Title

Models of interacting photosynthetic light harvesting complexes constructed from modified tobacco mosaic virus capsid proteins

Author

Bischoff, Amanda Joy

Publication Date

2023

Peer reviewed|Thesis/dissertation

Models of interacting photosynthetic light harvesting complexes constructed from modified tobacco mosaic virus capsid proteins

by

Amanda Joy Bischoff

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 Michelle C. Chang
Professor Anastasios Melis

Summer 2023

Models of interacting photosynthetic light harvesting complexes constructed from modified tobacco mosaic virus capsid proteins

Abstract

Models of interacting photosynthetic light harvesting complexes constructed from modified tobacco mosaic virus capsid proteins

by

Amanda Joy Bischoff

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Matthew B. Francis, Chair

Energy transfer in photosynthesis is orchestrated via dynamic networks of pigment arrays embedded within protein complexes that often span a lipid membrane. This process has been imitated in models that seek both to elucidate the structural characteristics leading to efficient energy transfer in photosynthetic organisms and to use these principles to design more efficient or "greener" solar technology. Pigment conjugates of assemblies of the tobacco mosaic virus capsid protein (TMVCP) have been extensively used as model light harvesting systems due to their synthetic tractability and structural similarity to many photosynthetic light harvesting complexes. This dissertation demonstrates how changes to the TMVCP sequence, including permutation and point mutations, affect the assembly state structure of TMVCP. In doing so, the utility of charge detection mass spectrometry for the characterization of heterogeneous assemblies of large protein complexes is demonstrated. In addition, several methods for dual protein functionalization are developed, which expand the TMVCP light harvesting model to capture interactions between adjacent light harvesting complexes. A site-selective, asymmetric, protein–protein conjugation using two expressed non-canonical amino acids and an oxidative coupling reaction is used to covalently link an engineered TMVCP assembly containing donor pigment arrays to an engineered TMVCP assembly containing acceptor pigment arrays. This model demonstrates directional energy transfer from the donor to acceptor complexes with 21% efficiency, measured by a decrease in donor fluorescence lifetime when donor complexes are in the presence of acceptor complexes versus donor complexes alone. Separately, an engineered TMVCP assembly is site-selectively labeled with a donor pigment, followed by labeling with a non-expressed His tag at an engineered cysteine residue catalyzed by the enzyme tyrosinase. This His tag is associated with a supported lipid bilayer, and the TMVCP light harvesting complexes associated with bilayers demonstrate lateral mobility, imitating the movement of natural photosynthetic light harvesting complexes within phospholipid bilayers. This is the first example of a photosynthetic membrane mimic using entirely synthetic protein, lipid, and pigment components.

In memory of Jiayue (Joy) Feng
who shared my love of chemistry and inspires me to always remain curious

Contents

List of Figures

List of Tables

Acknowledgments

I feel immensely grateful to have spent the last six years doing research to break new ground in the field of chemical biology. This time has been full of challenges, frustrations, and self-doubt, but I have also learned so much, developed confidence, and made wonderful friends along the way. There are many people who made parts or all of this thesis possible. My advisor, Prof. Matt Francis, has been unfailingly kind, compassionate, and encouraging. He has worked with me to come up with novel strategies to address difficult research problems, has shown endless patience as I struggled with a project, and helped boost my spirits when I was despairing about whether any of my projects would work out. He has given me the space to learn, explore, and sometimes fail, and I have learned so much by being a student in his lab. It truly was the best place I could've done my PhD research.

I would like to thank the members of my cohort in chemistry, who were the first people to make me feel that perhaps I truly belonged in the UC Berkeley Chemistry Department. Audrey Reeves has been an amazing friend from the beginning, who was always willing to share advice and seemed to always know the answers to my questions! I feel so lucky to have led the Transfer Student Mentorship Program and publish a manuscript on mentorship together with her last year. Vanha Pham has been a wonderful ski buddy and has been a great example to me of remaining positive in stressful situations. I loved being on the Expanding Your Horizons planning committee with her and appreciate her advice through the years as well. Joy Wang really helped me feel at home during my first year and has helped me broaden my perspective when things were difficult. Anneliese Gest has also been a wonderful friend in and out of lab, who is always willing to chat or give me good advice. My day is always a bit better after talking to Anneliese!

I would also like to thank the people who mentored me throughout my PhD. Shixian Lin and Alec Christian were my first mentors at Berkeley, who helped me learn and showed me grace during a stressful and difficult time. Prof. Chris Chang and the members of his lab took a chance on me right when I started graduate school and provided a lot of support throughout my first year. Rob Nichols taught me how to be a biologist. Prof. Dave Savage was always enthusiastic about science and willing to talk about my projects and give me ideas even after I was done rotating in his lab. Prof. Savage's lab taught me that people can have scientific debates and challenge one another's ideas while being respectful to each another. Ariel Furst read and edited my NSF GRFP proposal before she even met me and has given me amazing support and advice over the years. She also taught me how to work with DNA and reminded me to have fulfilling hobbies outside of the lab. Jing Dai taught me so much about expressing and purifying proteins, bioconjugation, and light harvesting models. Jing was always willing to think deeply about the research I was doing in lab, let me know about opportunities like internships, and help me when I was feeling stuck. I feel so lucky to have been on team TMV with Jing.

The former members of the Francis lab have been incredible colleagues, and I feel privileged to have worked with every one of them. In particular, I would like to thank Kristin Wucherer, who mentored me during my first year and was always willing to troubleshoot

with me. She has given me a lot of wonderful advice and is the rare type of person who offers to help others without them even having to ask. I am so grateful for her advice on protein expression and experiments and all the support she's given reading and helping with applications to fellowships, internships, and postdoc positions. Emily Hartman is a great example of a science communicator, and she has helped me become a better scientist through her example and our conversations. Celine Santiago is an amazing scientist who was always happy to talk about research. She also could always make me laugh and really helped me through difficult times in and out of lab! I appreciate my wonderful labmates from Latimer 738: Sarah Klass, who is a brilliant scientist, great ski buddy, and helped me have a grounded view of research and life; and Tahoe Fiala and Anthony Omole, whose enthusiasm for chemistry was contagious and who helped me feel at home in the Francis lab. Marco Lobba always had something nice or encouraging to say and could make everyone feel important. His example has shown me that I can accomplish difficult things if I believe in myself. Joel Finbloom was and continues to be someone I can always go to for good advice. Laura Sofen has been a great example of doing research with integrity and with consideration of the research's impact, and I have learned a lot from seeing how she engages her audience in research presentations. Vanessa Yang always makes time to check in on how I'm doing and offer encouragement, and has inspired me to jump in and try new things. Nick Dolan's cheerfulness and optimism have helped me feel less stressed. Alan Marmelstein is constantly curious about his research and the world, and has inspired me to do the same. Johnathan Maza has been a wonderful mentor who has given me great advice about my research and career trajectory. Alex Ramsey is someone I could talk to about anything, and I am so glad we got to work on a project together. I learned a lot from Alex's meticulous approach to research, design, and scientific writing. Dan Brauer has given good scientific advice and started the Transfer Student Mentorship Program, which helped me feel more connected to the Berkeley College of Chemistry. I really appreciated conversations with Jamie Gleason, who was willing to talk about difficult topics regarding how scientific research intersects with life.

I am endlessly grateful to the students who allowed me to mentor them on projects in the lab. Through failures and successes, I loved being part of their scientific journeys, and they helped me maintain motivation when I felt discouraged. My first undergraduate mentee, Nicole Chang, helped me regain enthusiasm after my qualifying exams in my second year. Amanda Li, also an undergraduate mentee, has had great and insightful discussions with me about science and life. She showed an amazing ability to adapt from experimental research to virtual research and back again during the COVID-19 pandemic. Undergraduate Cardy Nguyen was enthusiastic in our scientific discussions and always maintained a positive attitude. I am so proud of all of them and know they will continue to do great things in their chosen fields. Avery Tytla has always asked thoughtful questions and has been a great example to me of taking initiative, developing new ideas and setting up interesting collaborations. Deborah Zhuang jumped into the light harvesting project with enthusiasm and has fearlessly taken it in exciting new directions using native phycobilisome proteins. She has inspired me to try new things and has been a great person to talk to, or commiserate

with, about the light harvesting project (depending on the day). The future of Francis lab light harvesting couldn't be in more capable hands.

The current members of the Francis lab have made it a pleasure to go into lab every day. Wendy Cao is always willing to chat about research or life, and talking to her invariably has made my days in lab better! Arya Lall has been an incredible lab manager who removed a lot of daily stress. Derek García-Almedina was always willing to help out with a protein sample or kind word, and I'm grateful to have collaborated with him on a paper. Paul Huang's passion for his research has helped me have more energy for mine. Paige Pistono was a great GRC and ski buddy, and has taught me a lot about protein expression and given good internship advice. Hannah Martin has always been kind and open to discussing science. I am so glad Kate Dolph joined the lab and was a co-director of the Transfer Student Mentorship Program last year. She has integrity in science, cares deeply for others, and her kindness helped me through my most difficult time in grad school. I know the Transfer Student Mentorship Program is in good hands with her and Sheridon Kelly.

All of my work for this thesis was collaborative with other students in the Francis group and with students, postdocs, and professors in groups with complementary expertise. I have acknowledged them specifically at the end of each chapter. I would particularly like to thank Kiera Wilhelm, who worked with me to overcome many hurdles, and who I learned a lot from about fluorescence microscopy and troubleshooting confusing data. Kiera is an amazing scientist who also taught me a lot about the writing and publishing process. I also would like to thank Conner Harper, who was enthusiastic throughout our collaboration together and helped me figure out how to put together a really interesting story from our data. His example has helped me become a better scientist and collaborator. I would also like to thank Prof. Anne Baranger, who worked with me on mentorship research and tirelessly worked to ensure our data analysis was clear, convincing, and well presented. I appreciate her time and learned a lot from her that will be beneficial across research fields. Prof. Anastasios Melis and Prof. Kris Niyogi taught me so much about native light harvesting systems through their seminar course, and I am grateful to have learned from them.

I also would not have been able to start this thesis without the mentors and friends who encouraged me to come to Berkeley and supported me along the way. I wouldn't have pursued a PhD in chemistry if not for Prof. Matt Sigman, who gave me an opportunity to do research in his lab as an undergraduate. Matt had high expectations but was gracious when I fell short, which helped me learn and achieve more than I could've on my own. I am grateful to all the members of the Sigman lab, who taught me so much about chemistry and research. Specifically, I would like to thank Benjamin Stokes, Elizabeth Bess, Margaret Hilton, and Zachary Niemeyer for their mentorship in the Sigman lab. I would also like to thank Christine Nervig, who helped me through difficult times during my undergraduate research and has continued to give me perspective through the ups and downs of grad school. And outside of chemistry, my neighbor Brenda Miyafuji has always been excited to hear about my research, which reminds me to love what I get to do every day in the lab.

There are two very important people in my life who did not get to see me finish my PhD. One is my mom, Sheri Joy Bischoff, who told me often that I could be anything and

do anything I wanted to. Her cancer inspired me to pursue a career in scientific research, and her belief in me helped me have the confidence to try. I know she would be proud that I followed in her footsteps to get a PhD. The other is my friend Joy Feng, who passed away from cancer a few months before I submitted my thesis. I look fondly on our many conversations about chemistry, research, and finding meaning in life. She showed me how much a person can achieve when they are committed to solving a problem. I am so grateful for the time I had with Joy and my mom.

Finally, I am grateful to my family and partner, who have been with me through it all. My partner, Ingo Walter, gave feedback on so many practice presentations, patiently listened as I explained my research and its problems, and picked up the slack during the busy times (and it seemed like it was always a busy time)! His belief in me and daily support has helped me immensely through the grad school years. My dad, Jon Bischoff, has been my greatest cheerleader throughout my life, who has always been willing to help me at a moment's notice. His help and pride in my work have been a powerful motivator to me in grad school. My sibling, Sarah Bischoff, has been incredibly supportive, reminded me to take time for myself, and helped me remember how amazing it is that I got to work with microbes and proteins every day. My stepmom, Kathi, has helped buoy me through her care and pride. My aunt, Vicki Bushnell, showed me through her example that I could do anything I set my mind to, and continues to inspire me to try new things. I am proud to have followed in the footsteps of my aunt, Ellen Summers, to get a PhD and appreciate her pride and support. My aunt, Anne Cox, and her family have helped me embrace my inner nerd and share my passion for science. And to the many others who I haven't mentioned by name, but who have been a part of my journey- I am so grateful for you all.

Chapter 1

Protein-based artificial light harvesting systems

1.1 Abstract

Photosynthetic organisms absorb and harvest light energy using dynamic networks of pigment–protein complexes. This elegant strategy for capturing solar energy has inspired the construction of artificial systems that seek to mimic the light harvesting ability of photosynthetic organisms, including inorganic, organic, and biomolecular systems. The closest structural mimics of light harvesting complexes are synthetic pigment–protein complexes. Assemblies of virus capsid proteins are an optimal scaffold for artificial light harvesting complexes due to their resistance to degradation and composition of repeating monomeric subunits in large, ordered assemblies. This structure allows pigments to be constrained at uniform distances, similar to those found in natural light harvesting complexes, and allows for energy transfer across large distances. Synthetic light harvesting proteins can also be imbued with photocatalytic abilities through the choice of cofactors installed into the protein scaffolds. Here an overview of protein-based artificial light harvesting complexes is presented, and avenues for improving these models to more closely mimic natural light harvesting systems are suggested.

1.2 Introduction

Photosynthesis is a complex process within plants and some bacteria to convert solar energy to chemical energy. This occurs through a series of electronic energy transfer and electron transfer steps between cofactors embedded within protein assemblies. These protein assemblies usually span a lipid membrane, allowing for charge separation which ultimately drives the production of adenosine triphosphate (ATP) for energy storage. Due to the complexity of this process, numerous artificial light harvesting systems have been constructed to model natural photosynthesis. These models seek both to clarify the mechanisms of energy trans-

fer in photosynthetic organisms and how the light harvesting exhibited by photosynthetic organisms can be emulated in solar harvesting technology.

1.3 Light harvesting in photosynthetic organisms

The first step of photosynthesis is the absorption of light energy by a photosynthetic chromophore, which is promoted to an excited state. The excited state energy is then passed among chromophores until it reaches a reaction center, where electron transfer and charge separation occurs. Individual energy transfer steps can occur on extremely fast timescales, with time constants as fast 100 fs between pigments within photosystem 1 (PS1) in photosynthetic algae.[1–3] This energy transfer can occur so quickly due to the precise organization of chromophores within protein cavities of photosynthetic light harvesting complexes (LHCs). Many LHCs follow a general motif of identical protein subunits forming circular arrays, each containing embedded tetrapyrrole chromophores which absorb light in the visible range (Figure 1.1). Binding pockets within LHCs hold chromophores in precise orientations and at close distances, often under 1 nm to 4 nm separation between nearest neighbor chromophores within a single light harvesting complex and several nm between chromophores in adjacent light harvesting complexes.[2,4] This chromophore positioning allows for rapid and efficient energy transfer to reaction centers described by Förster theory at distances > 1.5 nm and Redfield theory at shorter distances, with a combination of the two mechanisms providing a good description of energy transfer mechanisms within and between photosynthetic complexes.[2,5,6] The organization of pigments within light harvesting complexes minimizes energetic loss to alternate relaxation pathways under low light conditions, while allowing for energy dissipation and photoprotection under high light conditions.[7–9] While the protein scaffold holds chromophores in optimal positions with respect to one another, the chromophores are often integral to the protein structure, assembly state, and localization as well.[10–12]

In purple photosynthetic bacteria such as *R. sphaeroides*, the primary light harvesting complexes are Light Harvesting Complex 2 (LH2) and Light Harvesting Complex 1 (LH1), which contains a Reaction Center (RC) within it (Figure 1.1a-b). LH2 is composed of nine αβ heterodimers which form a circular array. Each αβ heterodimer is associated with three bacteriochlorophyll a (Bcl a) pigments, one B800 and two closely associated B850 pigments named for their differing absorption maxima, and one carotenoid pigment.[13] LH1 is similarly oriented in an array of 14 αβ heterodimers surrounding an RC and containing a gap for formation of an RC-LH1 dimer. Each αβ heterodimer of LH1 contains two BChl a pigments (B875) and two carotenoid pigments, with the exception of the heterodimers near the gap, one of which lacks a single carotenoid and the other lacking both carotenoids (Figure 1.1c-d).[14,15] Light energy is first absorbed by LH2 antenna complexes, then transferred from LH2 to LH2 until the energy reaches LH1. Energy is then transferred between LH1 and the RC embedded within the center of LH1, where a special pair of pigments in the RC initiates charge separation. Energy transfer between BCl a pigments in single LH2 or LH1 complexes

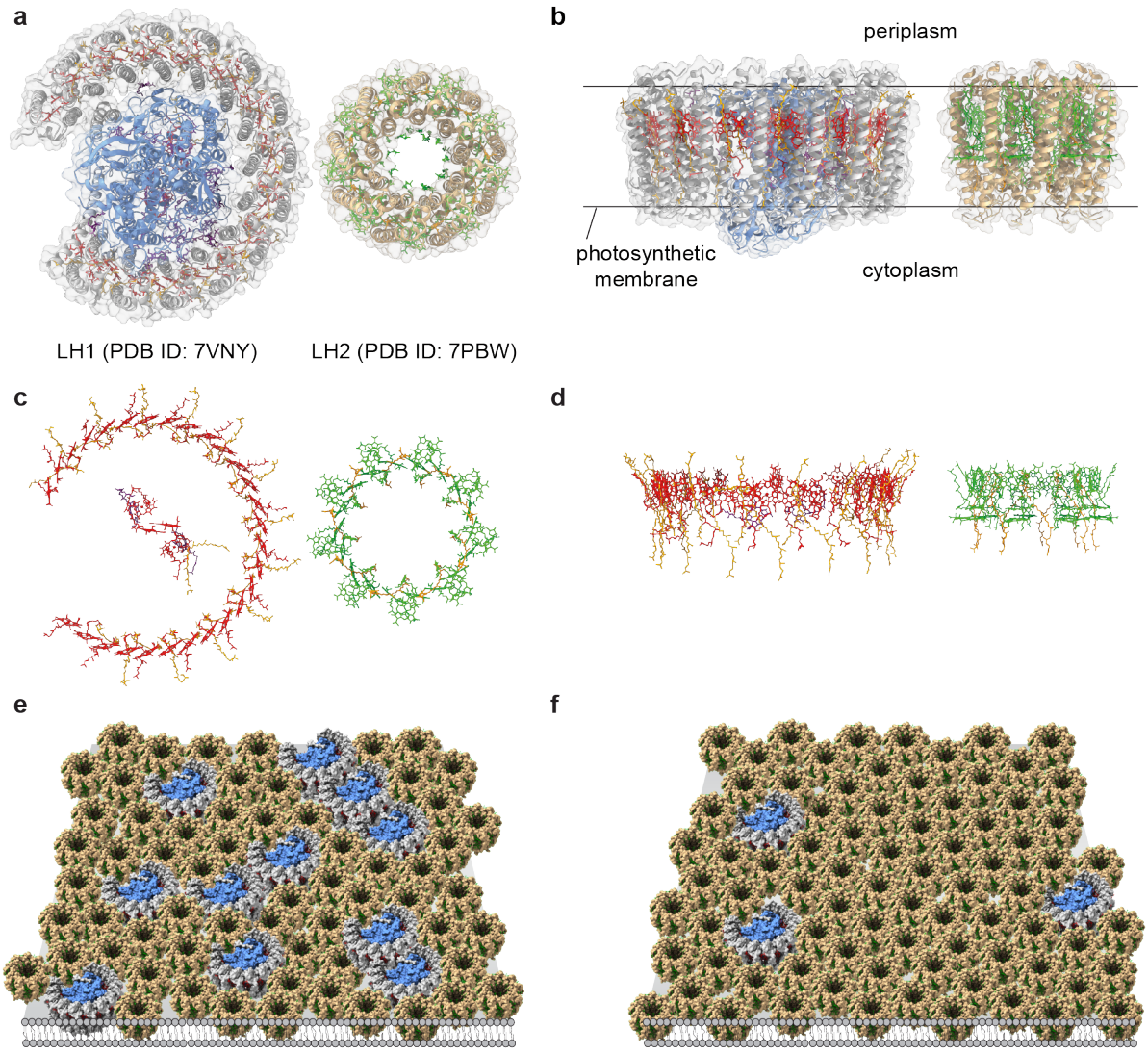


Figure 1.1: Light harvesting complexes from purple photosynthetic bacteria. (a) Light Harvesting Complex 1 (LH1, PDB ID: 7VNY) and Light Harvesting Complex 2 (LH2, PDB ID: 7PBW) from *R. sphaeroides* are both composed of repeating monomeric subunits in a circular or semi-circular array containing bacteriochlorophyll and carotenoid pigments. LH1 surrounds a reaction center (RC) containing the dimer special pair at which charge separation occurs. (b) A side view of LH1-RC and LH2 shows their orientation within a photosynthetic membrane. (c) Pigments are held within a precise orientation within LH1 and LH2, as shown in structures with the protein scaffold removed. (d) A side view shows the precise orientation of pigments within LH1 and LH2, shown in structures with the protein scaffold removed. (e) Under high light conditions, nearly all LH2 complexes are adjacent to LH1-RC complexes. (f) Under low light conditions, aggregates of LH2 and a lower proportion of LH1-RC maximize photon absorbance for each RC.

occurs on timescales of 500–700 fs, while energy transfer between separate LH2 complexes or between LH2 and LH1 occurs on timescales of 1.5–5 ps.[5,16–19] The slowest step in the energy transfer process is from LH1 to the RC, which occurs on a timescale of 35–40 ps.[5,20,21]

With some exceptions, most LHCs are embedded within photosynthetic membranes to separate the cell into compartments, such as chromatophores in bacteria or thylakoids in plants. This membrane allows for charge separation and provides a fluid two-dimensional surface within which LHCs can diffuse. This flexibility allows the many protein assemblies of photosynthesis to quickly assemble and disassemble in response to environmental conditions. For example, under low light conditions, the photosynthetic membranes of *R. sphaeroides* contain large patches of LH2 surrounding low quantities of LH1 to optimize the utilization of absorbed photons by an available LH1-RC complex. In contrast, under high light conditions, LH2 arranges in florettes around LH1-RC with fewer LH2–LH2 contacts and a larger proportion of LH1-RC to LH2, resulting in fewer absorbed photons reaching an available RC but greater absolute conversion of light to chemical energy (Figure 1.1e-f).[22,23] Modeling the process of energy transfer between two different LHCs, and constructing models that capture the complex lipid and protein environment where photosynthetic energy transfer occurs, are the goals of this dissertation.

1.4 Artificial light harvesting systems

There are many approaches for elucidating the physical properties of photosynthetic systems that lead to their function, including through constructing models mimicking the spatial relationship between light-absorbing chromophores that allow for energy transfer. Some strategies include studying pigment dimers, dendrimers, and controlled aggregates (Figure 1.2a);[24–32] embedding pigments in supramolecular scaffolds, such as nanotubes (Figure 1.2b)[33] and polymeric frameworks,[34,35] and embedding pigments in biomolecular frameworks.[36,37] All of these strategies have led to increased understanding of how light harvesting works, but biomolecular frameworks most closely simulate the context of energy transfer in photosynthetic organisms. Oligonucleotides can be arranged in precise nanoarchitectures, making them a scaffold to which chromophores can be covalently bound or non-covalently associated with good control over inter-chromophore distances (Figure 1.2c).[38–42] Peptides have also been appended to chromophores to construct artificial light-harvesting systems that allow for examination of chromophore interactions with amino acid sidechains[43–46] and that can aggregate into highly ordered mesoscale structures (Figure 1.2d).[47,48] Proteins are the most structurally similar to photosynthetic systems and have the added benefit of providing an opportunity for studying the effect of protein–pigment interactions.

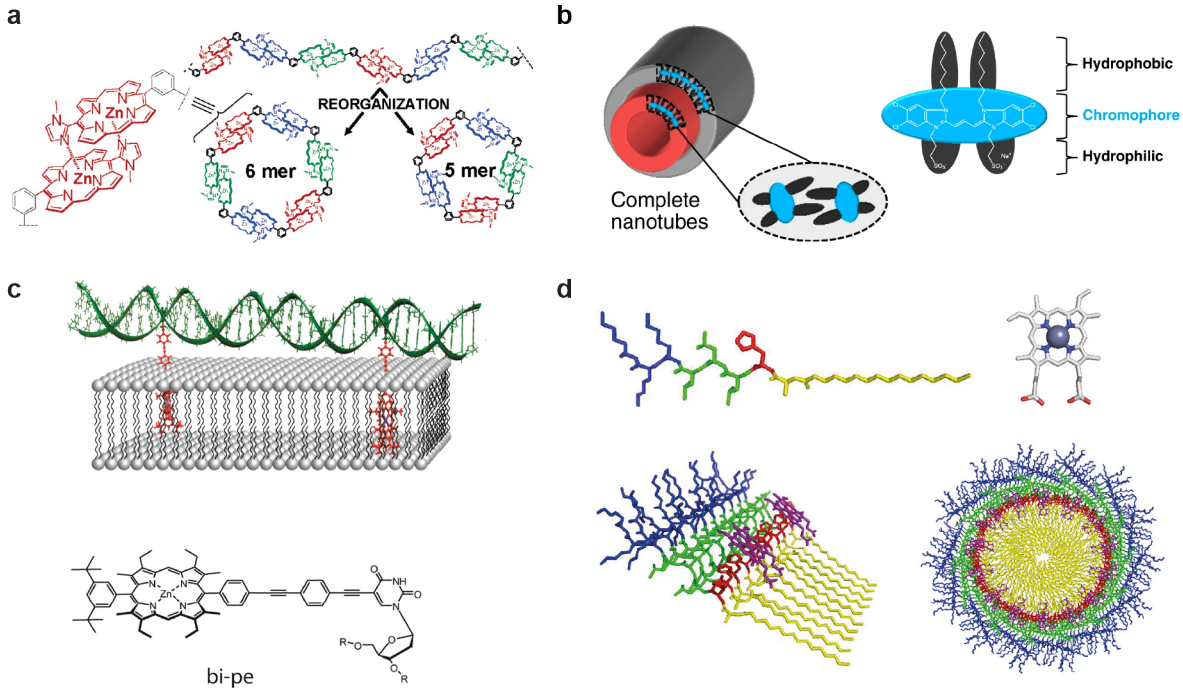


Figure 1.2: Strategies for imitating photosynthetic light harvesting. (a) Zinc porphyrins are arranged in circular arrays through zinc-imidazoyl coordination and π-π stacking. Reprinted with permission.[24] Copyright 2005, American Chemical Society. (b) Double-walled nanotubes are constructed from a small molecule containing a fluorescent core and hydrophobic and hydrophilic sections, for long-range and directional energy transport.[33] Adapted with permission under CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/). (c) DNA with porphyrin-modified bases is used as a scaffold to insert porphyrins into a supported lipid bilayer at controlled distances. Reprinted with permission.[42] Copyright 2010, American Chemical Society. (d) An amphiphilic peptide designed to coordinate zinc protoporphyrins via a histidine residue allows for the formation of ordered supramolecular structures producing close coupling between the pigments. Adapted with permission.[47] Copyright 2012, American Chemical Society.

