlled positions that facilitate both maximal absorption and optimum energy transfer to localize the excitation at a reaction center for conversion to chemical potential. Depending on the particular organism and type of photosynthetic structure used, this can be a wide variety of geometries, but for illustration we will introduce one of the more well-known light harvesting

system – LHII from purple bacteria. The crystal structure of LHII is shown in **Figure 1**, along with the internal chromophoric components. This is the antenna system of the purple bacteria, *Rhodospseudomonas acidophila*, and its role is to capture sunlight and transfer it to its reaction center, LHI.¹⁷ Within LHII are housed two different rings of bacteriochlorophyll *a*, known as the B800 and B850 rings. They are named after their primary absorption, with B800 having a λ_{max} of 800 nm, and B850 a λ_{max} of 850 nm.¹⁸

Both B800 and B850 are composed of the same pigment, and the spectral difference between them is attributed to the different geometries and inter-pigment coupling within each ring. B800 contains a single chromophore per protein monomer in the 8- or 9-fold symmetric LHII, and are oriented with the chlorophyll macrocycle in the same plane as the membrane. They have interchromophore distances of *ca* 2 nm and form a flat ring of chromophores aligned within the same plane. On the other hand, B850 contains two chromophores per protein monomer, and are oriented vertically with respect to the membrane containing the protein assembly. This produces a ring of chromophores with much closer spacing of $\sim 5 \text{ \AA}$ and increased inter-chromophore coupling.¹⁹

The physical origin of the B800 and B850 spectral signatures has been the subject of much research,^{20,21} and non-classical descriptions of light collection and energy transfer have been implemented to explain this phenomenon.^{22,23,24,25} One aspect that is particularly important is the delocalization of excitation energy within the chromophore rings. Through both experimental and theoretical methods, it has been shown that pigments within these arrays do not act as individual chromophores, but rather collectively as a coupled system.²⁶ The arrays display strong excitonic character, which influences the energetics and timescale of energy collection and transfer. Without going into further detail (this subject has been heavily reviewed²⁷), we will emphasize the effect on inter-chromophore coupling and geometrical relationships on spectral properties and energy transfer of LH systems.

Of the three important parameters mentioned above, the least studied is the effect of the local protein microenvironment on light harvesting. This is primarily due to the experimental difficulty in achieving this. The interior of proteins is typically considered to be hydrophobic, and an average dielectric is commonly employed to model the interior environment of proteins. Due to the complex interactions between amino acids, it is difficult to find accurate descriptions of the microenvironments that buried chromophores experience.²⁸ Natural systems are also primarily large membrane-bound protein-pigment complexes, and notoriously hard to express and study. The finely tuned structures do not lend themselves to easy manipulation while still retaining native folding and stability. Nonetheless, there have been some attempts to elucidate these questions.^{29,30,31}

There is a need to understand further the underlying physical basis for the astounding light collection efficiency observed in nature. Due to the complicated structures, in addition to direct experimental probing of natural systems, it is often necessary to create simpler synthetic models in order to probe specific aspects of the LH process directly. There have been many synthetic structures produced to study the individual processes of light collection and energy conversion, including but certainly not limited to: organometallic charge transfer molecules,^{32,33} porphyrin arrays,³⁴ chromophores linked in soluble donor-acceptor systems,^{35,36} large aggregates of organic chromophores,³⁷ polymeric based solid state materials,³⁸ and quantum dot based materials.³⁹ Indeed, there has been a wealth of data gained from studying these structures.

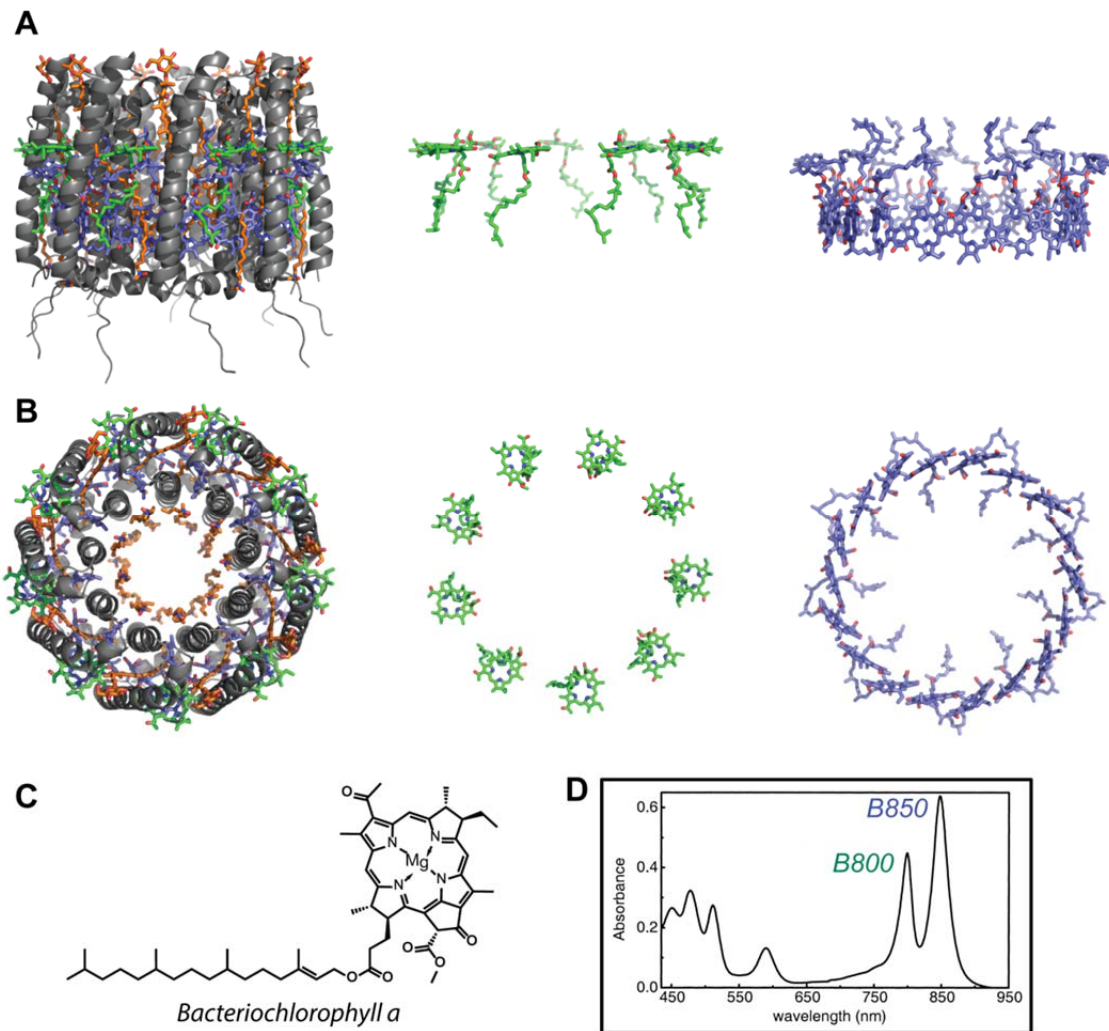


Figure 1. (A) Side view of LHI from purple bacteria viewed in the plane of the membrane. Shown with protein and all chromophores (B800 = green; B850 = blue; carotenoids = orange). Also shown are the isolated B800 and B850 rings in the same perspective. (B) Top view of structures shown in A. (C) Structure of bacteriochlorophyll *a*. (D) Absorption spectrum of LHI showing the distinct B800 and B850 absorptions (Taken from ref. 19).

However, an often overlooked and crucial aspect of photosynthesis is how the protein environment influences the processes. A handful of attempts have created some extremely elegant biomimetic light harvesting systems^{40,41,42} producing enlightening insight into the physical processes of photosynthetic light harvesting, but the need for protein templated biomimetic light harvesting systems is still present.⁴³

Previous work in our lab has shown that a protein-templated light harvesting system can be constructed by using the TMVcp as a scaffold, where organic chromophores are arranged into a regular array through covalent attachment to the self-assembling protein monomer. This was described in Chapter 2, and will be expanded upon in the following sections with the description of new protein-pigment light harvesting systems templated by RR-TMVcp.

3.3 TMV templated light harvesting systems – structure and design principles

The underlying motivation for the work described herein is the need for creating biomimetic protein-templated chromophore arrays. Simple models are needed in order to fully understand and characterize the intricate workings of natural photosynthetic systems. In order to create a protein-templated biomimetic light harvesting system, the RR-TMVcp was chosen as a primary scaffold for multiple reasons. It represents a self-assembling system capable of templating chromophores into well-defined arrays, while retaining protein-pigment interactions reminiscent of natural systems. Also, the stability of the disk isoform of RR-TMVcp allows discrete populations of protein-dye structures to be studied without artifacts stemming from equilibria to other TMV assemblies like helical rods or 4S species. The lower-dimensionality of the disks relative to the helical form is also appealing in that it allows a simpler model system to investigate energy transfer and/or delocalization within the chromophores arrays.

Experimentally, the identity of chromophores introduced is easy to control, and we are only limited to molecules that are water-soluble and carry a functional group capable of bioconjugation. Depending on the desired optical properties of the system, various chromophores can be incorporated and compared. Protein-chromophore interactions are not so easily tuned, but can be rationally engineered with knowledge of the protein microenvironment surrounding the pigment as well as attachment chemistry. It should be noted that in our synthetic LH systems, all protein-chromophore attachments are covalent, and not directly analogous to natural systems. Finally, the chromophore-chromophore interactions can be tuned through control of site-specific introduction of the chromophores into the protein. Based on the location of attachment within the protein monomer, different inter-chromophore distances and geometries are attained.

Based on the RR-TMVcp scaffold, we set about creating templated 17-fold symmetrical circular arrays of organic chromophores. The general scheme is shown in **Figure 2**, and depicts the locations within the protein that we chose to modify. There were two particularly interesting sites on the RR-TMVcp monomer that were ultimately chosen to compare, which are the 104 and 123 sites. Modification at the 123 site results in a chromophore being inserted on the top face of the monomer and within the empty RNA groove when the protein is in the helical formation. In the disk structure, the 123 position lies between the A and B rings, resulting in two solvent-accessible rings on the outer surfaces of the c2 symmetric A-A ring pair discussed in Chapter 2. This position places the chromophores within ~1.5 nm of each other, allowing facile energy transfer within the arrays. Inter-chromophore distances can vary slightly depending on shape and size of the actual chromophore, but as measured from the cysteine residue at the site of attachment is 1.8 nm. The 104 position of RR-TMVcp, on the other hand, resides on the inner loop section of the monomer corresponding to the pore region. This particular position was not resolved in the published crystal structures, but is near enough to the rigid α -helical secondary structure that with modeling can be accurately pictured. The 104 position is near the top surface of the monomer, and therefore will produce an extremely close packed array lining the pore of the disk. These positions are much closer than the 123 positions, and when modified with chromophores, produce aggregates that are essentially in Van der Waals contact due to the close proximities of attaching residues. These two sites of attachment were our primary systems of study, however, other sites of attachment on the RR-TMVcp will be addressed as they arise.

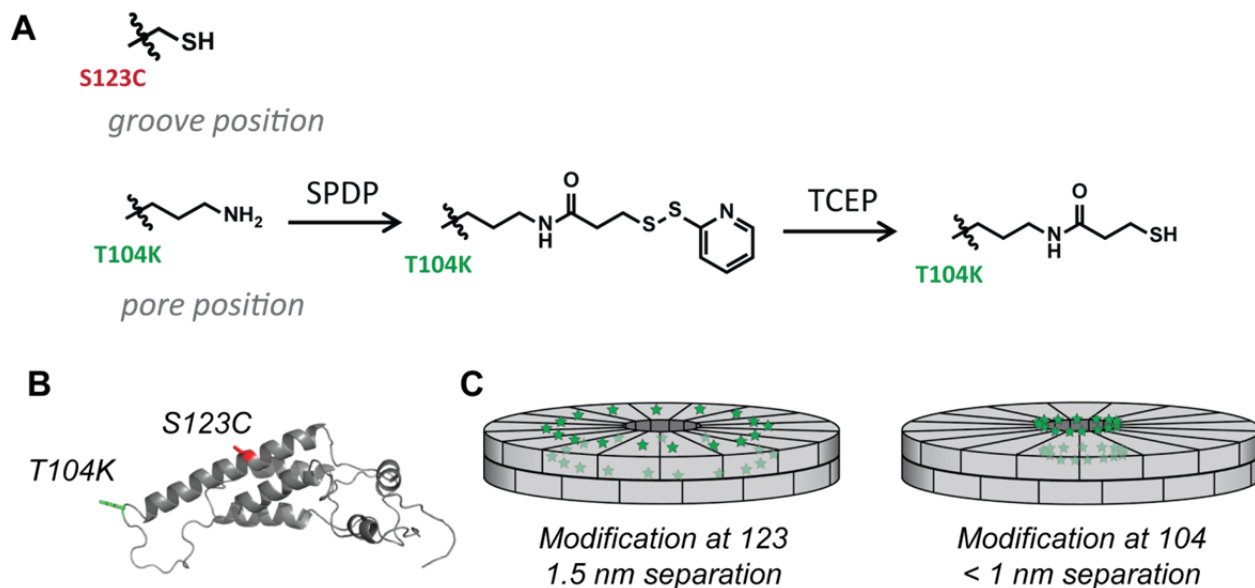


Figure 2. Synthetic scheme for dual modification of the RR-TMVcp monomer. (A) Installation of reactive thiol is accomplished through mutation at S123C and installation of the linker SPDP at the T104K position. These can then undergo subsequent reaction with maleimide reagents. (B) Structure of RR-TMVcp and location of S123 and T104K within the protein monomer. (C) Cartoon depiction of RR-TMVcp disks modified at either the 123 or 104 positions.

At the 104 position, the reactive amino acid chosen for chromophore attachment was lysine. In the RR-TMVcp, this is the only lysine contained in the primary sequence, and therefore is the only nucleophilic amino group on the protein, provided one is careful not to force modification at the N-terminus. This can be accomplished by control of solution conditions (buffer, pH) or oxidation/transamination prior to chemical modification. The 123 position was mutated to a cysteine, continuing the work set forth by Miller *et al.* and previously described.⁴² This allows standard bioconjugation techniques of thiol modification, e.g. maleimide or iodoacetamide to be accomplished orthogonally to the K104 modification with activated esters. Through mild reaction conditions, it is possible to label 100% of the C123 and K104 positions without affecting the secondary structure of the protein or modifying the buried native cysteine.

With this strategy, we set about creating complementary arrays that were identical in all aspects except the position of the chromophores on the protein, and therefore their spatial relationship with each other. The K104 modified protein produces highly coupled arrays that are essentially molecular aggregates of the chromophore embedded around the pore of the disk. Modification at the C123 position produces arrays with >1 nm spacings that can participate in energy transfer between neighbors, but remain as isolated molecules with no physical interaction or contact. By comparing spectral signatures of these two arrays, we hope to gain insight into the coupling between chromophores and how this manifests itself in the resulting optical properties. Using chromophores with different shapes, sizes, and dipolar orientations, the effects of inter-chromophore coupling on the arrays can be further interrogated. Selection of chromophore was based on several considerations. Commercially available fluorophores were commonly used due to the abundance of readily available dyes containing protein-reactive

functional groups available. We also settled on dyes that would be desirable for energy transfer studies, choosing dyes that possess spectral overlap with each other for dual color experiments. Four chromophores were chosen – pyrene, coumarin 343, Oregon Green 488, and Alexa Fluor 594. Additionally, control over the attachment chemistry and linkage can add another parameter to tune the LH systems.

We created a variety of systems with chromophores installed at either the 104 or 123 position, as well as dual color systems with FRET pairs either contained in one ring or on separate positions. Spectral characterization of our LH systems was carried out with steady state methods, as well as time-resolved techniques. UV/Vis, fluorescence, fluorescence anisotropy, and time-resolve fluorescence spectra were collected in order to visualize energy absorption and transfer within the arrays. Linker chemistry between the dye and protein was varied in both length and composition to mediate the protein-dye interactions. In general, our method involved comparing spectra of full arrays of chromophores, a single chromophore attached to a single protein disk, and the isolated chromophore free in solution.

3.4 Construction and spectral analysis of RR-TMVcp templated LH systems

3.4.1 Choice of chromophore

In the majority of experiments, chromophores employed were based on 3 common classes of fluorophore. These are: pyrene, coumarin, and xanthene based dyes. They were chosen for their spectral characteristics, availability of protein-reactive derivatives, and size/shape. Each class of dye has particular advantages and disadvantages in the context of light harvesting, and were used depending on the particular experiment with all of these in mind. Pyrene was used to test the physical proximity of adjacent dyes by exploiting its propensity to form excimers. Coumarin 343, Oregon Green 488 (OG488), Alexa Fluor 594 (AF594) were fluorophores used in spectral analysis and spectroscopic measurements of the LH systems. These three dyes were picked for their spectral overlap with each another, and to include a diversity of structural shapes and dipolar character.

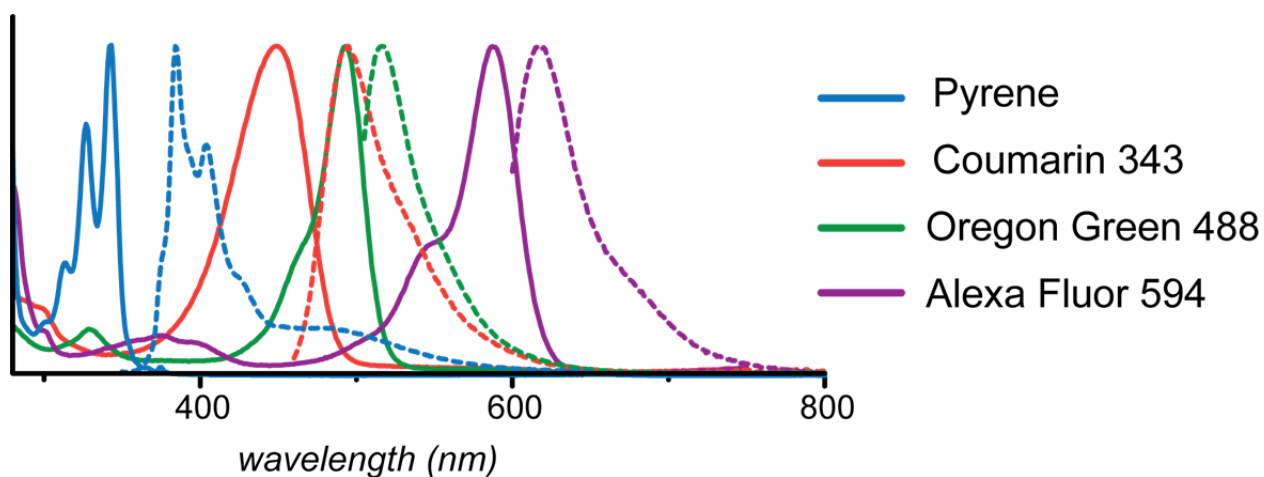


Figure 3. Spectral characteristics of commonly used fluorophores. Solid lines correspond to absorption spectra, and dotted lines fluorescence emission. All spectra normalized at peak wavelengths.

3.4.2 Modification at the C123 position

Installation of a chromophore is rapid in S123C RR-TMVcp. The C123 residue is the only solvent-accessible cysteine and readily reacts with maleimide functionalized chromophores. The protein can be modified to 100% with all chromophores we investigated. Typical solution conditions involved stirring the protein at high concentrations of 0.1 to 0.5 mM (in monomer) with maleimide-dye in 5-fold excess. The reactions were allowed to proceed at room temperature for 2 h, allowing complete consumption of maleimide reactive groups. If needed, to ensure complete modification of protein, the solutions were passed through a G-25 gel permeation column to remove unreacted dye, and the protein subjected to another 5-fold excess of chromophore. Unless otherwise noted, this procedure will be used for all cysteine modifications on S123C RR-TMVcp or T104K S123C RR-TMVcp. Consistent with the picture of an A-A pair active species in solution, these modifications proceeded smoothly and completely, implying an exposed C123 residue. If the A-B ring pair persisted in solution, the modification would be expected to occur much slower for the more buried C123 residues, and partially modified disks would be expected to be observed frequently, especially for larger chromophores such as Alexa Fluor 594.

3.4.3 Modification at the K104 position

Installation of chromophores at the 104 position of T104K RR-TMVcp or T104K S123C RR-TMVcp is also relatively straightforward, though the lysine residue is not as reactive as the cysteine positioned at 123 to molecules as large as visible chromophores. This stems simply from steric effects of the lysine side chains pointing directly into the middle of the pore region, presumably limiting accessibility of the reactive primary amines. Direct addition of chromophore linked activated esters yields partial modification of the reactive sites, but not full conversion. The reactivity of the K104 sidechain appears to depend heavily on the size of small molecule reagent. For larger N-hydroxysuccinimide functionalized chromophores, such as Texas Red NHS, reactivity is limited. No more than 10% of the monomers are modified under standard buffered conditions of 50 mM phosphate, pH 7.2 (data not shown). However, the installation of a smaller molecule can afford high levels of modification at the K104 site. Succinimidyl 3-(2-pyridylthio)propionate (SPDP) is small enough to access all of the K104 sidechains, and can modify 100% of them, though the reaction is a bit more sluggish, and extra equivalents of SPDP must be used to push the modification to full conversion. Typically, to a solution of 100 μ M protein in 50 mM phosphate buffer, pH 7.2 would be added 7 equivalents of SPDP, dissolved in DMF (keeping the organic solvent below 5% v/v). This reaction was allowed to proceed for 2 h, and another 7 equivalents of SPDP would be introduced. Often, this would yield nearly, but not completely, full conversion of the K104 sites, so a third and final round of SPDP would be added if needed. Once the SPDP is bound to the protein, the disulfide linkage can be reduced with tris(2-carboxyethyl)phosphine (TCEP), removing the pyridylthio group, and affording a free thiol for modification (See **Figure 2**). Subsequent reaction with maleimide chromophores in a similar fashion as described for modification of C123 RR-TMVcp can then afford covalent pigment attachment at the K104 site with full conversion of all the protein monomers. This procedure was used for all modifications at the 104 position unless otherwise noted.

3.4.4 Modification with pyrene maleimide

Pyrene is a small, rigid, hydrophobic fluorophore consisting of four fused benzene rings. It has a molar extinction coefficient of $34,700 \text{ M}^{-1} \text{ cm}^{-1}$ (in water)⁴⁴, and absorbs in the blue region of the visible spectrum. Emission from monomeric pyrene occurs as a set of structured bands within the 300 nm-350 nm range.⁴⁵ Pyrene is very well known to form excimers under conditions of high concentration or within aggregates, and these excimeric species emit characteristic fluorescence with a very wide emission profile centered at 475 nm. In order for excimers to form, two pyrene molecules must come into Van der Waals contact and interact electronically. This property was exploited to compare the degree of coupling between templated pyrene molecules and to visualize if any physical contact between adjacent chromophores occurs within RR-TMVcp light harvesting systems.

We created arrays of pyrene maleimide located at either the K104 or C123 positions of RR-TMVcp. **Figure 4** shows the fluorescence emission spectra associated with these two systems where excimer emission is observed in the K104 case, but not when the pyrene is located at C123. This was expected based on measurements from the crystal structure of the protein. Molecular modeling of the two corresponding structures shows that in the 123 position, the pyrene molecules are too far separated to physically be in contact and form excimers. Modification at the 104 position, on the other hand, places the pyrenes so close to each other that excimers should be possible to form. The presence of excimer emission at 104, but not 123, verifies the ability to generate arrays with controllable inter-chromophore distances which correlate to predicted behavior based on molecular modeling.

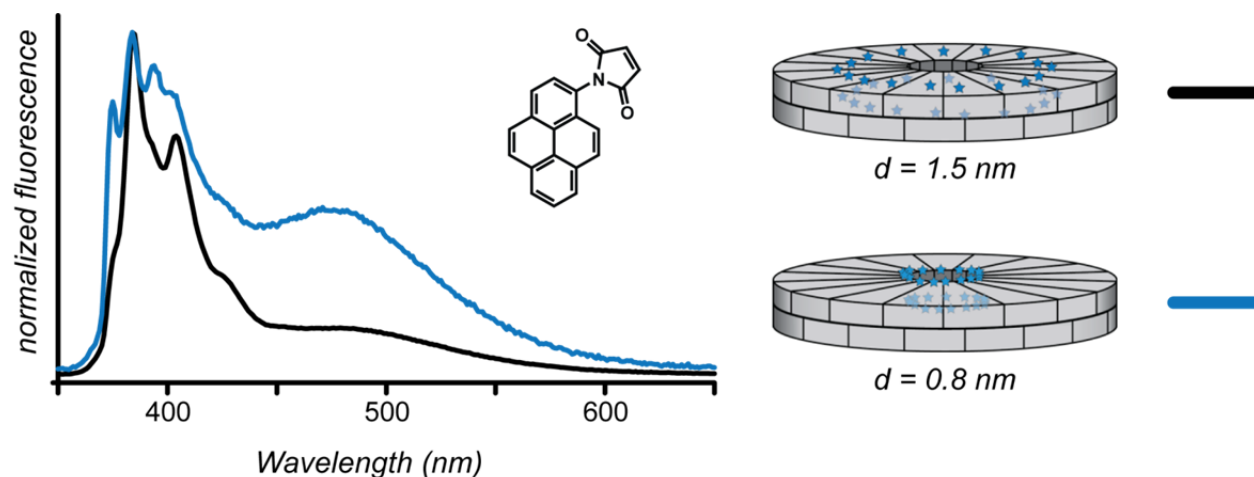


Figure 4. Pyrene excimer formation within RR-TMVcp. (A) Fluorescence emission spectra from RR-TMVcp modified with pyrene-maleimide (structure within inset) at either the S123C position or T104K. (B) Cartoon depictions of the assemblies and interchromophoric distances calculated from molecular modeling.

3.4.5 Modification with coumarin 343

After confirming that the different sites within the protein produced arrays with controllable distances and aggregation states, we moved to another fluorophore, coumarin 343. The structure of coumarin 343 in **Figure 5**, and its spectrum is shown in **Figure 4**. This chromophore was chosen primarily for its relatively small structure and simple electronic character. The coumarin class of chromophores has been extensively studied, and there is a lot of information available regarding the optical and electronic nature of these dyes. For examining the arrays from a modeling or theoretical standpoint, these dyes are ideal in that the transition dipole is well-defined, the major electronic transitions are known, and they contain small number of atoms.⁴⁶ Coumarin 343 in particular was used for to its spectral location within the visible region and ability to participate in energy transfer with other typical fluorophores. It also shows large solvchromatic character⁴⁷ with the maximum wavelength of both absorption and emission strongly dependent on the dielectric constant of its environment. Finally, it contains a free carboxylic acid functional group, which can be easily converted into protein reactive moieties – in particular maleimide groups.

Production of maleimide functionalized coumarin 343 was accomplished through a convergent approach by alkylating coumarin 343 with an amino-maleimide small molecule. The synthetic scheme is shown in **Figure 5**, and coumarin 343-maleimide product can be obtained in good yields over a 4-step route. By first starting with ethylenediamine, selective boc-protection of one amine is carried out, followed by alkylation of the free amine with maleic acid. Subsequent closing of the maleimide ring produces **1**, which can be stored as the solid boc-protected species. Deprotection provides a molecule containing the maleimide group for bioconjugation as well as a free amine for coupling to coumarin 343 **2**. This is then dissolved in DMF and stored as a concentrated solution at -20 °C for subsequent bioconjugation. In addition to the advantage of a simple 4-step process of this synthesis, many other bis-amino small molecules can be used in the place of ethylenediamine to vary the linkage of the coumarin to the protein both in composition and in length.

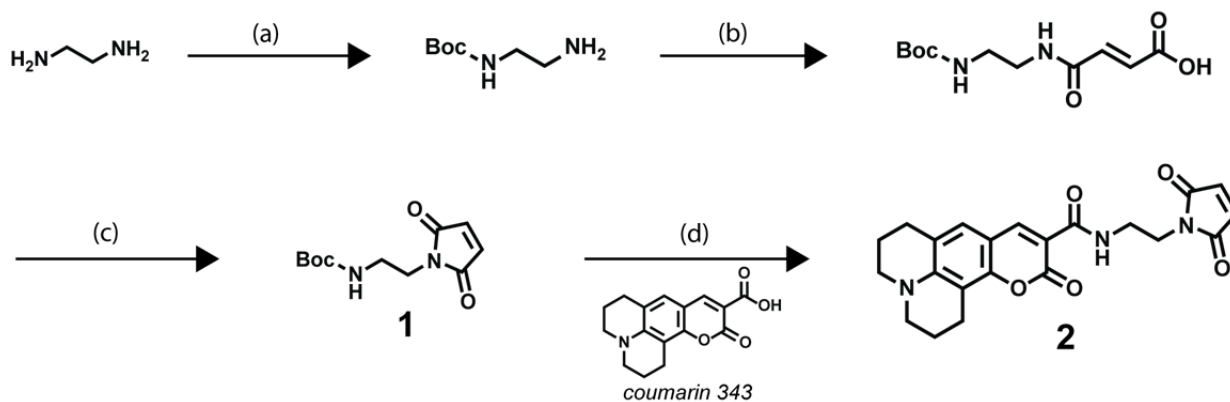


Figure 5. Synthesis of a coumarin 343 derivative for bioconjugation. (a) Di-*t*-butyl carbonate, DCM, 0 °C to rt, 5 h. (b) Maleic anhydride, MeCN, 95 °C, 5 h. (c) Acetic anhydride, NaHCO₃, 90 °C, 1.5 h. (d) 1.) 1:1 TFA:DCM, rt, 20 min. 2.) Coumarin 343, **1**, HCTU, DiPEA, DMF, rt, 2 h.

The ethyl linkage was the most commonly employed version of coumarin 343 used, and **2** was installed at both the 104 and 123 positions in the same fashion as described above for pyrene-maleimide. For the 123 site, the coumarin-maleimide was directly coupled to the cysteine residue, and at the 104 site the linker SPDP was used to install a thiol to which the coumarin-maleimide was covalently bound. UV/Vis spectra for the arrays are displayed in **Figure 6**, overlaid with the free coumarin spectrum. Also included are data for singly-modified disks, where the protein is labeled in such a low concentration that statistically each protein disk contains one or zero chromophores. Comparing all four systems (free dye, singly modified disks, full arrays at the 104 site, and full arrays at the 123 site) separates the protein-dye interaction from the chromophore-chromophore interactions.

Free coumarin 343 has an absorption maxima of 447 nm in the buffered aqueous solution used in all protein measurements. Prior to spectral analysis of free coumarin, the maleimide ring was reacted with β -mercaptoethanol to remove any electronic effects that the aromatic maleimide could impart onto the coumarin fluorophore. This spectrum will be used as a baseline with which to compare the singly modified disks and full protein templated arrays to determine protein-chromophore and chromophore-chromophore effects. Upon binding to the protein, the singly modified disks showed a bathochromic shift of 8 nm. The lineshape also changed, producing an absorption that appeared to split into two overlapping bands, with the higher energy peak appearing as a shoulder upon the lower energy peak centered at 453 nm. The arrays templated at the 104 position, on the other hand, showed a 16 nm blue shift relative to the free dye. This geometry also produced an absorption peak with lower symmetry than free coumarin 343, but now with a shoulder on the red edge of the maximum peak at 431 nm. Interestingly, the systems templated at 123 showed the same absorption maximum as the free dye, but with a broadened blue edge of the absorption profile.

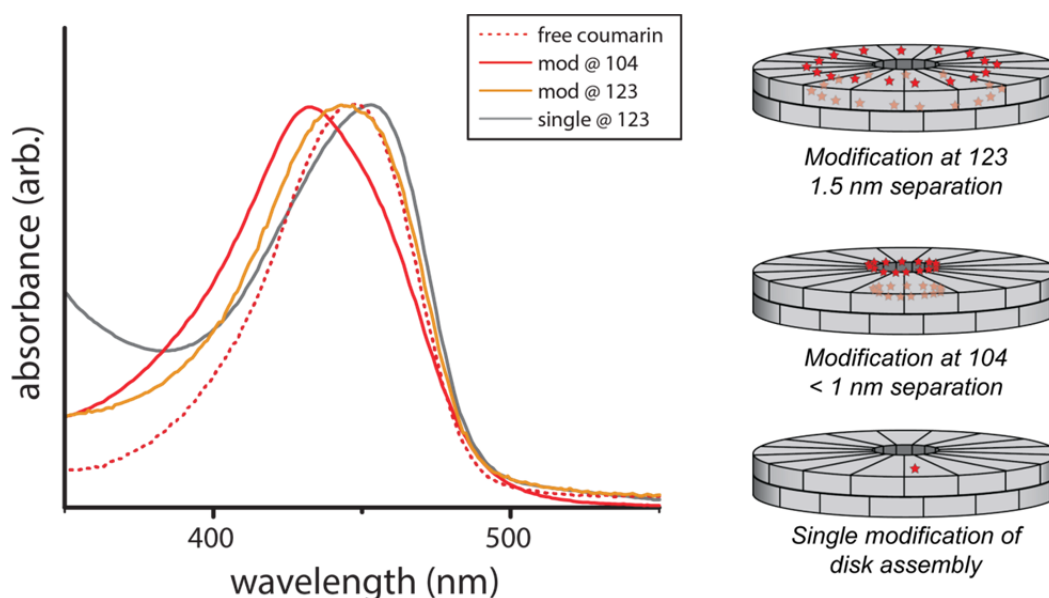


Figure 6. UV/Vis analysis of RR-TMVcp templated coumarin 343 arrays. Full modification at C123 and K104 are compared to spectra taken of singly modified disks and the free chromophore.

3.4.6 Modification of RR-TMVcp with xanthene dyes – OG488 and AF594

Similarly to coumarin 343, arrays were constructed at the 104 and 123 positions using OG488 and AF594 and compared to singly modified disks and free chromophore. These two xanthene based dyes were chosen for multiple reasons. First, they will complement the coumarin data to determine what effect the composition of the chromophore has upon the spectral signatures of the arrays. They are two common fluorophores used in protein conjugation for other purposes, and previous work, especially in the rod vs. disk experiments carried out by Miller have used them in similar contexts. Finally, more pragmatic considerations such as commercial availability of maleimide-functionalized variants and the engineered solubility and resistance to photobleaching are important reasons these chromophores were utilized.

Figure 7 shows the relevant UV/Vis spectra for Oregon Green 488 modified RR-TMVcp disks. Spectral shifts are more subtle than in the case of coumarin 343, but there are still trends that appear. Free Oregon Green 488 maleimide (after reacting with β -mercaptoethanol) has an absorption maximum of 491 nm, and the singly modified disks show a red shifted absorbance with maxima at 502 nm. This is an 11 nm bathochromic shift, which is also observed in the arrays set at the 123 position. The more coupled arrays at the K104 site actually show less spectral shifting, with an absorption maximum at 496 nm, corresponding to a 6 nm redshift.

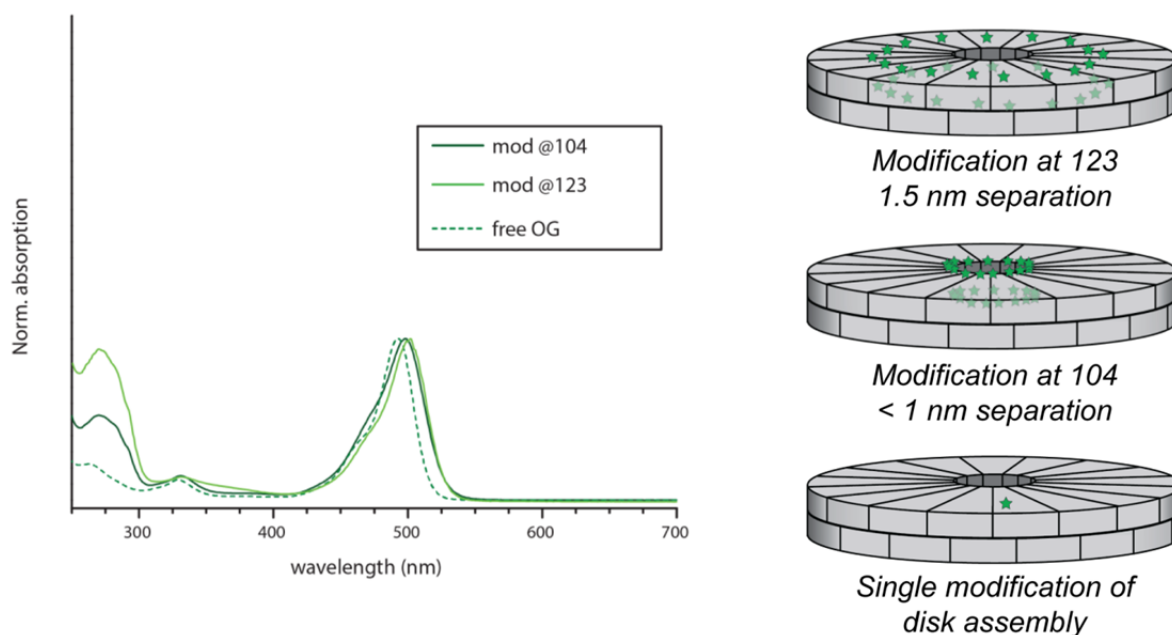


Figure 7. UV/Vis analysis of RR-TMVcp templated Oregon Green 488 arrays. Full modification at C123 and K104 are compared to spectra taken of singly modified disks and the free chromophore.

In the case of Alexa Fluor 594 maleimide, similar absorption effects are observed to those of Oregon Green 488 maleimide. Free Alexa Fluor 594 maleimide (after reacting with β -mercaptoethanol) in buffered solution absorbs with a maximum at 588 nm. The singly modified disks show red-shifted absorptions (relative to the free dye) of 6 nm. The C123 templated arrays

also show this same 6 nm redshift, while the K104 arrays similarly display a red-shifted absorption profile, but to a lesser degree, with a 4 nm shift. It is worth noting that in the case of Alexa Fluor 594 maleimide we always observe an increase in intensity of the vibronic shoulder centered at 549 nm as the interchromophore distance is decreased. The lineshape does not significantly change, but simply the intensity of the vibronic peak relative to the main absorption band. This effect was not observed in the case of Oregon Green 488, but was observed when the same modifications were performed with Texas Red c2 maleimide, which has a very large structural similarity to Alexa Fluor 594 c5 maleimide (data not shown).