1.4.1 Artificial light harvesting systems using proteins designed *de novo*

Proteins designed *de novo* are a compelling platform for tailoring the protein environment of individual chromophores and examining the effect of protein–chromophore and solvent–chromophore interactions on binding and excitation characteristics. Photosynthetic chromophores, such as chlorophyll and carotenoids, are often hydrophobic or have limited water solubility due to their localization within hydrophobic protein pockets or in contact with the hydrophobic interior of a lipid membrane. This presents challenges in designing chromophore-binding proteins that have strong associations with tetrapyrrole-based chromophores but also are easily synthesized and water-soluble. The primary basis for *de novo* proteins that bind these chromophores are bundles of four α-helices, with the helices bundled in a square formation and accomodating one or multiple chromophores in the region between the helices (Figure 1.3a-b).[49] By introducing linkers between the helices (dashed

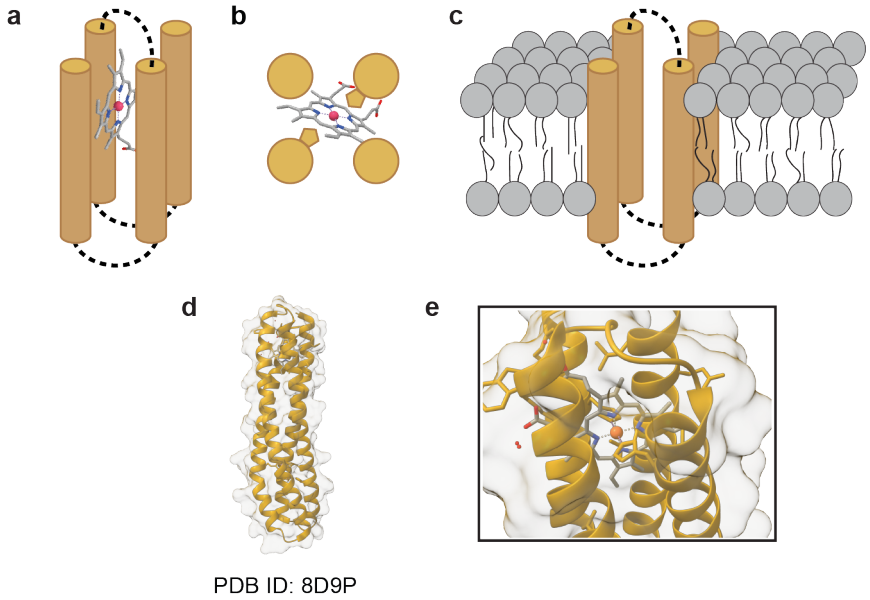


Figure 1.3: Protein maquettes are designed to interact with light-absorbing chromophores. (a) A schematic shows a *de novo* designed bundle of four parallel α-helices (with dotted black lines indicating amino acid linkers) designed to associate with metal porphyrins. (b) A side view of the four-helix bundle and metal porphyrin in (a) is shown. (c) By engineering hydrophobic residues at specific sites on the α-helices, four-helix bundles may be embedded within a lipid bilayer. (d) A crystal structure of a maquette (PDB ID: 8D9P) shows the orientation of a maquette binding Heme B. (e) A magnified view of the Heme B binding site from (d) shows that Heme B is associated with two engineered histidine residues on the interior of the maquette.

lines in Figure 1.3a-c), asymmetric functionalization of each helix and a mix of parallel and antiparallel orientations of the helices can be achieved.[50] Maquettes are often also designed with high thermostability, which can be increased upon the addition of tetrapyrrole cofactors or modulated by changing the polarity of the tetrapyrrole's substituents.[51] Within the hydrophobic core of the four-helix bundles, metal-coordinating amino acid side chains such as histidine residues can be introduced to ligate light- and redox-active cofactors such as hemes via single[52] or double ligation (Figure 1.4a).[53] Covalent methods for installing chromophores within four-helix bundles have also been developed, such as through the conjugation of an aldehyde-derivatized Zn-methylpheophorbide to an aminoxyacetyl-functionalized lysine residue[54] or thioether formation between a bilin and a cysteine residue.[55,56]

Cofactors associated with water-soluble maquettes exhibit altered excitation properties. Maquettes have been engineered to accommodate multiple chromophores within a single maquette, including pairs of chromophores with overlapping absorbance and excitation spectra that can engage in energy transfer. Energy transfer has been demonstrated from Zn pheophorbide a to Zn bacteriopheophorbide bound within a single maquette.[51] Maquettes have also been designed to bind dimers of 13^2-OH-Zn-bacteriochlorophyllide-a in an arrangement where the pigments self-quench.[57] Chromophore pairs have also been tightly constrained within maquettes at near-orthogonal orientations which limit energy transfer even at close distances.[56]

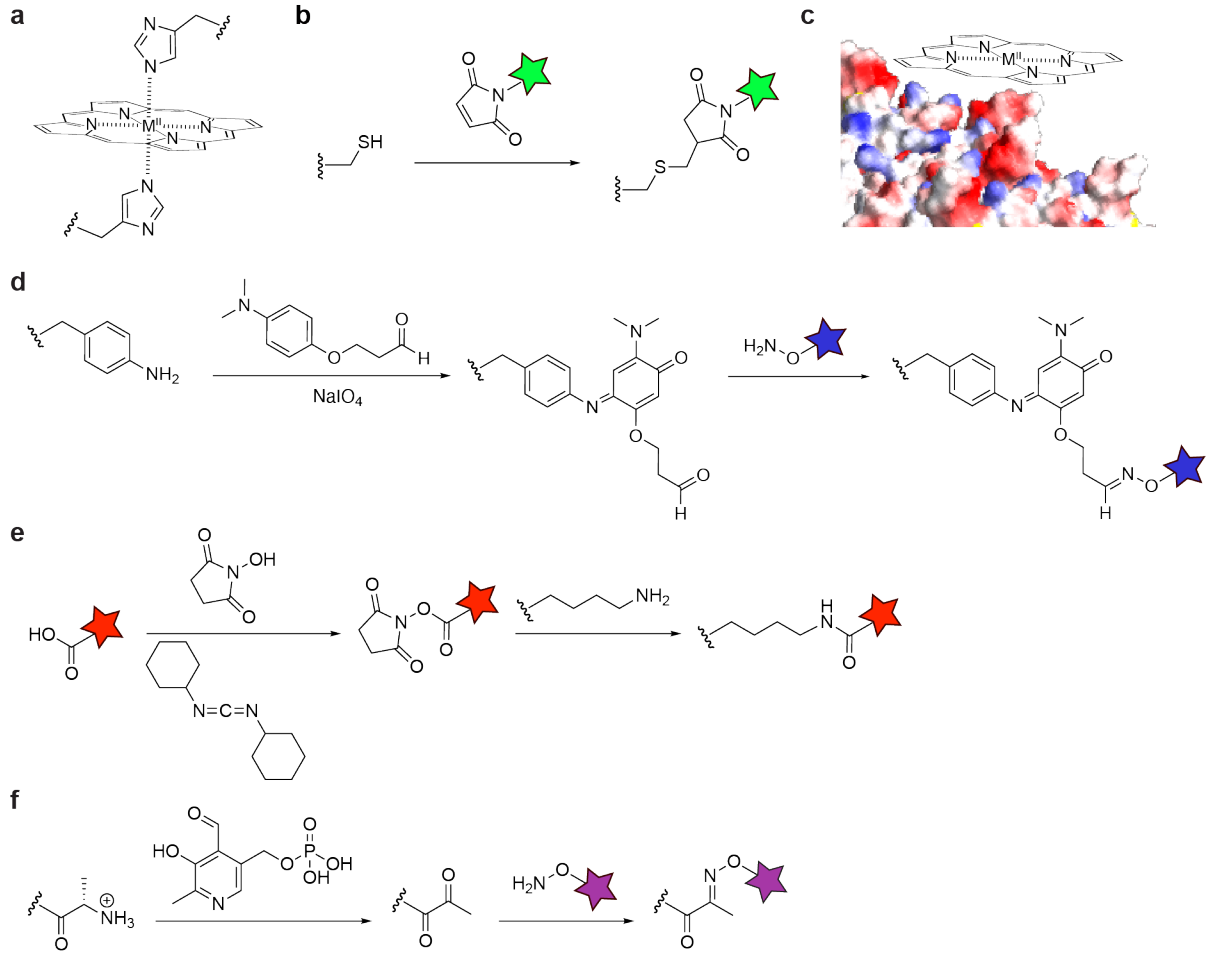


Figure 1.4: Strategies for associating chromophores with proteins. (a) Metal-containing tetrapyrroles, such as hemes or Zn porphyrins, can be ligated by one or two histidine residues within a protein binding pocket. (b) Native or engineered cysteine residues can be covalently modified with maleimide-functionalized chromophores. (c) Chromophores can associate with the protein surface via hydrophobic or electrostatic interactions, such as by installing cationic groups onto the chromophores and associating with a negatively-charged portion of the protein surface. (d) The non-canonical amino acid *p*-aminophenylalanine can be installed during protein expression or later functionalization for reaction with *N*,*N*-dimethylanisidine and an oxidant to create an irreversible covalent linkage. Derivatives of the *N*,*N*-dimethylanisidine can contain useful functional groups for further modification, such as an aldehyde for reaction with a hydroxylamine-containing chromophore. (e) Carbodiimide coupling can be used to convert a carboxylic acid to an *N*-hydroxysuccinimide ester, which is reactive with natural or engineered lysine residues on a protein surface. (f) The N-terminus of a protein may be converted to a diketone using pyridoxal-5′-phosphate, which then can react with a hydroxylamine-containing dye.

Maquettes have also been designed with outward-facing lipophilic sequences for solubilization in membrane or hydrophobic environments. This synthetic tunability is an advantage for mimicking light harvesting proteins, most of which are embedded within lipid membranes but some of which exhibit aqueous solubility, such as phycobilisomes that are found in the stroma of thylakoids.[58] Early examples engineering lipophilic sequences on the exterior of

maquettes binding bacteriochlorophyll analogues or hemes allowed for solubilization in detergent micelles, with maquettes spanning the micelle interface.[59] Maquettes have also been designed with a hydrophilic exterior for aqueous solubility or with bands of hydrophobic and hydrophilic residues on the exterior for insertion into a lipid membrane (Figure 1.3c).[60,61] A maquette incorporating porphyrin cofactors in both the hydrophilic and membrane-spanning domains suggests the future possibility of light-induced charge separation across a bilayer.[62] Membrane-spanning maquettes show promise for constructing the simplest possible model systems with photosynthetic functions and may provide insight into the evolutionary origins of photosynthetic proteins.

Redox-active cofactors can also be introduced into maquette binding sites for the use of light energy to enact charge transfer. By varying the coordination site, protein charge state, and ligating residues of a maquette, the redox midpoint potential of bound heme B was shown to span a range of >100 mV.[63] In addition, excitation of Zn protoporphyin IX bound within a maquette also binding heme resulted in electron transfer from Zn protoporphyin IX to heme, providing a simple and engineerable model system of photosynthetic light harvesting and charge separation.[64] Maquettes have also been engineered to contain three binding sites, to which were bound: an excitable Zn porphyrin pigment ligated by two histidine residues; a di-Manganese electron donor coordinated by a tyrosine residue; and a heme electron acceptor ligated by two histidine residues. This structure, shown in Figure 1.3d-e without the central Zn porphyrin, demonstrated electron transfer upon excitation with light and a charge separated state spanning hundreds of milliseconds, providing a model for a photosynthetic reaction center.[65,66]

Alone, *de novo* proteins are limited in their size and ability to organize into structures with control over large networks of pigments separated at close, defined distances ($\leq$1 nm) and orientations compatible with efficient energy transfer.[49,67] Using maquettes as components of broader networks or materials enhances their light-harvesting utility. A genetic fusion of a tetrapyrrole-binding maquette with the phycocyanin α subunit (CpcA) recombinantly expressed in *E. coli* showed energy transfer from CpcA to the maquette.[68] Similar fusions may function within photosynthetic organisms to harvest light and confer unique cellular functions, such as light-driven catalysis. Maquettes bound to Zn-mesoporphyrin have also been fused with a water-soluble domain of the light harvesting protein ApcEδ, with these maquettes transferring absorbed light energy to ApcEδ.[69] Maquettes binding Zn metalloporphyrins have also been immobilized on TiO_2 electrodes via electrostatic interactions, resulting in a stable photoelectrochemical device with enhanced photovoltage and increased electron lifetimes versus TiO_2 electrodes bound to Zn metalloporphyrins alone.[70] These examples highlight the potential of *de novo* protein–pigment complexes to increase the efficiency of photovoltaic and photoelectronic devices using regenerative and/or biodegradable components. For protein-based synthetic models of larger photosynthetic complexes such as LH1 and LH2, or interactions between complexes, a more practical approach involves repurposing naturally-occurring large protein assemblies through protein engineering and modification.

1.4.2 Incorporating fluorescent proteins into artificial light harvesting systems

Proteins are found in nature that naturally absorb and emit photons, such as Green Fluorescent Protein (GFP) and Blue Fluorescent Protein (BFP). Harnessing the energy absorbed by these fluorescent proteins, which are water-soluble and have been extensively engineered for optimal stability and fluorescence, is a promising method for constructing protein-based photovoltaic devices. A significant challenge in constructing efficient light-harvesting devices is positioning chromophores closely enough for efficient energy transfer while preventing chromophores from forming energetic sinks and self-quenching. Holding chromophores within a protein scaffold is a feasible method for orienting pigments in arrangements optimal for energy transfer. This was demonstrated via the construction of a microscale nanosheet composed of a single layer of covalently-linked Enhanced Green Fluorescent Protein (EGFP) and Enhanced Blue Fluorescent Protein 2 (EBFP2) (Figure 1.5a).[71] EBFP2 and EGFP were organized into single-layer nanosheets by engineering four thiols around the periphery of each protein, followed by conjugation with a difunctionalized maleimide linker (Figure 1.4b).[71] This allowed for tuning of the distance between the chromophores and energy transfer from EBFP2 to EGFP, with an intermediate distance using 3 PEG linkers between maleimides found to be optimal for the construction of the largest nanosheets. Energy transfer was similar between 3-PEG- and 2-PEG-linked fluorescent proteins, and a maximum energy transfer efficiency of 33.2% was achieved. Close packing of chromophores was achieved using this method with minimal fluorescent quenching, likely due to the protein shell surrounding each chromophore.

EGFP has also been fused with cytochrome b_{562}, a protein native to *E. coli*. Cytochrome b_{562} ligates a heme and has electron transfer capability, although its physiological function is unknown.[72] Energy transfer between EGFP and cytochrome b_{562} in the protein chimera occured with an efficiency of 65% as determined by fluorescence lifetime measurements.[73] The EGFP-cytochrome b_{562} fusion was also associated with an Au(111) substrate in a self-assembled monolayer via Au-thiol interactions.[74] This hybrid material generated a photocurrent upon illumination at the EGFP excitation wavelength of 488 nm, resulting from sequential electron transfer from EGFP to cytochrome b_{562} and from cytochrome b_{562} to the Au substrate. These examples demonstrate the utility of using naturally fluorescent proteins in artificial light-harvesting devices.

1.4.3 Replacing heme cofactors in hemoproteins with light-absorbing cofactors

Artificial light-harvesting complexes have also been constructed by repurposing the binding sites of natural hemoproteins to coordinate light-absorbing pigments. In one example, the hexameric tyrosine-coordinated heme protein (HTHP) was reconstituted with Zn protoporphyrin IX and Zn chlorin e_6 ($ZnCe_6$), resulting in a circular array of six Zn tetrapyrroles (Figure 1.5b).[75] Installation of a single Zn tetrapyrrole into the intact hemoprotein was

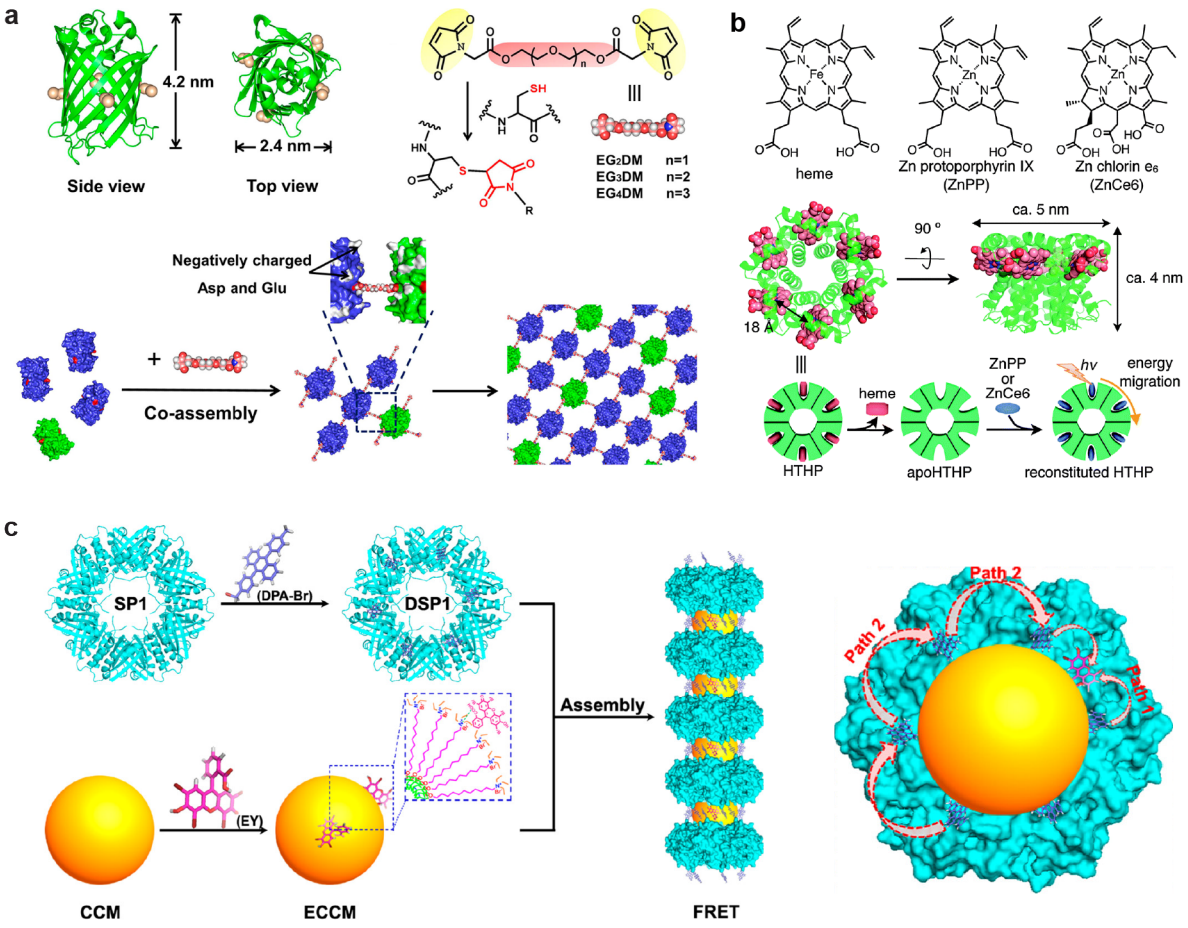


Figure 1.5: Models of photosynthetic light harvesting complexes using proteins not naturally involved in photosynthesis. (a) Single-layered nanosheets of Enhanced Green Fluorescent Protein (EGFP) and Enhanced Blue Fluorescent Protein 2 (EBFP2) capable of EBFP2 to EGFP energy transfer. Nanosheets were constructed through engineered cysteine residues on the peripheries of EBFP2 and EGFP linked together through a polyethylene glycol (PEG) linker capped by two maleimides. Adapted with permission.[71] Copyright 2019, American Chemical Society. (b) Hemes were removed from hexameric tyrosine-coordinated heme protein and reconstituted with Zn protopoporphyrin IX or Zn chlorin e_6 in the heme binding site. The Zn porphyrin analogues were capable of transferring energy within the circular array upon excitation. Adapted with permission under CC BY-NC 3.0 license (https://creativecommons.org/licenses/by-nc/3.0).[75] (c) An engineered cysteine residue in cricoid stable protein 1 (SP1) can be used to attach a bromine-functionalized fluorescent dye, 9-[4-(bromomethyl)phenyl]-10-(4-methylphenyl)anthracene (DPA-Br), to create a circular array of DPA. The DPA-modified SP1 can then associate with core-cross-linked micelles associated with Eosin Y disodium salt (EY) to form ordered rods, which exhibit energy transfer between DPAs and between DPA and EY. Adapted with permission.[76] Copyright 2016, American Chemical Society.

compared to full occupation of all six monomers by measuring the steady state fluorescence quenching of each construct with methyl viologen dication. The efficiency of quenching was 2.3–2.6-fold higher for the fully substituted complexes, indicating energy migration between tetrapyrroles in the complexes and mimicking energy transfer within a single light harvesting complex such as LH1.

Substituting light-absorbing tetrapyrroles for hemes in hemoproteins has also produced constructs with light-driven catalytic functions. Bacterioferritin (BFR) from *E. coli* natively contains a di-iron binding site and a potentially redox-active tyrosine residue, and dimerizes with a heme occupying the homodimer interface. By substituting Mn^{II} for Fe^{II} in the di-iron binding site and $ZnCe_6$ for heme in the inter-monomer heme binding pocket, BFR was imbued with the ability to harvest light.[67] Upon light illumination, this model system demonstrated the oxidation of Mn^{II} to Mn^{III}, mimicking water oxidation functions in photosynthetic reaction centers.

Hemoproteins have also been incorporated into microscale photocatalytic materials. Apoferritin (aFT) forms a 12 nm-diameter protein cage, which has been associated via electrostatic interactions (Figure 1.4c) with a small molecule mixture of a cationic Zn phthalocyanine (Zn Pc) and anionic pyrene to form a network of pigments on aFT's exterior surface.[77] These pigment–protein complexes were able to assemble into ordered crystals up to 100 μm in length. The Zn Pc excitation and emission properties were unaffected by complexation into the crystal structure, with no spectral shift observed. Additionally, the ability of Zn Pc to oxidize a small molecule reporter was evaluated, and little difference in the efficiency of oxidation was observed between free Zn Pc and that complexed with the protein crystal. The protein crystal also was stable under illumination conditions causing dye photobleaching and reactive oxygen species generation, with no degradation observed at up to 600 s irradiation. This protein-based crystal shows promise for designing well-controlled and robust light-harvesting materials that are also biodegradable.

1.4.4 Artificial light harvesting proteins using self-assembling, thermostable proteins

Artificial light-harvesting complexes have also been constructed using a protein scaffold binding no native metal tetrapyrrole cofactors. Cricoid stable protein 1 (SP1) is a water-soluble dodecamer composed of two C_2-symmetric rings of six monomers each, which retains its assembly state up to 107 °C.[78] Through protein engineering and chemical modification, SP1 has been organized into ordered hybrid structures for long-range energy transport. In one example, a thiol-reactive donor chromophore, 9-[4-(bromomethyl)phenyl]-10-(4-methylphenyl)anthracene (DPA-Br), was bound via an engineered cysteine residue on the SP1 surface to form a circular array of chromophores.[76] This protein–pigment complex was then electrostatically associated with core-cross-linked micelles (CCMs) containing acceptor chromophore Eosin Y disodium salt (EY) to form nanowires consisting of alternating SP1 and CCM up to 200 nm long (Figure 1.5c). Measuring the fluorescence emission of these assemblies indicated energy transfer between DPA and EY in the complexes as well as donor-to-donor transfer between DPAs within an SP1 assembly. Rods of SP1 capable of energy transfer have also been assembled via association with CdTe quantum dots.[79]

SP1 assemblies have also been linked at their peripheries to form 2-dimensional uniform nanosheets consisting of single layers of SP1. This has been done via reversible oxidation

to form disulfide bonds at engineered cysteine residues[80] or by conjugation of engineered tyrosine residues on the SP1 disk peripheries using horseradish peroxidase.[81] These sheets were used as a scaffold for the arrangement of controlled ratios of positively charged donor and acceptor CdTe quantum dots or carbon dots, which were bound via noncovalent interactions to the negatively charged protein surface. Energy transfer was demonstrated between the donor and acceptor dots by measuring the fluorescence intensity increase of the acceptor and decrease of the donor. This structure provides a model for measuring energy transport in a 2D sheet with redundant energy transfer pathways, mimicking the potential energy transfer pathways present in photosynthetic light harvesting complexes such as LH2 and LH1 embedded within photosynthetic membranes.

1.4.5 Virus capsid-based artificial light harvesting systems

Most virus capsids are composed of highly ordered, self-assembling protein structures that can reach >400 nm in diameter[82] and can exhibit high thermostability and resist degradation. These highly ordered structures present an opportunity for constructing artificial light-harvesting systems with a high degree of control over inter-chromophore distances and long-range energy transport capabilities.

1.4.5.1 Hepatitis-B-based artificial light harvesting systems

The hepatitis B virus capsid is composed of 120 copies of the hepatitis B virus capsid protein (HBV) dimer, which assemble into an icosahedron.[83] Engineering hexahistidine tags at the HBV N-terminus (yellow in Figure 1.6a) results in aggregates of three hexahistidine tags at 80 sites on the HBV exterior. The imidazole groups of these tags were able to associate with 80 hemes on the HBV surface to create a densely heme-decorated, 30 nm particle.[84] The hemes associated with HBV exhibited spectral and electrochemical properties consistent with each being site-isolated on the HBV surface and ligated by two axial imidazole groups, in a similar environment as hemes found in b-type cytochromes.

1.4.5.2 MS2-based artificial light harvesting systems

The coat protein of bacteriophage MS2 is a 27 nm-diameter icosahedron composed of 180 identical monomers (Figure 1.6b). Like HBV, it has been selectively engineered for modification with light-absorbing chromophores. To achieve a well-templated system for directional electron transfer, the exterior of MS2 was fully modified with acceptor chromophores in the form of zinc porphyrins, and the interior of MS2 was fully modified with donor chromophores, Oregon Green 488 and Alexa Fluor 350.[85] This required the site-selective modification of an internal cysteine to install the donor chromophores and of a non-canonical amino acid, *p*-aminophenylalanine (pAF), for functionalization with a bifunctional aldehyde-*N*,*N*-dimethylanisidine linker to install aminooxy-containing porphyrins (Figure 1.4d). Energy transfer across the 2 nm thick protein shell was observed for both donors. Additionally, the

ability of the porphyrin excited by a donor pigment to sequentially reduce methyl viologen dication was examined. Compared to a system without donor chromophores, a 3.5-fold increase in reduction potential was shown for the fully assembled system containing protein-bound donor and acceptor chromophores and methyl viologen dication, excited at the donor excitation wavelength of 505 nm. The ability to use light energy to excite a donor chromophore, transfer the excitation energy, and initiate electron transfer, is a model for photosynthetic energy and electron transport. Because the donor chromophores are on the opposite side of the protein surfaces from acceptor chromophores, this model also may prevent undesirable contact quenching. The chromophore positioning across a protein shell also imitates chromophore interactions in photosynthetic light harvesting complexes, which require energy transfer between chromophores embedded within protein cavities.

MS2 assemblies have also been used to mediate energy transfer between a small molecule dye (Alexa Fluor 488) and 10 nm gold nanoparticles (AuNPs).[86] MS2 was assembled around AuNPs associated with DNA strands (to provide negative charge and electrostatically associate with the positively-charged interior of MS2). An engineered pAF residue on each MS2 monomer was conjugated to 5′-aminophenol, 3′-fluorophore-containing hairpin DNA. The fluorophore was spaced 3, 12, or 24 bp from the bioconjugation site, and differences in fluorescence intensity in the presence and absence of AuNPs was measured using total internal reflection fluorescence (TIRF) microscopy. The greatest fluorescence intensity increase was observed for the shortest distance, indicating an increase in the rate of radiative decay due to interaction with the AuNP. A control with the DNA-fluorophore conjugated directly to the AuNP resulted in a drastic decrease in fluorescence intensity. This demonstrates the utility of using the protein shell between the nanoparticle and dye to prevent contact quenching.

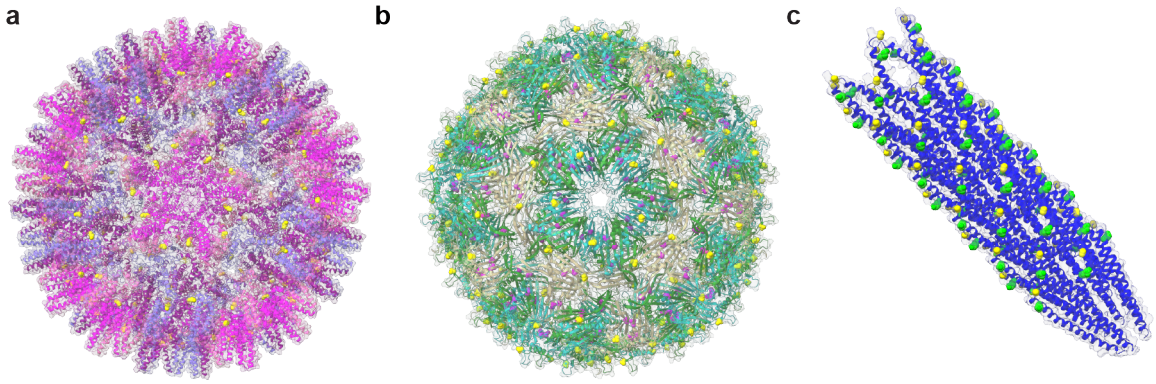


Figure 1.6: Virus capsid proteins used in artificial light-harvesting and their modification sites. (a) The Hepatitis B virus capsid protein (PDB ID: 6VZP)[83] was engineered with hexahistidine tags at its N-terminus (yellow) for association with hemes. (b) The MS2 virus capsid protein (PDB ID: 2MS2)[87] was engineered with cysteine at position 87 (magenta) and *p*-aminophenylalanine at position 19 (yellow) for attachment to maleimide- or aminooxy-functionalized chromophores. (c) The M13 virus capsid protein (PDB ID: 2C0W)[88] was modified at its native N-terminus (yellow) and lysine residues (green), as well as an engineered lysine residue, with NHS-ester functionalized choromophores. A segment of the full rod assembly is shown.

1.4.5.3 M13-based artificial light harvesting systems

While icosahedral assemblies of virus capsid proteins allow for precise templating of chromophores at defined distances, natural photosynthetic light harvesting complexes are not icosahedral but often have a disk morphology, which can then associate with other disks laterally within a membrane or stack into rods for long-range energy transport.[89] Under low light conditions, some cyanobacteria have also been shown to express an allophycocyanin paralog with improved absorption of far-red light that assembles into short helices of roughly 12 heterodimers.[90] Viruses that self-assemble into rods or helices, therefore, structurally mimic the morphology present in natural photosynthetic complexes. The M13 bacteriophage is composed of helical repeating monomeric subunits assembling into a hollow rod, which is 6 nm in width and reaches up to 2 μm in length.[88,91] Because M13 has a negatively-charged surface, it has been demonstrated to electrostatically associate with cationic porphyrins.[91] By engineering tryptophan residues into the M13 coat protein, energy transfer has been demonstrated between tryptophan and the cationic porphyrins. While the porphyrins were not bound to specific sites on the protein in this case, this shows the potential for organizing chromophores along M13 rods to achieve long-range energy transfer.

M13 assemblies have also been covalently modified with porphyrins attached via native residues (Figure 1.6c).[92] Zn(II) deuteroporphyrin IX 2,4-bis(ethylene glycol) was conjugated to the M13 virus using carbodiimide coupling of the pendant carboxylic groups with the N-terminus and native lysine residues (Figure 1.4e). At the densest modification levels, interpigment distances were estimated at 3.5 nm, and the Soret band of the Zn porphyrin was shown to broaden. Additionally, the excited state lifetime of individual pigments shortened, indicating energy transfer. Fluorescence emission also decreased dramatically, indicating that nonradiative pathways increased and suggesting that some of the Zn porphyrins bound to M13 were electronically coupled to form trap sites.

Through engineering the M13 coat protein to contain additional lysine residues and modifying the coat with donor and acceptor chromophores, M13 has been able to model the complex mechanisms present in photosynthetic energy transfer.[93] The chromophores used, AlexaFlour 488 NHS ester and Alexa Fluor 594 NHS ester, were attached to M13 monomers via the N-terminus and a pre-existing lysine for one mutant (M13CF) and to an additional engineered lysine residue for the other mutant (M13SF). Varying the ratio of donors to acceptors and analyzing the acceptor fluorescence allowed the diffusion length to be calculated for each mutant. The diffusion length was found to be larger for the M13SF mutant by 68%, indicating an increase in efficiency for the mutant with additional modification sites. Transient absorption spectroscopy measurements and modeling indicated that the energy transfer from donors to acceptors cannot just be attributed to Förster resonance energy transfer, but there also must be contributions from quantum coherent energy transfer. Quantum coherence contributes to the energy transfer efficiency in this M13 model, and a synergistic combination of coherent and Förster resonance energy transfer also likely account for the high energy transfer efficiency achieved in photosynthetic organisms.[94]

1.4.6 Tobacco mosaic virus-based artificial light harvesting systems

Like M13, the tobacco mosaic virus coat protein (TMV) assembles into a helical array of identical monomers forming a hollow rod. The wild type virus is 300 nm in length and 18 nm in diameter.[95] At higher salt concentrations and lower pH values, TMV can also assemble into aggregates of even-numbered disk stacks, with 17 monomers per disk.[96] Both the rod structure and stacked disk structure have potential advantages as templates for constructing artificial light harvesting complexes. While the rod structure can emulate the rods present in intact phycobilisomes and potentially engage in long-range energy transport over 100s of nm, the stacked disk structure is a good structural analogue and thus a useful model for many photosynthetic complexes, such as LH1 and LH2.

TMV has been engineered with reactive amino acid side chains to which chromophores can be attached. One of the earliest examples modifying TMV with chromophores involved introducing an individual cysteine residue at positions 98, 99, 100, and 101, within the pore of four different TMV constructs, for attachment to a maleimide-functionalized pyrene (blue, Figure 1.7a-b).[97] The TMV-attached pyrenes engaged in π stacking within the TMV pore, which stabilized the TMV rod and allowed the Q99C and A100C mutants to assemble to lengths >2 μm. Additionally, measurement of pyrene fluorescence of these mutants indicated that the pyrenes were strongly coupled as dimers and excimers, as evidenced by an increase in fluorescence intensity at the excimer emission wavelength of 465 nm. This demonstrates the interactivity between the attached chromophores and protein structure that is also present in photosynthetic complexes; pyrene's attachment to TMV influences its excitation properties, while the attached pyrene also influences the assembly state of TMV.

Arrays of Zn and free base porphyrins have also been templated within the pore of TMV through attachment of maleimide-functionalized porphyrins to an engineered cysteine residue at position N127 (purple, Figure 1.7a-b).[98] Energy transfer was then compared at varying ratios of the Zn porphyrin donor:free base porphyrin acceptor, as well as between the rod and disk forms of TMV. The absorbance spectrum of Zn porphyrin fully labeled on TMV did not shift, indicating that the porphyrins did not form aggregates in the rod or disk forms. Energy transfer between donor and acceptor porphyrins occurred in both the disk and rod structures, with an accelerated rate in the disk structure compared to the rod structure. This indicated that the porphyrins may be held in more rigid and distinct conformations in the rods and have a greater degree of flexibility and potential to be in closer proximity with neighboring porphyrins in the disks.

TMV has been modified with donor and acceptor chromophores within the central pore to reach an energy transfer efficiency of up to 90%.[99] Oregon Green 488 maleimide as a donor dye and Alexa Fluor 594 maleimide as an acceptor dye were attached to TMV via an engineered cysteine residue at position S123 near the TMV pore (green, Figure 1.7a-b). By taking advantage of differing assembly conditions, by which monomers, double stacked disks, and rods can each be interchanged, motifs containing 2:1 or 16:1 donors:acceptors were constructed. While energy transfer in double-stacked disks seemed less efficient than rods in the

16:1 case, this difference was eliminated in the 2:1 case. This likely occured because a large portion of the double-layered disks, containing 34 monomers each, may proportionally contain no acceptor dyes in the 16:1 case. Energy transfer efficiency using just Oregon Green 488 and Alexa Fluor 594 was 34–47%, but increased dramatically when a chromophore with an intermediate excitation and emission wavelength, tetramethylrhodamine, was added to the system. By constructing a population of double-stacked disks with only donors and another population with 4:1 intermediate pigment:acceptors and then assembling these two disk populations into rods, 90% energy transfer efficiency was achieved. This system demonstrates high control over chromophore arrangement in a protein environment by taking advantage of the reversible assembly state changes of TMV while using a single bioconjugation reaction to install three different dyes. This highlights the potential for model systems to take advantage of the range of the solar spectrum using a mixture of chromophores, analogous to the mixture of chromophores in photosynthetic systems. The same TMV rod donor–acceptor model system without the intermediate chromophore was also studied using global lifetime analysis to characterize the energy transfer pathways present.[100] This analysis indicated that energy transfer between donors occured at a rate of 70 ps, and donor to acceptor transfer occurred at a rate of 187 ps. While the rates are still slower than rates of energy transfer in natural photosynthetic systems, they are $>10\times$ faster than the fluorescence lifetime of Oregon Green 488 at 4.1 ns and Alexa Fluor 594 at 3.9 ns (SPEX Fluorescence Group, Horiba Jobin Yvon Inc.). The rapid energy transfer rates in this model promote efficient energy transfer and avoid energy loss to competing chromophore relaxation pathways.

The sensitivity of energy transfer in donor- and acceptor-labeled TMV rods and disks to defects introduced by photobleaching has been examined.[102] Defects were introduced into the system by using an Ir-based reducing agent, which was shown to deactivate chromophores without harming the intrinsic protein structure or assembly ability. Disks or rods were assembled with varying concentrations of monomers bearing active and deactivated donors and either 16:1 or 2:1 donor:acceptor. In each case, rods were much more efficient at energy transfer than disks, and the efficiency decreased linearly with % donor bleaching. In the disk assembly state, chromophores occupy the top and bottom faces of the disk, so energy transfer occurs primarily between donors and acceptors adjacent to each other in a 17-chromophore circular array. In this setup, a single donor losing functionality will impact the ability for energy to be transferred to the acceptor. In TMV rods, there are multiple options for energy transfer, not just laterally but also vertically between stacked monomers in the helix, even if one donor is deactivated. These findings indicate that chromophore-labeled TMV rods act more like a bulk material with many redundant energy transfer pathways, and thus may have more utility for designing photovoltaic materials. In contrast, TMV disks behave more like a discrete system and therefore may be a more informative model for how defects affect energy transfer in individual photosynthetic light-harvesting complexes such as LH1 or LH2.

A circular permutant of TMV (cpTMV) has also been engineered, which forms stable, C_2-symmetric disks at a large pH and ionic strength range (Figure 1.7c-d).[101] This permutant moves the N- and C-termini to the pore rather than the periphery of the disk. Modification of an engineered cysteine residue at the S23 position (red in Figure 1.7c) on

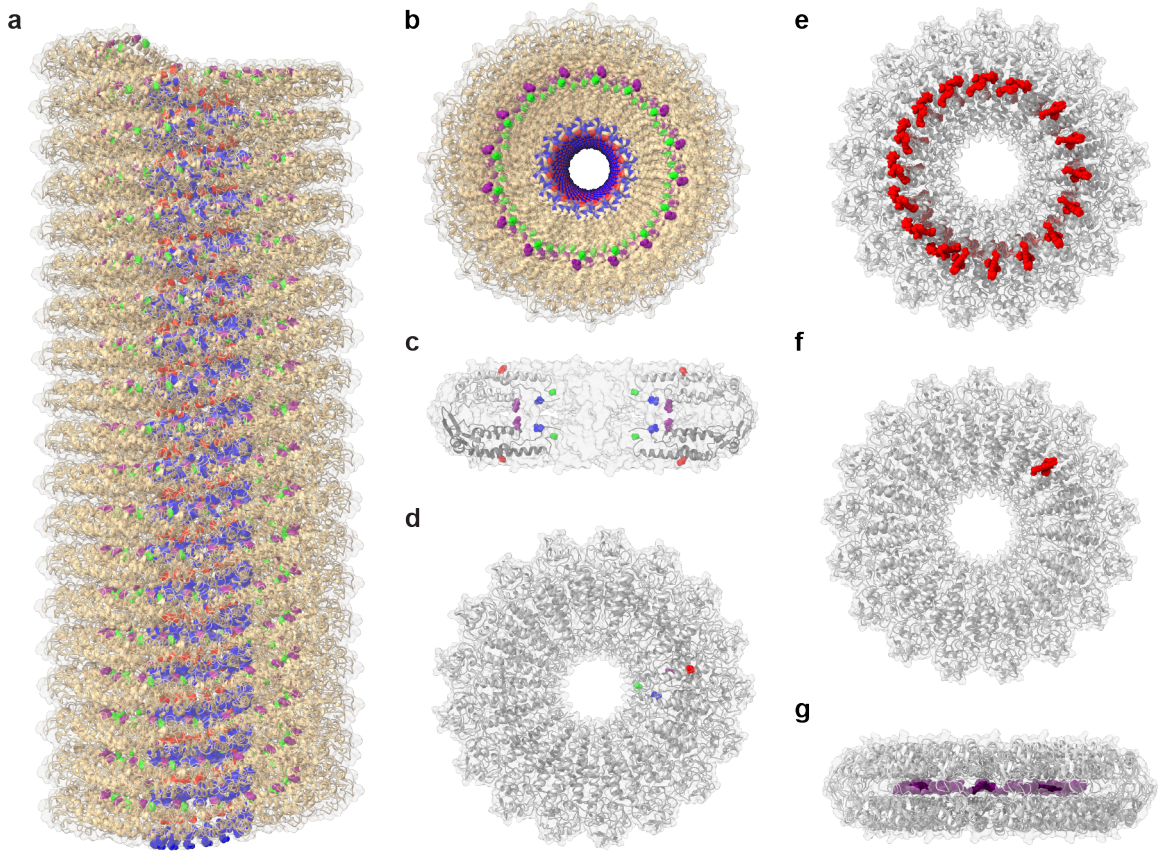


Figure 1.7: The tobacco mosaic virus as an artificial light harvesting assembly. The rod assembly state of TMV (PDB ID: 3J06)[95] is shown from side (a) and face (b) views with residues mutated to cysteine for chromophore attachement shown in blue (residues 98-101), red (residue 104), green (residue S123), and purple (residue N127). A circular permutant of TMV, cpTMV (PDB ID: 3KML),[101] is shown from side (c) and face (d) views with residues mutated for chromophore functionalization shown in green (N-terminus), red (residue 23), purple (residue 101), and blue (residue 157). Because cpTMV assemblies are stable in aqueous solution, they may be modified with arrays of chromophores to examine chromophore interactions with the protein, solvent, and other chromophores (e), or with a single chromophore to examine chromophore–protein and chromophore–solvent interactions in isolation (f). (g) Hemes have been ligated within the cavity between opposing cpTMV disks in a chiral, constrained environment.

the surface (Figure 1.7e) and the N-terminus in the pore of cpTMV (green in Figure 1.7c) resulted in energy transfer between two synthetic dyes (Texas Red and Alexa Fluor 647). The N-terminal modification was achieved using pyridoxal-5′-phosphate to install a ketone at the N-terminus, followed by reaction with an aminooxyacetamide-functionalized Alexa Fluor 647 (Figure 1.4f). The cpTMV structure also has a cavity region between the two disks, which provides an opportunity to position chromophores in a protein pocket with less solvent accessiblity.[103] Oregon Green 488 and an indocarbocyanine dye (Cy5) have been attached to two engineered cysteine residues within this cavity region, at positions 101 (in the protein cavity, further from the central pore, purple in Figure 1.7c) and 157 (within the

central pore, blue in Figure 1.7c). Using a combination of transient absorption and fluorescence measurements and MD simulations, the flexibility of the chromophores and solvent environment was examined. Chromophore and water molecule rearrangement was found to occur more slowly when the chromophore was attached to position 101, within the protein cavity, than position 157, adjacent to the cpTMV pore. In addition, the modification had a greater impact on Cy5, which has a flexible core, than on Alexa Fluor 647, which has a rigid core. The conformational rigidity also impacted the fluorescence lifetime of the chromophores, increasing the fluorescence lifetime of Cy5 from 1.24 ns at position 157 to 1.56 ns at position 101. The ability to increase the excited state lifetime of chromophores as a function of their attachment site to cpTMV demonstrates the possibility of increasing energy transfer efficiency in protein-based artificial light-harvesting systems.

The effect of chromophore rigidity and environment has been further explored by varying both the chromophore attachment site to cpTMV and the composition of the protein–chromophore linker.[104] Four sulforhodamine B (SRB) maleimide analogues have been prepared with linkers of differing lengths and stereochemistry, and these were conjugated to either the surface at position 23 or cavity at position 101 of cpTMV. Each cpTMV assembly was singly labeled (Figure 1.7f), and the excited state dynamics of SRB were probed using transient absorption spectroscopy and molecular dynamics simulations. Greater chromophore constraint was correlated to a longer excited state lifetime, and this constraint was affected both by the bioconjugation site and the length and stereochemistry of the bioconjugation linker. The more constrained chromophores also slowed the dynamics of the surrounding water molecules. For artificial light-harvesting applications, this demonstrated that not only the modification site, but also the attachment method and linker flexibility contribute to the excitation properties of the chromophores. Additionally, the ability to examine the interaction of a single chromophore with its protein and solvent environment allows the protein and solvent effects to be decoupled from interactions with neighboring chromophores. This ability to modify protein complexes with single chromophores contrasts with many natural photosynthetic light harvesting complexes, where removal of the chromophores can alter the protein's structure and assembly state. Detailed molecular mechanics simulations of full arrays of chromophores attached to positions 104 and 123 of non-circularly permuted TMV, which similarly position chromophores either within the pore or within a protein pocket, have also been performed.[105] These similarly indicated that chromophore identity and modification site have a substantial impact on chromophore rigidity, and showed that substantial disorder in chromophore geometry existed in the fully-labeled systems.

To introduce electron transfer function into the TMV model, the cavity region of cpTMV has been engineered to accomodate hemes via engineered histidine residues at position 101 (Figure 1.7g).[106] This scaffold was able to accomodate hemes via bis-his ligation, placing the hemes in a chiral environment. Hemes bound within the TMV cavity exhibited a lowered midpoint potential compared to free heme and could be reversibly oxidized and reduced. The cpTMV-bound heme midpoint potential was comparable to solvent-exposed hemes bound to cytochromes. The ability to bind arrays of redox-active porphyrins within the cpTMV pocket, combined with the ability to attach chromophores to the protein's exterior, demon-

strates the possibility of incorporating light harvesting and charge separation abilities into the cpTMV model.

TMV has also been incorporated into a material with photocatalytic function.[107] A water-soluble, highly charged, synthetic derivative of Zn phthalocyanine was constructed which promoted the assembly of TMV into bundles of rods held together by electrostatic interactions between the highly-charged phthalocyanine derivative and negatively-charged TMV surface. Addition of a dye sensitive to reaction oxygen species (ROSs), 2′,7′-dichlorofluorescein, followed by irradiation of the TMV bundles at the excitation wavelength of the phthalocyanine derivative revealed that ROSs were generated in the TMV bundles, but not outside of them. This process was repeatable when the TMV bundles were put into a microfluidic device, with little loss in ROS yield after 10 cycles. Irradiation at higher power resulted in disassembly and the ability of the rods to be removed as waste from the microfluidic device. The assembly of TMV into a material with photocatalytic function demonstrates its utility not only as a platform for understanding photosynthetic light harvesting systems, but also as a biodegradable component of light-harvesting devices.

1.5 Conclusion

Artificial light harvesting systems seek to emulate the complex and interdependent photosynthetic pigment–protein complexes that efficiently harvest light energy and convert it to chemical energy. Artificial light harvesting systems have been constructed from a wide range of inorganic, organic, and biomolecular materials, with the closest mimic of photosynthetic complexes constructed from protein scaffolds. *De novo* designed proteins, naturally fluorescent proteins, heme proteins, virus capsid assemblies, and other protein assemblies have been used as artificial light harvesting models. These models can template pigments for optimized, long-range energy transfer, charge separation, and catalytic functions, or be used to better understand the structural principles leading to efficient energy transfer in natural photosynthesis by closely mimicking the structure of photosynthetic complexes. For the latter purpose, the majority of research has focused on mimicking a single chromophore–protein interaction, or a single light harvesting complex. This dissertation will address expanding the TMV light harvesting model to capture interactions between photosynthetic light harvesting assemblies.

1.6 References

(1) Du, M.; Xie, X.; Jia, Y.; Mets, L.; Fleming, G. R. *Chemical Physics Letters* **1993**, *201*, 535–542.

(2) Croce, R.; van Amerongen, H. *Photosynthesis Research* **2013**, *116*, 153–166.

(3) Nelson, N.; Junge, W. *Annual Review of Biochemistry* **2015**, *84*, 659–683.

(4) Xie, H.; Lyratzakis, A.; Khera, R.; Koutantou, M.; Welsch, S.; Michel, H.; Tsiotis, G. *Proceedings of the National Academy of Sciences* **2023**, *120*, e2216734120.

(5) Scholes, G. D.; Fleming, G. R. In *Adventures in Chemical Physics*; Advances in Chemical Physics, 2005, pp 57–129.

(6) Yang, M.; Fleming, G. R. *Photoprocesses in Multichromophoric Molecular Assemblies* **2002**, *275*, 355–372.

(7) Fraser, N. J.; Hashimoto, H.; Cogdell, R. J. *Photosynthesis Research* **2001**, *70*, 249–256.

(8) Derks, A.; Schaven, K.; Bruce, D. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2015**, *1847*, 468–485.

(9) Pan, X.; Liu, Z.; Li, M.; Chang, W. *Current Opinion in Structural Biology* **2013**, *23*, 515–525.

(10) Preiss, S.; Thornber, J. P. *Plant Physiology* **1995**, *107*, 709–717.

(11) Frigaard Niels-Ulrik; Maresca Julia A.; Yunker Colleen E.; Jones A. Daniel; Bryant Donald A. *Journal of Bacteriology* **2004**, *186*, 5210–5220.

(12) Paulsen, H. In *The Photochemistry of Carotenoids*, Frank, H. A., Young, A. J., Britton, G., Cogdell, R. J., Eds.; Springer Netherlands: Dordrecht, 1999, pp 123–135.

(13) Qian, P.; Swainsbury, D. J. K.; Croll, T. I.; Castro-Hartmann, P.; Divitini, G.; Sader, K.; Hunter, C. N. *Biochemistry* **2021**, *60*, 3302–3314.

(14) Cao, P.; Bracun, L.; Yamagata, A.; Christianson, B. M.; Negami, T.; Zou, B.; Terada, T.; Canniffe, D. P.; Shirouzu, M.; Li, M.; Liu, L.-N. *Nature Communications* **2022**, *13*, 1977.

(15) Nagarajan, V.; Parson, W. W. *Biochemistry* **1997**, *36*, 2300–2306.

(16) Hess, S.; Chachisvilis, M.; Timpmann, K.; Jones, M. R.; Fowler, G. J.; Hunter, C. N.; Sundström, V. *Proceedings of the National Academy of Sciences* **1995**, *92*, 12333–12337.

(17) Sundström, V.; Pullerits, T.; van Grondelle, R. *The Journal of Physical Chemistry B* **1999**, *103*, 2327–2346.

(18) Agarwal, R.; Yang, M.; Xu, Q.-H.; Fleming, G. R. *The Journal of Physical Chemistry B* **2001**, *105*, 1887–1894.

(19) Shreve, A.; Trautman, J.; Frank, H. A.; Owens, T.; Albrecht, A. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1991**, *1058*, 280–288.

(20) Bergström, H.; van Grondelle, R.; Sundström, V. *FEBS Letters* **1989**, *250*, 503–508.

(21) Visscher, K. J.; Bergström, H.; Sundström, V.; Hunter, C. N.; Van Grondelle, R. *Photosynthesis Research* **1989**, *22*, 211–217.

(22) Niederman, R. A. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2016**, *1857*, 232–246.

(23) Scheuring, S.; Sturgis, J. N. *Science* **2005**, *309*, 484–487.

(24) Takahashi, R.; Kobuke, Y. *The Journal of Organic Chemistry* **2005**, *70*, 2745–2753.

(25) Satake, A.; Kobuke, Y. *Tetrahedron* **2005**, *61*, 13–41.

(26) Hashimoto, H.; Uragami, C.; Cogdell, R. J. In *Carotenoids in Nature: Biosynthesis, Regulation and Function*, Stange, C., Ed.; Springer International Publishing: Cham, 2016, pp 111–139.

(27) Aratani, N.; Kim, D.; Osuka, A. *Accounts of Chemical Research* **2009**, *42*, 1922–1934.

(28) Sessler, J. L.; Wang, B.; Harriman, A. *Journal of the American Chemical Society* **1995**, *117*, 704–714.

(29) Kleiman, V. D.; Melinger, J. S.; McMorrow, D. *The Journal of Physical Chemistry B* **2001**, *105*, 5595–5598.

(30) Neuwahl, F. V. R.; Righini, R.; Adronov, A.; Malenfant, P. R. L.; Fréchet, J. M. J. *The Journal of Physical Chemistry B* **2001**, *105*, 1307–1312.

(31) Yang, J.; Yoon, M.-C.; Yoo, H.; Kim, P.; Kim, D. *Chemical Society Reviews* **2012**, *41*, 4808–4826.

(32) Hu, Y.-X.; Li, W.-J.; Jia, P.-P.; Wang, X.-Q.; Xu, L.; Yang, H.-B. *Advanced Optical Materials* **2020**, *8*, 2000265.

(33) Kriete, B.; Lüttig, J.; Kunsel, T.; Malý, P.; Jansen, T. L. C.; Knoester, J.; Brixner, T.; Pshenichnikov, M. S. *Nature Communications* **2019**, *10*, 4615.

(34) Brixner, T.; Hildner, R.; Köhler, J.; Lambert, C.; Würthner, F. *Advanced Energy Materials* **2017**, *7*, 1700236.

(35) Haedler, A. T.; Kreger, K.; Issac, A.; Wittmann, B.; Kivala, M.; Hammer, N.; Köhler, J.; Schmidt, H.-W.; Hildner, R. *Nature* **2015**, *523*, 196–199.

(36) Spillmann, C. M.; Medintz, I. L. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2015**, *23*, 1–24.

(37) Gong, C.; Sun, S.; Zhang, Y.; Sun, L.; Su, Z.; Wu, A.; Wei, G. *Nanoscale* **2019**, *11*, 4147–4182.

(38) Albinsson, B.; Hannestad, J. K.; Börjesson, K. *Coordination Chemistry Reviews* **2012**, *256*, 2399–2413.

(39) Dutta, P. K.; Varghese, R.; Nangreave, J.; Lin, S.; Yan, H.; Liu, Y. *Journal of the American Chemical Society* **2011**, *133*, 11985–11993.

(40) Stein, I. H.; Steinhauer, C.; Tinnefeld, P. *Journal of the American Chemical Society* **2011**, *133*, 4193–4195.

(41) Hemmig, E. A.; Creatore, C.; Wünsch, B.; Hecker, L.; Mair, P.; Parker, M. A.; Emmott, S.; Tinnefeld, P.; Keyser, U. F.; Chin, A. W. *Nano Letters* **2016**, *16*, 2369–2374.

(42) Börjesson, K.; Wiberg, J.; El-Sagheer, A. H.; Ljungdahl, T.; Mårtensson, J.; Brown, T.; Nordén, B.; Albinsson, B. *ACS Nano* **2010**, *4*, 5037–5046.

(43) Brea, R. J.; Pérez-Alvite, M. J.; Panciera, M.; Mosquera, M.; Castedo, L.; Granja, J. R. *Chemistry – An Asian Journal* **2011**, *6*, 110–121.

(44) Razeghifard, R. *Journal of Photochemistry and Photobiology B: Biology* **2015**, *152*, 416–424.

(45) Nandy, A.; Mukherjee, S. *The Journal of Physical Chemistry Letters* **2022**, *13*, 6701–6710.

(46) Zou, Q.; Liu, K.; Abbas, M.; Yan, X. *Advanced Materials* **2016**, *28*, 1031–1043.

(47) Fry, H. C.; Garcia, J. M.; Medina, M. J.; Ricoy, U. M.; Gosztola, D. J.; Nikiforov, M. P.; Palmer, L. C.; Stupp, S. I. *Journal of the American Chemical Society* **2012**, *134*, 14646–14649.

(48) Zou, Q.; Zhang, L.; Yan, X.; Wang, A.; Ma, G.; Li, J.; Möhwald, H.; Mann, S. *Angewandte Chemie International Edition* **2014**, *53*, 2366–2370.

(49) Noy, D.; Moser, C. C.; Dutton, P. L. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2006**, *1757*, 90–105.

(50) Ennist, N. M.; Mancini, J. A.; Auman, D. B.; Bialas, C.; Iwanicki, M. J.; Esipova, T. V.; Discher, B. M.; Moser, C. C.; Dutton, P. L. In *Photosynthesis and Bioenergetics*; World Scientific: 2017, pp 1–33.

(51) Kodali, G.; Mancini, J. A.; Solomon, L. A.; Episova, T. V.; Roach, N.; Hobbs, C. J.; Wagner, P.; Mass, O. A.; Aravindu, K.; Barnsley, J. E.; Gordon, K. C.; Officer, D. L.; Dutton, P. L.; Moser, C. C. *Chemical Science* **2017**, *8*, 316–324.

(52) Koder, R. L.; Valentine, K. G.; Cerda, J.; Noy, D.; Smith, K. M.; Wand, A. J.; Dutton, P. L. *Journal of the American Chemical Society* **2006**, *128*, 14450–14451.

(53) Solomon, L. A.; Kodali, G.; Moser, C. C.; Dutton, P. L. *Journal of the American Chemical Society* **2014**, *136*, 3192–3199.

(54) Rau, H. K.; Snigula, H.; Struck, A.; Robert, B.; Scheer, H.; Haehnel, W. *European Journal of Biochemistry* **2001**, *268*, 3284–3295.

(55) Sheehan, M. M.; Magaraci, M. S.; Kuznetsov, I. A.; Mancini, J. A.; Kodali, G.; Moser, C. C.; Dutton, P. L.; Chow, B. Y. *Biochemistry* **2018**, *57*, 6752–6756.

(56) Mancini, J. A.; Sheehan, M.; Kodali, G.; Chow, B. Y.; Bryant, D. A.; Dutton, P. L.; Moser, C. C. *Journal of The Royal Society Interface* **2018**, *15*, 20180021.

(57) Cohen-Ofri, I.; van Gastel, M.; Grzyb, J.; Brandis, A.; Pinkas, I.; Lubitz, W.; Noy, D. *Journal of the American Chemical Society* **2011**, *133*, 9526–9535.

(58) Shively, J.; Cannon, G.; Heinhorst, S.; Fuerst, J.; Bryant, D.; Gantt, E.; Maupin-Furlow, J.; Schüler, D.; Pfeifer, F.; Docampo, R.; Dahl, C.; Preiss, J.; Steinbüchel, A.; Federici, B. In *Encyclopedia of Microbiology (Third Edition)*, Schaechter, M., Ed.; Academic Press: Oxford, 2009, pp 404–424.

(59) Discher, B. M.; Noy, D.; Strzalka, J.; Ye, S.; Moser, C. C.; Lear, J. D.; Blasie, J. K.; Dutton, P. L. *Biochemistry* **2005**, *44*, 12329–12343.

(60) Noy, D.; Discher, B. M.; Rubtsov, I. V.; Hochstrasser, R. M.; Dutton, P. L. *Biochemistry* **2005**, *44*, 12344–12354.

(61) Korendovych, I. V.; Senes, A.; Kim, Y. H.; Lear, J. D.; Fry, H. C.; Therien, M. J.; Blasie, J. K.; Walker, F. A.; DeGrado, W. F. *Journal of the American Chemical Society* **2010**, *132*, 15516–15518.

(62) Goparaju, G.; Fry, B. A.; Chobot, S. E.; Wiedman, G.; Moser, C. C.; Leslie Dutton, P.; Discher, B. M. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2016**, *1857*, 503–512.

(63) Solomon, L. A.; Witten, J.; Kodali, G.; Moser, C. C.; Dutton, P. L. *The Journal of Physical Chemistry B* **2022**, *126*, 8177–8187.

(64) Farid, T. A. et al. *Nature Chemical Biology* **2013**, *9*, 826–833.

(65) Ennist, N. M.; Stayrook, S. E.; Dutton, P. L.; Moser, C. C. *Frontiers in Molecular Biosciences* **2022**, *9*.

(66) Ennist, N. M.; Zhao, Z.; Stayrook, S. E.; Discher, B. M.; Dutton, P. L.; Moser, C. C. *Nature Communications* **2022**, *13*, 4937.

(67) Conlan, B.; Cox, N.; Su, J.-H.; Hillier, W.; Messinger, J.; Lubitz, W.; Dutton, P. L.; Wydrzynski, T. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **2009**, *1787*, 1112–1121.

(68) Mancini, J. A.; Kodali, G.; Jiang, J.; Reddy, K. R.; Lindsey, J. S.; Bryant, D. A.; Dutton, P. L.; Moser, C. C. *Journal of The Royal Society Interface* **2017**, *14*, 20160896.