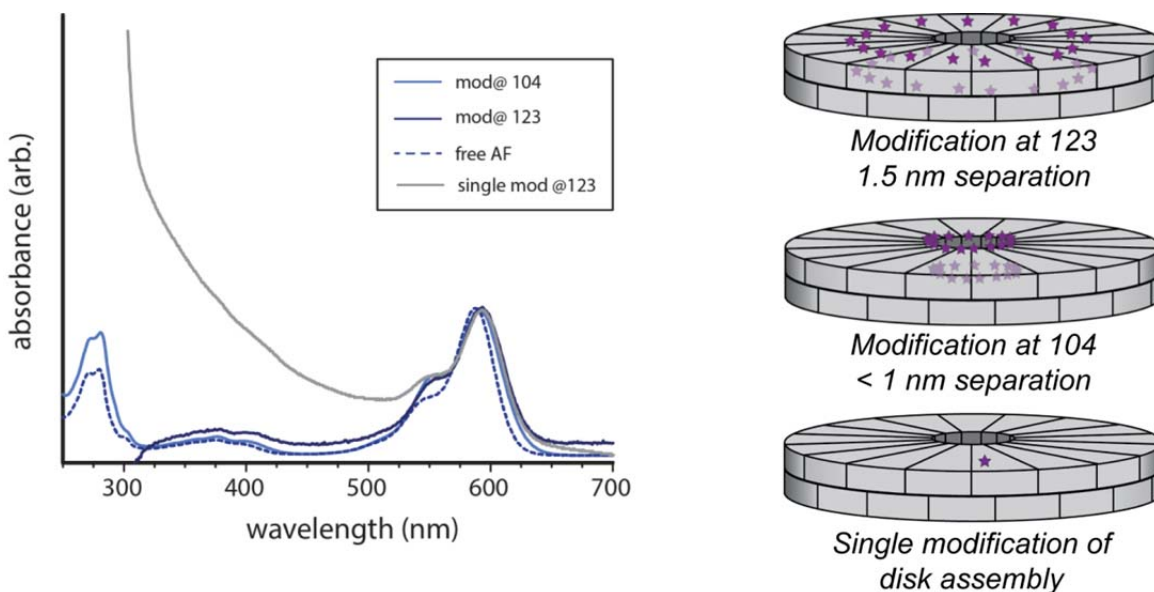


Figure 8. UV/Vis analysis of RR-TMVcp templated Alexa Fluor 594 arrays. Full modification at C123 and K104 are compared to spectra taken of singly modified disks and the free chromophore.

3.4.7 Fluorescence emission spectra of the templated arrays

In addition to the spectral changes in the absorption profiles, we also analyzed the fluorescence from the LH systems relative to that of free dyes in solution. Excitation and emission fluorescence spectra were collected for the free dyes and compared to spectra from purified protein-pigment constructs. Excitation spectra in arrays with one color were unremarkable, in that they all showed no change in the absorption profiles. The emission spectra showed subtle shifts in emission maxima, but not with the trends observed in the absorption spectral changes. There does not seem to be any reasonable order to the emission shifts observed. With the coumarin-based systems, the only construct that deviated from the free dye's emission profile was the C123 templated arrays. They showed a red-shifted emission of 10 nm relative to the other systems. OG488-based systems showed emission maxima of 522 nm, 526 nm, 538 nm, and 523 nm for free dye, singly modified disks, C123 templated arrays, and K104 templated arrays, respectively. As with coumarin, the largest departure from the isolated chromophore's emission comes from the dyes at the C123 location. In the case of AF594, the singly modified species showed the largest change in emission, a blue shift of 8 nm from the free

dyes emission maximum at 616 nm. The protein-templated arrays emit with maxima at 619 nm and 622 nm for the C123 and K104 sites, respectively.

3.4.8 Fluorescence quenching within the RR-TMVcp templated arrays

Another physical parameter of the systems that is altered upon protein conjugation and introduction of adjacent chromophores is the fluorescence quantum yield of the dyes. Upon photoexcitation, there are multiple pathways through which the energy can dissipate. Most commonly, these consist of non-radiative dissipation into heat, contact quenching with species in solution, and fluorescence emission.⁴⁵ By tethering the chromophore to a large protein assembly, there is a drastic change in the pigment's microenvironment, changing the local effective dielectric, changing the pigment's diffusion and rotation in solution, and introducing new vibrational modes of the protein that can couple into the pigment. These effects along with other possible unknown or sample-specific effects will no doubt change the excited state character and dynamics. We set out to measure and compare the effect on quantum yield that the various geometries had on the chromophores. For these experiments, we used a simplified and self-normalized parameter, "brightness". This metric is a measure of the emission counts obtained divided by the absorption for a particular system. For each sample, an absorption spectrum was obtained, and the raw fluorescence emission counts were collected after exciting at the peak wavelength. The emission counts were then divided by the sample OD, providing the brightness value. Simply put, we measured "photons out" relative to "photons in". Free dyes in solution were normalized to 100% brightness, and all other systems of the same dye were compared to these. In all cases, we observed a moderate to high level of quenching upon conjugation to protein. In all three cases (coumarin 343, OG488, and AF594), the singly modified disks showed on the order of 50% brightness. The protein itself thus quenches all the fluorophores by about half. Installing the full arrays and introducing neighboring chromophores brings the brightness down further, though by different amounts in all cases. One theme that arose from all of this work is that the chromophores in these systems undergo a very high degree of quenching, especially in the full arrays regardless of the position of modification on the protein.

Quenching of the coumarin 343 systems occurred in a fashion that at first glance seemed expected. Binding of one coumarin 343 maleimide to the protein resulted in an overall brightness of 57% relative to the free dye. It is not uncommon for fluorophores to be quenched upon conjugation to a macromolecule.⁴⁸ For arrays located at the C123 position, the brightness was reduced to 48% - a small decrease from the singly modified sample, implying that the inter-chromophore effects are not strongly quenching these systems. When the array was formed at K104, the brightness fell to 7% of the free dye. It seemed that by increasing the inter-chromophoric coupling (decreasing distance), the degree of quenching increased.

Similarly, OG488 also was quenched by around two-fold upon binding to the protein, but did not undergo significant quenching due to adjacent chromophores in either the K104 or C123 case. The brightness values measured for OG488 systems were 42%, 33%, and 50% for singly modified disks, C123 arrays, and K104 arrays, respectively. Interestingly, the closely packed arrays showed *less* quenching than the singly modified sample. The C123 templated arrays displayed a 20% decrease from the singly modified sample, but still retaining a third of the free dye's brightness.

Measurements done on singly modified AF594 samples also showed about half of the brightness of free AF594 maleimide. These had a brightness value of 55%. However, unlike the

previous cases, both of the arrays were quenched an order of magnitude from that of the free dye with brightnesses of 9% and 12% for the C123 and K104 positions, respectively. These findings do not show any clear pattern to the quenching behavior of the fluorophores within the protein framework. If we approximate values, the following general statements can be made: 1. All chromophores experienced quenching by half upon protein binding; 2. Relative to the singly modified samples, the C123 position modifications did not further quench coumarin 343, showed slight quenching with OG488, and showed strong quenching AF594 arrays; 3. Again, relative to the singly modified samples, the K104 position modifications quenched both coumarin 343 and AF594 to a high degree, but did not affect the fluorescence quantum yield of the OG488 systems.

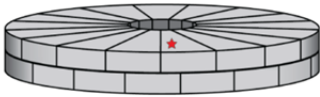
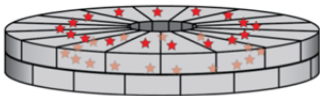
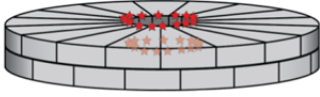
	Coumarin 343	OG488	AF594
<i>free chromophore</i>	100%	100%	100%
 <i>singly modified disk</i>	57%	42%	57%
 <i>array at C123</i>	48%	33%	9%
 <i>array at K104</i>	7%	50%	12%

Figure 9. Brightness values from fluorescence emission spectra. Measures the degree of fluorescence quenching relative to the free chromophore in solution, normalized to sample OD.

3.4.9 Fluorescence lifetime analysis

The previous spectroscopic analysis described was all done in the steady-state. However, there are many faster processes that can be masked by only interrogating the system with steady-state methods. To investigate the excited state dynamics on much faster timescales, we measured the fluorescence lifetimes using a time correlated single photon counting (TCSPC) system. This instrument delivered an excitation pulse of <200 ps and measured the emission as a function of time. The data collected were then deconvolved from the instrument response function using commercial software and fitted for decay constants. This pulse width is small enough to measure fluorescence lifetimes, which are typically on the order of a few nanoseconds, but not fast enough to capture ultrafast behavior on the fs to ps timescales. Values obtained from

the literature were compared to the measured lifetimes of the free chromophores for validation of the fitting and proper experimental conditions.

The literature value for the fluorescence lifetime of coumarin 343 is 4.6 ns. In our measurements, we obtained a value of 4.7 ns. The singly modified sample showed a very similar lifetime of 4.9 ns – a slight increase, but not substantial. However, when the arrays were formed, both positions produced systems with significantly lower lifetimes of 3.0 ns and 3.2 ns for the C123 and K104 positions, respectively.

OG488 showed less change in its lifetime upon protein conjugation, though the measured value of 3.5 ns for the free chromophore is lower than the published value of 4.1 ns. This is due to the different solution conditions. The literature value of 4.1 ns was collected in an alkaline (pH 9) solution, and our experiments were carried out at pH 7. It is not surprising that the lifetimes vary, for fluorescence is known to be very solution dependent. Both the arrays positioned at C123 and K104 gave lifetimes near our measured free chromophore of 3.7 ns and 3.8 ns, respectively. Unfortunately, the singly modified OG488 lifetime was not collected.

Lifetimes for AF594 samples followed no more trends than the other chromophores. The value collected for free AF594 maleimide corresponded exactly to the literature value of 3.9 ns. Singly modified disks gave increased lifetimes of 4.7 ns, while the arrays at C123 and K104 gave values of 4.9 ns and 3.6 ns, respectively.

3.4.10 Fluorescence anisotropy analysis of RR-TMVcp systems – linker effects

One aspect of our LH systems that needs to be addressed is the flexibility and disorder of the chromophores. In natural systems, the chromophoric components are held in fixed positions, but retain a degree of disorder, both homogeneous and inhomogeneous.⁴⁹ The degree of disorder can have large implications in the LH capabilities of chromophore arrays and energy dynamics therein. To investigate the rigidity of the chromophores after conjugation, we turned to the technique of fluorescence anisotropy. This is a method that determines the rotational time constant of the excited state, and therefore measures the degree of rotation of the molecule in the excited state. Fluorescence anisotropy has been applied to a vast number of systems, and is reviewed in depth here.⁴⁵ For the present, a brief explanation of the experimental technique and application to our LH systems is in order.

Initially, the sample is irradiated with linearly polarized light. The emission is collected as in a typical fluorimeter, but through a polarizer set either parallel or antiparallel to the excitation polarization. Thus, if the molecule has a very fast rotation within the excited state, the radiation will be isotropically radiated, and not carry the excitation polarization with it. If, however, in the other extreme, the molecule is completely rigid, the excitation polarization will be retained. By comparing the ratios of light intensities collected through the parallel and antiparallel emission polarizers, the anisotropy can be calculated. The anisotropy depends on the rotational time constant, but also is a function of the fluorescence lifetime as well as analyte specific parameters such as the transition dipole. However, there is a general theoretical window into which all anisotropy values must fall. On one extreme, in completely isotropic solutions, the lower limit for fluorescence anisotropy values is -0.2. The upper limit, in an ideally rigid solution is 0.4. These are used to gauge the degree of rigidity in the excited state – higher anisotropy corresponds to a more rigid species.

We set out investigate the rigidity of the previously described coumarin 343 protein-dye conjugates bound through the ethyl linker **2**. Additionally, a series of coumarin 343 analogs with

variable linker lengths and chirality was synthesized to create systems with identical protein and chromophoric components, but with different linker length, chirality, and composition. The various coumarin 343 maleimide dyes were synthesized in identical fashion as **2** in **Figure 5**, but starting with different diamine precursors. Within the set of coumarin 343 species measured was the original ethyl linker to compare to an ethylene glycol spacer in the attempt to generate a more flexible linkage. Also used were two very rigid and chiral cyclohexyl linkages to investigate how a rigid linkage with no rotatable bonds affected the chromophore. These molecules were conjugated to the RR-TMVcp at the C123 position through thiol-maleimide chemistry as discussed above.

Because the chromophores are well within distances where energy transfer is expected, singly modified double disks were used in the anisotropy measurements. Energy transfer within the chromophore arrays will inherently scramble the polarization, and must be eliminated to minimize artifacts. If one imagines a single chromophore within the 17-chromophore rings to be excited, after transferring energy and emission from neighboring dyes, the emission will come from a chromophore in a different orientation than the excited species, and produce radiation with changed polarization that did not stem from rotation, but from absorption and emission from different species.

The measured anisotropy of free coumarin 343 maleimide in solution produces a value of 0.03. All measurements are relative, and this value will be used for the limit of freely diffusing and rotating coumarin 343 maleimide and a baseline of which to compare anisotropies once attached to the protein assembly. Once the RR-TMVcp disks are singly modified, the ethyl linkage produces coumarin 343 with fluorescence anisotropies of 0.30. While this is a relative measurement, it is clear that it is much higher than 0.03 observed for the free dye. It is also near the theoretical upper limit of 0.4 for any chromophore, and can thus be interpreted as being very rigid on the protein surface. Surprisingly, the ether linkage of **2** shows the same rigidity once attached to the protein, producing an anisotropy value of 0.29. This molecule was made to produce a more flexible linker, hypothesizing that the extra length and ether linkage would promote rotation, flexibility, and increased solvation. This was not observed through monitoring anisotropy. With the intention of forming very rigid, non-rotatable protein-chromophore linkages, the chiral cyclohexyl linkers were introduced in the same manner for singly-modified RR-TMVcp disks. These, also surprisingly, showed the same behavior as the ethyl and ether containing coumarin 343 systems. Fluorescence anisotropy values of 0.32 and 0.29 were measured for the S,S and R,R chiralities, respectively.

From these measurements, we concluded that regardless of the linkage identity, each of the coumarin 343 derivatives displayed very rigid character once bound to the protein assembly. The four different linkages did not appear to impose much influence on the rotational ability of the dyes, which is primarily seen to be a product of the decreased tumbling and diffusion of the large protein assembly.

When the full arrays are constructed and anisotropies examined, we observed a decrease in fluorescence anisotropy relative to the singly modified species, but still high values compared to the 0.03 measured for free coumarin 343. Full modification of the RR-TMVcp with the ethyl linked coumarin 343 gives an anisotropy value of 0.15, which is directly in the middle of the free dye value and that of the singly modified disks. The chiral linkers were also used to generate arrays on the protein at the C123 position, and produced values of 0.18 and 0.21 for the S,S and R,R chiralities, respectively. These are also relatively high values, implying that the

chromophores are still rigid with full modification of the protein, but due to energy transfer within the assemblies, the anisotropy values drop from those on the singly modified proteins.

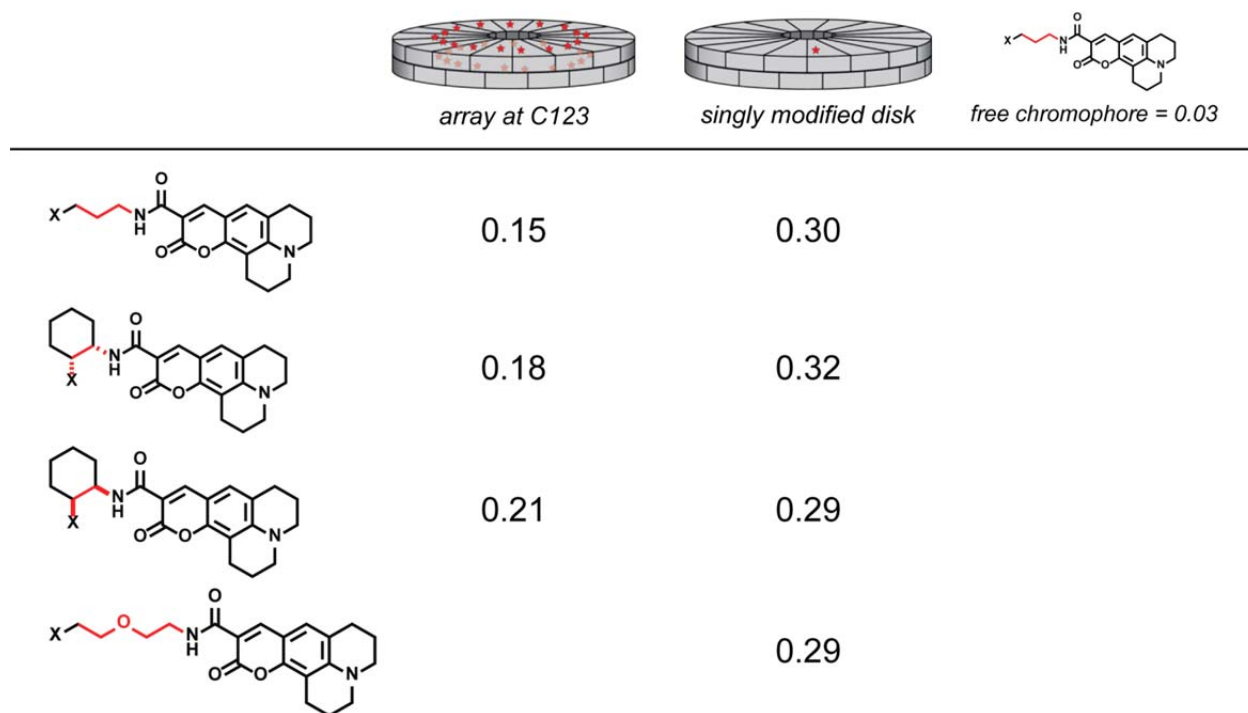


Figure 10. Fluorescence anisotropy values for a series of linkers. Each variable linker was installed on coumarin 343 and attached to the C123 position of RR-TMVcp. All resulting anisotropies were compared to that of free coumarin 343.

3.5 Dual color systems – energy transfer within the RR-TMVcp assemblies

3.5.1 Construction and design of donor:acceptor systems

Extending the analysis of the single color chromophore arrays discussed in this section, we produced RR-TMVcp templated systems containing two chromophores capable of energy transfer. The purpose of these studies was to investigate the ability of coupled arrays to participate in energy transfer, a property that is of utmost important in natural systems such as LHII. In the bacterial light harvesting antenna LHII, it is well known that the chromophores within a particular ring of B800 or B850 act not as isolated chromophores, but as excitonically coupled systems where the excited state is delocalized over many chromophores.²⁵ These rings of chromophores where the excitation is spread over multiple sites achieve energy transfer within a particular LHII complex, between adjacent LHII rings, and also from LHII to LHI in a very fast and efficient manner.⁴⁹

We created dual modified RR-TMVcp systems with a ring of donor chromophores located at the either the C123 or K104 position and the other site containing the acceptor. Modification of the protein at both sites was accomplished to investigate the energy transfer

between rings of dyes. To generate these structures, we used the same modification chemistry as described for the single color constructs, but in a sequential manner. First, the C123 position was modified completely through thiol-maleimide chemistry. Once the modification was validated by mass spectrometry and the protein purified, SPDP was attached to the K104 position. Reduction of the disulfide afforded a free thiol to then participate in conjugation to a second maleimide functionalized chromophore. In this way, we could orthogonally modify each site with the desired species, and generate donor/acceptor pairs located at the C123 and K104 positions, and vice versa. The D:A pairs we chose to investigate were Coumarin 343 - OG488 and OG488 - AF594. From the spectra shown in **Figure 3**, each of these pairs has excellent overlap and are ideal energy transfer partners. For each donor:acceptor pair, both geometries were produced to compare the effect of interchromophore distances on the energy transfer ability.

3.5.2 Absorption spectra of donor:acceptor systems

UV/Vis analysis of the dual color systems showed no change from spectra of the single color arrays. We observed the same site-dependent spectral shifts for each chromophore as when they were templated in the single color systems, and when overlaid, they mimic the two color spectrum exactly. Note that the intensities of the absorption peaks are not what one would predict based on relative extinction coefficients of the dyes, but this is simply due to incomplete modification of the protein. When multiple modifications are attempted on a single monomer, especially in a 4-step sequence (maleimide at 123, SPDP at 104, TCEP cleavage, maleimide reaction at 104), it is very difficult to push all modifications to 100%. That being said, all chromophore conjugations were > 80% based on MS peak integration.

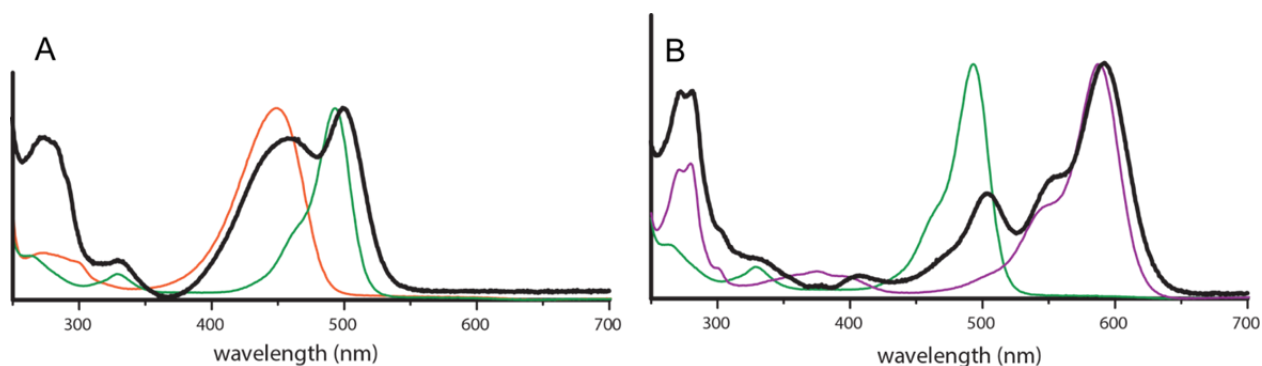


Figure 11. UV/Vis analysis of dual color RR-TMVcp templated LH systems. (A) Coumarin 343@123 : OG488@104. Absorption of free coumarin 343 and free OG488 are shown in orange and green, respectively. Note the shifting of the individual peaks in the two-color absorption, which correspond exactly to those observed in single color experiments. (B) Identical to A, but with a OG488@123 : AF594@104 pair. Relative peak intensities are not equal due to incomplete modification of some sites.

More interesting are the fluorescence data collected from the donor:acceptor systems. This will give an insight into the energy transfer occurring between the rings and how different inter-chromophore coupling affects energy transfer ability. We wished to monitor both the emission spectra and the excitation spectra for all wavelengths, so the data are plotted in 2-

dimensional heat maps, with the emission along the x-axis and excitation along the y-axis. This is a somewhat nontraditional way to present fluorescence spectra, where a horizontal slice corresponds to a typical emission spectrum, and vertical slices excitation spectra. When examining these spectra, we need to take into account all of the effects that perturb the fluorescence of the individual chromophoric arrays in the absence of a partner (homo-FRET), in addition to how the two colors interact with each other (hetero-FRET).

3.5.3 Fluorescence analysis of donor:acceptor systems

The first system produced was a coumarin 343:OG488 donor:acceptor pair. The coumarin 343 emission profile overlaps virtually exactly the OG488 absorption spectrum, and the fluorescence data are shown in **Figure 12**. **Figure 12A** shows the system with coumarin 343 (donor) in the C123 position and OG488 (acceptor) in the K104 site. From this plot, we can see no direct coumarin 343 fluorescence. In the case of coumarin 343 at this position with no acceptor, its fluorescence is only quenched to 48% brightness, so this lack of any detectible fluorescence in the donor:acceptor system would typically be attributed to transfer into the acceptor. However, the fluorescence spectrum does not show this. A vertical slice (excitation spectrum) taken at the OG488 emission at 520 nm shows no significant contribution from the coumarin 343 absorption at 445 nm. The excitation energy from the coumarin 343 is therefore being quenched by the introduction of the donor, but not with the classic FRET signatures.

Even more surprising is the emission profile of the OG488 in this case. When the emission spectrum is analyzed from direct excitation of OG488 at 490 nm, we see significant quenching of the OG488 emission relative to the array of OG488 templated at K104 in the absence of coumarin 343. From direct OG488 excitation at 490 nm, we measured a brightness value for OG488 at 535 nm emission of 2.3%. This is over a factor of 20 lower than when there is no donor present (50% brightness). This counter-intuitive finding suggests that the donor is in fact *quenching* the acceptor, not predicted by classical FRET theory.

If we compare **Figure 12A** to **Figure 12B**, we see similar but not identical results. In this case, the positions of the donor and acceptors are switched, and energy transfer is expected to occur from the pore region radially outward. Here, the coumarin is located at the K104 position, and is not completely quenched. In general, we observe quenching of the entire spectrum. It shows both a direct coumarin 343 emission at 490 nm, and also a peak that corresponds to inter-ring energy transfer with excitation at 445 nm (coumarin 343) and emission at 535 nm (OG488). Excitation spectra collected at the OG488 emission shows a contribution from the coumarin 343, but with low efficiency as overlaid with the absorption profile for the two color system. As with the other orientation, the OG488 was quenched by a large degree by the introduction of the coumarin 343 ring. Compared to a brightness value of 33% when OG488 is located at the C123 position in the absence of donor, the OG488 fluorescence has a brightness of 0.8% to that of the free chromophore when coumarin 343 is present at the K104 position. This represents a 40-fold decrease in fluorescence of the acceptor due to the presence of the donor.

Next, we constructed identical systems, but now using OG488 as a donor and AF594 as an acceptor. These donor:acceptor systems displayed similar spectral character as with coumarin 343 and OG488. **Figure 13** shows the fluorescence spectrum of the two geometries for this dye pair. In the case of donor (OG488) at the C123 position and acceptor (AF594) in the pore position at K104, we see that the primary emission comes from AF594 at 620 nm. Donor emission due to OG488 ($\lambda_{\text{ex}} = 490 \text{ nm}$; $\lambda_{\text{em}} = 520 \text{ nm}$) is barely distinguishable, but present

nonetheless. A small peak corresponding to OG488 to AF594 energy transfer is observed ($\lambda_{\text{ex}} = 490$ nm; $\lambda_{\text{em}} = 620$ nm), but not possessing nearly the intensity suggesting efficient energy transfer (see **Figure 13** excitation vs. abs plots). Again, the donor excitation energy is being quenched due to the acceptor's presence, but not appearing significantly in the acceptor's fluorescence excitation spectrum. Also observed was the quenching of the direct acceptor fluorescence by introducing the donor species. In this case, the AF594 has a brightness of 1.4% relative to the free dye (measured at $\lambda_{\text{ex}} = 590$ nm and $\lambda_{\text{em}} = 620$ nm), an order of magnitude lower than the 12% brightness observed for the K104 modified AF594 systems in the absence of OG488.

The other geometry of OG488 and AF594 was analyzed in **Figure 12B**, and shows very similar activity to that observed with the coumarin 343 and OG488 system. Here, as in the previously discussed system, the donor located in the pore region shows incomplete energy transfer or quenching, and in fact the most intense emission from this system occurs from direct OG488 excitation and emission. Energy transfer is observed due to the presence of the peak at $\lambda_{\text{ex}} = 490$ nm and $\lambda_{\text{em}} = 620$ nm, though this is at a lower intensity than the direct OG488 emission. Direct AF594 emission is observed to be quenched relative to the single color AF594 at that position, as was earlier observed with the other chromophores. Again, these systems appeared even further quenched at all wavelengths. Monitoring direct OG488 emission ($\lambda_{\text{ex}} = 490$ nm and $\lambda_{\text{em}} = 520$ nm), we see a brightness value of OG488 to be 1.6% that of the free dye. This is compared to a brightness of 50% without any acceptor present. In a typical FRET analysis, this decreased donor fluorescence would be attributed to energy transfer to the acceptor chromophore, but the low intensity of the emission from $\lambda_{\text{ex}} = 490$ nm and $\lambda_{\text{em}} = 620$ nm does not support this classic FRET picture. Additionally, the AF594 is quenched upon introducing the donor OG488 to a brightness of 1.8% relative to the free dye. This is contrasted to the system with only AF594 at the C123 position, which has a brightness of 9%, a 5-fold increase.

3.6 Discussion and interpretation of spectral data

Although an abundance of spectroscopic information was obtained from the RR-TMVcp templated light harvesting arrays, the observed results did not appear to correlate pigment location and inter-chromophoric coupling to any obvious trend in spectral properties. Each of the previously described systems displays spectral signatures upon protein binding and integration into templated arrays, though these effects appear to be very chromophore dependent. For the purpose of using these systems as tools to gain insight into the overall effects upon protein templating of chromophores and application for the future construction of efficient devices, it is important to consolidate the information gained thus far in order to extract general patterns from the observed results. This section will outline the results and data presented from the above experiments and attempt to organize them into a coherent picture of the spectral and physical effects of chromophores imposed upon formation of protein-templated arrays.

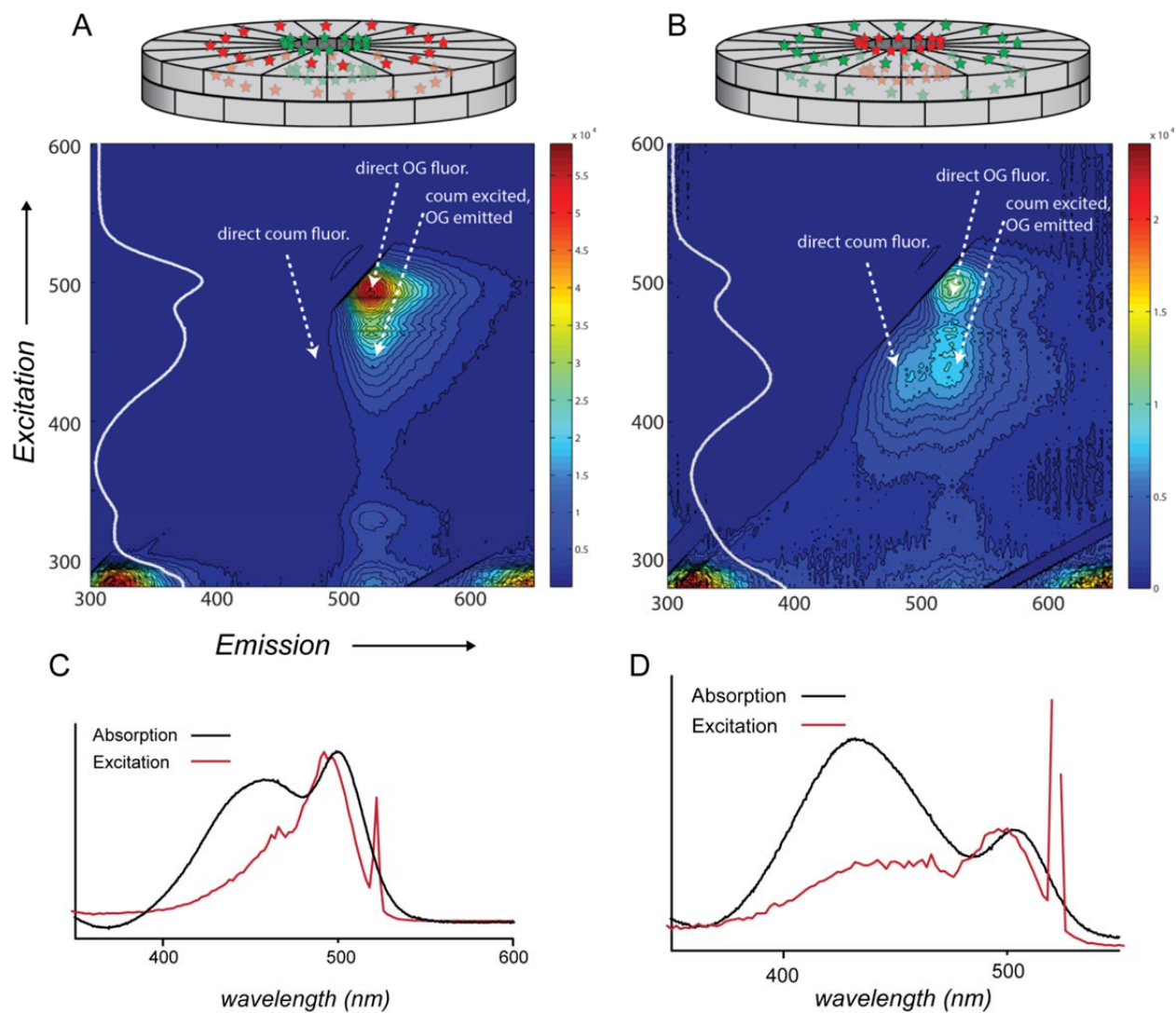


Figure 12. Fluorescence analysis of two-color donor:acceptor systems for coumarin 343 : OG488. (A) and (B) Fluorescence emission and excitation plotted as a heat map for the two respective geometries at the C123 and K104 positions. (C) and (D) overlay the absorption spectrum of the system with a fluorescence excitation spectrum obtained by taking a vertical slice of (A) or (B) at the acceptor's emission wavelength. These two are normalized at the acceptor absorption maximum.

3.6.1 Physical nature of RR-TMVcp templated chromophore arrays

The assembly state and structure of the RR-TMVcp disks did not undergo any observable change upon attachment of any of the chromophores used in these experiments at either of the positions investigated. By monitoring the assembly state through size-exclusion chromatography (SEC) and transmission electron microscopy (TEM), all constructs appeared as stable disk-form TMV. TEM images taken before and after chromophore conjugation were identical, and well as the solution size determined through SEC and light-scattering methods. From these

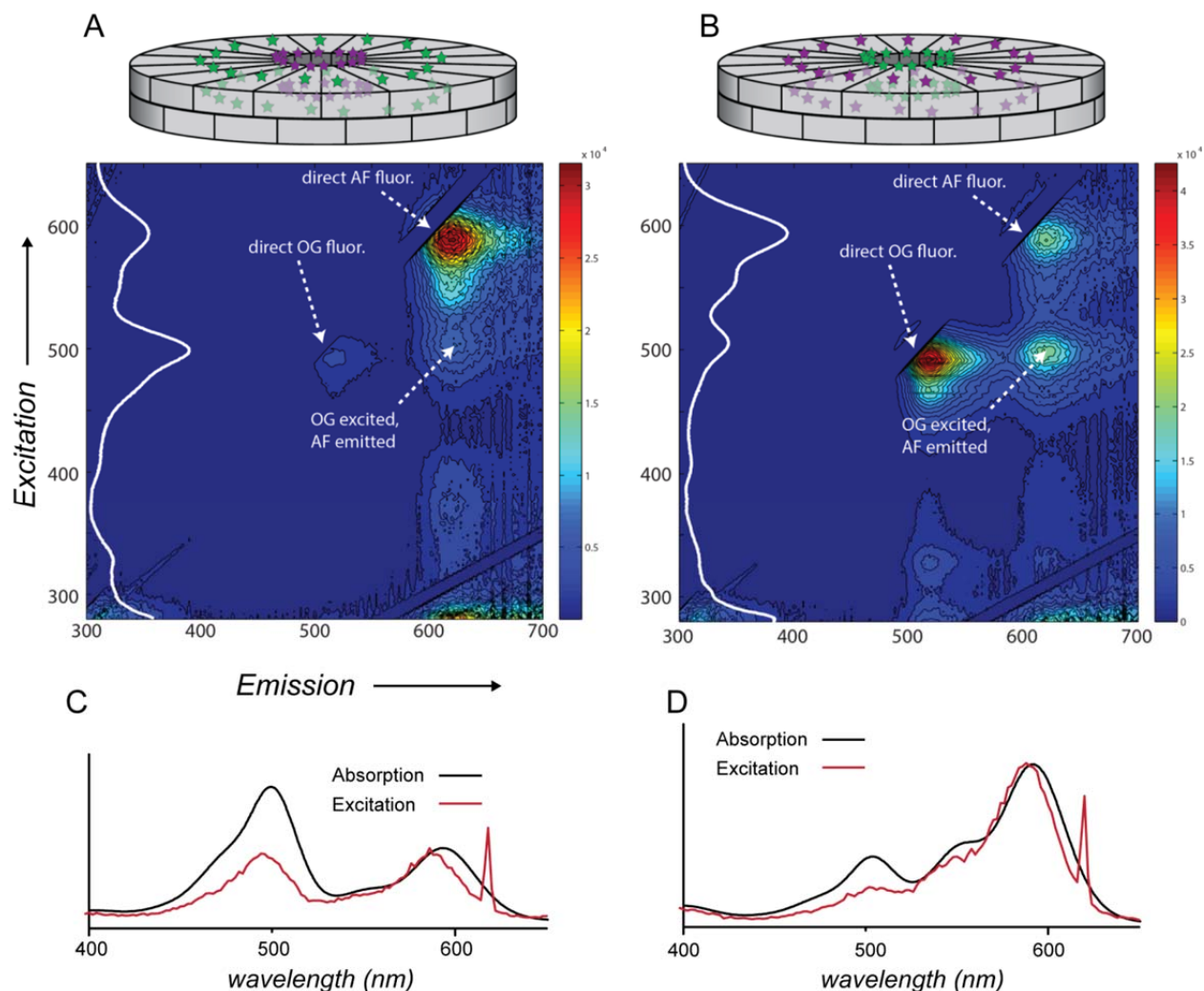


Figure 13. Fluorescence analysis of two-color donor:acceptor systems for OG488: AF594. (A) and (B) Fluorescence emission and excitation plotted as a heat map for the two respective geometries at the C123 and K104 positions. (C) and (D) overlay the absorption spectrum of the system with a fluorescence excitation spectrum obtained by taking a vertical slice of (A) or (B) at the acceptor's emission wavelength. These two are normalized at the acceptor absorption maximum.

observations, we can be confident that the integration of chromophores does not significantly perturb the quaternary structure of the protein assembly, and all subsequent modeling of the systems based on the crystallized structures are valid.