(69) Zeng, X.-L.; Tang, K.; Zhou, N.; Zhou, M.; Hou, H. J. M.; Scheer, H.; Zhao, K.-H.; Noy, D. *Journal of the American Chemical Society* **2013**, *135*, 13479–13487.

(70) Hobbs, C. J.; Roach, N.; Wagner, P.; van der Salm, H.; Barnsley, J. E.; Gordon, K. C.; Kodali, G.; Moser, C. C.; Dutton, P. L.; Wagner, K.; Officer, D. L. *Materials Advances* **2020**, *1*, 1877–1885.

(71) Li, X.; Qiao, S.; Zhao, L.; Liu, S.; Li, F.; Yang, F.; Luo, Q.; Hou, C.; Xu, J.; Liu, J. *ACS Nano* **2019**, *13*, 1861–1869.

(72) Yanase, T.; Okuda-Shimazaki, J.; Mori, K.; Kojima, K.; Tsugawa, W.; Sode, K. *Biochemical and Biophysical Research Communications* **2020**, *530*, 82–86.

(73) Takeda, S.; Kamiya, N.; Arai, R.; Nagamune, T. *Biochemical and Biophysical Research Communications* **2001**, *289*, 299–304.

(74) Lee, B.; Takeda, S.; Nakajima, K.; Noh, J.; Choi, J.; Hara, M.; Nagamune, T. *Biosensors and Bioelectronics* **2004**, *19*, 1169–1174.

(75) Oohora, K.; Mashima, T.; Ohkubo, K.; Fukuzumi, S.; Hayashi, T. *Chemical Communications* **2015**, *51*, 11138–11140.

(76) Sun, H.; Zhang, X.; Miao, L.; Zhao, L.; Luo, Q.; Xu, J.; Liu, J. *ACS Nano* **2016**, *10*, 421–428.

(77) Mikkilä, J.; Anaya-Plaza, E.; Liljeström, V.; Caston, J. R.; Torres, T.; de la Escosura, A.; Kostiainen, M. A. *ACS Nano* **2016**, *10*, 1565–1571.

(78) Dgany, O.; Gonzalez, A.; Sofer, O.; Wang, W.; Zolotnitsky, G.; Wolf, A.; Shoham, Y.; Altman, A.; Wolf, S. G.; Shoseyov, O.; Almog, O. *Journal of Biological Chemistry* **2004**, *279*, 51516–51523.

(79) Miao, L.; Han, J.; Zhang, H.; Zhao, L.; Si, C.; Zhang, X.; Hou, C.; Luo, Q.; Xu, J.; Liu, J. *ACS Nano* **2014**, *8*, 3743–3751.

(80) Li, Y.; Xia, C.; Tian, R.; Zhao, L.; Hou, J.; Wang, J.; Luo, Q.; Xu, J.; Wang, L.; Hou, C.; Yang, B.; Sun, H.; Liu, J. *ACS Nano* **2022**, *16*, 8012–8021.

(81) Zhao, L.; Zou, H.; Zhang, H.; Sun, H.; Wang, T.; Pan, T.; Li, X.; Bai, Y.; Qiao, S.; Luo, Q.; Xu, J.; Hou, C.; Liu, J. *ACS Nano* **2017**, *11*, 938–945.

(82) Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; La Scola, B.; Suzan, M.; Claverie, J.-M. *Science* **2004**, *306*, 1344–1350.

(83) Zhao, Z.; Wang, J. C.-Y.; Segura, C. P.; Hadden-Perilla, J. A.; Zlotnick, A. *ACS Chemical Biology* **2020**, *15*, 3124–3132.

(84) Prasuhn, D. E.; Kuzelka, J.; Strable, E.; Udit, A. K.; Cho, S.-H.; Lander, G. C.; Quispe, J. D.; Diers, J. R.; Bocian, D. F.; Potter, C.; Carragher, B.; Finn, M. *Chemistry & Biology* **2008**, *15*, 513–519.

(85) Stephanopoulos, N.; Carrico, Z. M.; Francis, M. B. *Angewandte Chemie International Edition* **2009**, *48*, 9498–9502.

(86) Capehart, S. L.; Coyle, M. P.; Glasgow, J. E.; Francis, M. B. *Journal of the American Chemical Society* **2013**, *135*, 3011–3016.

(87) Golmohammadi, R.; Valegård, K.; Fridborg, K.; Liljas, L. *Journal of Molecular Biology* **1993**, *234*, 620–639.

(88) Marvin, D.; Welsh, L.; Symmons, M.; Scott, W.; Straus, S. *Journal of Molecular Biology* **2006**, *355*, 294–309.

(89) Croce, R.; van Amerongen, H. *Nature Chemical Biology* **2014**, *10*, 492–501.

(90) Gisriel, C. J.; Elias, E.; Shen, G.; Soulier, N. T.; Flesher, D. A.; Gunner, M. R.; Brudvig, G. W.; Croce, R.; Bryant, D. A. *Science Advances* **2023**, *9*, eadg0251.

(91) Scolaro, L. M.; Castriciano, M. A.; Romeo, A.; Micali, N.; Angelini, N.; Lo Passo, C.; Felici, F. *Journal of the American Chemical Society* **2006**, *128*, 7446–7447.

(92) Nam, Y. S.; Shin, T.; Park, H.; Magyar, A. P.; Choi, K.; Fantner, G.; Nelson, K. A.; Belcher, A. M. *Journal of the American Chemical Society* **2010**, *132*, 1462–1463.

(93) Park, H. et al. *Nature Materials* **2016**, *15*, 211–216.

(94) Engel, G. S.; Calhoun, T. R.; Read, E. L.; Ahn, T.-K.; Mančal, T.; Cheng, Y.-C.; Blankenship, R. E.; Fleming, G. R. *Nature* **2007**, *446*, 782–786.

(95) Ge, P.; Zhou, Z. H. *Proceedings of the National Academy of Sciences* **2011**, *108*, 9637–9642.

(96) Bhyravbhatla, B.; Watowich, S. J.; Caspar, D. L. *Biophysical Journal* **1998**, *74*, 604–615.

(97) Endo, M.; Wang, H.; Fujitsuka, M.; Majima, T. *Chemistry – A European Journal* **2006**, *12*, 3735–3740.

(98) Endo, M.; Fujitsuka, M.; Majima, T. *Chemistry – A European Journal* **2007**, *13*, 8660–8666.

(99) Miller, R. A.; Presley, A. D.; Francis, M. B. *Journal of the American Chemical Society* **2007**, *129*, 3104–3109.

(100) Ma, Y.-Z.; Miller, R. A.; Fleming, G. R.; Francis, M. B. *The Journal of Physical Chemistry B* **2008**, *112*, 6887–6892.

(101) Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. *Nano Letters* **2010**, *10*, 181–186.

(102) Miller, R. A.; Stephanopoulos, N.; McFarland, J. M.; Rosko, A. S.; Geissler, P. L.; Francis, M. B. *Journal of the American Chemical Society* **2010**, *132*, 6068–6074.

(103) Noriega, R.; Finley, D. T.; Haberstroh, J.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *The Journal of Physical Chemistry B* **2015**, *119*, 6963–6973.

(104) Delor, M.; Dai, J.; Roberts, T. D.; Rogers, J. R.; Hamed, S. M.; Neaton, J. B.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *Journal of the American Chemical Society* **2018**, *140*, 6278–6287.

(105) Lee, J.; Lee, D.; Kocherzhenko, A. A.; Greenman, L.; Finley, D. T.; Francis, M. B.; Whaley, K. B. *The Journal of Physical Chemistry B* **2018**, *122*, 12292–12301.

(106) Dai, J.; Knott, G. J.; Fu, W.; Lin, T. W.; Furst, A. L.; Britt, R. D.; Francis, M. B. *ACS Nano* **2021**, *15*, 8110–8119.

(107) Anaya-Plaza, E.; Aljarilla, A.; Beaune, G.; Nonappa; Timonen, J. V. I.; de la Escosura, A.; Torres, T.; Kostiainen, M. A. *Advanced Materials* **2019**, *31*, 1902582.

Chapter 2

Characterizing heterogeneous mixtures of assembled states of mutants of the tobacco mosaic virus

2.1 Abstract

The tobacco mosaic virus capsid protein (TMV) is a frequent target for derivatization for myriad applications including drug delivery, biosensing, and light harvesting. However, solutions of the stacked disk assembly state of TMV are difficult to characterize quantitatively due to their large size and multiple assembled states. Charge detection mass spectrometry (CDMS) addresses the need to characterize heterogeneous populations of large protein complexes in solution quickly and accurately. Using CDMS, previously unobserved assembly states of the circular permutant of TMV (cpTMV), including 16-monomer disks and odd-numbered disk stacks, have been characterized. We additionally employed a peptide–protein conjugation reaction in conjunction with CDMS to demonstrate that modified cpTMV proteins do not redistribute between disks. CDMS was also used to discriminate between protein complexes of near-identical mass but different configurations. Molecular dynamics simulations of point mutants of intact cpTMV double disk assemblies demonstrated how point mutations affecting the protein's surface charge impact the assembly state. Additionally, removing the flexible N- and C-terminal regions of cpTMV was demonstrated to increase the number of cpTMV disks per assembly. These measurements have led to a greater understanding of the behavior of TMV, a protein used across a broad variety of fields and applications, in the solution state.

This chapter is based on the following publication:

Bischoff, A. J.; Conner, C. H.; Williams, E. R.; Francis, M. B. Characterizing Heterogeneous Mixtures of Assembled States of the Tobacco Mosaic Virus Using Charge Detection Mass Spectrometry. *J. Am. Chem. Soc.* **2022**, *144*, 23368–23378.

This article can be accessed via the following link:
https://pubs.acs.org/doi/full/10.1021/jacs.2c09160

2.2 Introduction

Particles derived from the tobacco mosaic virus (TMV) have broad potential in drug and vaccine development,[1–3] sensory assays,[4,5] filtration,[6] and as mimics of photosynthetic light harvesting complexes,[7–9] among myriad other uses.[10,11] This is partially due to TMV's unique assembly state as a 300 nm helix in its native form or as stacked disks under varied environmental conditions.[12] TMV also has a remarkable capacity for assembling into disk or rod structures when subject to a vast array of mutations to the protein surface, in the pore, and within the cavity between the disks.[8,13–16] The ability of a variety of mutants and conjugates of TMV to maintain ordered assemblies has enabled its functionalization with therapeutics for drug delivery applications,[17,18] or dyes or nanoparticles in synthetic light harvesting systems.[7,8,19] TMV can also be assembled into nanoscale materials with morphologies that do not occur naturally, such as nanosheets exhibiting filtration ability[6] and spherical nanoparticles used for drug delivery.[20,21]

In the absence of a nucleic acid template, TMV can form multiple assembly states sensitive to environmental conditions. At lower pH and ionic strength, TMV monomers assemble into helical rods with 16 $\frac{2}{3}$ monomers per helical turn, similar to the native structure, but assemblies of stacked disks consisting of 17 monomers per disk are formed at neutral pH and higher ionic strengths (PDB ID 1EI7, Figure 2.1a-c).[22] Mutations made to the TMV capsid can also bias the structure toward forming rods or stacked disks. One example of this is a circular permutant, cpTMV, which favors two-disk stacks under a wider variety of pH and ionic strength conditions than native TMV (Figure 2.1d-e).[13] In its stacked disk form, solutions of TMV often consist of heterogeneous assembled states of even-numbered disk stacks,[13,15] which can make detailed characterization challenging. Progress has been made recently in low-resolution characterization of TMV particles of various size in the solution state using small angle x-ray scattering.[23] However, probing how single mutations may effect subtle structural changes within a single disk or change the equilibrium of multiple assembly populations present in solution remains a challenge.

TMV conjugates have been studied extensively for therapeutic applications, and we have previously made use of TMV conjugates as a tunable platform for modeling energy transfer processes in photosynthetic light harvesting complexes. These models consist of TMV labeled covalently with synthetic dyes in circular arrays on the assembly surface or within the region between individual disks.[7,8] In measurements of energy transfer in these systems, a detailed understanding of the interchromophore distances and relative orientations, as well as protein environment, is crucial to interpreting spectroscopic data.[24] In addition, a concrete picture of the timescales of disassembly and reassembly of the complexes in solution is necessary for experiments involving energy transfer between pairs of donor and acceptor complexes. In the context of TMV conjugate drug and vaccine candidates, a clear analysis

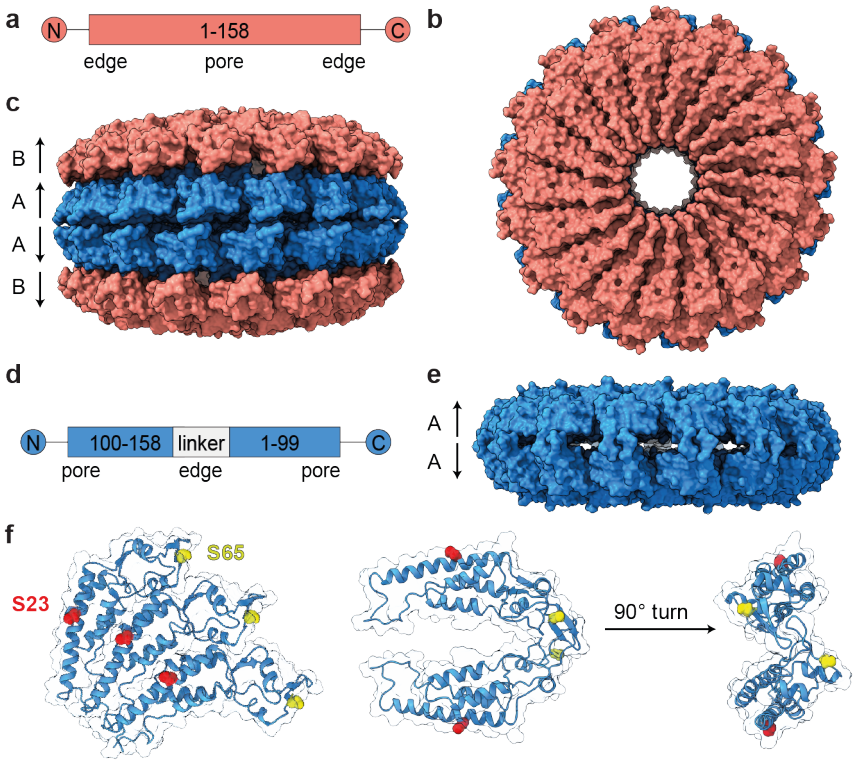


Figure 2.1: Structures of TMV used as light harvesting complex mimics. (a) A gene map, (b) face view, and (c) side view of a four-layer stack of wtTMV disks displays C_2 symmetry about the horizontal axis, with A disks being identical to one another and B disks identical to one another but with slight structural variation from A disks (PDB ID: 1EI7). (d) A gene map and (e) side view of a homology model of cpTMV show the favored structure consisting of two disks, containing 17 monomers each, with C_2 symmetry (PDB ID: 3KML). (f) Engineered residues used for attachment to chromophores or surfaces in light harvesting mimics are shown for three monomers on one cpTMV disk, with the S23 site shown in red and S65 site in yellow. The S23 and S65 sites are also shown for two monomers on opposing disks from a cut-away and edge-on view.

of all species present and their assembly kinetics is also vital to understanding the behavior of these conjugates in vivo.[25–27] The use of high-resolution structural characterization techniques, including x-ray crystallography, cryo-electron microscopy, and NMR spectroscopy, to characterize protein samples at the unique intersection of large (MDa+) size and sample heterogeneity characteristic of TMV conjugates is typically challenging and labor-intensive.[28] In contrast, native electrospray ionization charge detection mass spectrometry (ESI-CDMS) is a promising method for providing a detailed and quantitative analysis of heterogeneous TMV assembly states in solution and their dynamics.

Traditional electrospray ionization mass spectrometry (ESI-MS) is limited in its ability to measure the mass of large protein complexes. The ionization process cleaves non-covalent bonds, so that only information about the mass of monomers can be recovered, and information about quaternary structure cannot be obtained.[29] Additionally, traditional MS is

limited in its ability to measure high mass structures due to the mass reconstruction process. Traditional MS measures the mass versus charge (m/z) of an analyte solution, which produces a charge ladder from which the analyte mass can be constructed. This charge ladder is obscured at higher masses by an abundance of charged species and by salt adduction, for example, with Na^+ ions.[30–32] Sample heterogeneity can further complicate mass reconstruction at high masses.[31] Native ESI uses gentler ionization conditions that preserve non-covalent interactions between monomers in protein complexes.[33] CDMS measures the charge in addition to m/z of individual ions, so that the mass can be directly calculated from a population of charge and m/z data.[34–36]

Salt ions and non-volatile buffers provide additional challenges to measuring the mass of large complexes. This problem is usually mitigated by exchanging samples into a volatile buffer prior to measurement; however, the choice of buffer can affect the higher order structure of protein complexes.[37] The use of submicron electrospray emitter tips addresses this problem by limiting the number of salt ions present in a single droplet.[38] The measurement of individual ions additionally helps to overcome the hurdle of cluttered charge ladders to allow for the measurement of protein complexes at multi-MDa sizes. CDMS has been used to measure the intact mass of many large biomolecules including large macromolecules[39–41] and virus capsid assemblies.[42,43] This chapter will demonstrate the utility of CDMS in measuring highly heterogeneous populations of TMV assembly states and their dynamics in solution. This chapter will also probe the correlation between the TMV sequence and observed assembly states through molecular dynamics simulations and transmission electron microscopy (TEM) of cpTMV mutants.

2.3 Results and Discussion

2.3.1 Effects of point mutations on cpTMV disk subunit stoichiometry

The circular permutant of TMV, cpTMV, maintains its ability to fold and assemble when subject to a variety of mutations at sites on the protein surface and within the central cavity.[8] This ability is crucial to its use as a mimic of photosynthetic light harvesting complexes, as it allows for the attachment of chromophores in sites with varying protein, solvent, and adjacent chromophore environments that can then be compared. A primary location for TMV functionalization to light harvesting materials in addition to nanosheets is at residue S3 of recombinant TMV (rTMV) and the analogous S65 position of cpTMV. This site, on the disks' periphery, has been mutated to cysteine for the formation of nanosheet filters and to *p*-aminophenylalanine (pAF) for attachment to nanoparticles (Figure 2.1f).[6,19] Another widely utilized location for TMV functionalization is position S23 on the face of cpTMV and its analogous position on rTMV, which is readily mutable to cysteine and has been used for attachment of synthetic dyes and small molecule chemotherapeutics (Figure 2.1f).[7,8,18] However, the effects of point mutations on the assembly state of cpTMV have not previously

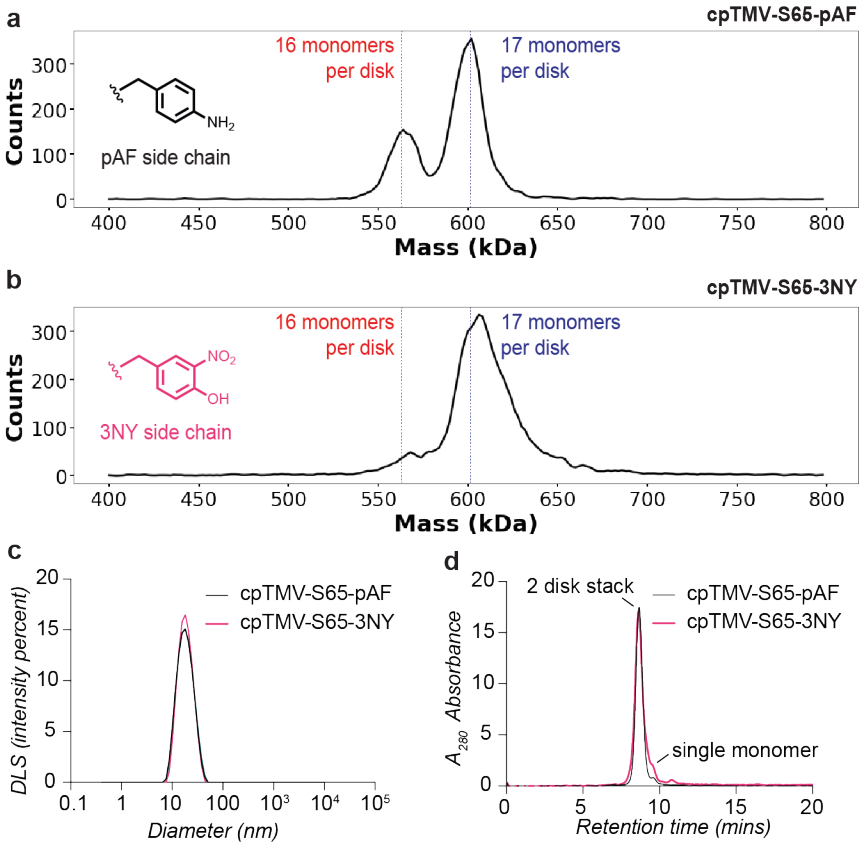


Figure 2.2: Mutants of TMV displaying disk assemblies with different symmetries. (a) A mass histogram of a circular permutant of TMV with a peripheral *p*-aminophenylalanine (pAF) mutation (cpTMV-S65-pAF) in 100 mM ammonium acetate solution clearly differentiates between 16-fold symmetric and 17-fold symmetric stacks of two disks. (b) A mutant of cpTMV with 3-nitrotyrosine (3NY) at the same site, cpTMV-S65-3NY, in 100 mM ammonium acetate solution contains a much smaller ratio of 16:17-mers. Distinguishing between 16- and 17-mers is challenging by other common characterization methods, with no differentiation observed using (c) dynamic light scattering and (d) size exclusion chromatography.

been studied in detail. Reliably and efficiently measuring small differences in macromolecular assemblies is a strength of CDMS that we sought to apply to understand how point mutations affect both stacking behavior and the intradisk architecture of cpTMV.

A peripheral mutation from serine to the non-canonical amino acid *p*-aminophenylalanine (cpTMV-S65-pAF) of cpTMV was of interest due to its position at the interface of adjacent monomers within a single disk and its use as a handle for nanoparticle attachment.[19] The expected mass for disks consisting of 17 subunits is 302 kDa, resulting in a mass of 604 kDa for the anticipated double disk assembly state. However, the CDMS mass histogram for this mutant displayed a high proportion of 16-monomer disks, with a mass of 570 kDa for a two-disk stack, in addition to the 607 kDa species observed for 17-monomer disks (Figure 2.2a). Although the helical turn found in wtTMV is made up of 16 $\frac{2}{3}$ monomers per turn, previously determined crystal structures have shown that disks in discrete stacks consist of 17 monomers per disk.[22] However, by electrospray ionization from a 100 mM AA solution and

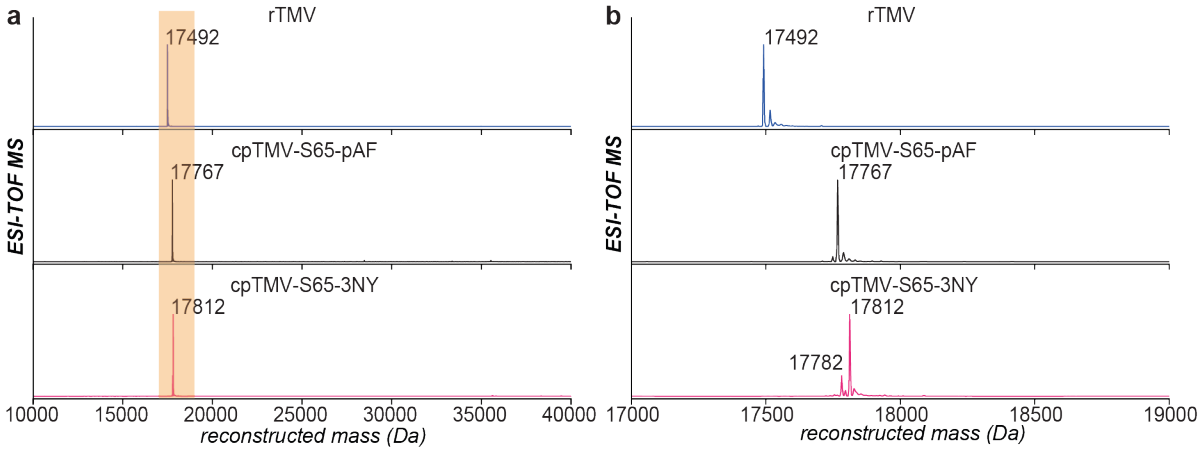


Figure 2.3: Denaturing mass spectra of TMV mutants assembling into odd-numbered disk stacks or 16-monomer disks. (a) The mass spectra of rTMV (expected MW: 17493 Da), cpTMV-S65-pAF (expected MW: 17767 Da), and cpTMV-S65-3NY (expected MW: 17813 Da) are shown, indicating high purity and homogeneity of each sample. The orange bar indicates an enlarged window that is depicted in (b). (b) An enlarged window of the same spectra in (a) shows a small proportion of the reduced version of cpTMV-S65-3NY in which the nitrophenol is spontaneously reduced to an aminophenol (expected MW: 17783 Da). This mass difference would not be expected to significantly affect the assembled mass of a 34-monomer complex (at an expected mass of 605 kDa for a 34-monomer complex of 17783 Da monomers versus an expected mass of 606 kDa for a 34-monomer complex of 17812 Da monomers.)

CDMS, two distinct populations of 16- and 17-monomer disks could be measured in tandem, indicating that both are present in solution. No peak consisting of hybridized 16- and 17-mer disks at 584 kDa was observed, indicating that if a hybridized species of 16 and 17-mer disks exists, its abundance is minimal. Instead, the bias toward forming stacks of disks of like number (e.g. two 16-mers or two 17-mers) suggests that there is a strong pairwise interaction between individual monomers on opposing disks in these cpTMV assemblies. Several other cpTMV variants with mutations in the S65 position were also assessed for possible assembly into 16-monomer disks. A mutant with 3-nitrotyrosine (3NY) at position S65, but otherwise identical to cpTMV-S65-pAF, is structurally similar but has a decreased charge with respect to the pAF mutant. The cpTMV-S65-3NY mutant still displayed a small population of 16-monomer disks, but at a much lower proportion of 16:17 monomers than for the pAF mutant (Figure 2.2b). Mass spectra under denaturing conditions of cpTMV-S65-pAF and cpTMV-S65-3NY showed that the monomers of each population have homogeneous masses, indicating that the presence of multiple assembled states is not due to monomer heterogeneity (Figure 2.3). Other mutants, including cpTMV-S65C and cpTMV without any mutation at the S65 position, also displayed small proportions of 16-monomer populations at lower levels than for the pAF mutant (Figure 2.4). This fraction of 16-monomer disks is barely distinguishable from noise in several of the mutants; however, in all mass histograms of cpTMV mutants, a small number of 16-monomer disks were detected.

Detecting the subtle difference in size between the 16-monomer and 17-monomer assemblies is difficult using other methods typically used for measuring the mass of biomolecules

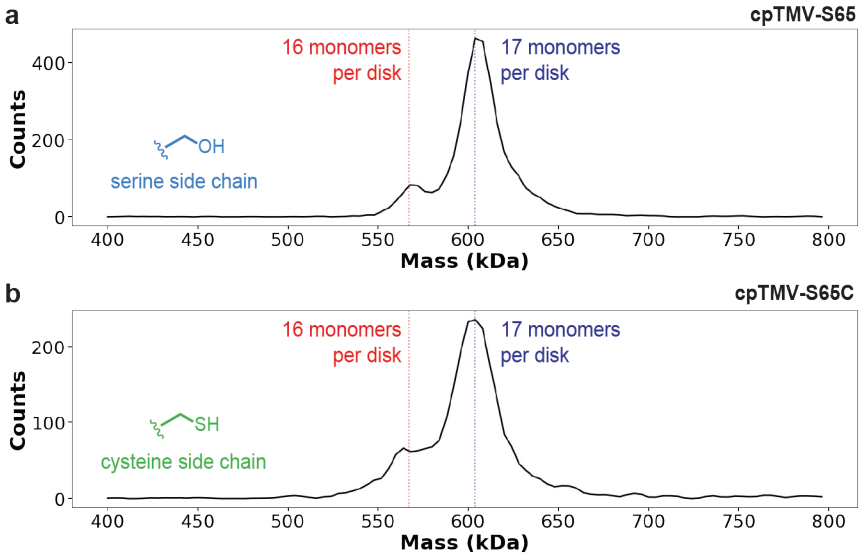


Figure 2.4: Ratio of 16:17-mers of cpTMV point mutants at position S65. (a) A CDMS histogram of cpTMV with the native serine residue at position S65 shows a small proportion of 16-monomer disks alongside 17-monomer disks in a 2-disk stack. (b) A CDMS histogram of cpTMV-S65C also shows a small proportion of 16-monomer disks in a 2-disk stack.

in the size range of these TMV assemblies (>600 kDa). For example, a measurement of cpTMV-S65-pAF using dynamic light scattering (DLS) results in a single, uniform peak with no obvious asymmetries that might be attributable to stacks of both 16-monomer and 17-monomer disks (Figure 2.2c). A comparison of the cpTMV-S65-pAF mutant to the cpTMV-S65-3NY mutant using size exclusion chromatography-high performance liquid chromatography (SEC-HPLC) yielded essentially identical retention times and only a single resolvable peak in the 600 kDa range (Figure 2.2d). Assessing and understanding the uniformity of the disk assemblies are crucial to the materials-based applications for which TMV is used, such as components of drug[18] and vaccine candidates[44] and as models for light harvesting systems.