Molecular modeling and energy minimization of the systems showed that while in the C123 position, chromophores appeared as isolated chromophores held rigidly by the protein. Chromophores located at the 123 position fell well within inter-chromophore distances to facilitate energy transfer, but were separated enough to disallow physical contact and aggregation. In the K104 position, on the other hand, chromophores were positioned within such proximity that they preferred to aggregate and form clusters of 2-4 chromophores around the ring. We attribute this aggregation to the tendency of the aromatic chromophores to form π -stacked aggregates, limited by the geometrical constraints imposed by the circular dimensions of

the array. The selective formation of excimers in the pyrene experiments at K104 vs. C123 corroborate this picture.

Through fluorescence anisotropy, the rigidity of the chromophores at C123 was apparent. The determining factor for chromophore flexibility and rotational ability was primarily the protein scaffold. The high fluorescence anisotropy values collected for chromophores attached through the four linkers with variable flexibility suggested that regardless of the linker, the chromophores were held rigidly in place by the large protein assembly. Even when more flexible linkers were engineered such as the diethyl ether linkage, the dyes displayed high rigidity on the protein surface.

3.6.2 Comparison of chromophores

All spectral analysis revealed specific effects upon conjugation for each chromophore used, though not all of these properties share commonality between the dyes investigated. One property that did occur with all three chromophores (coumarin 343, OG488, and AF594) was a red-shift in the absorption spectrum maximum of the singly modified disks. In the case of coumarin 343, this was recorded as an 8 nm bathochromic shift. For OG488 and AF594, the shifts were 9 nm and 6 nm, respectively. Synthetic dyes have been observed to undergo slight redshifts upon conjugation to macromolecules,⁵⁰ so this is not unexpected. The origin of this effect is most likely due to extra vibrational modes brought upon by the protein and/or the reduced effective dielectric constant seen by the fluorophore.

When the full arrays were constructed at K104 and C123, however, the dyes behaved quite differently from one another. The coumarin 343 displayed the most drastic spectral changes, showing a blue-shift of 8 nm at C123 and 24 nm at K104 relative to the singly modified sample. OG488 had no change when at C123, and a moderate blue shift of 5 nm when located at K104 (relative to singly modified sample). AF594 showed even less change, with no change at C123, and a negligible blue-shift of 2 nm when at K104. Emission maxima did not alter in any noticeable pattern, with shifts widely varying.

Quenching of fluorescence was observed in all cases to some degree, with the protein imparting about two-fold decrease in fluorescence brightness. Upon formation of arrays at C123 and K104, introduction of chromophore-chromophore interactions introduced additional quenching, where the site-dependence was different in all cases. With coumarin 343, the degree of quenching scaled with interchromophore distance, where the more closely packed arrays at 104 displayed higher quenching than the more separated arrays at 123, which were about as bright as the singly modified samples. OG488 did not alter its brightness significantly in any of the 3 geometries compared (singly modified vs. 123 vs. 104). AF594 was observed to be quenched 5-fold in both positions relative to the singly modified construct.

Measurement of fluorescent lifetimes gave insight into the excited state dynamics of the systems, but once again did not reveal any overarching trends to the behavior of chromophores as a function of position on the protein. Coumarin 343 did not display significant change in lifetime upon protein conjugation, but once the complete arrays were constructed the lifetime dropped by a third, and were similar in both the 123 and 104 positions. This behavior did not track with either the absorbance shifts (progressive blue-shifting with increasing interchromophore distance) or the brightness values (singly modified and 123 very similar, but different from 104). OG488, again, showed the least drastic change in spectral properties. In both cases of 123 and 104 constructed arrays, the OG488 only showed slight increases in lifetime

relative to the free dye, to comparable extents. Results from AF594 samples had 1 ns increased lifetimes in the case of the singly modified sample as well as the array at the 123 position, but in the case of 104 modification, the lifetime was decreased to slightly below that of the free dye. Comparing these data to the absorption and fluorescence data, where the arrays at both 123 and 104 show similar behavior but act much differently from the singly modified sample yields no clearer picture than with coumarin 343 or OG488.

3.6.3 Donor:acceptor systems

Even more puzzling than the single color experiments described above are the results from the two color donor:acceptor (D:A) systems. The absorption profiles of the D:A assemblies emulated that of the individual parts exactly, but the fluorescence and energy transfer properties behaved anomalously. In general, as in the single-color experiments, all systems underwent fluorescence quenching to some degree. In the case of two color systems, each chromophore was quenched even further upon introduction of the energy transfer partner - both donors and acceptors. Since the donors and acceptors are well within expected FRET radii, quenching of the donor is expected with transfer of the energy to the acceptor, which then is capable of fluorescence emission. The observed quenching of the *acceptor*, on the other hand, is unexplained. We observed in all cases that the introduction of the donor array in fact further quenched the acceptor emission significantly when monitored at the acceptor peak absorption and emission. This implies that the donors and acceptors are coupling in a way in which the excitation on the acceptor is dissipated through interaction with the donor, contrary to the typical Jablonski picture of energy transfer. Although all the fluorescence spectra in **Figures 12 and 13** display signatures of energy transfer, the directionality (123 \rightarrow 104 vs. 104 \rightarrow 123) appears very important. In both systems, peaks corresponding to d \rightarrow a transfer are more apparent in the cases of 104 \rightarrow 123 transfer than in the 123 \rightarrow 104 geometry. Thus, the energy transfer seems more facile occurring radially outward than inward. Again, this asymmetric efficiency of energy transfer is not predicted based on a typical FRET picture, where the interchromophore distance is the major factor in determining energy transfer efficiency. The resulting geometry of the chromophore upon protein conjugation may play a role in this asymmetry of energy transfer, with preferred dipolar orientations emerging from the donor@104 and acceptor@123 systems. However, this is not likely to be the root of observed spectral effects due to the similar behavior but vastly different structures and dipole directions in the three chromophores investigated.

3.6.4 Conclusion and perspectives

One observation that emerges from investigation into these protein-templated light harvesting systems is that the chromophores are highly coupled and no longer are behaving as isolated dipoles, but rather as collective ensembles. In all systems investigated, and through a variety of experimental techniques, we have observed interesting and sometimes perplexing effects of both attachment to protein and incorporation of dyes into highly coupled arrays. While we are still developing an explanation into the precise physical origin of our observed spectral changes, these systems indeed represent a step towards creating functional biomimetic light harvesting systems. Natural systems orient highly coupled chromophores as energy transfer networks that also deviate significantly from the sum of the parts. This seems to be a theme in photosynthetic light harvesting, and in order for lessons to be learned and put towards

construction of functional synthetic systems, the origin and individual aspects of these coupled systems must be understood. Additionally, the mechanism in which the coupled nature of the chromophores is manifested as energetic delocalization, excitonic character, and non-classical behavior will ultimately need to be investigated.

3.7 Materials and methods

General methods

Unless otherwise noted, all chemicals and solvents were of analytical grade and used as received from commercial sources without further distillation or purification. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F₂₅₄ plates with visualization by ultraviolet (UV) irradiation at 254 nm and/or potassium permanganate stain. Purifications by flash chromatography were performed using EM silica gel 60 (230-400 mesh). The eluting system for each purification was determined by TLC analysis. All organic solvents were removed under reduced pressure using a rotary evaporator. Water (dd-H₂O) used as reaction solvent was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Spin concentration steps were performed using 100 kDa molecular weight cutoff spin concentrators from Millipore (Billerica, MA). Centrifugations were performed with an Eppendorf 5415 benchtop centrifuge, (Eppendorf, Hauppauge, NY).

Instrumentation and sample analysis

Spectroscopy. UV-Vis spectroscopic measurements were conducted on a Cary 50 Scan benchtop spectrophotometer (Varian Inc., USA). All measurements were done with sample OD between 0.1-1.0. Steady-state fluorescence measurements were obtained on a Fluoromax-4 spectrofluorometer (Horiba Scientific, USA). All fluorescence spectra were taken with sample OD between 0.05-0.1. Fluorescence anisotropy measurements were also taken on the same instrument, using the built in polarizers. The *anisotropy* parameter was collected in all cases. Time-resolved fluorescence spectroscopy was obtained on a NanoLog spectrofluorimeter (Horiba Scientific, USA) equipped with light emitting diodes (LEDs) for Time Correlated Single Photon Counting (TCSPC) lifetimes measurement. LEDs of 455 nm, 485 nm, or 590 nm were used as needed. The resulting TCSPC traces were deconvolved and fitted with Origin 8.0 software (Originlab, USA). Protein concentration was routinely determined by UV/Vis analysis on a Nanodrop 1000 instrument (Nanodrop, USA) by monitoring absorbance at 280 nm.

NMR. ¹H and was measured with a Bruker AVQ-400 (400 MHz) spectrometer. ¹H NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to CDCl₃ (δ 7.26, singlet). Multiplicities are reported as follows: s (singlet), d (doublet), t (triplet), dd (doublet of doublets), br (broad) or m (multiplet). Coupling constants are reported as a J value in Hertz (Hz). The number of protons (n) for a given resonance is indicated as nH and is based on spectral integration values.

Mass Spectrometry. Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DE system (PerSeptive Biosystems, USA) and data were analyzed using Data Explorer software. Samples were co-crystallized with 10 mg/mL α -cyano-4-hydroxycinnamic acid or sinipinic acid in 1:1 acetonitrile (MeCN) to H₂O with 0.1% trifluoroacetic acid (TFA). Protein bioconjugates were analyzed using an Agilent 1200 series

liquid chromatograph (Agilent Technologies, USA) that was connected in-line with an Agilent 6224 Time-of-Flight (TOF) LC/MS system equipped with a Turbospray ion source. Spectra were quantified by peak area integration using Origin 8.0 software (Originlab, USA).

High Performance Liquid Chromatography. HPLC was performed on Agilent 1100 Series HPLC Systems (Agilent, USA). Sample analysis for all HPLC experiments was achieved with an in-line diode array detector (DAD) and in-line fluorescence detector (FLD). Analytical reverse-phase HPLC of small molecules was accomplished using a C18 stationary phase and a H₂O/MeCN with 0.1% TFA gradient mobile phase. Size exclusion chromatography (SEC) was performed using a Phenomenex polysep GFC-P-5000 column (4.6 x 250 mm) (Phenomenex, USA) at 1.0 mL/min using a mobile phase of 10 mM phosphate buffer, pH 7.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus (Bio-Rad, Hercules, CA), using a 10-20% precast linear gradient polyacrylamide gel (Bio-Rad). The sample and electrode buffers were prepared according to Laemmli.⁵¹ All protein electrophoresis samples were heated for 5-10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of disulfide bonds. Gels were run for 70 min at 120 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad). Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA). ImageJ was used to determine the level of modification by optical densitometry.

Transmission Electron Microscopy. TEM images were obtained at the UC Berkeley Electron Microscope Lab (www.em-lab.berkeley.edu) using a FEI Tecnai 12 transmission electron microscope with 120 kV accelerating voltage. Protein samples were prepared for TEM analysis by pipetting 5 µL of the samples onto Formvar-coated copper mesh grids (400 mesh, Ted Pella, Redding, CA) and allowed to incubate 3 min. The samples were then wicked with filter paper and rinsed with dd-H₂O. The grids were stained with 5 µL of a 1% (w/v) aqueous solution of uranyl acetate for 90 s as a negative stain. After excess stain was wicked away, the grids were allowed to dry in air and imaged within 1 day of preparation.

Small molecule synthesis

Synthesis of *N*-boc aminoethylene maleimide (1). This synthesis was adapted from reference 52. To 35 mL of freshly distilled DCM was added di-*t*-butyl carbonate (2 g, 9.2 mmol). To 35 mL of freshly distilled DCM was added ethylenediamine (5.5 g, 92 mmol). At 0 °C the two solutions were combined and stirred. The solution was allowed to warm to room temperature and stirred for an additional 5 h. The crude reaction mixture was extracted into acidic aqueous media. The pH was then raised to 14 with 20% v/v NaOH, and back extracted into DCM. Solvent was removed *in vacuo* affording 1.15 g of a pale yellow oil to be used directly in the next step without further purification. The oil was then taken up in 35 mL of dry MeCN, to which was added maleic anhydride (1.06 g, 10.9 mmol). The solution was refluxed for 5 h and cooled to room temperature. Removal of solvent *in vacuo* resulted in a flaky white solid. The solid was then suspended in 13 mL of acetic anhydride, and NaHCO₃ (0.91 g, 10.9 mmol) added.

This mixture was allowed to reflux 6 h. The light brown solid was purified on silica gel using a gradient of 0-60% EtOAc in hexanes. The solvent was removed under reduced pressure and the product dried *in vacuo* (1.2 g, 79%). ¹H NMR (400 MHz, CDCl₃): δ 6.70 (s, 2H), 4.65 (br s, 1H), 3.65 (m, 2H), 3.33 (m, 2H), 1.4 (s, 9H).

Synthesis of coumarin 343 maleimide (2). The Boc protected **1** was first deprotected by stirring in 1:1 TFA:DCM for 20 min, followed by removal of solvent under reduced pressure. The solid was stored *in vacuo* overnight to remove any residual TFA. A 10 mL round bottom flask was then charged with 1 mL of DMF. Coumarin 343 was added (50 mg, 0.18 mmol). HCTU (79 mg, 0.193 mmol) and DiPEA (45 mg, 0.35 mmol) were added and allowed to stir for 10 min to promote activation of the carboxylic acid of coumarin 343. This mixture was then added to the solid **1**, and the reaction stirred under nitrogen for 1.5 h at room temperature. The solvent was removed under reduced pressure, and the product purified on silica gel. After removal of solvent under reduced pressure, the product was dried *in vacuo* (30 mg, 42%). ¹H NMR (400 MHz, CDCl₃): δ 8.93 (br t, 1H), 8.56 (s, 1H), 6.99 (s, 1H), 6.70 (s, 2H), 3.78 (m, 2H), 3.65 (m, 2H), 3.32 (m, 4H), 2.87 (t, 2H), 2.76 (t, 2H), 1.97 (m, 4H).

Protein expression and modification

Construction of TMVcp expression plasmids. The starting point for the RR TMV mutants was a gene for the coat protein of the TMV U1 strain optimized for the codon usage of *E. coli* (Genscript, Piscataway, NJ). This sequence contained the mutation S123C. Site-directed mutagenesis was performed using QuikChange mutagenesis (Stratagene, Santa Clara, CA).

Buffer A: 20 mM TEA, pH 7.2

Buffer B: 10 mM phosphate, pH 7.0

Protein expression and purification. Tuner DE3pLysS competent cells (Novagen) were transformed with the vector described above, and cultured in Terrific Broth with 100 µg/L ampicillin. When cultures reached an optical density of 0.6 to 0.8, IPTG was added to a final concentration of 30 µM. Cultures were grown 24 h at 30 °C, harvested by centrifugation, and stored at -80 °C. Cells (from a 1 L expression batch) were thawed, resuspended in 20 mL of 20 mM TEA pH 7.2 containing 10 µg/mL DNase and RNase, and lysed by sonicating with a 2 s on, 6 s off cycle for a total of 20 minutes using a standard disruptor horn at 90% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared by ultracentrifugation for 45 min at 11,000 rpm using a SLA-600TC rotor in a Sorvall RC5C Plus centrifuge (Waltham, MA). The clarified lysate was decanted, warmed to rt, and stirred while adding a saturated solution of ammonium sulfate dropwise to a final concentration of 50% (v/v). After 5 min, the white ppt that formed was pelleted by ultracentrifugation (45 min at 11,000 rpm using a SLA-600TC rotor), washed with deionized water, and resuspended in 5-10 mL buffer A. The resulting protein solution was next loaded onto a DEAE column equilibrated with buffer A and eluted with a 0 – 300 mM NaCl gradient (in buffer A). Purity was confirmed by SDS-PAGE

and HPLC/ESI-MS. Pure fractions were pooled, and fractions containing desired TMV in addition to impurities were re-run in identical DEAE conditions. This preparation provided pure RR-TMVcp in yields up to 100 mg/L culture. The resulting RR-TMVcp was only observed to be in the disk form under all laboratory conditions (pH 6-9). Aliquots of 10 mg/mL RR-TMVcp were flash frozen and stored at -20 °C for further use.

General procedure for chromophore modification of RR-TMVcp. Frozen aliquots of RR-TMVcp were thawed and exchanged into buffer B by using NAP-5 G-25 gel filtration columns (GE Healthcare, USA) followed by spin concentration to produce the desired protein concentration for bioconjugation.

Modification of RR-TMVcp at Cys123. To 100 μ L of RR-TMVcp at a concentration of 200 μ M was delivered 5-10 equivalents of maleimide functionalized chromophore from a concentrated (20-100 mM) stock solution in DMF. Reactions were allowed to proceed in 1.5 mL Eppendorf tubes at room temperature under very gentle magnetic stirring using triangular stir bars. After 2 h, the crude reactions were passed through a NAP-5 gel filtration column and subjected to spin concentration. When flow-through during spin concentration ceased to contain any visible residual color, the protein conjugates were analyzed with HPLC-SEC for assessment of purity and validation of assembly state.

Modification of RR-TMVcp at Lys104. To 100 μ L of RR-TMVcp at a concentration of 200 μ M was delivered 7 equivalents of succinimidyl 3-(2-pyridyldithio)propionate (SPDP) as a concentrated stock solution in DMF. Reactions were allowed to proceed in 1.5 mL Eppendorf tubes at room temperature under very gentle magnetic stirring using triangular stir bars. After 2 h, another 7 equivalents of SPDP were added in the same fashion, and the resulting solution allowed to react for another 2 h. Crude reactions were passed through a NAP-5 gel filtration column and subjected to spin concentration. When flow-through during spin concentration ceased to contain any small molecule reagent, the protein conjugates were analyzed with HPLC-SEC for assessment of purity and validation of assembly state. HPLC-ESI-MS verified the degree of modification, and if the level of protein modification was lower than desired, the protein sample was subjected to another 7 equivalents of SPDP, followed by purification as above. The protein bioconjugates were then stirred in 10 mM TCEP for reduction of the disulfide, affording a free thiol for alkylation. Alkylation of the thiol with maleimide-functionalized chromophores was accomplished in an identical fashion as described for C123 modification.

Dual modification of RR-TMVcp. Modification of RR-TMVcp with two chromophores was accomplished in identical fashion as described for the singly modified protein, but was performed sequentially. Initially, the C123 position was modified with thiol-maleimide chemistry as described above. The protein conjugates were analyzed by LCMS and confirmed to have the desired level of modification, and remaining small molecule reagents removed through application to a NAP-5 column and spin concentration. Next, the K104 position was modified through SPDP/TCEP/chromophore attachment as previously described. After SPDP addition as well as TCEP reduction, the conjugates were checked *via* LCMS and small molecules removed

with spin concentration. Once the final doubly chromophore modified sample was produced, they were once again run through a NAP-5 column and spin concentrators to the desired concentration.

Purification of dye-protein conjugates. All chromophore labeled proteins were subjected to HPLC purification prior to spectral analysis. SEC was performed, and the peak corresponding to the disk isoform of TMVcp was collected from 8-9 min under the elution conditions of 1 mL/min with 10 mM phosphate, pH 7.2. The fractions were then combined and subjected to spin concentration to obtain the desired protein concentration. Samples were then stored away from light at 4 °C.