To understand the structural features leading to the cpTMV-S65-pAF mutant forming a greater proportion of 16-monomer assemblies than other cpTMV mutants, the structures of cpTMV with the native serine as well as mutations to pAF and 3NY at position 65 were analyzed. This analysis presented challenges because there was a lower proportion of 16-monomer disk populations versus 17-monomer disk populations in all mutants measured, although the proportions differed between mutants. Visual examination of the inter-monomer interface between adjacent monomers and monomer pairs on two different disks did not indicate that the serine at position 65 was strongly associated with residues on either adjacent or opposing monomers. Molecular dynamics (MD) simulations of the mutants were therefore conducted to provide insight into why the mutations influence cpTMV disks' monomer stoichiometry. Due to the large size of the cpTMV double disk, a model using only two monomers of cpTMV was first used to analyze inter-monomer interactions over a 100 ns simulation. While using a cpTMV dimer reduced the computational time versus using intact assemblies,

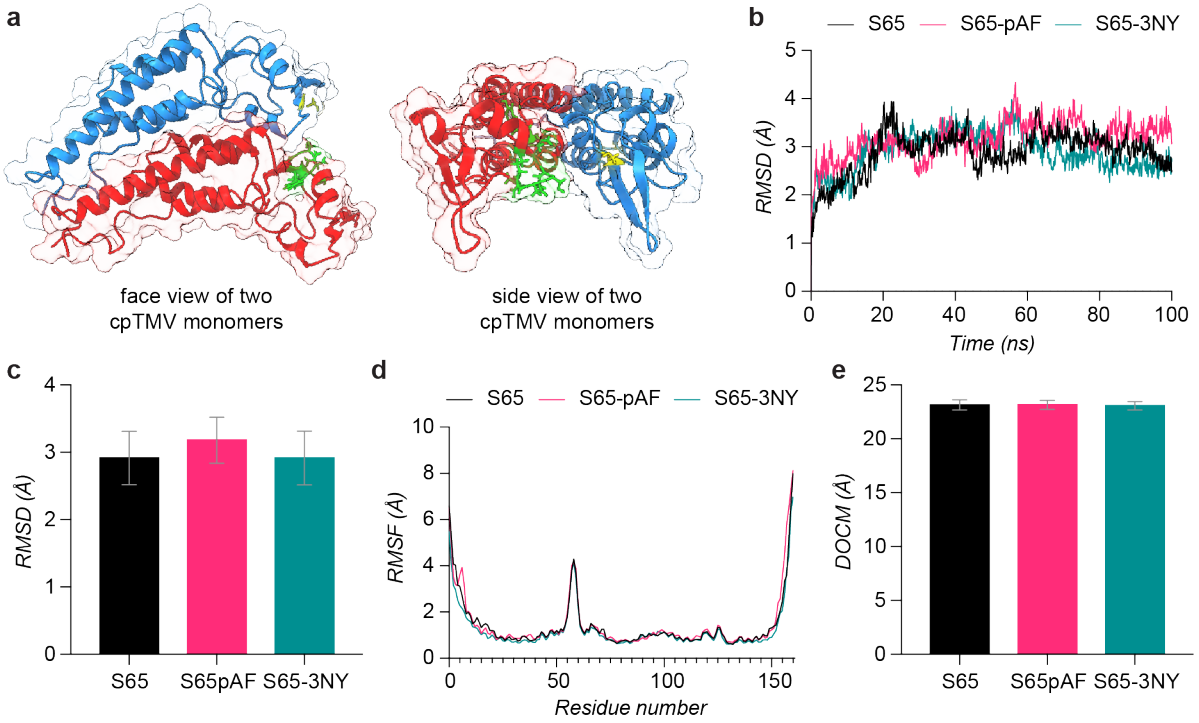


Figure 2.5: Molecular dynamics simulations of cpTMV dimers. (a) Two adjacent cpTMV monomers are shown in red and blue, with the mutation site S65 shown in yellow. Residues closest to the mutation site are shown in green, demonstrating that no direct interactions are present. (b) The root-mean-square deviation (RMSD) over the 100 ns simulation for each mutant are compared. (c) The average RMSD ± SD from 20–100 ns for each mutant are compared. (d) The root-mean-square fluctuation (averaged across the two monomers) over the 100 ns simulation shows flexibility at the termini and loop region containing residue 65. (e) The average distance of the center of mass ± SD between the two monomers from 20–100 ns for each mutant are compared.

these simulations could only provide insight into interactions between monomers and would not capture features dependent on the complete assembly structure, such as the change in the assembly's charge state upon making point mutations to all monomers. Two adjacent monomers within the same disk were initially used as an input structure for MD simulations (Figure 2.5a), containing either the native serine at position 65 or the non-canonical amino acid residues pAF and 3NY.

For each mutant, the two monomers maintained their basic tertiary structure and remained associated with one another over the course of the 100 ns simulation. The root-mean-square deviation (RMSD) of the C_α carbons, an indication of the extent a protein's structure changes over the course of a simulation, was analyzed for each mutant. The RMSD across the 100 ns trajectory for each cpTMV dimer did not result in any clear differences between each mutant (Figure 2.5b), with each mutant showing an initial spike and then random fluctuations centering around 3 Å. Averaging the RMSD across the 20–100 ns time period indicated slightly more movement in the pAF mutant than in the other two mutants

(Figure 2.5c, Table 2.1), suggesting that the protein's structure may be disrupted more by the S65-pAF mutation than by S65-3NY. The 0–20 ns time period was not included in the averaging to avoid skewing the results based on large fluctuations that can occur during initial relaxation of the structure at early time points. A comparison of the root-mean-square fluctuation (RMSF) of each C_α carbon (averaging the same residue on both monomers) showed that the greatest flexibility occurred at the N- and C-termini, as well as in the loop region between residues 56–65 (Figure 2.5d). However, no clear differences between the the RMSF of each mutant were observed. Finally, the distance of the center of mass (DOCM) between the two monomers of each dimer was averaged across the 20–100 ns time period (Figure 2.6e, Table 2.1). A smaller DOCM may indicate a stronger association between monomers in a certain mutant, which may increase the mutant's propensity to form 16-monomer disks. No significant difference in the DOCM was observed between each mutant dimer.

Table 2.1: Molecular dynamics trajectory analysis of cpTMV dimers.

Mutant	Average RMSD[a,c]	Average DOCM[b,c]
S65	2.9 ± 0.4	23.1 ± 0.5
S65-pAF	3.2 ± 0.3	23.1 ± 0.4
S65-3NY	2.9 ± 0.4	23.1 ± 0.4

[a] root-mean-square deviation [b] distance of the center of mass [c] average ± standard deviation

While analysis of the MD simulations of the cpTMV dimers suggested that the S65-pAF mutant may be slightly more flexible than the other two mutants, no significant differences were observed among the mutants. Thus, MD simulations were conducted, and the trajectories analyzed, of intact double disks with either 16 or 17 monomers per disk (Figure 2.6a) and either the native serine or pAF or 3NY at position 65. At ∼0.5 million atoms and an 18 nm size, double disks of cpTMV are nearing a limiting size for all-atom MD simulations.[45] Nevertheless, these simulations proceeded without issue and without great disruption to the protein's tertiary structure and double disk assembly state. Averaging the RMSD of both 16- and 17-monomer disks over the 20–100 ns time period resulted in similar results for each mutant (Figure 2.6b-c, Table 2.2), with the exception of the 16-monomer mutant with the native serine at position 65. CDMS results indicated that the S65 mutant had an intermediate proportion of 16-monomer disks between the pAF and 3NY mutants; therefore, this result did not provide insight into what may cause a higher propensity to form 16-monomer disks. An individual comparison of the RMSD of the 16- and 17-monomer disks across the 100 ns trajectory for each mutant (Figure 2.6d) also demonstrates that most mutants show normal behavior of an increasing RMSD that steadily levels off over time. However, the 16-monomer pAF mutant shows more irregular behavior over the course of the simulation, with two large jumps.

Differences in the RMSF of the C_α carbons of each mutant were also examined (Figure 2.6e). This was done by averaging the RMSF of the same residue across all monomers of the disks, 32 for the 16-monomer disks and 34 for the 17-monomer disks. The averaged

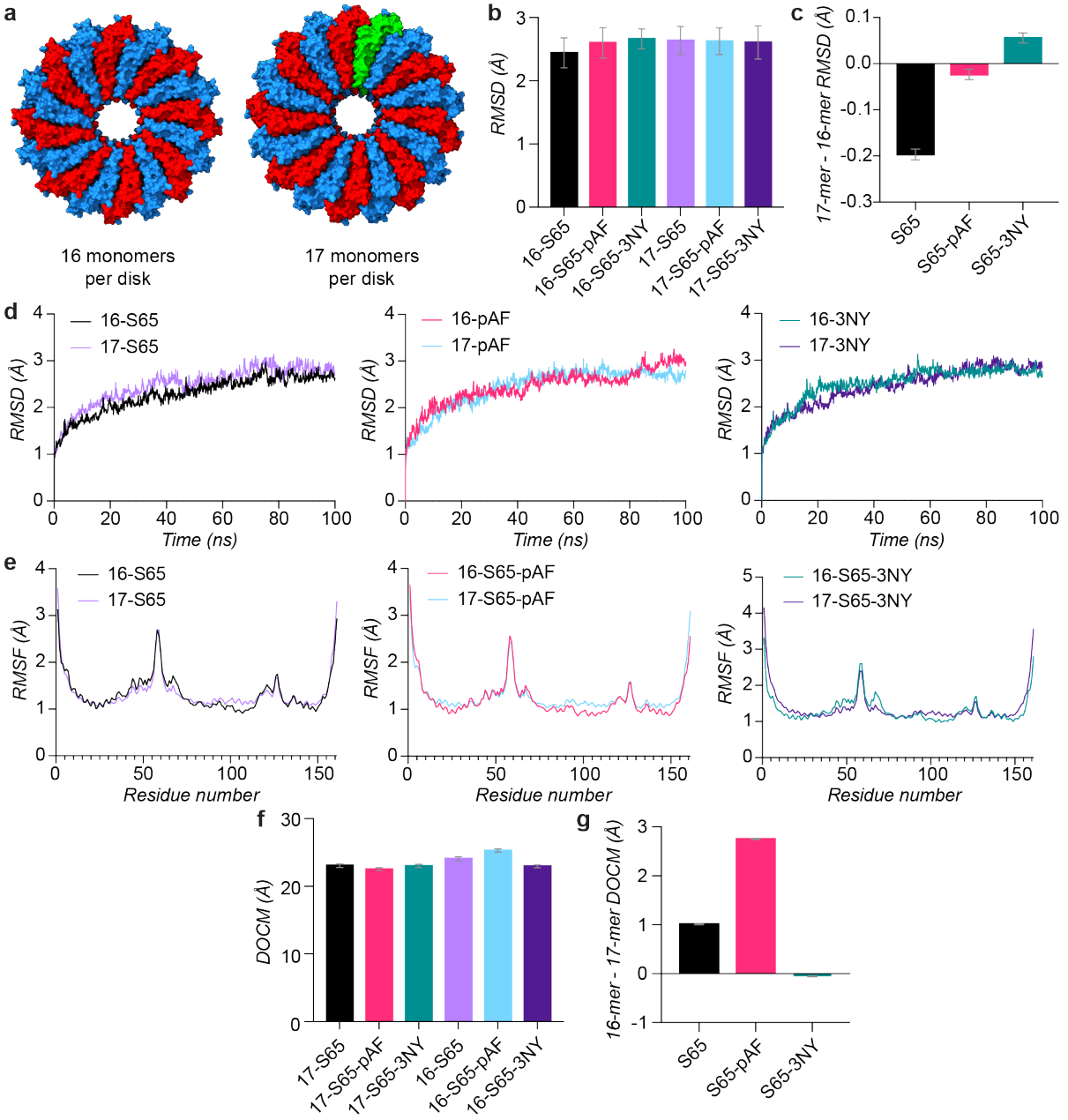


Figure 2.6: Molecular dynamics (MD) simulations of intact cpTMV disks. (a) Double disks with 16 monomers per disk (32 monomers total) and 17 monomers per disk (34 monomers total) used as input structures for the MD simulations are shown. (b) The average root-mean-square deviation (RMSD) ± SD from 20–100 ns for each mutant are compared. (c) The difference between RMSD ± SE between the 17-monomer disks and 16-monomer disks are compared. (d) The RMSD over the 100 ns simulation for each mutant are compared, grouped by S65 mutation. (e) The root-mean-square fluctuation (averaged across the all monomers) over the 100 ns simulation, grouped by S65 mutation, shows flexibility at the termini and loop region containing residue 65. (f) The average distance of the center of mass (DOCM) ± SD between two adjacent monomers from 20–100 ns for each mutant are compared. (g) The difference between DOCM ± SE between the 16-monomer disks and 17-monomer disks are compared.

Table 2.2: Molecular dynamics trajectory analysis of cpTMV double disk assemblies RMSD.

Mutant	17-mer RMSD[a,b]	16-mer RMSD[a,b]	17-mer -16-mer RMSD[a,c]
S65	2.6 ± 0.2	2.4 ± 0.2	-0.19 ± 0.01
S65-pAF	2.6 ± 0.2	2.6 ± 0.2	-0.02 ± 0.01
S65-3NY	2.6 ± 0.3	2.7 ± 0.2	0.10 ± 0.01

[a] root-mean-square deviation [b] average ± standard deviation [c] average ± standard error

RMSF values showed similar trends as in the cpTMV dimer simulations, with the largest fluctuations occurring at the N- and C-termini and the loop region between residues 56–65. No large differences were observed between the 16- and 17-monomer disks of each mutant, although the 3NY mutant appeared to have slightly more fluctuation in the 16-monomer than in the 17-monomer case. The DOCM was also calculated between two adjacent monomers chosen at random for each structure (Figure 2.6f, Table 2.3). The DOCM appeared highest for the 16-monomer S65-pAF mutant, followed by the 16-monomer S65 mutant, and similar for all other mutants. Taking the difference between the DOCM of the 16-monomer and 17-monomer disks for each mutant (Figure 2.6g, Table 2.3) further demonstrates the increased inter-monomer distance between the 16-monomer and 17-monomer pAF disks. In future work, averaging across the DOCM of all adjacent monomers in each structure would more rigorously demonstrate differences in adjacent monomer proximity.

Table 2.3: Molecular dynamics trajectory analysis of cpTMV double disk assemblies DOCM.

Mutant	17-mer DOCM[a,b]	16-mer DOCM[a,b]	16-mer -17-mer DOCM[a,c]
S65	23.1 ± 0.3	24.1 ± 0.3	1.01 ± 0.01
S65-pAF	22.5 ± 0.2	25.3 ± 0.3	2.75 ± 0.01
S65-3NY	23.0 ± 0.3	23.0 ± 0.2	-0.06 ± 0.01

[a] distance of the center of mass [b] average ± standard deviation [c] average ± standard error

Based on the increased DOCM in the 16-monomer S65-pAF assembly, a structural comparison of the S65-pAF and S65-3NY mutants was conducted for the final, lowest-energy frame of each MD trajectory. An observation of the side view of each mutant suggests that the 16-monomer S65-pAF assembly (in red) has greater structural disruption than the other assemblies, including the 17-monomer S65-pAF assembly (Figure 2.7a). To gain a visual understanding of the structural differences, these assemblies were also superimposed on one another using the align action in PyMol (Figure 2.7b). An overlay of the two 17-monomer disks (pink and yellow) did not reveal any significant differences between the disks. In contrast, an overlay of the two 16-monomer disks (cyan and red) shows that the pAF mutant appears to take up a greater volume than the 3NY mutant, evidenced by the pAF disk having a greater radius in top and side views of the overlays. A superimposition of the 3NY disks shows that the 16-monomer disk (cyan) is smaller than the 17-monomer disk (pink), evidenced by the smaller pore size of the cyan disk. Finally, a superimposition of the pAF disks shows that the 16-monomer disk (red) appears to take up the same volume as the

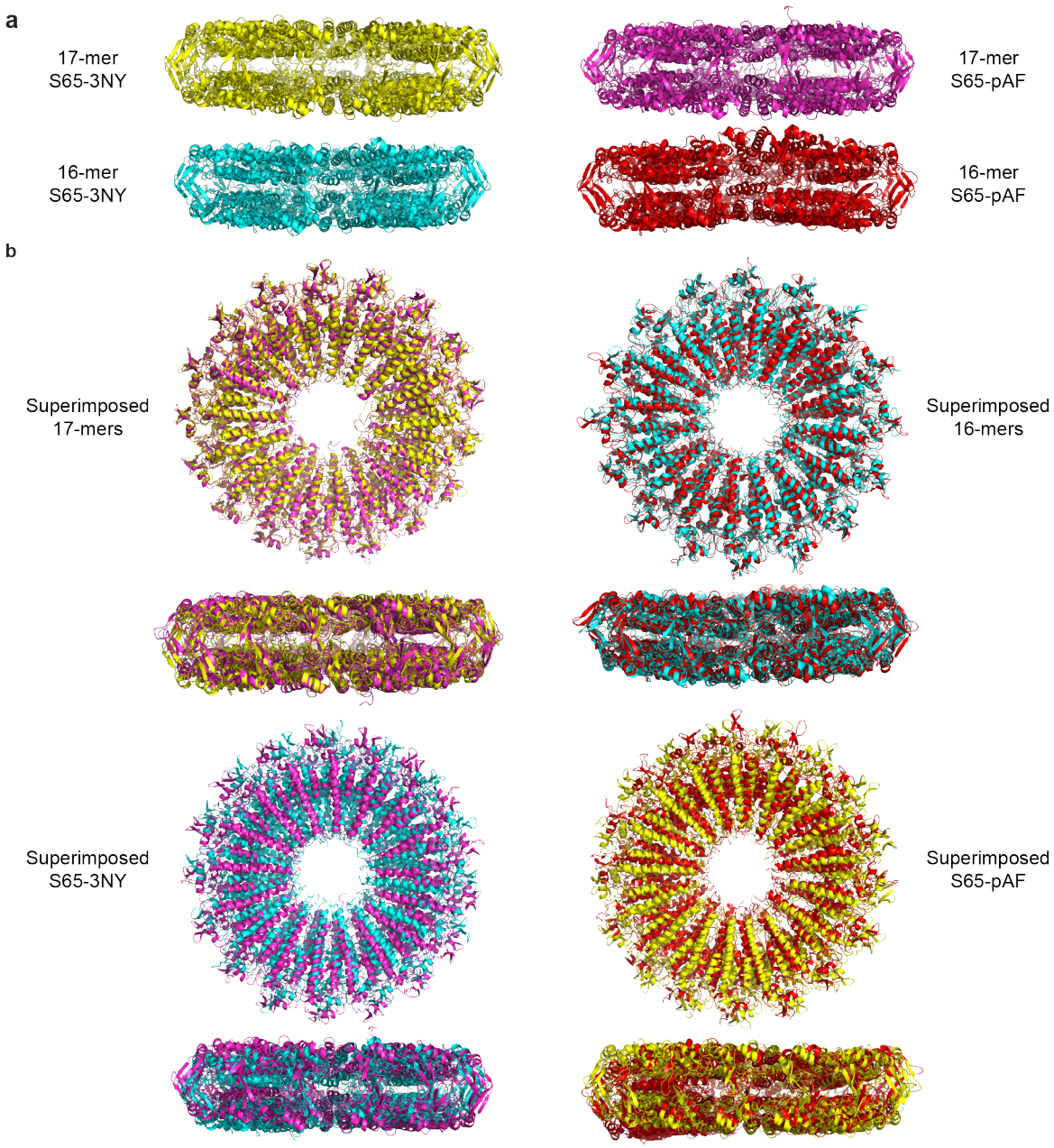


Figure 2.7: Superimposition of the final frame of 100 ns molecular dynamics simulations of cpTMV mutants. (a) The side views of the final frame of 100 ns molecular dynamics simulations of 16- and 17-monomer cpTMV disks with pAF and 3NY mutations at position 65 are shown. Significant disorder is observed in the 16-monomer pAF disk in red. (b) Superimposition of various combinations of the structures in (a) suggest that the 16-monomer pAF mutant has a larger radius than the 16-monomer 3NY mutant.

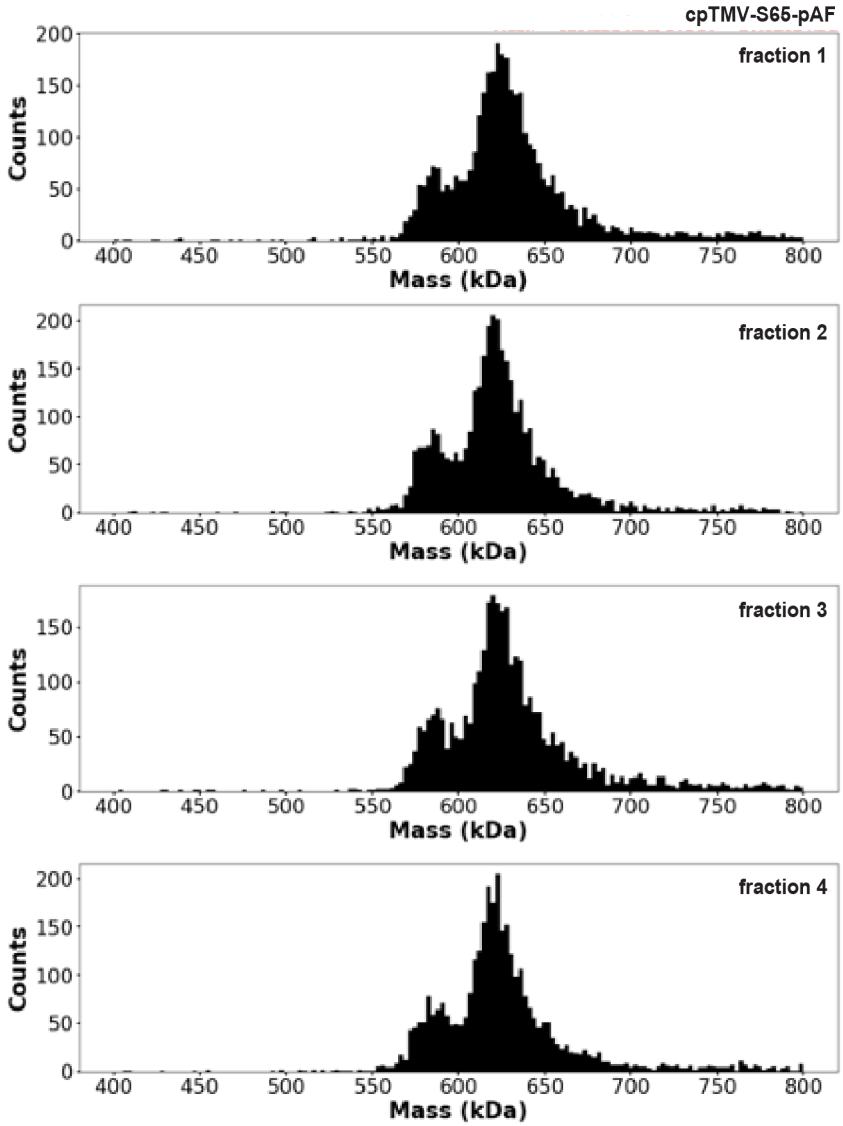


Figure 2.8: Anion exchange fractions of cpTMV-S65-pAF. No separation was observed in earlier or later fractions.

17-monomer disk (yellow). This visually supports the DOCM data and indicates that the 16-monomer S65-pAF disk occupies a greater volume than 16-monomer disks of the other mutants.

A causal link between the propensity to form 16-monomer disks and the residue at position 65 was then pursued, based on the DOCM and superimposition data. Monomers of cpTMV containing the native serine at position 65 have a theoretical isoelectric point of 4.92,[46] indicating that cpTMV assemblies are negatively charged at the neutral or close to neutral pH values used for CDMS measurements. The native amino acid serine at position 65 is a weakly hydrophilic amino acid.[47] By comparison, 3NY is expected to be 50% deprotonated at pH 7.2 based on the 2-nitrophenol pKa of 7.2,[48] although the extent of deprotonation

may be altered by the protein environment. In contrast, pAF is more hydrophobic than both 3NY and serine. The increased flexibility and volume of the 16-monomer pAF variant may provide a slight advantage based on the hydrophobic effect, allowing for the pAF residues to more closely associate with other hydrophobic regions on adjacent monomers. The difference in the polarity of the protein surface is therefore the most likely reason for the difference in proportion of 16- to 17-monomer disks for the different point mutants. However, the exact interactions leading to these differences are still unclear. Attempts to separate the 16-monomer disks from 17-monomer disks using anion exchange were also made. This method should separate assemblies of different charge states, with assemblies of greater negative charge binding more strongly to the resin and eluting later. This separation method was unsuccessful, with the same proportion of 16 to 17 monomer disks observed in earlier and later fractions (Figure 2.8). It is possible that the resin used was not sensitive enough to capture small differences between the 16- and 17-monomer disk charge states, or that the charge differences are negligible. Due to their similarity in size, size exclusion chromatography is also unlikely to separate the two populations. In future work, it is possible that a careful selection of disassembly and reassembly conditions may be able to favor one assembly state over the other.

2.3.2 Equilibration of monomeric cpTMV and cpTMV disk assemblies

In TMV-based artificial light harvesting systems, donor and acceptor chromophores can be embedded within protein complexes in defined locations for the measurement of directional energy transfer.[7] Inter-chromophore energy transfer is likely to be affected by the sensitivity of non-covalent interactions holding the assembled TMV together, and on what scale monomers may be exchanged from one complex to another (a schematic is shown in Figure 2.9). We investigated a potential equilibrium between intact TMV disk complexes and monomers in solution using CDMS. Often, a small proportion of monomeric TMV is present in purified samples as evidenced by size exclusion chromatography (Figure 2.2d), but it is difficult to determine whether the disk and monomer states are interchanging or whether the two populations remain unequilibrated. To probe this potential equilibrium, we distinguished one cpTMV population from another through irreversible labeling of each monomer with a 1788 Da peptide, *N*-acetyl-α-endorphin. We then mixed the α-endorphin-labeled population with unlabeled cpTMV and monitored the mass distributions within the mixture. Convergence to a single mass over time would indicate the exchange of monomers between the unlabeled and labeled species.

Labeling of cpTMV-S23C-S65-pAF with α-endorphin was performed using the enzyme tyrosinase from *Agaricus bisporus*, which activates exposed tyrosine residues toward nucleophilic addition.[49,50] The N-terminal tyrosine of α-endorphin was oxidized to an *o*-quinone with tyrosinase, to which either the engineered cysteine or aniline of pAF in cpTMV-S23C-S65-pAF could add (Figure 2.10a). This labeling resulted in a highly heterogeneous sample

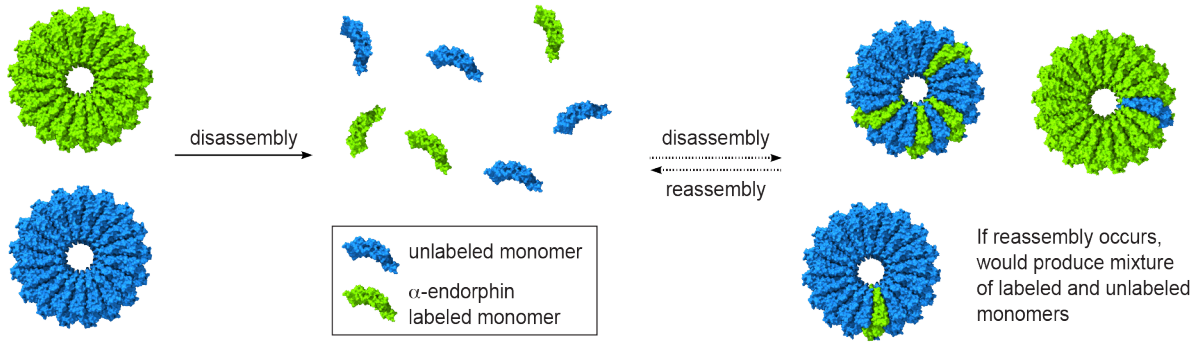


Figure 2.9: A schematic of the experiment mixing labeled and unlabeled cpTMV populations. This experiment was used to determine whether disassembly of cpTMV disks into monomers is reversible under a specific set of conditions (in this work, incubating the protein at room temperature in solution at a concentration of 0.5 mg/mL in 100 mM ammonium acetate buffer). One cpTMV disk population at a higher molecular weight, labeled with α-endorphin (shown in chartreuse) was mixed with an equivalent concentration of unlabeled cpTMV, at a lower molecular weight (shown in blue). If complete or incomplete disassembly and reassembly occured in solution, disks containing mixtures of unlabeled and labeled disks would be expected to form. If disassembly is irreversible, only uniform disks would be expected to be observed.

consisting of double disk complexes incorporating 32+ monomers with between 1–3 attached peptides per monomer. The sample was further complicated by a truncation that occurred during purification, leading to final complexes containing monomers with eight different possible masses (Figure 2.10b). Additionally, both 32- and 34-monomer double disk assemblies were formed, further increasing the heterogeneity of the sample. These modifications and other molecular variables result in a mixture of 6540 possible unique assembly masses- a number that does not include contributions from stable isotopes and adducts, which increases the possibilities substantially. Statistical analysis based on the relative abundances of each monomer type was used to calculate expected theoretical mass centroids of 590 kDa for the unlabeled 34-mer and 680 kDa for the labeled 34-mer, resulting in an expected mean difference of 89.9 kDa between α-endorphin-labeled and unlabeled samples. The calculated centroid values match relatively well with the observed mass histograms shown in Figure 2.10c of the unmodified and modified samples, respectively (Figure 2.11), slightly underestimating the mass of the modified sample. Despite the expected increase in peak width for these samples due to heterogeneity, the two different populations are clearly resolvable, making it possible to monitor any monomer exchange.