3.8 Literature cited

1. Olson, J.M.; Blankenship, R.E. *Photosynth. Res.* **2004**, *80*, 373-386.
2. Blankenship, R.E. *Antonie van Leeuwenhoek* **1994**, *65*, 311-329.
3. Nelson, N.; Yocum, C. F. *Ann. Rev. Plant Biol.* **2006**, *57*, 521-565.
4. Khatypov, R. A.; Khmel'nitskiy, A. Y.; Leonova, M. M.; Vasilieva, L. G.; Shuvalov, V. A. *Photosynth. Res.* **2008**, *98*, 81-93.
5. Law, C. J.; Cogdell, R. J. *Comprehensive Series in Photochemical & Photobiological Sciences* **2008**, *8*, 205-259.
6. Kerfeld, C. A.; Krogmann, D. W. *Ann. Rev. Plant Phys. And Plant Mol. Biol.* **1998**, *49*, 397-425.
7. Hohmann-Marriott, M.F.; Blankenship, R.E. *Annu. Rev. Plant Biol.* **2011**, *62*, 515-48.
8. Eberhard, S.; Finazzi, G.; Wollman, F.A. *Ann. Rev. Gen.* **2008**, *42*, 463-515.
9. Moore, G. F.; Brudvig, G. W. *Annu. Rev. Condens. Matter Phys.* **2011**, *2*, 303-327.
10. Murray, J. W.; Barber, J. J. *Struct. Biol.* **2007**, *159*, 228-237.
11. Schlau-Cohen, G. S.; Calhoun, T. R.; Ginsberg, N. S.; Read, E. L.; Ballottari, M.; Bassi, R.; van Grondelle, R.; Fleming, G. R. *J. Phys. Chem. B* **2009**, *113*, 15352-15363.
12. Chew, A. G. M.; Bryant, D. A. *Ann. Rev. Microbiology*, **2007**, *61*, 133-129.
13. Bollivar, D. W. *Photosynth. Res.* **2006**, *90*, 173-194.
14. Glazer, A.N. *J. Biol. Chem.* **1989**, *264*, 1-4.
15. Beale, S. I. *Photosynth. Res.* **1999**, *60*, 43-73.
16. Blankenship, R. E.; Matsuura, K. *Adv. Photosynth. & Resp.* **2003**, *13*, 195-217.
17. Papiz, M. Z.; Prince, S. M.; Howard, T.; Cogdell, R. J.; Isaacs, N. W. *J. Mol. Biol.* **2003**, *326*, 1523-1538.
18. Pullerits, T.; Sundstroem, V. *Acc. Chem. Res.* **1996**, *29*, 381-389.
19. Papagiannakis, E.; Kennis, J. T.; van Stokkum, I. H. M.; Cogdell, R. J. *PNAS* **2002**, *99*, 6017-6022.
20. Krueger, B. P.; Scholes, G. D.; Fleming, G. R. *J. Phys. Chem. B* **1998**, *102*, 5378-5386.
21. Scholes, G. D.; Fleming, G. R. *J. Phys. Chem. B* **2000**, *104*, 1854-1868.
22. Richter, M. F.; Baier, J.; Cogdell, R. J.; Oellerich, S.; Koehler, J. *Springer Series in Chemical Physics* **2010**, *96*, 513-533.
23. Pullerits, T.; Chachisvilis, M.; Sundstroem, V. *J. Phys. Chem. B* **1996**, *100*, 10787-10792.
24. Chin, A. W.; Huelga, S. F.; Plenio, M. B. *Phil. Trans. Series A* **2012**, *370*, 3638-3657.
25. Jang, S.; Newton, M. D.; Silbey, R. J. *J. Phys. Chem. B* **2007**, *111*, 6807-6814.
26. Cheng, Y. C.; Silbey, R. J. *Phys. Rev. Lett.* **2006**, *96*, 028103/1-028103/4.
27. Novoderezhkin, V.; van Grondelle, R. *J. Phys. Chem. B* **2013**, *117*, 11076-11090.
28. Isom, D. G.; Castaneda, C. A.; Cannon, B. R.; Velu, P. D.; Bertrand, G. M. E. *PNAS* **2010**, *107*, 16096-16100.
29. Leonova, M. M.; Fufina, T. Y.; Vasilieva, L. G.; Shuvalov, V. A. *Biochemistry (Moscow)* **2011**, *76*, 2355-2363.
30. Yu, D.; Huang, G.; Xu, F.; Wang, M.; Liu, S.; Huang, F. *Photosynth. Res.* **2014**, *120*, 311-321.
31. Pflock, T.; Dezi, M.; Venturoli, G.; Cogdell, R. J.; Köhler, J.; Oellerich, S. *Photosynth. Res.* **2008**, *95*, 291-298.

32. Kobuke, Y. *Eur. J. Inorg. Chem.* **2006**, *12*, 2333-2351.
33. Guldi, D. M. *Chem. Soc. Rev.* **2002**, *31*, 22-36.
34. Nakamura, Y.; Aratani, N.; Osuka, A. *Chem. Soc. Rev.* **2007**, *36*, 831-845.
35. D'Souza, F.; Maligaspe, E.; Ohkubo, K.; Zandler, M. E.; Subbaiyan, N. K.; Fukuzumi, S. *JACS* **2009**, *131*, 8787-8797.
36. Satake, A.; Kobuke, Y. *Org. Biomol. Chem.* **2007**, *5*, 1679-1691.
37. Choi, M. S.; Yamazaki, T.; Yamazaki, I.; Aida, T. *Angewandte Chemie, Int. Ed.* **2003**, *43*, 150-158.
38. Frechet, J. M. J.; Adronov, A.; Robello, D. R. *Polymer Preprints* **2000**, *41*, 851-852.
39. Kamat, P. V. *J. Phys. Chem. C* **2008**, *112*, 18737-18753.
40. Nam, Y. S.; Shin, T.; Park, H.; Magyar, A. P.; Choi, K.; Fantner, G.; Nelson, K. A.; Belcher, A. M. *JACS* **2010**, *132*, 1462-1463.
41. Endo, M.; Fujitsuka, M.; Majima, T. *Chem. Eur. J.* **2007**, *13*, 8660-8666.
42. Miller, R. A.; Presley, A. D.; Francis, M. B. *J. Am. Chem. Soc.* **2007**, *129*, 3104-3109.
43. Hammarstrom, L.; Hammes-Schiffer, S. *Acc. Chem. Res.* **2009**, *42*, 1859-1860.
44. Siu, H.; Duhamel, J. *J. Phys. Chem. B* **2008**, *112*, 15301-15312.
45. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd Ed. Springer, USA, **2010**.
46. Lee, D.; Greenman, L.; Sarovar, M.; Whaley, K. B. *J. Phys. Chem. A* **2013**, *117*, 11072-11085.
47. Liu, X.; Cole, J. M.; Low, K. S. *J. Phys. Chem. C* **2013**, *117*, 14731-14741.
48. Chen, H.; Ahsan, S. S.; Santiago-Berrios, M. B.; Abruña, H. D.; Webb, W. W. *J. Am. Chem. Soc.* **2010**, *132*, 7244-7245.
49. Van Grondelle, R.; Novoderezhkin, V. I.; Dekker, J. P. In *Photosynthesis in silico*; Laisk, A.; Nedbal, L.; Govindjee, B., 1st Ed.; Springer: The Netherlands, 2009; Vol. 29; 33-53.
50. Panchuk-Voloshina, N.; Haugland, R. P.; Bishop-Stewart, J.; Bhalgat, M. K.; Millard, P. J.; Mao, F.; Leung, W. Y.; Haugland, R. P. *J. Histochem. Cytochem.* **1999**, *47*, 1179-1188.
51. Laemmli, U. K. *Science* **1970**, *227*, 680-685.
52. Garimella, P. D.; Datta, A.; Romanini, D. W.; Raymond, K. N.; Francis, M. B. *J. Am. Chem. Soc.* **2011**, *133*, 14704-14709.

Chapter 4

Synthesis of a water-soluble phthalocyanine for bioconjugation

4.1 Abstract

This chapter represents a slight departure from the previous discussion of protein assemblies for light harvesting, and focuses on the synthesis of a small molecule sensitizer capable of bioconjugation. Phthalocyanines are a class of chromophores that have had much use in many applications, both industrial and technological. In particular, we are interested in using phthalocyanines as protein conjugates due to their NIR absorption, as well as their desirable excited-state electron and triplet sensitization properties. The syntheses of water-soluble phthalocyanines for attachment to proteins will be described, with the ability to both extract energy from the previously described TMVcp templated light harvesting systems, as well as to generate singlet oxygen in photodynamic therapy applications using spherical viral capsids. Synthesis and characterization of symmetric and asymmetric phthalocyanines will be presented, in addition to preliminary investigations into their singlet oxygen generation properties once conjugated to the bacteriophage MS2 viral capsid.

4.2 Introduction

At the heart of all solar harvesting processes lies a charge-separation event, converting an initial photoexcitation into an electron-hole pair. Biological systems accomplish this within their reaction centers, where a “special pair” of strongly coupled chlorophylls is oxidized.¹ The high energy electrons released are stored as chemical energy in the form of NADPH and ATP for further use in cellular processes.² Oxidized chlorophylls are then subsequently re-reduced through, for instance, water oxidation in plants, to reset the reaction center for another turnover. In photovoltaic materials, the photoexcitation is directly converted into electrical potential through various mechanisms. Dye-sensitized solar cells use an adsorbed photosensitizer to inject high energy electrons into a semiconductor interface,^{3,4} whereas organic photovoltaics achieve charge separation through charge separation at a donor/acceptor interface.^{5,6}

The light harvesting systems described in Chapter 2 contained light absorption and energy transfer functionality, but lacked a component for charge separation and collection. We wished to extend the scope of these synthetic systems and develop a water-soluble photosensitizer suitable for protein bioconjugation that is capable of both solar harvesting as well as electron transfer processes. The introduction of such a molecule will allow extraction of the absorbed energy and create a truly biomimetic artificial light harvesting system.

While the study of these light harvesting systems for academic purposes is a major aspect of our research, our group is also very interested in applying our viral templated light harvesting systems toward producing functional materials. In particular, our lab has put much effort towards developing viral capsid based materials for medical diagnostics and therapeutic purposes. The many examples outlined in Chapter 1 highlight the utility of these materials, constructed from a variety of typically spherical viral capsids. Previous work in our lab has primarily focused on the bacteriophage MS2 viral capsid, a spherical assembly 27 nm in diameter. In previously described work, strategies have been developed for the differential modification of the interior and exterior surfaces of the hollow structures. This allows transport of internal cargo for drug delivery, medical imaging, or photodynamic therapy (PDT) agents while simultaneously containing external targeting groups.^{7,8,9,10,11,12}

We were inspired by the elegant experiments done by Stephanopolous and coworkers, who integrated anionic porphyrins into the MS2 scaffold, producing both light harvesting structures as well as conjugates capable of singlet oxygen generation and targeted PDT.^{13,11} In these cases, they synthesized a porphyrin containing three sulfonate groups to impart water solubility, and a functional group for protein conjugation. They then functionalized the interior with a donor fluorophore capable of energy transfer to the porphyrin residing on the exterior of the capsid surface and monitored the electron transfer properties to methyl viologen in solution. Alternatively, a porphyrin could be installed on the interior surface of MS2, and a DNA aptamer targeting group attached to the exterior. They showed that the aptamer could effectively target specific cancer cell lines, and upon irradiation at the porphyrin Soret band, induce singlet oxygen generation and ultimately cell death through triplet-triplet sensitization. This system proved to be effective, although in a medical context the blue 420 nm light needed to excite porphyrins is less than ideal. Tissue penetration at this wavelength is much lower than in the red region of the visible spectrum. For this reason, we sought to construct a similar system, but rather with phthalocyanines, a class of molecules having much in common with porphyrins structurally, but with primary absorbances near 700 nm. This chapter will outline our experimental strategy and

the synthesis of phthalocyanines for protein conjugation and use in light harvesting and PDT applications.

When comparing photosensitizers, we became interested in the phthalocyanines for their desirable absorption in the NIR region of the visible spectrum, along with their electron transfer abilities and triplet sensitization properties.¹⁴ Phthalocyanines, or tetrabenzotetraazoporphyrins, represent an important class of pigments, whose value has been known for decades.¹⁵ Today, they are produced on the industrial scale as dyes for such diverse applications as car paint, printer ink, and textile pigments. Aside from their abundant commercial uses, phthalocyanines are also an increasingly important class of molecules as photodynamic therapy (PDT) agents.^{16,17} Recently, Pcs have also played a large role in nanotechnology applications,¹⁸ being used in dye-sensitized solar cells,¹⁹ organic photovoltaics,²⁰ small molecule charge transfer systems,^{21,22,23} nonlinear optical materials,²⁴ and magnetic materials.²⁵ The basic phthalocyanine (Pc) structure is shown in **Figure 1**, along with the conventional numbering of carbons. Through variation of both the metal center, as well as substitutions around the periphery of the ring, Pcs can be produced with a range of photophysical properties. Peripheral substituents lend electronic effects to the macrocycle ring, while the metal center strongly interacts with the ring system through spin orbit coupling, dependent on the size and spin state of the metal. A commonly used Pc contains Zn as the metal center, and for the entirety of this work, all Pcs discussed will be zinc phthalocyanine (ZnPc). ZnPc has a characteristic UV/Vis spectrum with an intense Q-band absorption at 680 nm ($\epsilon > 10^5 \text{ M}^{-1} \text{ cm}^{-1}$). ZnPc has moderate fluorescence quantum yield of 0.3, and a long lived triplet lifetime of *ca* 100 μs .¹⁴

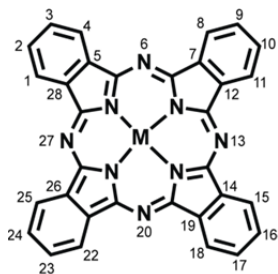


Figure 1. Structure of generic metallophthalocyanine with the common numbering scheme

Phthalocyanines are notoriously insoluble in most solvents, in particular polar solvents such as water. The hydrophobic nature of the aromatic core readily promotes aggregation into columnar stacks through face-to-face interactions of the macrocycles. This aggregation may be desirable in some cases, such as formation of liquid crystals or solid state materials,¹⁴ but for the purposes of triplet sensitization, the monomeric forms of Pcs are preferable.²⁶ Our intention was then to produce a phthalocyanine derivative that contained both a functional group for bioconjugation as well as solubilizing groups for compatibility in aqueous environments and to inhibit aggregation. While there have been reports in other groups of the production of Pcs containing both of these properties, we endeavored to develop a more straightforward and synthetically concise route to generate Pcs capable of bioconjugation.^{27,28}

4.3 Synthesis of phthalocyanines

Typically, phthalocyanines are produced from one of two initial building blocks. Either 1-2, dicyanobenzenes (phthalonitriles) or phthalic acids are condensed to form the basic Pc core structure.²⁸ In the case of phthalonitriles, this is a base-catalyzed process that occurs at elevated temperatures and requires four phthalonitrile units to cyclize and form one Pc. In the phthalic acid route, a nitrogen source (commonly urea) is introduced, and the temperature raised to over 200 °C to promote urea decomposition and in-situ conversion of the phthalic acids to diiminoisoindolines, which then readily cyclize to the Pc, especially when in the presence of a molybdenum catalyst. Both of these methods are accelerated by a metal center to template the ring formation, but can also be accomplished without a metal center to produce the free base of Pc (H₂Pc).

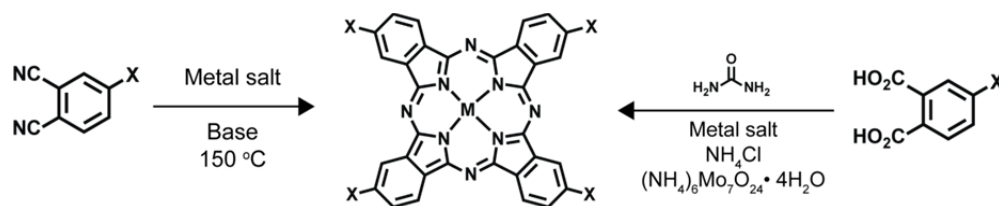


Figure 2. Common general synthetic routes toward phthalocyanine synthesis. The left scheme is done in solution at elevated temperatures and in the presence of base. The right route requires high temperatures of 220 °C, and is done solvent-free allowing all reagents to melt and react in the molten state.

The substituents of either starting material can be virtually any functionality, providing that the group survives the harsh conditions of cyclization and any subsequent purification/workup required. The first requirement of our synthesis was to produce Pcs with water-solubilizing groups to impart compatibility with aqueous environments. The most straightforward way to impart water solubility on this type of hydrophobic molecule is to install electrostatically charged functional groups. From previous work in our lab on synthesis of porphyrin bioconjugates we have observed that cationic porphyrins tended to form strong non-specific interactions with the TMV and MS2 protein capsids, and due to the structural similarity between porphyrins and Pcs, we selected anionic moieties to serve as solubilizing groups. Initially, carboxylic acids were our choice, but we eventually moved on to sulfonic acids to fill this role. Next, we required that there be at least one functional group capable of covalent protein attachment present. This can take on a variety of identities, and will be addressed further as the need arises.

4.3.1 Synthesis of tetra-carboxylic acid functionalized Pcs

The first type of Pc synthesized were symmetric Pcs containing four carboxylic acid groups. These served both to solubilize the Pc, as well as to act as protein conjugation groups. The anionic nature of the acid at neutral pH provided one static negative charge per acidic moiety, and with coupling reagents these can be directly coupled to lysine residues on proteins through formation of amide bonds. Alternatively, small molecule reagents could be coupled to

one of the free acids, producing an asymmetric Pc with one protein reactive functional group, and the other three acids free for solubilization.

To begin this synthesis, we started with 4-hydroxy phthalonitrile as a substrate. The phenolic hydroxyl group can be easily deprotonated with a weak base, allowing a strong nucleophile to be formed capable of S_N2 displacement of primary halides. We found that condensation of 4-hydroxy phthalonitrile with methyl 6-bromo hexanoate proceeded smoothly to produce **1**. This reagent can then undergo cyclization to the Pc **2**, providing a symmetric phthalocyanine with four ester groups connected through ether linkages. This Pc is still compatible with organic solvents, and the synthesis was carried out with the esters intact to facilitate the workup and purification in organic solvents. Flash chromatography was used for purification, and the resulting Pc **2** was obtained in high purity. Quantitative saponification of the esters afforded **3**, a Pc containing four free carboxylic acid groups.