A solution of the mixture of labeled and unlabeled complexes was monitored for changes in mass distribution over time. Directly after mixing the two samples, two distinct populations were seen for both double-layered disks (with centroids for the 34-mer at 590 and 690 kDa, respectively) and for 4-layer disk stacks composed of 68 monomers (Figure 2.10d). The two cpTMV populations in the mixed sample remained distinct after 3 h and 72 h (Figure 2.10e-f). This indicates that, on the timescales used for typical excited state energy transfer experiments, with samples being measured at room temperature for several hours (or kept at 4 °C or frozen for longer periods), there is no significant exchange of monomers between

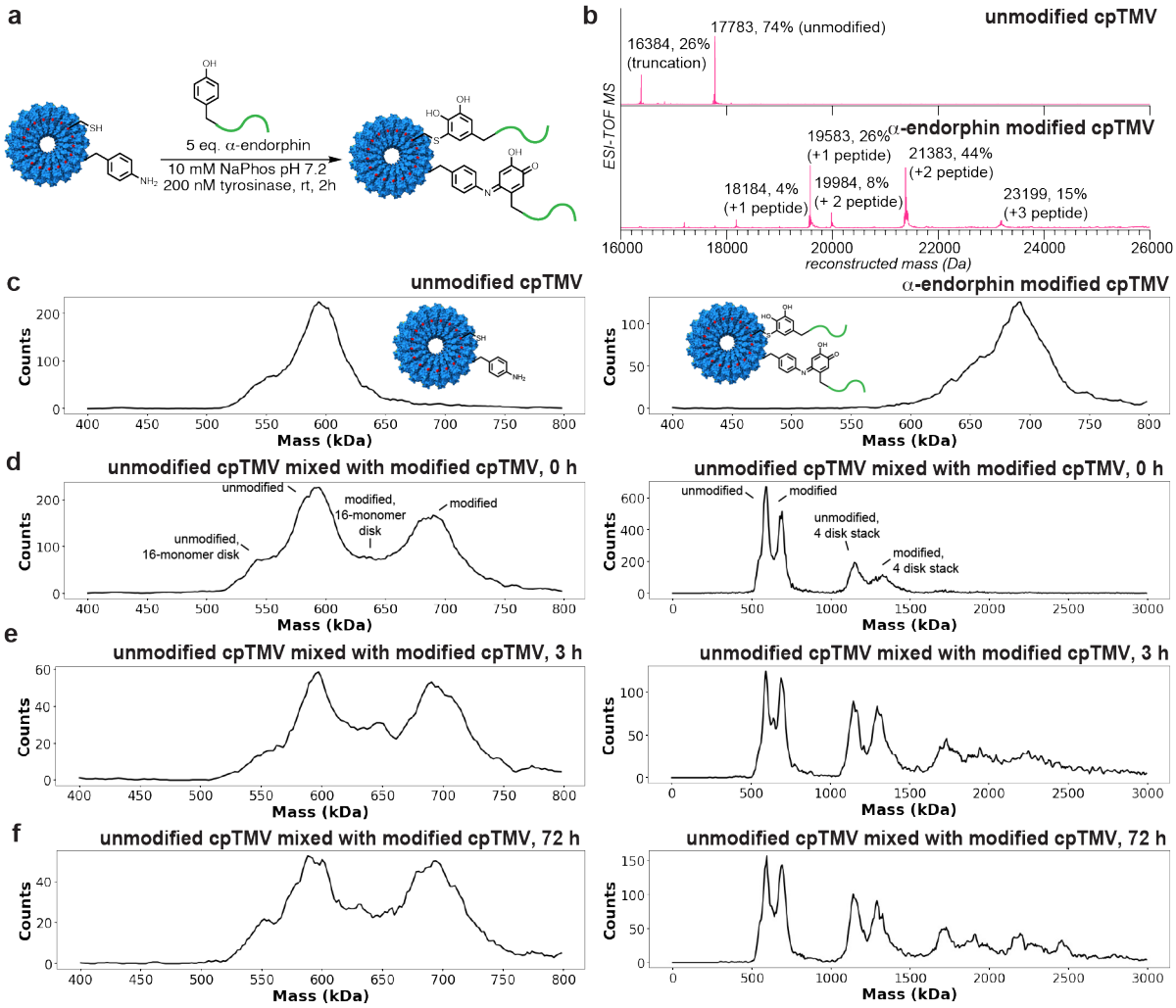


Figure 2.10: Monitoring protein assembly dynamics using charge detection mass spectrometry. (a) cpTMV-S23C-S65-pAF was modified with α-endorphin, a 1788 Da peptide, using the enzyme tyrosinase from *Agaricus bisporus*. (b) One to two copies of α-endorphin were attached per cpTMV-S23C-S65-pAF monomer as assessed by ESI-TOF mass spectrometry. (c) A comparison of the mass histograms of unmodified and α-endorphin-modified cpTMV-S23C-S65-pAF disks in 100 mM ammonium acetate solution shows a clear difference in size of approximately 90 kDa. A mixture of the unmodified and α-endorphin-modified disks in 100 mM ammonium acetate solution (d) immediately after mixing, (e) incubated at rt 3 h after mixing, and (f) incubated at rt 72 h after mixing displays two distinct populations that do not equilibrate over time. For (d), (e), and (f), both a window showing only the two-disk stack and a larger window showing higher stacking stoichiometries are displayed. A statistical analysis of the intact disk distributions based on the monomer % modification is shown in Figure 2.11.

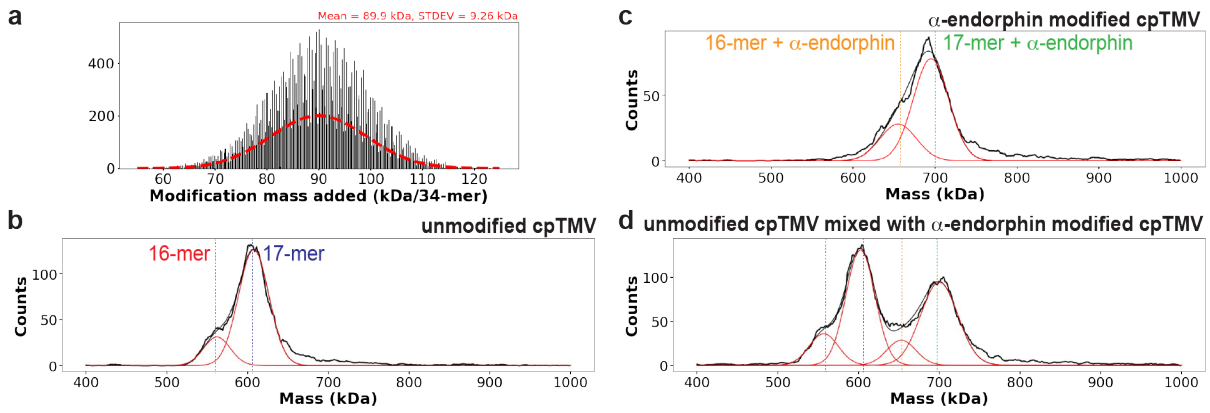


Figure 2.11: Predicted masses based on % modification and truncation levels of unlabeled and α-endorphin-labeled cpTMV-S23C-S65-pAF. (a) A synthetic mass histogram and overlaid smoothed curve demonstrating that the mean difference between modified and unmodified monomers should be 89.9 kDa. (b) An overlay of the unmodified cpTMV-S23C-S65-pAF CDMS histogram (black) and predicted distributions (red). (c) An overlay of α-endorphin-labeled cpTMV-S23C-S65-pAF CDMS histogram (black) and predicted distributions (red). (d) An overlay of mixed unlabeled and α-endorphin-labeled cpTMV-S23C-S65-pAF CDMS histogram (black) and predicted distributions (red).

disks. A secondary result of this experiment monitoring the assembly state of cpTMV over time indicates that cpTMV may have a propensity to stack into larger complexes the longer it is in solution, as evidenced by the appearance of higher MW species after 3 h and 72 h than after initial mixing. While it is possible that variation in the electrospray ionization process may account for some of the observed increase of higher MW species, the label-specific stacking preference is visible even for larger stacks, suggesting that the formation of these structures is favored under the conditions and timescales used in this study. The larger stacks that are formed in solution after 72 h also show a label-specific preference, indicating that their formation in solution is favored over structures consisting of both labeled and unlabeled subunits on this timescale. This intrinsic stability may also contribute to the observation of no exchange for the smaller assemblies even over several days.

2.3.3 Distinguishing between cpTMV conformational states of similar mass

One goal of the TMV-based light harvesting model is to measure energy transfer between laterally joined cpTMV disks (Figure 2.12a), imitating adjacent light harvesting complexes within the membranes of photosynthetic organisms. We sought to distinguish between this peripherally joined configuration and four disk stacks of cpTMV protein complexes of nearly identical molecular weight using CDMS. To create an assembly with peripherally connected cpTMV disks, we engineered a cysteine residue onto a surface-exposed position on the periphery of the cpTMV double disk (red residues in Figure 2.12a-b). When exposed to air, these thiols can form disulfide bonds with neighboring disk assemblies, as has been observed for a cysteine mutation in an analogous position on rTMV.[16] While these conjugated as-

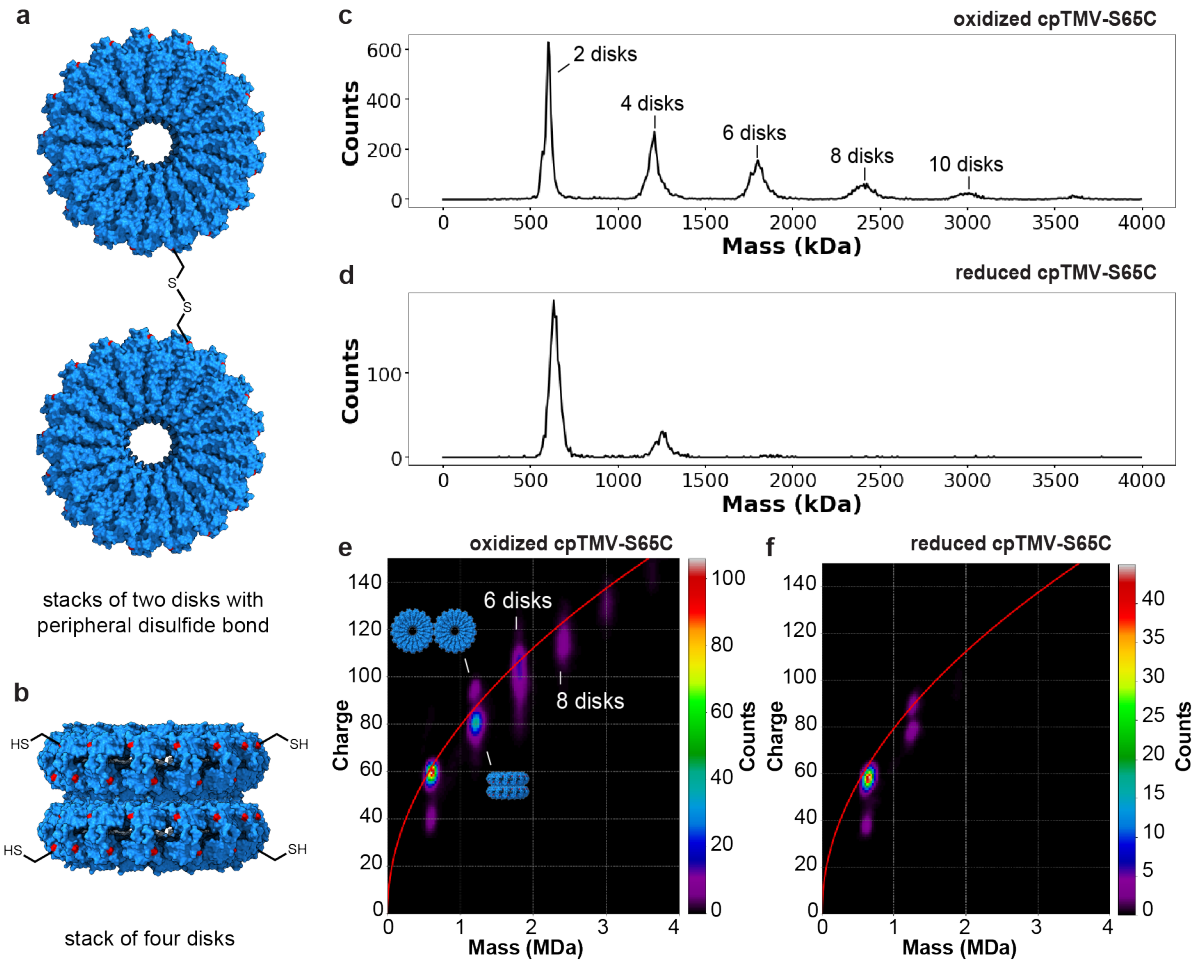


Figure 2.12: Distinguishing between assembly states of nearly identical mass using CDMS. (a) The cpTMV-S65C mutant can both form disulfide bonds at the disk's periphery and (b) assemble into stacks of four disks, resulting in two structures with a nearly equal mass at 1.2 MDa. The location of the S65C mutation is shown in red. (c) A mass histogram of cpTMV-S65C in 100 mM ammonium acetate solution oxidized in air displays assemblies with even-numbered disk stoichiometries varying from two to twelve disks per assembly. (d) After reduction with 1 mM tris(2-carboxyethyl)phosphine (TCEP) for 1 h, there is a decrease in the proportion of larger-sized assemblies of cpTMV-S65C. (e) A two-dimensional mass vs. charge histogram of air-oxidized cpTMV-S65C in 100 mM ammonium acetate solution displays two distinct populations at 1.2 MDa above and below the Rayleigh limit (red line). It also shows broad charge distributions for species with greater than four disks which indicate a mixture of different stacked and edge-on conformers. (f) After treatment with TCEP for 1 h, populations larger than four disks were no longer visible, and only a small population of the higher charged species at 1.2 MDa remained, indicating that this is the peripherally disulfide bonded species.

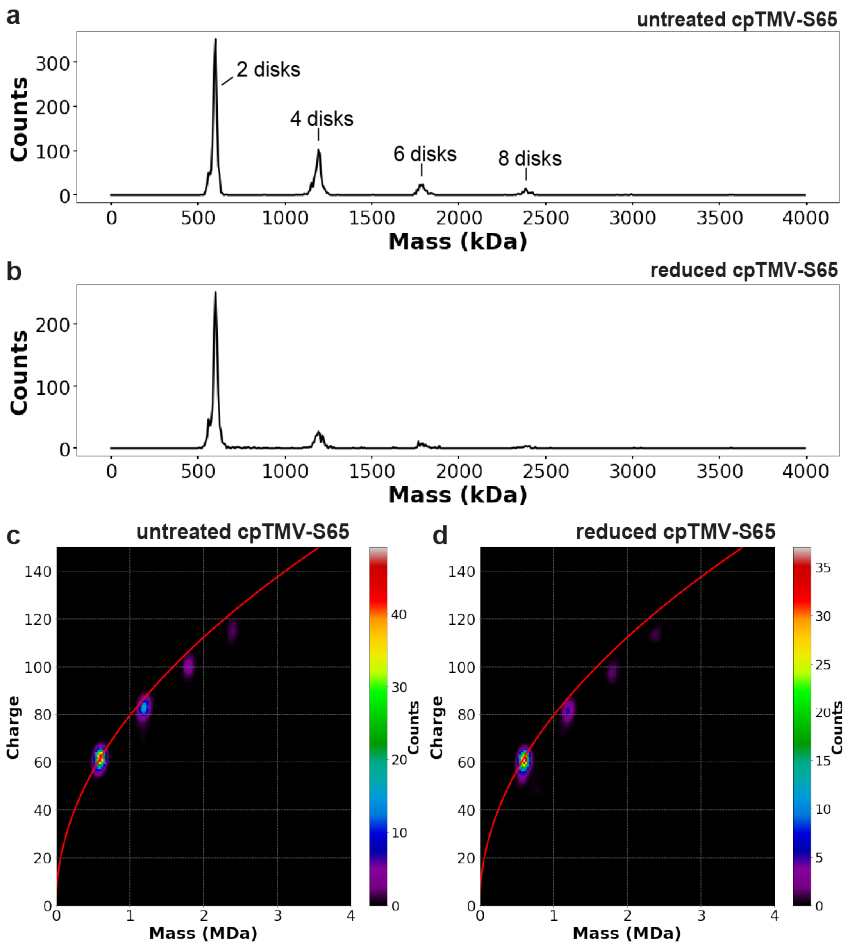


Figure 2.13: The effect of reduction of cpTMV without an engineered cysteine residue on assembly state. A comparison of cpTMV-S65 without a cysteine mutation at position 65 was made under identical conditions as those used for the oxidation and reduction of cpTMV-S65C shown in Figure 2.12. (a) A mass histogram of cpTMV-S65 shows even-numbered disk stacks, with a lower proportion of higher molecular weight stacks compared to cpTMV-S65C. (b) After reduction with 1 mM tris(2-carboxyethyl)phosphine (TCEP), most of the higher MW species remain. A mass versus charge diagram comparing cpTMV-S65 (c) before and (d) after reduction with TCEP shows minimal differences between the two conditions, with neither condition showing multiple species with the same mass but different charge observed for cpTMV-S65C.

semblies are depicted as parallel to one another in Figures 2.12a and 2.12e to distinguish them from the stacked assembly state, it is likely that there is some flexibility in the disulfide linkage joining the two protein assemblies such that the two bonded complexes are not necessarily parallel to one another. When two double-disk cpTMV assemblies are linked via these peripheral disulfide linkages, they would be expected to form an assembly of 1.2 MDa with a difference of only a few Da between disulfide-bonded complexes and stacks of four disks as shown in Figure 2.12b. This small difference relative to the large complex mass makes the two species impossible to differentiate on the basis of mass alone. Because CDMS measures both the m/z and charge of each individual analyte ion independently, species with the

same mass can be differentiated based on their charge state. While the two conformations of four disks are nearly identical in molecular weight, their charge states are expected to differ significantly because the apposed assembly of two disulfide-bonded disks (with approximate dimensions of 5 nm x 18 nm x 36 nm) has a significantly larger surface area available for charging than vertical stacks of four disks (with approximate dimensions of 10 nm x 18 nm x 18 nm). A CDMS histogram for the cpTMV-S65C mutant exhibits a greater proportion of larger disk stoichiometries per assembly than other cpTMV mutants, indicating that the mutation either promotes disk stacking and/or peripheral disulfide bonding (Figure 2.12c). To support that the species with higher charge in the cpTMV-S65C sample was the expected disulfide-bonded complex, we treated the sample with a reducing agent, TCEP, and monitored the charge vs. mass ratio of the sample over time. After 1.5 h, nearly all of the assemblies containing greater than four disks had disappeared (Figure 2.12d). Additionally, a sample identical but without the peripheral cysteine mutation, cpTMV-S65, did not have as high a proportion of higher mass species as cpTMV-S65C and did not have an appreciable depletion of higher mass species upon treatment with TCEP (Figure 2.13a-b).

While the two populations of peripherally bonded versus stacked complexes cannot be distinguished by mass alone, a two-dimensional mass versus charge histogram of a cpTMV-S65C sample in Figure 2.12e shows two distinct populations at 1.2 MDa that are differentiated by ∼13 charges. The stacked disk structure likely corresponds to the population with charge states below the Rayleigh charging limit (red line in Figure 2.12e), and the peripherally bonded disk structure to the population with charge states above the Rayleigh limit, a characteristic that typically indicates an assembly that deviates from a globular (spherical) structure. The higher-charged species at 1.2 MDa was also depleted after treatment with TCEP, though some of this species was still visible (Figure 2.12f). In contrast, cpTMV-S65 did not exhibit multiple charge states at 1.2 MDa and higher masses, either before or after treatment with TCEP (Figure 2.13c-d). This provides evidence for the identity of the higher charged species as two double-layered disks bonded via disulfides at their periphery and shows how multiple assembly configurations can be distinguished using CDMS. There is also a bimodal charge-state distribution for the 602 kDa complex. The distribution centered at ∼40 charges is too far below the Rayleigh limit for a globular structure formed directly from solution. This indicates that this population is likely formed by gas-phase dissociation of a higher-order structure of the peripherally bonded complexes, consistent with the absence of this population in the cpTMV-S65 data (Figure 2.13c-d). The observation of some dissociation products despite very soft instrument conditions indicates that the interactions between the two double disks are weak, consistent with a limited contact patch between the assemblies. The dissociation occurs with symmetric charge partitioning between products, consistent with prior results where symmetric dissociation occurs for weakly bound complexes under low energy excitation conditions.[51]

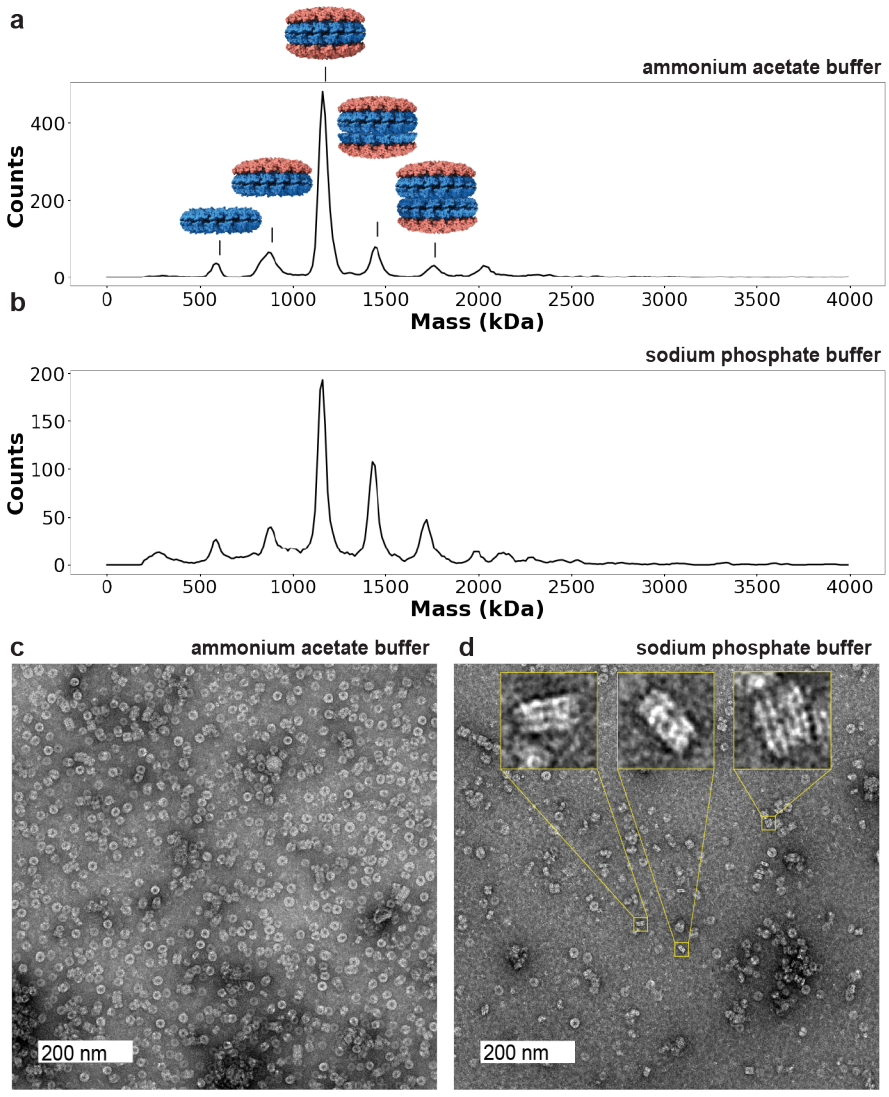


Figure 2.14: Odd-numbered stacks of disks in recombinant tobacco mosaic virus capsids (rTMV). Mass histograms of rTMV capsid ions formed by electrospray from (a) 18.2 mM ammonium acetate pH 7 and (b) 10 mM sodium phosphate pH 7.2 solutions display different ratios of number of disks per stack, with a single disk having a molecular weight of 300 kDa. (c) A representative TEM image of rTMV in 18.2 mM ammonium acetate pH 7 shows many disks on their faces with a few larger stacks. (d) A representative TEM image of rTMV in 10 mM sodium phosphate pH 7.2 shows several examples of odd-numbered stacks of rTMV disks. Inlays show stacks with odd numbers of layers (3 and 5 disks per stack). A collection of images can be found in Figure 2.15.

2.3.4 TMV stacking stoichiometries in different buffer solutions

Due to the sensitivity of TMV's assembly state distribution to buffer conditions, we additionally sought to understand whether the assembly state distribution of TMV particles was affected by buffer identity. The 10 mM NaPhos, pH 7.2 buffer is typically used for the chemical modification of proteins and for the purification and storage of TMV particles due to

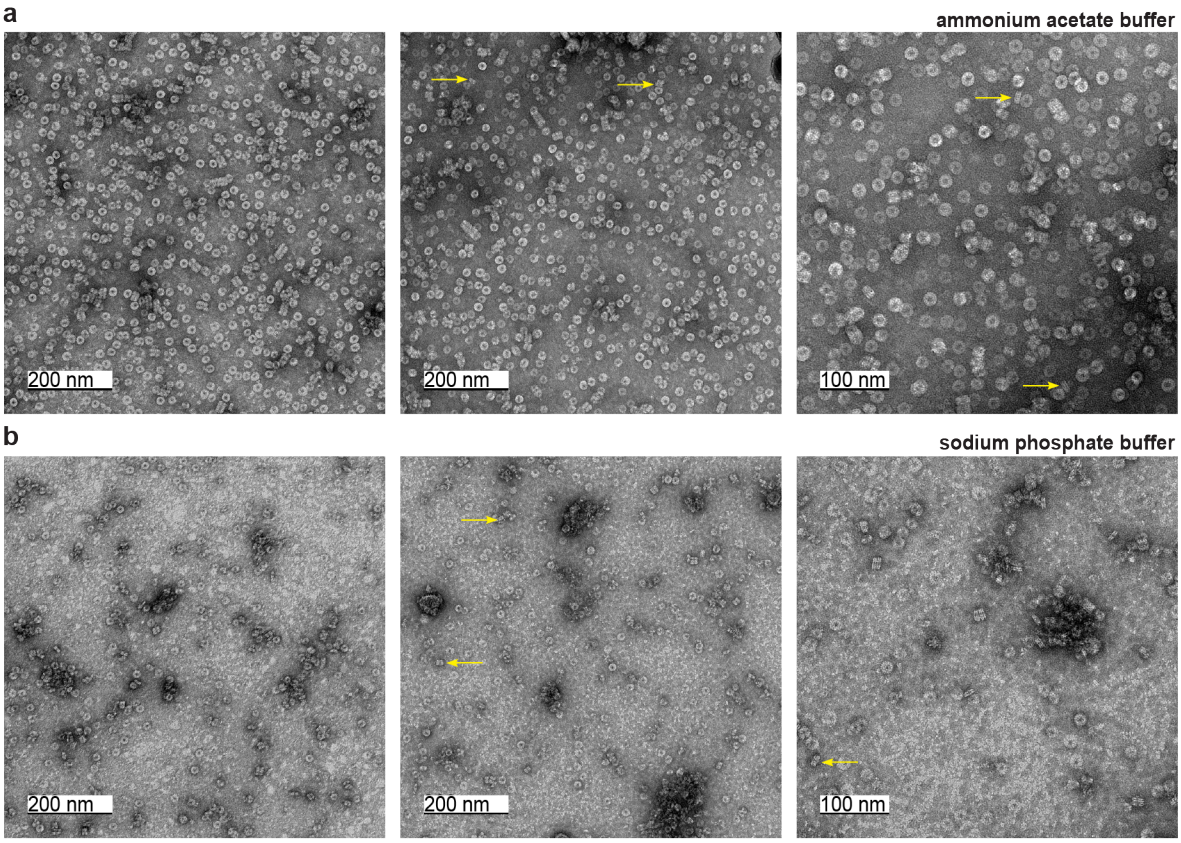


Figure 2.15: Additional TEM images of rTMV. (a) TEM images of rTMV in 100 mM ammonium acetate show face down views of rTMV and some examples of short stacks of disks. (b) TEM images of rTMV in 10 mM sodium phosphate buffer pH 7.2 show both face and side views of short stacks of disks. Several examples of odd-numbered stacks of disks are identified with yellow arrows.

its buffering capacity and biological compatibility.[8,52] Recombinant TMV particles (rTMV) consisting of monomers with a sequence identical to wtTMV, but lacking N-terminal acylation, were expressed and assembled in *E. coli*. rTMV particles in 10 mM sodium phosphate (NaPhos) were then compared to rTMV exchanged into a volatile buffer of the equivalent ionic strength preferred for mass spectrometry, 18.2 mM ammonium acetate (AA). The versatility of CDMS with respect to sample heterogeneity combined with the use of submicron ESI emitters that reduce the extent of non-volatile adducts[38] made it possible to acquire CDMS mass histograms from each of these solutions (Figure 2.14a-b). In each of these solutions, rTMV forms even-numbered stacks as expected based on previously determined structures. However, significant populations of stacks of 3 or 5 disks at ∼900 and ∼1500 kDa were also present, a stacking stoichiometry that has not been observed previously for rTMV. Differences in relative ion count between numbers of disks per stack were observed between the protein solutions in AA and NaPhos, with the solutions in NaPhos appearing to have a greater proportion of higher-order stacks, particularly for the 5-disk stack.

These differences illustrate the utility and importance of CDMS and submicron emitters in analyses from non-volatile buffers and indicate that buffer identity can influence TMV stacking stoichiometries. These data were collected on the same day under identical instrumental conditions to preclude differences in mass distribution due to mass and instrumental biases; however, some variation between the two spectra may be due to environmental factors. Salt adduction and baseline noise are also more pronounced in NaPhos buffer than AA solution, indicating that AA can be preferable for resolving finer differences in mass distributions, such as those shown in Figures 2.2 and 2.10. Under denaturing conditions, a mass spectrum of the rTMV monomer showed that the mass is homogeneous (Figure 2.3), indicating that the different populations observed using CDMS were due to differing assembly states of identical monomers rather than monomer heterogeneity.

In order to determine whether the odd-numbered stacking behavior occurred as a result of non-specific aggregation during the electrospray ionization (ESI) process or in solution before ESI, we evaluated rTMV in each buffer solution using negative staining and transmission electron microscopy (TEM). TEM images of rTMV were consistent with CDMS observations, with even- and odd-numbered stacks of disks of varying stoichiometry observed in side views of rTMV assemblies (Figure 2.14c-d; a collection of images can be found in Figure 2.15). Multiple clear examples of odd-numbered disk stacks were observed in NaPhos buffer (Figure 2.14d, inlays), illustrating that these structures were present in solution and did not form as a result of non-specific aggregation in the ESI process. A crystal structure of the four-layered aggregate of wtTMV indicates that the four-disk stacks are C_2-symmetric about their center,[22] and solved stacked disk structures of additional TMV mutants also exhibit C_2 symmetry.[13,53] While neither CDMS nor TEM was able to resolve the facial directionality of the individual disks comprising the odd-numbered stacks, they cannot be in fully C_2-symmetric arrangements.

2.3.5 Removing the flexible regions of cpTMV

The protein environment has been shown to have a significant impact on the excitonic properties of chromophores attached to the surface of cpTMV.[8] The crystal structure of cpTMV is not resolved for the N-terminal residues 1–12 or C-terminal residues 154–161, indicating that these residues are flexible or have non-uniform conformations. These residues face the pore of cpTMV, and due to the length of these unresolved sequences, these regions have the potential to interact with chromophores attached to cpTMV in unpredictable ways. In order to have a more minimal and predictable system, mutants of cpTMV with these flexible regions truncated were expressed and purified to examine whether cpTMV could maintain its double disk assembly state with truncations to the N- and C-termini. This resulted in a series of truncation mutants that expressed and formed monomers of the correct mass (2.16a-b), although two mutants with 9 and 10 residues removed from the N-terminus contained an extra N-terminal methionine due to a lack of methionine cleavage during expression. A mutant with 12 residues removed from the N-terminus expressed at low levels and was puri-

fied, showing a band of the expected mass on an SDS-PAGE gel (Figure 2.16c). However, a mass spectrum of this mutant was not successfully attained, likely due to a lack of material.

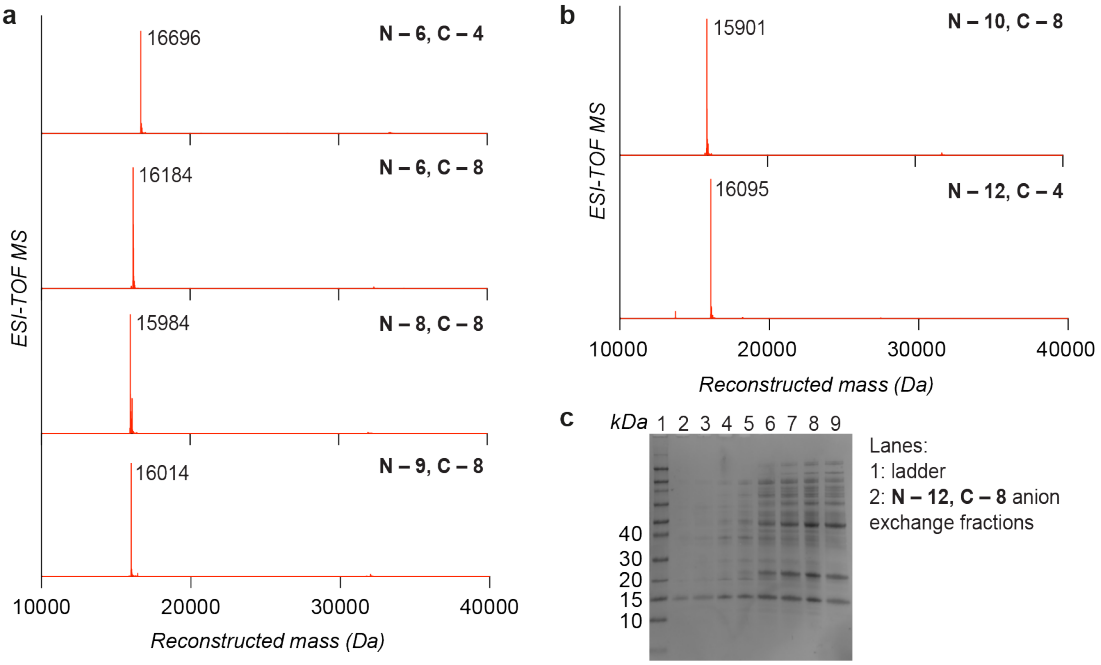


Figure 2.16: Masses of cpTMV with the N- and C-termini truncated. (a) Masses of several truncation mutants, with 6 residues removed from the N-terminus and 4 residues removed from the C-terminus (expected mass: 16696 Da), 6 residues removed from the N-terminus and 8 residues removed from the C-terminus (expected mass: 16184 Da), 8 residues removed from the N-terminus and 8 residues removed from the C-terminus (expected mass: 15984 Da), and 9 residues removed from the N-terminus and 8 residues removed from the C-terminus (expected mass: 16013 Da without N-terminal methionine cleavage), are shown. (b) Masses of truncation mutants with 10 residues removed from the N-terminus and 8 residues removed from the C-terminus (expected mass: 15901 Da without N-terminal methionine clevaage) and 12 residues removed from the N-terminus and 4 residues removed from the C-terminus (expected mass: 16095 Da) are shown. (c) An SDS-PAGE gel shows anion exchange fractions of the full truncation mutant with 12 residues removed from the N-terminus and 8 residues removed from the C-terminus, with a mass close to 15 kDa (expected mass: 15567 Da).

Truncation of the flexible regions of the N- and C-termini up to 6 amino acids from the N-terminus and 8 amino acids from the C-terminus resulted in little disruption to the cpTMV assembly state as shown by TEM (Figure 2.17a-b). These mutants displayed a large proportion of disks on their faces, which likely indicates a double disk structure due to the dimensions of the double disk, with an 18 nm face and 6 nm depth. Some of the examples of disks on their face may also be larger stacks, but this is increasingly unlikely as the length of the stack increases. Some short or interrupted disk stacks are also visible on their sides. As the truncation from the N-terminus continues to increase to 9 amino acids, the stacking behavior becomes slightly more pronounced, with disk stacks >100 nm in length observed, although a large proportion of disks on their face are also visible (Figure 2.17c). Further increasing to 10 amino acids from the N-terminus also results in a mixture of double disks

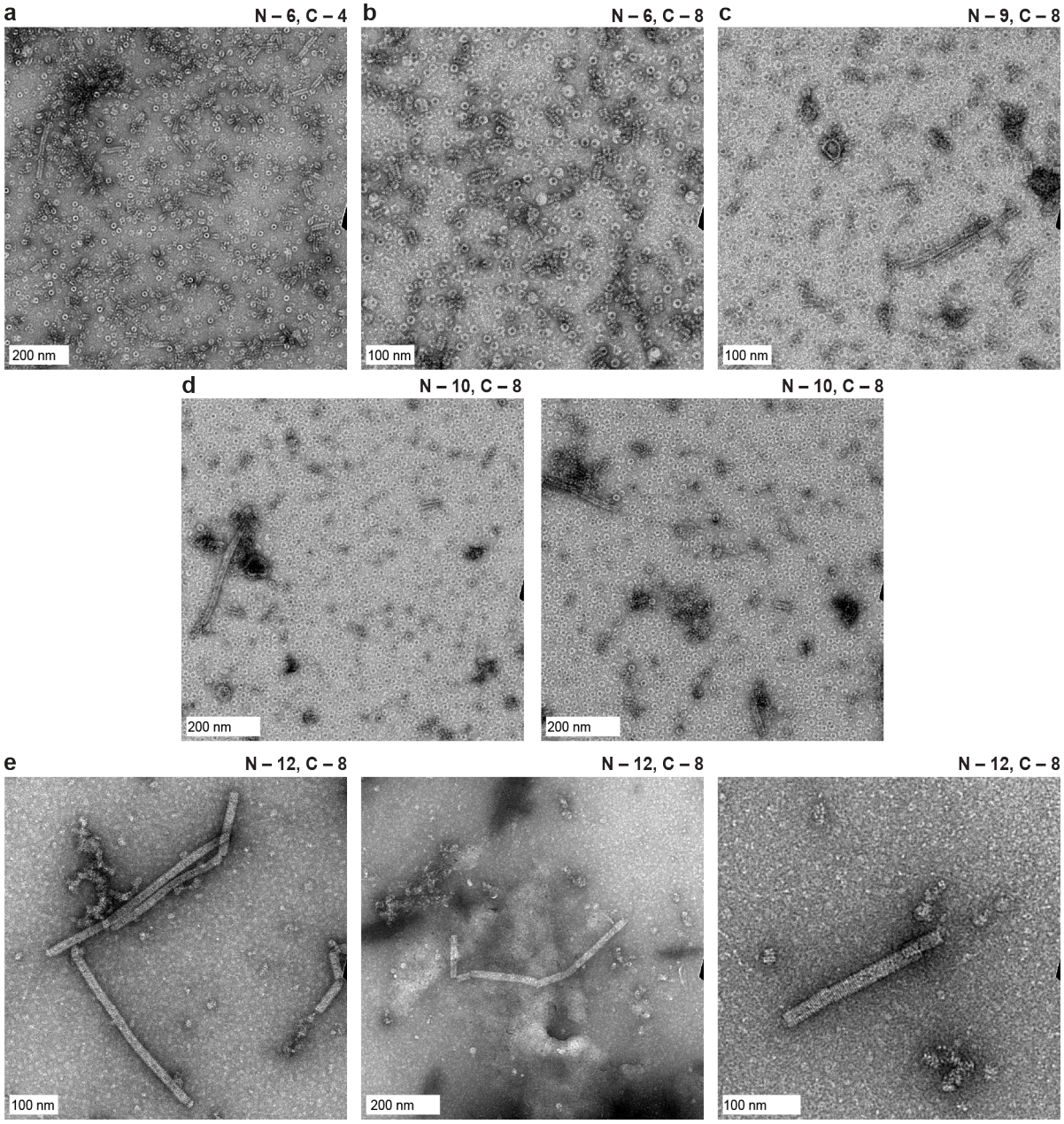


Figure 2.17: TEM images of cpTMV with the N- and C-termini truncated. Mutants with 6 residues removed from the N-terminus and 4 residues removed from the C-terminus (a) and 6 residues removed from the N-terminus and 8 residues removed from the C-terminus (b) still form double-disk structures with some short disk stacks. (c) A mutant with 9 residues removed from the N-terminus and 8 residues removed from the C-terminus shows a mixture of double disks and some disk stacks >100 nm in length. (d) A mutant with 10 residues removed from the N-terminus and 8 residues removed from the C-terminus shows a mixture of double disks and some disk stacks >200 nm in length. (e) A mutant with 12 residues removed from the N-terminus and 8 residues removed from the C-terminus shows few examples of double disks, consisting primarily of disk stacks 100s of nm in length.

and long disk stacks 100s of nm in length (Figure 2.17d). Truncation of the full flexible region, 8 amino acids from the C-terminus and 12 amino acids from the N-terminus, results in an assembly state with few to no double disks or images of cpTMV on its face, instead forming disk stacks up to 400 nm in length (Figure 2.17e). This is notably longer than the 300 nm rods formed by TMV virus particles. The images of the fully truncated mutant appear to be stacks of disks rather than helical rods, although this is difficult to determine with certainty based on the images (Figure 2.17e).

The formation of increasingly longer disk stacks at greater N- and C-terminal truncations indicate that the less structured N- and C-termini of cpTMV within the cpTMV pore likely interfere with the formation of larger stacks through steric interference. This suggests a mechanism for the greater stability of cpTMV double disks under a wider range of buffer and ionic strength conditions than wtTMV or rTMV. While residues 1–12 and 154–161 are flexible termini in cpTMV, the corresponding residues 92–110 of wtTMV and rTMV consist of a well-defined loop region. This loop region may interfere less with disk stacking than the flexible corresponding residues in cpTMV. In future experiments, the full truncation mutants with 12 amino acids removed from the N-terminus and 8 amino acids removed from the C-terminus may be a useful construct for studying long-range energy transfer over 100s of nm. In contrast, truncations greater than 6 amino acids from the N-terminus should be avoided when the double disk structure of cpTMV is desired.

2.4 Conclusion and Outlook

This work has identified previously unrecognized assembly states of particles derived from TMV, including stacks of disks containing 16 monomers and odd-numbered stacks of disks that necessarily cannot exhibit C_2 symmetry. It has also further emphasized the importance of developing techniques for studying large protein complexes in their native buffers, as buffer exchange can influence the ratio of assembly states present in heterogeneous protein complex solutions. In addition to providing precise mass measurements and proportional estimates of multiple assemblies present in heterogeneous solutions, measuring charge and m/z in tandem allowed for differentiation between species of near identical mass but with different configurations. Molecular dynamics simulations of intact cpTMV complexes suggested that the surface charge of the cpTMV assembly affects whether it assemblies into 16- or 17-monomer disks. CDMS also allowed us to probe the possibility of an exchange of monomers between assembly structures in solution and was used to determine that dissociation of assembled TMV complexes into monomers is minimal under the conditions and time periods studied. Finally, truncation of the N- and C-terminal regions of cpTMV was demonstrated to affect their stacking stoichiometry, with greater truncations leading to longer disk stacks. This suggests that the flexible N- and C-termini within the pore of cpTMV prevent disk stacking, likely due to steric hindrance at the disk interfaces.

The precise morphology, stoichiometry, and arrangement of monomers within quaternary protein structures have broad implications for their function. In the context of TMV,

its morphology has been demonstrated to affect its drug delivery capability[18] and has implications for the development of biosensing or light harvesting materials utilizing TMV monomers. Measuring multiple assembled states of TMV present in solution can be accomplished quickly and with minimal protein preparation using CDMS, allowing for the rapid testing of multiple conditions and even time-course experiments. These observations will inform the continued development of light-harvesting models based on mutants of TMV. In addition, the techniques described herein using CDMS are expected to have broad utility for the detailed characterization of heterogeneous populations of large protein complexes.

2.5 Materials and Methods

General Methods and Instrumentation. Unless noted otherwise, all reagents were obtained from commercial sources and used without further purification. Tyrosinase isolated from *Agaricus bisporus* (abTyr) was purchased from Sigma-Aldrich (St. Louis, MO). *N*-Ac-α-endorphin was purchased from GenScript (Piscataway, NJ). Protein expression media and all buffers were prepared using doubly distilled water obtained from a NANOpure purification system (Barnstead, USA).

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

Gel Analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in a Mini cell tank apparatus (Life Technologies, Carlsbad, CA), using NuPAGE^{TM} Novex^{TM} 4–12% Bis-Tris Protein Gels (Life Technologies). The sample and electrode buffers were prepared according to the suggestions of the manufacturer. All protein electrophoresis samples were heated for 5–10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure the reduction of disulfide bonds. Gels were run for 30 min at 200 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for the assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad, Hercules, CA). Gel imaging was performed on a Gel Doc (Bio-Rad, Hercules, CA).

Denaturing Mass Spectrometry (MS). Proteins and their bioconjugates were analyzed using an Agilent 1260 series liquid chromatograph that was connected in-line with an Agilent 6530C Quadrupole Time-of-Flight (QTOF) LC/MS system (Santa Clara, CA). Protein samples were run with a Proswift RP-4H column (Dionex, USA). Protein mass reconstruction was performed on the charge ladder with Agilent Mass Hunter software, Qualitative Analysis Version B.10.0, Build 10.0, Agilent Technologies Inc.© 2020 (Agilent, USA).

High Performance Liquid Chromatography (HPLC). HPLC was performed on Agilent 1200 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). Size exclusion chromatography (SEC) was performed using a Polysep-GFC-P-5000 column (4.6 x 250 mm) (Phenomenex, USA) at 1.0 mL/min using a mobile phase of 10 mM sodium phosphate buffer, pH 7.2.

Dynamic Light Scattering (DLS). DLS was performed on a Zetasizer Nano Series (Malvern Instruments, UK). Measurements were taken in triplicate at protein concentrations of 0.2–1.0 mg/mL in 10 mM sodium phosphate buffer, pH 7.2, at 25 °C.

Synthesis of rTMV and cpTMV mutants. Synthesis of rTMV and cpTMV mutants was performed as described previously.[8,19,52,54,55] For rTMV and cpTMV mutants not containing non-canonical amino acids, BL21-Codonplus (DE3)-RIL cells were transformed with the appropriate pET vector, and colonies were selected for inoculation of Terrific Broth cultures. Cultures were induced with 30 μM isopropyl-β-D-thiogalactopyranoside (IPTG) at an OD_{600} of 0.6–0.8 and allowed to grow 14–18 h at 20 °C before harvesting cell pellets and storing them at -20 °C. For cpTMV mutants containing non-canonical amino acids, DH10B cells were co-transformed with the appropriate pBAD-cpTMV and pDule-pAF or pDule-3NY vectors, and colonies were selected for inoculation in autoinduction media. At an OD_{600} of 0.6–0.8, 1 mM *p*-amino-L-phenylalanine or 3-nitrotyrosine (pAF or 3NY) was added and the culture was allowed to grow 18 h at 37 °C. For both rTMV and cpTMV, cell pellets were collected at 8000 rpm for 30 min and stored at -20 °C until purification. For purification, cell pellets were resuspended in 10 mL lysis buffer, 20 mM triethanolamine (TEA) pH 7.2 and lysed by sonication with a 2 s on, 4 s off cycle for a total of 10 min using a standard disruptor horn at 65% amplitude. The resulting lysate was cleared at 14,000 rpm for 30 min. The supernatant was treated with 30–40% volume of saturated ammonium sulfate and allowed to rotate for 10 min at 4 °C. The precipitated protein was collected at 11,000 rpm for 30 min and resuspended in 10 mL lysis buffer, then dialyzed in 1 L lysis buffer overnight with at least one buffer exchange. The resulting protein solution was treated with 5 μL benzonase and 4 mg $MgCl_2$ at room temperature for 30 min and purified using a DEAE column with a 0–180 mM NaCl gradient elution in 20 mM TEA buffer, pH 7.2. Further purification was performed using a Sephacryl S-500 column in 10 mM NaPhos pH 7.2 elution buffer. Purity and general assembly state were confirmed by SDS-PAGE, ESI-TOF LC-MS, and HPLC-SEC.

Preparation of rTMV and cpTMV samples for CDMS analysis. Preparation of rTMV and cpTMV samples was conducted as described previously.[54] TMV assemblies were thawed and then exchanged into the appropriate buffer for subsequent analysis via repeated buffer exchange and concentration using 100 kDa MWCO centrifugal filters. Samples were adjusted to a protein concentration of 0.5 mg/mL protein and filtered through a 0.22 μm filter prior to analysis.

TMV mixing experiments. For experiments mixing populations of α-endorphin-labeled and unlabeled cpTMV, samples were mixed to a final concentration of 0.5 mg/mL protein (0.25 mg/mL of each individual sample) in 100 mM ammonium acetate. The sample was

then incubated at room temperature undisturbed indefinitely. CDMS measurements were recorded directly after mixing and at subsequent time intervals of 3 h and 72 h.

Procedure for labeling of cpTMV-S23C-S65-pAF with α-endorphin. Enzyme-catalyzed modification of cpTMV-S23C-S65-pAF with α-endorphin was conducted according to literature procedure with slight modifications.[19,50] The protein was first exchanged into the reaction buffer (10 mM sodium phosphate, pH 7.2). To 100 μL of α-endorphin (250 μM) was added 400 nM abTyr, followed by 50 μM protein. The reaction mixture was briefly agitated and then incubated in 1.5 mL Eppendorf tubes at room temperature. After 2 h, the reaction was quenched with 1 mM tropolone and incubated for 5 min. Excess peptide and abTyr were removed via repeated centrifugation through 100 kDa molecular weight cutoff filters. The protein conjugates were analyzed with MS to measure modification levels of monomers.

Reduction of disulfide bonds of cpTMV-S65C. To a 0.5 mg/mL (22 μM) cpTMV-S65C solution in 100 mM ammonium acetate was added 1 mM tris(2-carboxyethyl)phosphine (TCEP). The reaction was allowed to incubate at least 1 h at room temperature, and no subsequent buffer exchange was performed before CDMS analysis.

Charge detection mass spectrometry. Individual ion mass measurements were obtained using a home-built charge detection mass spectrometer that is described in detail elsewhere.[39,40] Ions are formed by nanoelectrospray ionization from borosilicate capillaries (1.0 mm outer diameter, 0.78 mm inner diameter) with tips pulled to an inner diameter of 0.5–2.0 μm using a Flaming/Brown P-87 micropipette puller (Sutter Instruments, Novato, CA) with smaller diameter tips (0.5–0.6 μm) used to reduce non-volatile salt adducts from solutions other than ammonium acetate.[38,56] Ions are introduced into the instrument through a modified Waters Z-Spray source (Waters Corporation, Milford, MA) and enter a region containing two rf-only quadrupole ion guides where they are accumulated for up to 1 s. Ions are then pulsed into an electrostatic ion trap containing a charge detection electrode at a pressure of 3×10^{-9} Torr where they are analyzed for 1 s. Typically, several ions are simultaneously trapped and analyzed to reduce the time required to acquire a statistically significant number of individual ions. All experiments in this work required <1 h of data acquisition, with most experiments requiring ∼15–20 min to accumulate ion counts. Trapped ions repeatedly induce charge on the detector and the individual oscillation frequencies and amplitudes are used to determine the m/z, charge, and mass of each individual ion. Ion signals are amplified by a CoolFET charge-sensitive preamplifier (Amptek, Bedford, MA) and pass through a custom-built filter stage that removes noise and further amplifies signals. Signals are analyzed using short-time Fourier transform (STFT) methods,[57] with 50 ms segment lengths stepped across the transients in 5 ms increments. Only ions trapped for the entire 1 s trapping period (>75% of all ions) were included in the analysis. The resolution expected for a 1 s data acquisition depends on the inherent sample heterogeneity and instrument performance; a detailed analysis of these factors for the same instrument is given elsewhere.[54]

Procedure for generating cpTMV homology model. For visualization purposes, a homology model of cpTMV containing all amino acid residues was generated based on the crystal structure of cpTMV (PDB code 3KML)[13] and the crystal structure of wtTMV (PDB code 1EI7),[22] both obtained from the Protein Data Bank. Initial structural preparation was conducted using Pymol, Version 2.4.2.[58] First, a monomer of wtTMV was superimposed on a monomer of cpTMV, and the residues resolved in the crystal structure of wtTMV but not cpTMV (residues 92–110 of wtTMV and 1–12, 154–161 of cpTMV) were fused to the unresolved N- and C-termini of cpTMV. The bond between residues 99 and 100 of wtTMV was then cleaved *in silico* and an N-terminal glycine was added to produce the N- and C-termini of cpTMV. Following this, cpTMV was symmetry expanded to create the double disk assembled structure consisting of two C_2-symmetric disks, each containing 17 monomers. The Schrödinger Maestro package (version 2022-1)[59] was used for subsequent structural preparation and molecular dynamics-based side chain minimizations. The Desmond system builder was used to solvate the double disk structure in an orthorhombic box with periodic boundaries at 10 Å from the protein of water molecules described using the TIP3P model,[60,61] and neutralized with sodium ions, with the addition of 150 mM NaCl in an OPLS4 force field.[62] A molecular dynamics simulation/minimization was performed with an NPT ensemble of T = 300 K, P = 1 bar, a Coulombic cutoff radius of 9.0 Å, and a 100 ps simulation time with sampling time of 5 ps. This short simulation time was chosen to relax sidechain and solvent interactions for the protein representations shown herein, but not alter the overall quaternary structure of the assembly. The pressure control was applied using the Martyna-Tobias-Klein barostat method[63] with a 2.0 ps relaxation time, and the temperature control was applied using the Nosé-Hoover thermostat method[64] with a 1.0 ps relaxation time. The trajectory was analyzed, and the lowest energy frame was used as the homology model.

Molecular dynamics simulations of cpTMV dimers. The homology model described above was used as an input structure for side-by-side (on the same disk, rather than monomers on opposing disks) dimers of cpTMV. Point mutations were introduced at position 65 to pAF or 3NY for comparison to the native serine prior to any structural processing for a total of three input structures: adjacent cpTMV dimers with serine, pAF, or 3NY at position 65. The Schrödinger Maestro package (version 2022-1)[59] was used for subsequent structural preparation and molecular dynamics simulations. The Desmond system builder was used to solvate the double disk structure in an orthorhombic box with periodic boundaries at 10 Å from the protein of water molecules described using the TIP4PEW model,[65] neutralized with sodium ions, with the addition of 18 mM NaCl in an OPLS4 force field.[62] A molecular dynamics simulation/minimization was performed with an NPT ensemble of T = 300 K, P = 1.01325 bar, a Coulombic cutoff radius of 9.0 Å, and a 100 ns simulation time with sampling time of 100 ps. The pressure control was applied using the Martyna-Tobias-Klein barostat method[63] with a 2.0 ps relaxation time, and the temperature control was applied using the Nosé-Hoover thermostat method[64] with a 1.0 ps relaxation time. The MD trajectories were analyzed for the protein root-mean-square deviation (RMSD, Å), root-mean-square fluctu-

ation (RMSF, Å) of each residue, and the distance of the center of mass between adjacent monomers (DOCM, Å).

Molecular dynamics simulations of intact cpTMV complexes. The homology model described above was used as an input structure for simulations of 17-monomer disks. For 16-monomer disks, a top-bottom monomer pair of cpTMV monomers from the homology model were symmetry expanded to form a 16-monomer disk structure. Point mutations were introduced at position 65 to pAF, or 3NY for comparison to the native serine prior to any structural processing for a total of six input structures: 32- and 34-monomer double disks with serine, pAF, or 3NY at position 65. The Schrödinger Maestro package (version 2022-1)[59] was used for subsequent structural preparation and molecular dynamics simulations. The Desmond system builder was used to solvate the double disk structure in an orthorhombic box with periodic boundaries at 10 Å from the protein of water molecules described using the TIP4PEW model,[65] neutralized with sodium ions, with the addition of 18 mM NaCl in an OPLS4 force field.[62] A molecular dynamics simulation/minimization was performed with an NPT ensemble of T = 300 K, P = 1.01325 bar, a Coulombic cutoff radius of 9.0 Å, and a 100 ns simulation time with sampling time of 100 ps. The pressure control was applied using the Martyna-Tobias-Klein barostat method[63] with a 2.0 ps relaxation time, and the temperature control was applied using the Nosé-Hoover thermostat method[64] with a 1.0 ps relaxation time. The MD trajectories were analyzed for the protein root-mean-square deviation (RMSD, Å), root-mean-square fluctuation (RMSF, Å) of each residue, and the distance of the center of mass between two adjacent monomers chosen at random from the assembly (DOCM, Å). The final frames of each trajectory were also compared via superimposition using the align action in Pymol, Version 2.4.2.[58]

Statistical analysis. For the molecular dynamics trajecotry analyses, the error in the difference between RMSD and DOCM were evaluated by taking the standard error of the difference between means.[66,67] Statistical analyses were performed in Microsoft Excel for Mac, Version 16.62.

2.6 Acknowledgements

Thank you to Dr. Conner Harper for discussions about experimental design, performing CDMS and analyzing CDMS data, and input into structuring and presenting this work. Thank you to Prof. Evan Williams for analyzing CDMS data, and input into structuring and presenting this work. Thank you to Dr. Kathleen Durkin and the Molecular Graphics and Computation Facility for help with protein modeling and advice on conducting molecular dynamics simulations. Thank you to Dr. Jing Dai for providing plasmids for the cpTMV truncation mutants. Thank you to the the staff at the University of California, Berkeley, Electron Microscope Laboratory for their advice and assistance in electron microscopy sample preparation and data collection.

2.7 References

(1) Smith, M. L.; Lindbo, J. A.; Dillard-Telm, S.; Brosio, P. M.; Lasnik, A. B.; McCormick, A. A.; Nguyen, L. V.; Palmer, K. E. *Virology* **2006**, *348*, 475–488.

(2) Venkataraman, S.; Hefferon, K. *Viruses* **2021**, *13*, 1697.

(3) Chung, Y. H.; Cai, H.; Steinmetz, N. F. *Advanced Drug Delivery Reviews* **2020**, *156*, 214–235.

(4) Koch, C.; Wabbel, K.; Eber, F. J.; Krolla-Sidenstein, P.; Azucena, C.; Gliemann, H.; Eiben, S.; Geiger, F.; Wege, C. *Frontiers in Plant Science* **2015**, *6*.

(5) Fan, X. Z.; Naves, L.; Siwak, N. P.; Brown, A.; Culver, J.; Ghodssi, R. *Nanotechnology* **2015**, *26*, 205501.

(6) Zhang, S.; Zhang, J.; Fang, W.; Zhang, Y.; Wang, Q.; Jin, J. *Nano Letters* **2018**, *18*, 6563–6569.

(7) Miller, R. A.; Presley, A. D.; Francis, M. B. *Journal of the American Chemical Society* **2007**, *129*, 3104–3109.