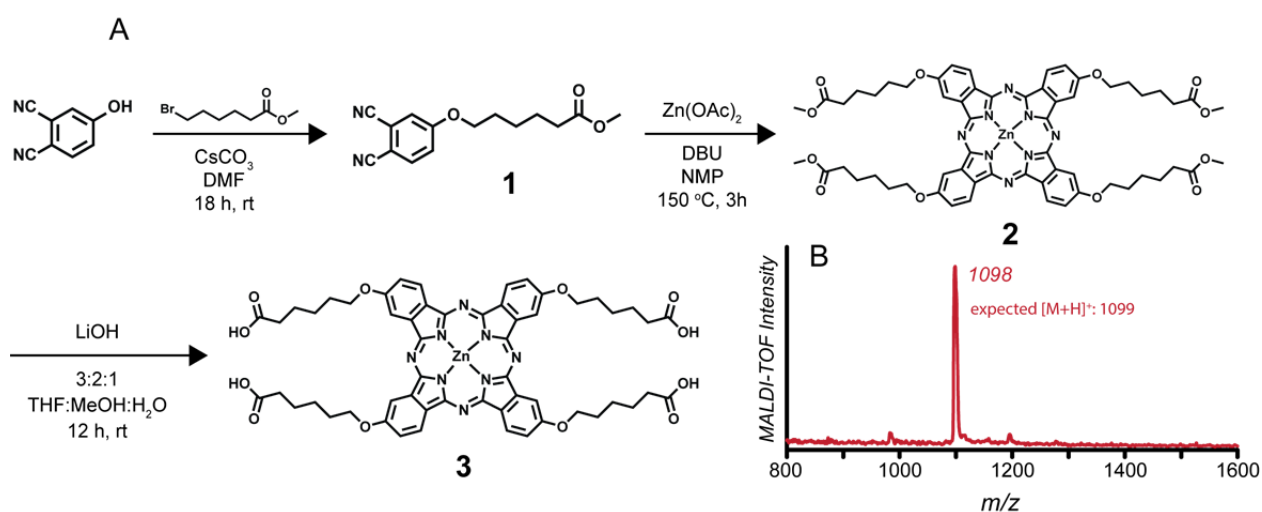


Figure 3. Synthesis of tetracarboxylate Pc **3**. (A) Description of synthetic route to produce **3**. (B) MALDI-TOF MS spectrum of **3** after saponification.

The Pc **3** was obtained in high purity as determined through mass spectrometry, and it was clear that the high temperatures of the ring cyclization did not affect the ester groups, as well as leaving the ether linkage between the phthalonitrile core and alkyl group intact. UV/Vis analysis of **3** in THF showed the typical phthalocyanine absorption character with an intense Q-band absorption at 680 nm and a Soret band absorption centered at 375 nm. However, when **3** was dissolved in H₂O, we observed a change in the Q-band absorption indicative of aggregation into dimeric species. The formation of face-to-face stacked dimers in aqueous solution is well known,²⁹ and results in a decreased intensity of the major peak at 680 nm (attributed to monomeric species) with concomitant increase in the relative intensity of a 630 nm absorption (associated with a dimer). Although the Pc **3** could successfully be conjugated to proteins, this did not appear to remove the dimer character all together, and we decided to move toward a stronger solubilizing group in the form of sulfonic acids.

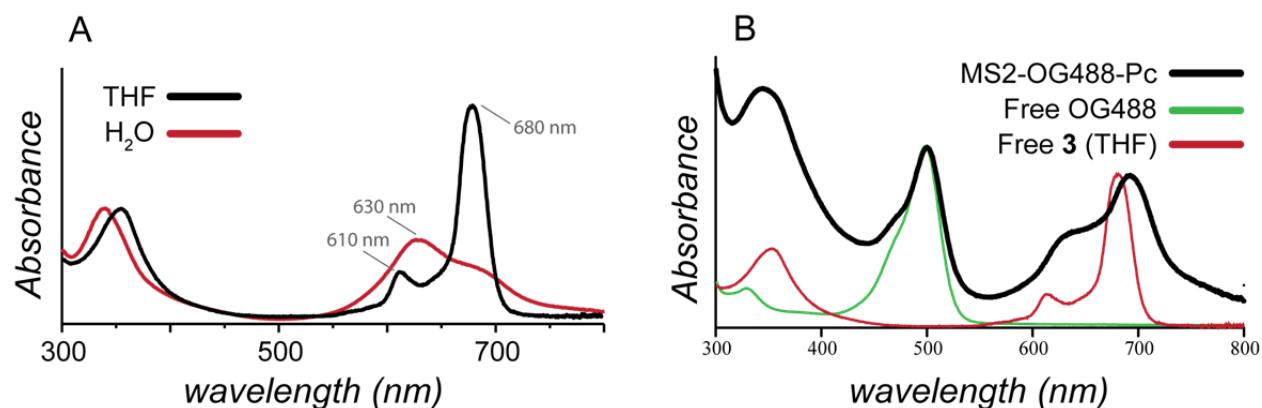


Figure 4. (A) Absorption spectra of **3** in THF vs. H₂O. The dimeric species is apparent from the decreased extinction at 680 nm and the increase of absorption at 630 nm. (B) Absorption spectra of **3** conjugated to pre-fluorescently labeled MS2 coat protein. This protein contains an N87C mutation on the interior of the capsid where OG488 was attached prior to Pc modification. The black trace is the spectrum of the entire construct, while the green and red lines represent absorption spectra of isolated OG488 and **3**, respectively. The Pc **3** was attached through amide bond formation to exterior native lysines of the MS2 capsid. Note the dimeric signatures still apparent, although to a lesser degree than **3** directly dissolved in water as in (A).

4.3.2 Synthesis of asymmetric sulfonated phthalocyanines

Some of the most highly charged and effective water-solubilizing functional groups are sulfonic acids. Tetra-sulfonate ZnPc is readily available, and has shown promise in PDT applications, though it cannot be coupled directly to proteins as is the case with tetra-carboxylate Pc **3**. For the molecule to be able to participate in protein conjugation, we needed to synthesize an asymmetric Pc, with one of the four units capable of bioconjugation and the other three units containing sulfonic acids. There are examples in the literature of asymmetric Pc formation using methods such as solid phase synthesis³⁰ or by going through subphthalocyanine intermediates.³¹ However, many of these methods are synthetically laborious, or simply not compatible with functional groups for bioconjugation. Our strategy instead involved the mixed condensation of phthalonitrile and phthalic acid precursors to produce a distribution of Pcs with varying isomeric compositions. After condensation, the mix of Pcs must then be separated based on the substitution pattern. This separation step is often the most difficult, but can be accomplished with careful selection of reverse-phase HPLC (RP-HPLC) conditions.

After much trial and error in determining compatible starting materials and reaction conditions, the outline of our synthetic scheme is shown in **Figure 5**. We began with a sulfo-phthalic acid and an amino phthalonitrile, in a 3:1 ratio. By using the urea melt method,³² we were able to produce a mixture of isomers with the major species as the tetra-sulfonate (TSZnPc) and the tri-sulfo, mono amino Pc **4**. The crude reaction mixture was then applied to a C18 reverse phase column and run on a preparatory scale HPLC to separate the different isomers. Due to the strong similarity between the different species, the separation was very difficult, but eventually after settling on a gradient of phosphate buffer and methanol, the tri-sulfonated mono amino Pc was obtained in reasonable purity. The HPLC trace is shown in **Figure 6**, and it is clear that the peaks do not display baseline separation, but still elute as individual species. In these conditions, the initial and largest peak corresponds to the TSZnPc and the subsequent

peaks are identified as the other structural isomers, as well as positional isomers inherent in the mixed condensation reaction. All fractions corresponding to Pcs were collected, and through MALDI analysis, the desired products were collected and combined. This process results in the collection of asymmetric tri-sulfonated Pc containing an aniline moiety directly attached to the Pc ring. We then wished to convert the aniline to a maleimide group capable of protein conjugation. The following steps are adapted from a synthesis of a similar anionic maleimide-containing porphyrin.³³

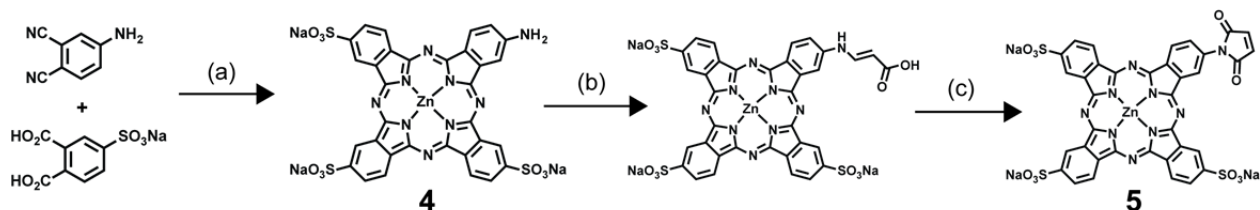


Figure 5. Synthesis of trisulfonated maleimide functionalized Pc. (a) Urea, NH_4Cl , $(\text{NH}_4)_6\text{Mo}_7\cdot 4\text{H}_2\text{O}$, $\text{Zn}(\text{OAc})_2$, 220°C , 3 h. (b) Maleic anhydride, MeCN, 85°C for 5 h, then rt for 12 h. (c) Acetic anhydride, NaOAc, 80°C , 1.5 h.

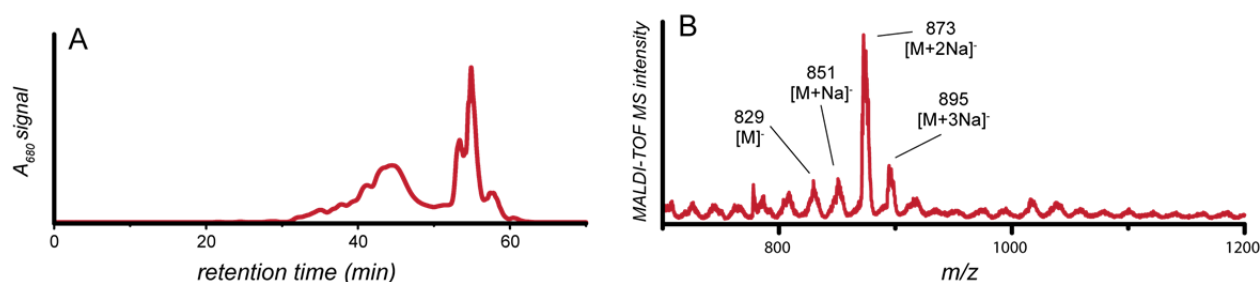


Figure 6. (A) Representative HPLC trace of **4**. Chromatography was accomplished on a C18 semipreparative column with a gradient of 0-100% MeOH in 10 mM phosphate buffer, pH 7.0 over 70 min. The Pcs elute with increasing hydrophobicity, with the tetrasulfonic acid containing Pcs appearing first, and the later narrow peaks corresponding to single and double amine containing Pcs. All fractions were collected, and MALDI-TOF MS was used for determination of identity of individual species. Along with the different compositional isomers in the crude reaction mixture, there are positional isomers that complicate the chromatogram further. (B) Negative-mode MALDI-TOF MS spectrum of **4** showing ionization primarily as the disodium adduct.

4.3.3 Installation of maleimide ring onto trisulfonated Pc

Pc **4** is insoluble in all solvents except for H_2O , MeOH, and DMF/DMSO. This is a problem when attempting to convert the amino group to a maleimide, so after purification of **4**, an ion pair was formed between the sulfonate groups and tetrabutylammonium cation (Bu_4N^+). This allowed phase transfer, and displayed high solubility in organic solvents. Typically, an aqueous solution of **4** was formed, and an excess of Bu_4NCl was added. The Pcs could then be extracted into CHCl_3 by simply shaking in a separatory funnel, and after evaporation of solvent, stored as the organic compatible **4** ion-paired with hydrophobic Bu_4N^+ . Next, this was reacted with maleic anhydride in MeCN to install a maleic acid group, followed by ring closing promoted by acetic anhydride to form **5**. Pc **5** was then stored as a concentrated DMF stock, poised for subsequent bioconjugation.

4.3.4 Characterization of trisulfonated maleimide Pc

Readily dissolving in water, we were confident that **5** would be a valuable sensitizer for both electron transfer and triplet sensitization applications. It showed the capability of conjugation to proteins through thiol-maleimide chemistry as expected, though not in optimal yields. Additionally, we observed primarily dimeric character to the absorption spectrum taken in water. This finding was discouraging, as it appears that even the sulfonated Pcs retained dimeric character in aqueous solution, and even after attachment to proteins. The absorption spectrum of **5** is shown in DMF and in H₂O, normalized to highlight the decrease in extinction upon dimerization. Also shown is the absorption profile of **5** after attachment to the N87C position of MS2. This position is on the interior of the protein, and based on the degree of labeling, would be expected to place the Pcs in isolated environments on the protein surface.

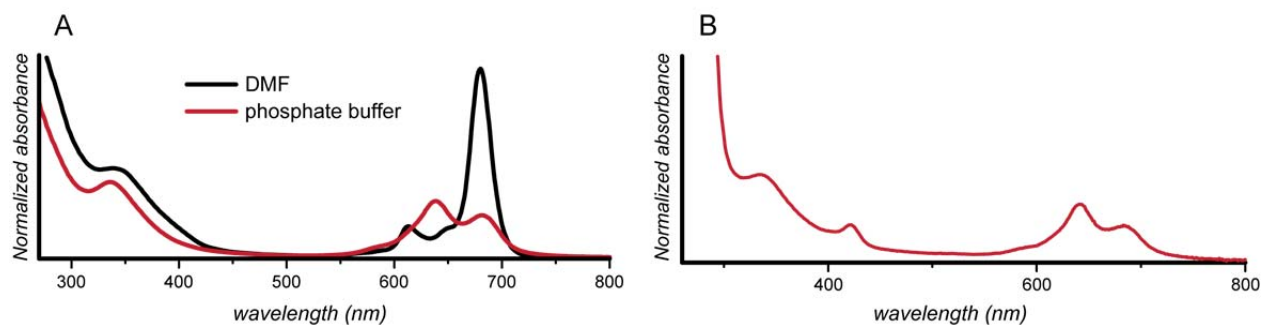


Figure 7. (A) Normalized absorption spectra of **5** in DMF vs. aqueous buffer. The two spectra contain equal concentrations of **5**, highlighting the solvent effect and spectral manifestation of dimer formation in aqueous environments. (B) Upon conjugation to the interior of MS2 capsid, the Pcs do not regain monomeric absorption and continue to resemble dimeric species.

To validate the degree of protein modification, we turned to inductively coupled plasma atomic absorption spectroscopy (ICP-AES) to quantify the number of Zn atoms, and therefore Pc molecules contained in the Pc-MS2 conjugates. Our primary source for protein modification quantification is HPLC-MS, and we have observed suppression of ionization by Pcs, skewing the MS data. Comparing the recorded MS quantification to ICP-AES measurements showed that indeed the Pc was suppressing ionization, and in fact **5** is much more reactive towards proteins than the initial MS data suggests. From the ICP-AES data, we observed that after attachment to MS2 and purification of the bioconjugate, MS2 capsids could be produced with 16% of the monomers modified with **9**, containing on average 29 Pcs per capsid.

4.4 Generation of singlet oxygen

Although the optical properties of **5** were not optimal for PDT applications due to the dimer character, we investigated the ability of **5** to generate singlet oxygen for future use in PDT. It is a well-known that the dimer species of Pcs are much less active in singlet oxygen generation, stemming from a combination of a lower overall extinction coefficient in conjunction

with a much shorter-lived triplet state.³⁴ Regardless, we continued our analysis and quantified the amount of singlet oxygen generated by using an assay based on the bleaching of anthracene-9,10-dipropionic acid (ADPA).

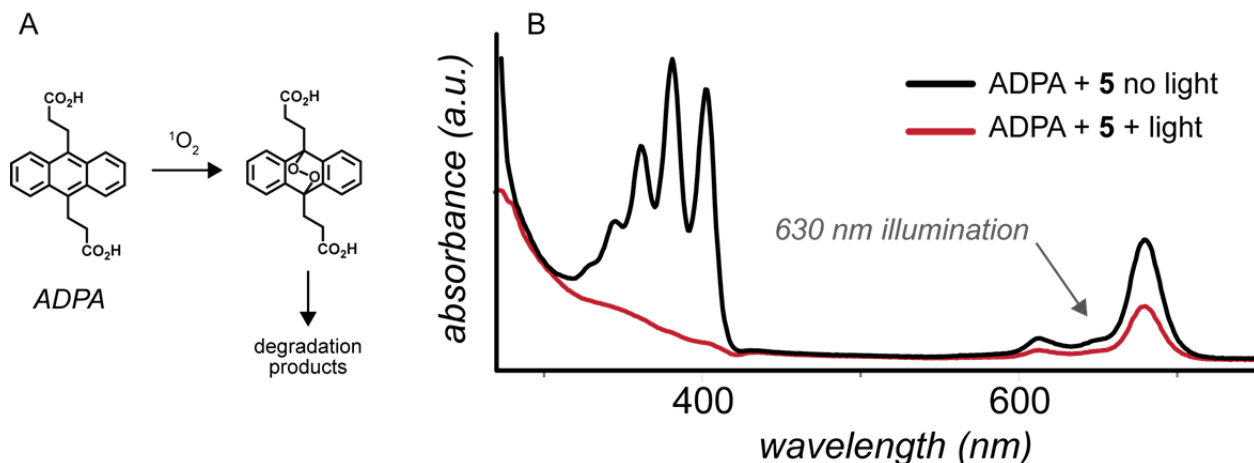


Figure 8. (A) Structural basis of singlet oxygen degradation of ADPA. (B) Results from singlet oxygen generation using **5**. ADPA and **5** were mixed in DMF at concentrations of 1 mM and 1 μ M, respectively. The red trace shows the absorbance immediately after illumination at 630 nm for 20 min. The black trace is the same solution kept in the dark with no 630 nm illumination. After 20 min of exposure to red light, there is no detectable ADPA remaining based on the absorbance at around 400 nm.

ADPA undergoes a Diels-Alder cycloaddition with singlet oxygen, producing a peroxo-intermediate, followed by rapid degradation (see **Figure 8**).³⁵ Monitoring the decrease in ADPA absorbance, it is possible to quantify the amount of singlet oxygen produced and consumed by ADPA. We first ran this assay in an organic solvent, DMF, in the absence of protein. A solution of **5** and excess ADPA was subjected to irradiation at the Pc Q-band with a red lamp obtained from a commercial source. This light source is composed of LEDs with an emission profile centered at 630 nm, and is well suited for excitation of Pcs. Upon illumination for 20 minutes, **5** degraded ADPA to undetectable levels, showing that in this solvent system, the Pc **5** displays a high degree of singlet oxygen generation. We then went on to monitor the singlet oxygen generation ability of MS2-**5** conjugates. Once again, an ADPA assay was employed, and solutions of either **5** or MS2-**5** were made with excess ADPA in buffered aqueous solution. Illumination with the 630 nm lamp provided the data presented in **Figure 9**. Qualitatively, it is clear that in both cases ADPA is consumed, but not nearly to the quantitative degree that was accomplished in DMF as a solvent. Through quantification of the reaction by monitoring the decrease in ADPA absorbance while knowing both initial concentrations of ADPA and Pc, it was found that **5** degraded 360 molecules of ADPA *per* Pc molecule. Similarly, the MS2-**5** conjugate was observed to destroy 355 molecules of ADPA *per* Pc. To put these numbers in perspective, the MS2-porphyrin conjugate produced in Reference 13 was observed to produce 2,000 molecules of singlet oxygen per porphyrin species when illuminated at the porphyrin Soret band.

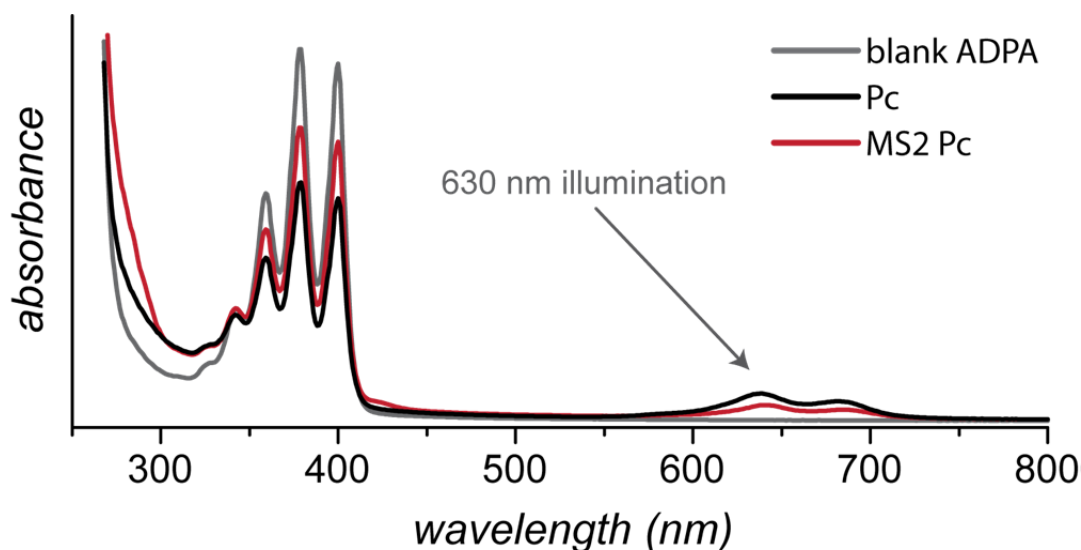


Figure 9. Degradation of ADPA by **5** and MS2-**5** in phosphate buffer, pH 9. Loss of ADPA was quantified by measuring the decrease in absorbance at 403 nm. All spectra were illuminated for 20 min with a 630 nm lamp, the grey trace containing no phthalocyanine, the green trace with ADPA + **5**, and the red trace with ADPA + MS2-**5**. From these measurements, each molecule **5** degraded 360 molecules of ADPA when alone in solution, and degraded 355 molecules of ADPA when attached to MS2 capsids.

4.5 Conclusion and perspectives

While the results described in our synthesis of phthalocyanines for bioconjugation and applications in PDT are still being optimized, these experiments represent a synthetically tractable route toward Pcs capable of bioconjugation and a step forward in the development of novel protein-based PDT materials. We successfully produced tetra-carboxylic acid functionalized phthalocyanines capable of covalent protein attachment, and also described a method for producing asymmetric sulfonated Pcs containing a maleimide group. The maleimide-Pc was shown to be an effective singlet oxygen generator in organic solvents based on an ADPA degradation assay. The maleimide-Pcs were coupled to the MS2 viral capsid to moderate levels of 16%, installing nearly 30 Pc molecules on the interior of a single MS2 capsid. Additionally, the MS2-Pc conjugates displayed low to moderate levels of activity in singlet oxygen assays. We feel that while these results do not presently show a system capable of efficient PDT in biological environments, they show promise for further optimization and improvement if the aggregation state of the phthalocyanines can be properly addressed.

4.6 Materials and methods

General methods

Unless otherwise noted, all chemicals and solvents were of analytical grade and used as received from commercial sources without further distillation or purification. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F₂₅₄ plates with visualization by ultraviolet (UV) irradiation at 254 nm and/or potassium permanganate stain. Purifications by flash chromatography were performed using EM silica gel 60 (230-400 mesh). The eluting system for each purification was determined by TLC analysis. All organic solvents were removed under reduced pressure using a rotary evaporator. Water (dd-H₂O) used as reaction solvent was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Spin concentration steps were performed using 100 kDa molecular weight cutoff spin concentrators from Millipore (Billerica, MA). Centrifugations were performed with an Eppendorf 5415 benchtop centrifuge, (Eppendorf, Hauppauge, NY).

Instrumentation and sample analysis

Spectroscopy. UV-Vis spectroscopic measurements were conducted on a Cary 50 Scan benchtop spectrophotometer (Varian Inc., USA). All measurements were done with sample OD between 0.1-1.0. Steady-state fluorescence measurements were obtained on a Fluoromax-4 spectrofluorometer (Horiba Scientific, USA). All fluorescence spectra were taken with sample OD between 0.05-0.1. Protein concentration was routinely determined by UV/Vis analysis on a Nanodrop 1000 instrument (Nanodrop, USA) by monitoring absorbance at 280 nm.

NMR. ¹H and was measured with a Bruker AVQ-400 (400 MHz) spectrometer. ¹H NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to CDCl₃ (δ 7.26, singlet). Multiplicities are reported as follows: s (singlet), d (doublet), t (triplet), dd (doublet of doublets), br (broad) or m (multiplet). Coupling constants are reported as a J value in Hertz (Hz). The number of protons (n) for a given resonance is indicated as nH and is based on spectral integration values.

Mass Spectrometry. Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DE system (PerSeptive Biosystems, USA) and data were analyzed using Data Explorer software. Samples were co-crystallized with 10 mg/mL α -cyano-4-hydroxycinnamic acid or sinipinic acid in 1:1 acetonitrile (MeCN) to H₂O with 0.1% trifluoroacetic acid (TFA). Protein bioconjugates were analyzed using an Agilent 1200 series liquid chromatograph (Agilent Technologies, USA) that was connected in-line with an Agilent 6224 Time-of-Flight (TOF) LC/MS system equipped with a Turbospray ion source. Spectra were quantified by peak area integration using Origin 8.0 software (Originlab, USA).

High Performance Liquid Chromatography. HPLC was performed on Agilent 1100 Series HPLC Systems (Agilent, USA). Sample analysis for all HPLC experiments was achieved with an

in-line diode array detector (DAD) and in-line fluorescence detector (FLD). Analytical, semi-preparatory or preparatory reverse-phase HPLC of small molecules was accomplished using a C18 stationary phase and a H₂O/ MeCN with 0.1% TFA gradient mobile phase. Size exclusion chromatography (SEC) was performed using a Phenomenex polysep GFC-P-5000 column (4.6 x 250 mm) (Phenomenex, USA) at 1.0 mL/min using a mobile phase of 10 mM phosphate buffer, pH 7.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus (Bio-Rad, Hercules, CA), using a 10-20% precast linear gradient polyacrylamide gel (Bio-Rad). The sample and electrode buffers were prepared according to Laemmli.³⁶ All protein electrophoresis samples were heated for 5-10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of disulfide bonds. Gels were run for 70 min at 120 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad). Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA). ImageJ was used to determine the level of modification by optical densitometry.

Small molecule synthesis

Synthesis methyl-6-hexanoate phthalonitrile (1). To a flame-dried 5 mL RBF were added 4-hydroxyphthalonitrile (14.4 mg, 0.1 mmol) and cesium carbonate (32.6 mg, 0.11 mmol). The solids were diluted with DMF prior to addition of methyl 6-bromohexanoate (25.4 mg, 1.1 mmol). The reaction was allowed to stir at room temperature for 18 h, at which time it was diluted with water and extracted with EtOAc. The solvent was removed under reduced pressure and the product purified on silica gel over a gradient of 10-50% EtOAc in hexanes. The solvent was removed under reduced pressure and the product was dried *in vacuo*. ¹H NMR (400 MHz, CDCl₃): δ 7.68 (d, 1H), 7.22 (d, 1H), 7.13 (dd, 1H), 4.12 (q, 2H), 4.02 (t, 2H), 2.32 (t, 2H), 1.83 (m, 2H), 1.7 (m, 2H), 1.5 (m, 2H), 1.2 (t, 3H).

Synthesis of tetra methyl ester phthalocyanine (2). Product **1** (20 mg, 73 μmol) was added to a 20 mL RBF along with zinc acetate (4 mg, 18.3 μmol). A combination of 1 mL of NMP and 100 μL of DBU were added, and the reaction heated to 150 °C for 3 h. The reaction was allowed to cool to room temperature and purified on silica gel using 10% THF in EtOAc. The colored green fraction was collected as pure **2**. MALDI-TOF MS: for C₆₀H₆₄N₈O₁₂Zn, calculated 1154.6 [M+H]⁺, found 1153 [M+4H]⁺.

Synthesis of tetra carboxylic acid phthalocyanine (3). Product **2** (20 mg, 1.7 μmol) was dissolved in 1 mL of a 3:2:1 mixture of THF:MeOH:H₂O. LiOH was added in excess, and the mixture stirred overnight at room temperature. The solvent was removed under reduced pressure to afford **3** in quantitative yield. MALDI-TOF MS: for C₅₆H₅₆N₈O₁₂Zn, calculated 1098.5 [M+H]⁺, found 1099 [M+4H]⁺.

Synthesis of mono amino trisulfonatophthalocyanine (4). Sulfophthalic acid (936 mg, 3 mmol), 4-aminophthalonitrile (143.2 mg, 1 mmol), urea (2.4 g, 40 mmol), ammonium chloride (214 mg, 4 mmol), ammonium molybdate (5 mg, 4 μ mol), and zinc acetate (219 mg, 1 mmol) were mixed into a 50 mL RBF, and the solid powder homogenized thoroughly. The mixture was then heated to 220 °C in a sand bath and allowed to form a molten state and react for 3 h. The mixture was cooled to room temperature and washed with EtOAc. The aqueous solvent was then removed under reduced pressure. Next, the crude reaction mixture was dissolved in a small amount of water and passed through a C18 sep-pak column (Waters, USA) as a rough purification step before using a preparatory scale HPLC equipped with a C18 reverse phase column. The phthalocyanines eluted from a 0-100% gradient of MeOH in phosphate buffer, pH 7. Fractions were analyzed by MALDI, and those containing the product **4** were collected and combined. MALDI-TOF MS: for $C_{32}H_{17}N_9O_9S_3Zn$, calculated 832.1[M-H]⁻, found 829 [M-4H]⁻; 851 [M+Na]⁻; 873 [M+2Na]⁻; 895 [M+3Na]⁻.

Synthesis of trisulfonato-maleimidophthalocyanine (5). Compound **4** was first made compatible with organic solvents through introduction of a phase transfer cation tetrabutylammonium. The water-soluble **4** was diluted with water, and excess tetrabutylammonium chloride was added. Extraction into $CHCl_3$ proved unproblematic, and the extraction was repeated until no color remained in the aqueous layer. Removal of the solvent under reduced pressure left an oily solid. The tetrabutyl ammonium capped sulfonated phthalocyanine (117 mg, 70 μ mol) was then diluted into 1.5 mL of dry MeCN. Maleic anhydride (10.4 mg, 106 μ mol) was then added, and the reaction was refluxed 5 h. After cooling to room temperature, the solid was taken on to the next step without further purification. Directly to the material was added 2 mL of acetic anhydride and NaOAc (9 mg, 106 μ mol). The resulting solution was heated to 80 °C for 1.5 h under a condenser and then cooled to room temperature. The solvent was removed under reduced pressure. This product is not compatible with silica gel, so purification entailed first extracting into DI water followed by back extraction into $CHCl_3$ with excess tetrabutylammonium chloride. A final pass through a C18 sep pak column (Conditioned with MeOH and water, eluted in water) removed any small molecules remaining from the synthesis. MALDI-TOF MS: for $C_{36}H_{17}N_9O_{11}S_3Zn$, calculated 912.1[M-H]⁻, major species found 935 [M+Na]⁻

Protein expression and modification

Expression and purification of T19pAF N87C MS2. The MS2 capsids used in these experiments contained an N87C internal mutation, as well as an external *p*-aminophenylalanine unnatural amino acid at the T19 position. Protein with the unnatural amino acid was expressed and purified as previously reported.³⁷ The plasmid containing the amber stop codon mutation in the MS2 coat protein was co-transformed into DH10B cells with the pDULE plasmid containing the pAF aminoacyl tRNA synthetase and tRNA. The expression was carried out in minimal media, following the published protocol. Cultures were grown 24 h at 37 °C, harvested by centrifugation, and stored at -80 °C. Cells (from a 1 L expression batch) were thawed, resuspended in 20 mL of 20 mM TEA pH 7.2 containing 10 μ g/mL DNase and RNase, and lysed by sonication with a 2 s on, 6 s off cycle for a total of 20 minutes using a standard disruptor

horn at 90% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared by ultracentrifugation for 45 min at 11,000 rpm using a SLA-600TC rotor in a Sorvall RC5C Plus centrifuge (Waltham, MA). The clarified lysate was decanted and applied to a DEAE-sephadex column (GE Healthcare) equilibrated with 20 mM taurine, pH 9.0. An isocratic gradient of 3 mL/min was used, and the first 60 mL of elution was collected. The eluent was then stirred while adding a saturated solution of ammonium sulfate dropwise to a final concentration of 50% (v/v) and placed on a rotator at 4 °C. After 12 h of precipitation at 4 °C, the white precipitate was pelleted by ultracentrifugation (45 min at 11,000 rpm using a SLA-600TC rotor) and further purified on a Sephacryl-500 gel filtration column (GE Healthcare). Purity was confirmed by SDS-PAGE and HPLC/ESI-MS. Pure fractions were pooled, and fractions containing desired MS2 in addition to impurities were subjected to a second gel filtration purification. Approximately 10 mg of purified protein was recovered per L of expression culture.

General procedure for cysteine modification of T19pAF N87C MS2. To 100 μ L of MS2 at 100 μ M was added 5 equivalents of maleimide functionalized chromophore from a concentrated (20 mM-100 mM) DMF stock. The reaction was stirred at room temperature for 2 h, and the unreacted chromophore was removed through application to a NAP-5 column and subsequent spin concentration.

General procedure for lysine modification of MS2. Carboxylic acid containing chromophore was initially activated by stirring with 5 equivalents HCTU and 10 equivalents DiPEA in 10 μ L of DMF. This mixture was allowed to activate for 10 min prior to addition to the protein. To 200 μ L of MS2 at 100 μ M was added the DMF solution of activated chromophore. The reaction was stirred at room temperature for 2 h, and the unreacted chromophore was removed through application to a NAP-5 column and spin concentration.

Purification of dye-protein conjugates. All chromophore labeled proteins were subjected to HPLC purification prior to spectral analysis. SEC was performed, and the peak corresponding to the MS2 was collected in conditions of 1 mL/min of 10 mM phosphate, pH 7.2. The fractions were then combined and subjected to spin concentration to obtain the desired protein concentration. Samples were then stored away from light at 4 °C.

4.7 Literature cited

1. Ermler, U.; Fritzsche, G.; Buchanan, S. K.; Michel, H. *Structure* **1994**, *2*, 925–936.
2. Barber, J. *Chem. Soc. Rev.* **2009**, *38*, 185–196.
3. Grätzel, M. *J. Photochem. Photobiol. C Photochem. Rev.* **2003**, *4*, 145–153.
4. Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. *Chem. Rev.* **2010**, *110*, 6595–6663.
5. Nelson, J. **2002**, *6*, 87–95.
6. Hains, A. W.; Liang, Z.; Woodhouse, M. a; Gregg, B. a. *Chem. Rev.* **2010**, *110*, 6689–6735.
7. Hooker, J. M.; Kovacs, E. W.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 3718–3719.
8. Kovacs, E. W.; Hooker, J. M.; Romanini, D. W.; Holder, P. G.; Berry, K. E.; Francis, M. B. *Bioconjug. Chem.* **2007**, *18*, 1140–1147.
9. Hooker, J. M.; Datta, A.; Botta, M.; Raymond, K. N.; Francis, M. B. *Nano Lett.* **2007**, *7*, 2207–2210.
10. Garimella, P. D.; Datta, A.; Romanini, D. W.; Raymond, K. N.; Francis, M. B. *J. Am. Chem. Soc.* **2011**, *133*, 14704–14709.
11. Stephanopoulos, N.; Tong, G. J.; Hsiao, S. C.; Francis, M. B. *ACS Nano* **2010**, *4*, 6014–6020.
12. Tong, G. J.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B. *J. Am. Chem. Soc.* **2009**, *131*, 11174–11178.
13. Stephanopoulos, N.; Carrico, Z. M.; Francis, M. B. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9498–9502.
14. Okura, I. *Photosensitization of Porphyrins and Phthalocyanines*, 1st Ed.; Gordon and Breach: The Netherlands, 2000.
15. De la Torre, G.; Claessens, C. G.; Torres, T. *Chem. Commun.* **2007**, 2000.
16. Hur, B.; Lier, V.; Lier, V.; Hur, B. **1991**, 53.
17. Allen, C. M.; Sharman, W. M.; Van Lier, J. E. *J. Porphyrins Phthalocyanines* **2001**, *05*, 161–169.
18. Elemans, J. a. a. W.; van Hameren, R.; Nolte, R. J. M.; Rowan, a. E. *Adv. Mater.* **2006**, *18*, 1251–1266.
19. Imahori, H.; Umeyama, T.; Ito, S. *Acc. Chem. Res.* **2009**, *42*, 1809–1818.
20. Walter, M. G.; Rudine, A. B.; Wamser, C. C. *J. Porphyr. Phthalocyanines* **2010**, *14*, 759–792.
21. Seitz, W.; Kahnt, A.; Guldi, D. M.; Torres, T. *J. Porphyr. Phthalocyanines* **2009**, *13*, 1034–1039.
22. Satake, A.; Kobuke, Y. *Org. Biomol. Chem.* **2007**, *5*, 1679–1691.
23. Ohkubo, K.; Fukuzumi, S. *J. Porphyr. Phthalocyanines* **2008**, *12*, 993–1004.
24. De la Torre, G.; Vázquez, P.; Agulló-López, F.; Torres, T. *J. Mater. Chem.* **1998**, *8*, 1671–1683.
25. Satake, A.; Kobuke, Y. *Org. & Biomol. Chem.* **2007**, *5*, 1679–1691.
26. Ogunsipe, A.; Chen, J.; Nyokong, T. *New J. Chem.* **2004**.
27. Dumoulin, F.; Durmuş, M.; Ahsen, V.; Nyokong, T. *Coord. Chem. Rev.* **2010**, *254*, 2792–2847.
28. Mack, J.; Kobayashi, N. *Chem. Rev.* **2011**, *111*, 281–321.
29. Ortí, E.; Brédas, J. L.; Clarisse, C. *J. Chem. Phys.* **1990**, *92*, 1228.

30. Erdem, S. S.; Nesterova, I. V; Soper, S. a; Hammer, R. P. *J. Org. Chem.* **2009**, *74*, 9280–9286.
31. Kudrevich, S.; Brasseur, N.; La Madeleine, C.; Gilbert, S.; van Lier, J. E. *J. Med. Chem.* **1997**, *40*, 3897–3904.
32. Sakamoto, K.; Ohno, E. *Prog. Org. coatings* **1997**, *9440*.
33. Endo, M.; Fujitsuka, M.; Majima, T. *Chem. Eur. J.* **2007**, *13*, 8660-8666.
34. Uslan, C.; Sesalan, B. Ş.; Durmuş, M. *J. Photochem. Photobiol. A Chem.* **2012**, *235*, 56–64.
35. Lindig, B. *JACS* **1980**, *5590–5593*.
36. Laemmli, U. K. *Science* **1970**, *227*, 680–685.
37. Carrico, Z. M.; Romanini, D. W.; Mehl, R. a; Francis, M. B. *Chem. Commun.* **2008**, 1205–1207.