(8) Delor, M.; Dai, J.; Roberts, T. D.; Rogers, J. R.; Hamed, S. M.; Neaton, J. B.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *Journal of the American Chemical Society* **2018**, *140*, 6278–6287.

(9) Endo, M.; Fujitsuka, M.; Majima, T. *Chemistry – A European Journal* **2007**, *13*, 8660–8666.

(10) Koch, C.; Eber, F. J.; Azucena, C.; Förste, A.; Walheim, S.; Schimmel, T.; Bittner, A. M.; Jeske, H.; Gliemann, H.; Eiben, S.; Geiger, F. C.; Wege, C. *Beilstein Journal of Nanotechnology* **2016**, *7*, 613–629.

(11) Chu, S.; Brown, A. D.; Culver, J. N.; Ghodssi, R. *Biotechnology Journal* **2018**, *13*, 1800147.

(12) Butler, P. J. G. *Journal of General Virology* **1984**, *65*, 253–279.

(13) Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. *Nano Letters* **2010**, *10*, 181–186.

(14) Kadri, A.; Maiß, E.; Amsharov, N.; Bittner, A. M.; Balci, S.; Kern, K.; Jeske, H.; Wege, C. *Virus Research* **2011**, *157*, 35–46.

(15) Zhang, J.; Zhou, K.; Wang, Q. *Small* **2016**, *12*, 4955–4959.

(16) Zhang, J.; Wang, X.; Zhou, K.; Chen, G.; Wang, Q. *ACS Nano* **2018**, *12*, 1673–1679.

(17) Lin, R. D.; Steinmetz, N. F. *Nanoscale* **2018**, *10*, 16307–16313.

(18) Finbloom, J. A.; Aanei, I. L.; Bernard, J. M.; Klass, S. H.; Elledge, S. K.; Han, K.; Ozawa, T.; Nicolaides, T. P.; Berger, M. S.; Francis, M. B. *Nanomaterials* **2018**, *8*, 1007.

(19) Ramsey, A. V.; Bischoff, A. J.; Francis, M. B. *Journal of the American Chemical Society* **2021**, *143*, 7342–7350.

(20) Bruckman, M. A.; VanMeter, A.; Steinmetz, N. F. *ACS Biomaterials Science & Engineering* **2015**, *1*, 13–18.

(21) Bruckman, M. A.; Czapar, A. E.; VanMeter, A.; Randolph, L. N.; Steinmetz, N. F. *Journal of Controlled Release* **2016**, *231*, 103–113.

(22) Bhyravbhatla, B.; Watowich, S. J.; Caspar, D. L. *Biophysical Journal* **1998**, *74*, 604–615.

(23) Ksenofontov, A. L.; Petoukhov, M. V.; Prusov, A. N.; Fedorova, N. V.; Shtykova, E. V. *Biochemistry (Moscow)* **2020**, *85*, 310–317.

(24) Miller, R. A.; Stephanopoulos, N.; McFarland, J. M.; Rosko, A. S.; Geissler, P. L.; Francis, M. B. *Journal of the American Chemical Society* **2010**, *132*, 6068–6074.

(25) Vugmeyster, Y.; Xu, X.; Theil, F.-P.; Khawli, L. A.; Leach, M. W. *World Journal of Biological Chemistry* **2012**, *3*, 73–92.

(26) Ducry, L. In *Therapeutic Proteins: Methods and Protocols*, Voynov, V., Caravella, J. A., Eds.; Humana Press: Totowa, NJ, 2012, pp 489–497.

(27) Vhora, I.; Patil, S.; Bhatt, P.; Misra, A. In *Advances in Protein Chemistry and Structural Biology*, Donev, R., Ed.; Academic Press: 2015; Vol. 98, pp 1–55.

(28) Heck, A. J. R. *Nature Methods* **2008**, *5*, 927–933.

(29) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. *Biochemical and Biophysical Research Communications* **2014**, *445*, 683–693.

(30) Weiss, V. U.; Bereszczazk, J. Z.; Havlik, M.; Kallinger, P.; Gösler, I.; Kumar, M.; Blaas, D.; Marchetti-Deschmann, M.; Heck, A. J. R.; Szymanski, W. W.; Allmaier, G. *Analytical Chemistry* **2015**, *87*, 8709–8717.

(31) Lössl, P.; Snijder, J.; Heck, A. J. R. *Journal of the American Society for Mass Spectrometry* **2014**, *25*, 906–917.

(32) McKay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V. *Journal of the American Chemical Society* **2006**, *128*, 11433–11442.

(33) Leney, A. C.; Heck, A. J. R. *Journal of the American Society for Mass Spectrometry* **2017**, *28*, 5–13.

(34) Keifer, D. Z.; Pierson, E. E.; Jarrold, M. F. *Analyst* **2017**, *142*, 1654–1671.

(35) Kafader, J. O.; Melani, R. D.; Senko, M. W.; Makarov, A. A.; Kelleher, N. L.; Compton, P. D. *Analytical Chemistry* **2019**, *91*, 2776–2783.

(36) Harper, C. C.; Elliott, A. G.; Oltrogge, L. M.; Savage, D. F.; Williams, E. R. *Analytical Chemistry* **2019**, *91*, 7458–7465.

(37) Harper, C. C.; Brauer, D. D.; Francis, M. B.; Williams, E. R. *Chemical Science* **2021**, *12*, 5185–5195.

(38) Susa, A. C.; Xia, Z.; Williams, E. R. *Analytical Chemistry* **2017**, *89*, 3116–3122.

(39) Elliott, A. G.; Harper, C. C.; Lin, H.-W.; Williams, E. R. *Analyst* **2017**, *142*, 2760–2769.

(40) Elliott, A. G.; Merenbloom, S. I.; Chakrabarty, S.; Williams, E. R. *International Journal of Mass Spectrometry* **2017**, *414*, 45–55.

(41) Lutomski, C. A.; Gordon, S. M.; Remaley, A. T.; Jarrold, M. F. *Analytical Chemistry* **2018**, *90*, 6353–6356.

(42) Lutomski, C. A.; Lyktey, N. A.; Zhao, Z.; Pierson, E. E.; Zlotnick, A.; Jarrold, M. F. *Journal of the American Chemical Society* **2017**, *139*, 16932–16938.

(43) Lutomski, C. A.; Lyktey, N. A.; Pierson, E. E.; Zhao, Z.; Zlotnick, A.; Jarrold, M. F. *Journal of the American Chemical Society* **2018**, *140*, 5784–5790.

(44) Royal, J. M. et al. *Vaccines* **2021**, *9*, 1347.

(45) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. *Current Opinion in Structural Biology* **2015**, *31*, 64–74.

(46) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. In *The Proteomics Protocols Handbook*, Walker, J. M., Ed.; Humana Press: Totowa, NJ, 2005, pp 571–607.

(47) Kyte, J.; Doolittle, R. F. *Journal of Molecular Biology* **1982**, *157*, 105–132.

(48) ElShafei, G. M.; Al-Sabagh, A.; Yehia, F.; Philip, C.; Moussa, N.; Eshaq, G.; ElMetwally, A. *Applied Catalysis B: Environmental* **2018**, *224*, 681–691.

(49) Maza, J. C.; Bader, D. L. V.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Krska, S. W.; Parish, C. A.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 3885–3892.

(50) Marmelstein, A. M.; Lobba, M. J.; Mogilevsky, C. S.; Maza, J. C.; Brauer, D. D.; Francis, M. B. *Journal of the American Chemical Society* **2020**, *142*, 5078–5086.

(51) Jurchen, J. C.; Williams, E. R. *Journal of the American Chemical Society* **2003**, *125*, 2817–2826.

(52) Hamerlynck, L. M.; Bischoff, A. J.; Rogers, J. R.; Roberts, T. D.; Dai, J.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *The Journal of Physical Chemistry B* **2022**, *126*, 7981–7991.

(53) Li, X.; Song, B.; Chen, X.; Wang, Z.; Zeng, M.; Yu, D.; Hu, D.; Chen, Z.; Jin, L.; Yang, S.; Yang, C.; Chen, B. *PLOS ONE* **2013**, *8*, e77717.

(54) Harper, C. C.; Miller, Z. M.; Lee, H.; Bischoff, A. J.; Francis, M. B.; Schaffer, D. V.; Williams, E. R. *Analytical Chemistry* **2022**, *94*, 11703–11712.

(55) Miller, Z. M.; Harper, C. C.; Lee, H.; Bischoff, A. J.; Francis, M. B.; Schaffer, D. V.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **2022**, *33*, 2129–2137.

(56) Jordan, J. S.; Xia, Z.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **2022**, *33*, 607–611.

(57) Harper, C. C.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **2019**, *30*, 2637–2645.

(58) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

(59) Schrödinger Release 2021-4: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.

(60) Mahoney, M. W.; Jorgensen, W. L. *The Journal of Chemical Physics* **2000**, *112*, 8910–8922.

(61) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(62) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. *Journal of Chemical Theory and Computation* **2021**, *17*, 4291–4300.

(63) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *The Journal of Chemical Physics* **1994**, *101*, 4177–4189.

(64) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *The Journal of Chemical Physics* **1992**, *97*, 2635–2643.

(65) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *The Journal of Chemical Physics* **2004**, *120*, 9665–9678.

(66) Altman, D. G., *Practical statistics for medical research*; Chapman and Hall: London, 1991.

(67) Kirkwood, B. R.; Sterne, J. A. C., *Essential Medical Statistics, 2nd Edition*; Wiley-Blackwell: Oxford, 2003.

Chapter 3

A protein-based model for energy transfer between photosynthetic light harvesting complexes is constructed using a direct protein–protein conjugation strategy

3.1 Abstract

Photosynthetic organisms utilize dynamic and complex networks of pigments bound within light harvesting complexes to transfer solar energy from antenna complexes to reaction centers. Understanding the principles underlying the efficiency of these energy transfer processes, and how they may be incorporated into artificial light harvesting systems, is facilitated by the construction of easily tunable model systems. Herein a protein-based model to mimic directional energy transfer between light harvesting complexes using a circular permutant of the tobacco mosaic virus coat protein (cpTMV), which self-assembles into a 34-monomer hollow disk, is developed. Two populations of cpTMV assemblies, one labeled with donor chromophores and another labeled with acceptor chromophores, were coupled using a direct protein–protein bioconjugation method. Using potassium ferricyanide as an oxidant, assemblies containing 3-aminotyrosine were activated towards the addition of assemblies containing *p*-aminophenylalanine. Both of these non-canonical amino acids were introduced into the cpTMV monomers through amber codon suppression. This coupling strategy has the advantages of directly, irreversibly, and site-selectively coupling donor with acceptor protein assemblies and avoids cross-reactivity with native amino acids and undesired donor–donor or acceptor–acceptor combinations. The coupled donor–acceptor model was shown to transfer energy from an antenna disk containing donor chromophores to a downstream disk containing acceptor chromophores. This model ultimately provides a con-

trollable and modifiable platform for understanding photosynthetic inter-assembly energy transfer and may lead to the design of more efficient functional light harvesting materials.

This chapter is based on the following publication:

Bischoff, A. J.; Hamerlynck, L. M.; Li, A. J.; Roberts, T. D.; Ginsberg, N. S.; Francis, M. B. Protein-Based Model for Energy Transfer between Photosynthetic Light Harvesting Complexes Is Constructed Using a Direct Protein–Protein Conjugation Strategy. *J. Am. Chem. Soc.* **2023**, *145*, 15827–15837.

This article can be accessed via the following link:
https://pubs.acs.org/doi/full/10.1021/jacs.3c02577

3.2 Introduction

Within the membranes of photosynthetic organisms lie dynamic networks of multiple types of light harvesting complexes, which are responsive to varying levels of light and environmental conditions.[1–3] Many of these light harvesting complexes are made up of identical repeating protein subunits embedded with photosynthetic pigments, which are often further embedded within lipid bilayers to form membrane-spanning complexes. Within photosynthetic membranes, light harvesting complexes dynamically organize to form heterogeneous macroassemblies composed of antenna complexes that absorb light and transfer excited state energy to complexes containing reaction centers. The rate and efficiency of processes in the energy transfer pathway, and where energetic bottlenecks may occur, differ among photosynthetic organisms. One such example is the pathway from light harvesting complex 2 (LH2), which absorbs light energy and funnels it to light harvesting complex 1 (LH1) and eventually to a reaction center (RC) in purple photosynthetic bacteria such as *R. sphaeroides* (Figure 3.1a). The inter-pigment distances between nearest neighbors are low (∼1.0 nm in LH1 and ∼0.9 nm in LH2) with well-aligned transition dipole moments, leading to rapid and efficient energy transfer within single complexes.[4–6] Energy transfer between light harvesting complexes, such as from LH2 to LH2 or from LH2 to LH1, is not as well constrained and can occur across a range of distances, as LH2 and LH1 are able to move laterally within the photosynthetic membrane. The lateral diffusion of LH2 and LH1 in the bilayer affects the inter-pigment distances between complexes and therefore energy transfer efficiency, but also allows for the reorganization of complexes to respond to differing environmental conditions such as varying light intensity. Numerous studies have shown that the slow step in this process of energy transfer in purple photosynthetic bacteria is the LH1-to-RC transfer, likely due to the relatively large distances involved between bacteriochlorophyll pigments (3–4 nm).[7–10] In contrast, 2D electronic spectroscopy has revealed that the transfer of energy between chlorosome antennae and the FMO complex in green sulfur bacteria occurs on a slower timescale than that between the FMO complex and RCs, and on a much slower timescale than energy transfer within a subunit.[2,11]

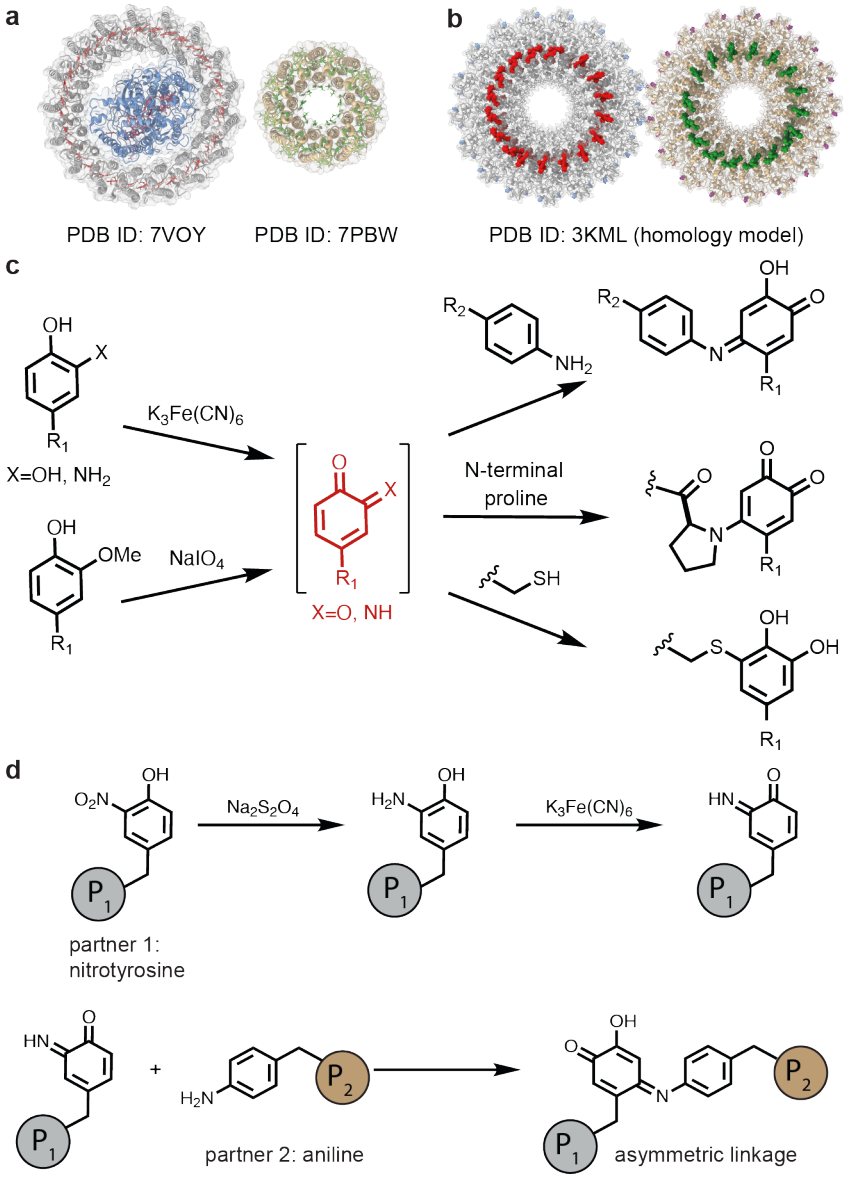


Figure 3.1: Strategy for site-selective protein–protein coupling. (a) A schematic of LH1 is shown next to LH2 from *R. sphaeroides.* Donor chromophores are shown in green, and acceptor Bcl chromophores are shown in red. (b) A mimic for LH2-to-LH1 energy transfer is envisaged using covalently conjugated cpTMV disks labeled with fluorescent dyes. Donor pigments are shown in green and acceptor pigments are shown in red. (c) Oxidative coupling strategies can effect the chemoselective bioconjugation of *o*-quinoid intermediates to anilines, N-terminal proline residues, and thiols. (d) A scheme of the oxidative coupling-based strategy is shown for the asymmetric conjugation of two distinct, engineered protein assemblies containing non-canonical amino acids.

These studies have made substantial progress toward understanding the timescales involved in photosynthesis. However, deconvolving the numerous processes influencing energy transfer efficiency *in vivo* or in isolated photosynthetic membranes presents a significant challenge due to additional components within the photosynthetic membrane and the environment, and difficulty in isolating functional networks of associated light harvesting complexes from their membranes. Previous studies using synthetic model systems have shed light on exciton transport across pigment compositions and distances analogous to those found in photosynthesis. This has been achieved using layered carbon nanotubes,[12] supramolecular polymers[13] and metallacycles,[14] multi-component self-assembling materials,[15] and biomolecular frameworks.[16] These self-assembling systems have demonstrated how light harvesting capabilities may have arisen in primitive organisms from simple peptide, porphyrin, and mineral components.[17] Providing a scaffold to confine chromophores within ordered structures in solution also can increase their photocatalytic ability while avoiding photodegradation.[18,19] However, few of these scaffolds imitate the precise, discrete, circular arrangement of chromophores present in many photosynthetic complexes such as LH1 and LH2. Understanding how these protein-bound chromophore assemblies interact to transfer energy across distances within photosynthetic membranes is an added challenge that is difficult to address in model systems.[12,20]

Herein, a model system is developed for the study of energy transfer between distinct light harvesting complexes (LHCs) of differing pigment compositions using a circular permutant of the tobacco mosaic virus coat protein (cpTMV). cpTMV assembles into double-layered disks with a hollow pore, with 17 monomers per disk and 34 monomers per assembly, and its assembly state is stable across a broad pH and ionic strength range.[21] Mutants and conjugates of cpTMV have been shown to maintain their assembly state after storage for several days at room temperature (see Chapter 2).[22] The structure of cpTMV resembles the structure of light harvesting complexes such as LH1 and LH2, which also assemble into flattened disks composed of identical monomers in circular arrays. cpTMV has previously been used as a tunable scaffold for the attachment of circular chromophore arrays, with varying constraint and inter-chromophore distance, to mimic single LHCs, providing insight into the physical characteristics underpinning energy transfer.[21,23–29] The inter-pigment distances in pigment-labeled single cpTMV assemblies can be varied based on the attachment site on the protein to between 1.6 and 3.2 nm on average, slightly larger than those found in LH2 and LH1.[29] Flexibility in the linkers between chromophores and the protein surface also introduce disorder in the orientation between chromophores attached to cpTMV, although this flexibility can be mitigated by attaching chromophores to the region between the two disks in single cpTMV assemblies.[28,29]

This chapter describes the use of a bioconjugation strategy to appose complete assemblies of cpTMV disks that contain donor and acceptor chromophores in a controlled fashion, imitating the spatial relationships that foster inter-complex energy transfer between non-identical light harvesting complexes such as LH1 and LH2 (Figure 3.1a-b). LH1 and LH2 are found in complex membrane environments, and LH2 to LH1 energy transfer is difficult to study in isolation. The coupled cpTMV complex model with a single acceptor complex

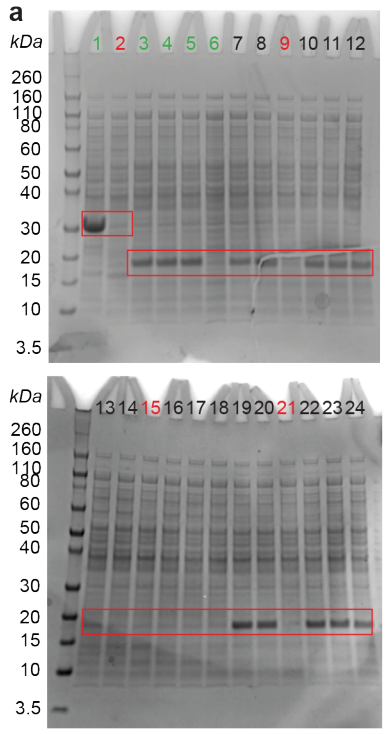
linked peripherally to donor complexes allows inter-complex energy transfer to be examined in aqueous solution, isolated from other photosynthetic components. The cpTMV inter-assembly energy transfer model was constructed by modifying an oxidative coupling strategy previously developed in the Francis lab to bind two large protein assemblies directly and covalently. The multi-component cpTMV-based systems demonstrate energy transfer between donor and acceptor complexes. Going forward, this strategy provides an adjustable, solution-state model for studying photosynthetic energy transfer.

3.3 Results and Discussion

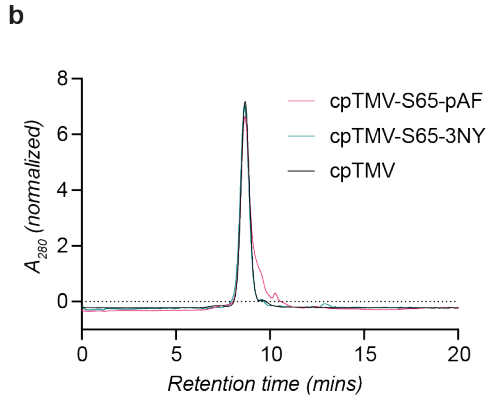
To construct the protein scaffold for inter-assembly energy transfer, a method was required for covalently coupling two distinct donor and acceptor protein assemblies, each of which was composed of 34 monomers assembled non-covalently into a double-layered hollow disk. A promising method would modify a strategy previously developed by the Francis lab in which a catechol, aminophenol, or methoxyphenol is oxidized to an *o*-quinone or *o*-iminoquinone, using potassium ferricyanide or sodium periodate as a small molecule oxidant, before coupling to thiol or amine-based nucleophiles (Figure 3.1c).[30–32] In this chapter, the strategy is expanded to form direct, site-selective, oxidative protein–protein linkages utilizing a pair of non-canonical amino acids (ncAAs), *p*-aminophenylalanine (pAF) and 3-nitrotyrosine (3NY), installed during protein expression (Figure 3.1d).

3.3.1 Engineering non-canonical amino acid residues into cpTMV assemblies

The first step in creation of the donor–acceptor complex assemblies was the installation of ncAAs in peripheral sites on cpTMV disks. To select the sites for amino acid mutagenesis, we took inspiration from the work of the Wang laboratory, which engineered recombinant TMV to assemble into thin nanosheets of conjugated assemblies through engineered cysteine or histidine residues on and near the N-termini of the disks.[33,34] We selected analogous sites on the cpTMV construct, which is a permutant of the wild type, for mutation to pAF and 3NY. These constructs were prepared using amber codon suppression as reported previously.[35–39] Screening across a variety of expression conditions and mutation sites revealed that the S65 site was most amenable to installation of these ncAAs, though there was also significant yield at the S63 site (Figure 3.2a). The cpTMV-S65-pAF mutant had also been previously shown to couple to phenol-modified gold nanoparticles using an enzymatic oxidative coupling method,[38] suggesting that this site may be optimal for the conjugation of two large protein assemblies. Incorporation of the ncAA was verified by mass spectrometry (MS). Because tyrosine and pAF differ by only 2 Da and are therefore difficult to distinguish using protein MS, further verification that the correct amino acid had been installed was performed by omitting the ncAA from the expression media and analyzing the cell lysate using gel electrophoresis. This showed that full-length cpTMV expression did not occur when the ncAA was omitted



lane	protein	mutation	arabinose concentration (% w/v)	3NY addition time (h)	arabinose addition time (h)
1	GFP	N150-3NY	0.05	3.5	0
2	GFP	N150-3NY	0.05	-	0
3	cpTMV	S23C	0.05	3.5	0
4	cpTMV	S23C	0.05	2.5	0
5	cpTMV	S23C	0.05	-	0
6	cpTMV	S23C	0.05	3.5	0.5
7	cpTMV	S63-3NY	0.05	3.5	0
8	cpTMV	S63-3NY	0.05	2.5	0
9	cpTMV	S63-3NY	0.05	-	0
10	cpTMV	S63-3NY	0.05	3.5	0.5
11	cpTMV	S63-3NY	0.1	3.5	0
12	cpTMV	S63-3NY	0.1	3.5	0.5
13	cpTMV	Y64-3NY	0.05	3.5	0
14	cpTMV	Y64-3NY	0.05	2.5	0
15	cpTMV	Y64-3NY	0.05	-	0
16	cpTMV	Y64-3NY	0.05	3.5	0.5
17	cpTMV	Y64-3NY	0.1	3.5	0
18	cpTMV	Y64-3NY	0.1	3.5	0.5
19	cpTMV	S65-3NY	0.05	3.5	0
20	cpTMV	S65-3NY	0.05	2.5	0
21	cpTMV	S65-3NY	0.05	-	0
22	cpTMV	S65-3NY	0.05	3.5	0.5
23	cpTMV	S65-3NY	0.1	3.5	0
24	cpTMV	S65-3NY	0.1	3.5	0.5



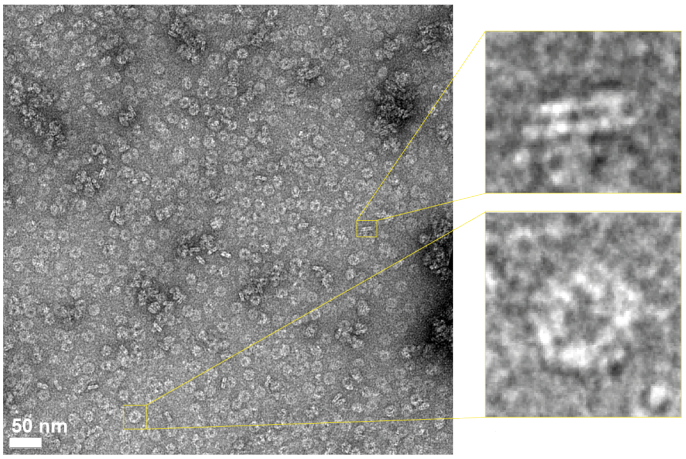


Figure 3.2: Screen of peripheral sites for expression of cpTMV containing non-canonical amino acid 3-nitrotyrosine (3NY) and assembly state of cpTMV-S65-3NY. (a) An SDS-PAGE gel of the soluble fraction of small scale expressions with the variations shown in the corresponding table shows that the S65-3NY mutant was expressed at the highest level of mutants screened. Positive controls (GFP-N150-3NY and cpTMV-S23C without a non-canonical amino acid) and their expression conditions are shown in green. Negative controls to which no 3NY was added are shown in red. Red outlines superimposed on the SDS-Page gels indicate the expected MW for GFP (28 kDa) and cpTMV (16 kDa). (b) The size of both cpTMV-S65-pAF and cpTMV-S65-3NY matches that of previously reported cpTMV as demonstrated by size exclusion chromotography. (c) TEM images of cpTMV-S65-3NY demonstrate that both constructs assemble into double-layered disks similar to other cpTMV mutants. Enlarged images show a disk on its side with two layers (top image) and on its face with a central pore (bottom image).

from the protein expression media (Figure 3.2a). The size of these constructs matched the previously reported cpTMV double-disk structures,[21] at approximately 600 kDa as verified by size exclusion chromatography (SEC, Figure 3.2b) and 18 nm × 5 nm with the double disk with pore morphology as confirmed by transmission electron microscopy (TEM) of the cpTMV-3NY construct (Figure 3.2c).

3.3.2 Asymmetric oxidative coupling of cpTMV assemblies

The cpTMV disks containing ncAAs were then subjected to potassium ferricyanide-mediated oxidative coupling conditions as previously reported by the Francis laboratory for pAF- and 3NY-containing proteins (Figure 3.3a).[30–32] The accessibility of the mutated amino acids in the cpTMV-S65-pAF and -3NY constructs was first examined through small molecule couplings. For the 3NY mutant, treatment with sodium dithionite was first required to reduce the nitrophenol to an aminophenol, producing cpTMV-S65-3-aminophenol (cpTMV-S65-3AY). This reduction proceeded at 83% yield (Figure 3.3a-b) as measured by MS peak integration; however, some re-oxidation of the aminophenol may have occurred in air or during elution prior to the MS measurement. After removal of sodium dithionite, cpTMV-S65-3AY was directly added to a solution of *p*-toluidine and potassium ferricyanide, resulting in a single modification of each monomer to complete conversion (Figure 3.3b). This suggests that reduction of cpTMV-S65-3NY to cpTMV-S65-3AY proceeded to full conversion prior to the oxidative coupling reaction, as no appreciable amount of unmodified protein was observed. Treating the pAF-containing variant with 2-amino-*p*-cresol and potassium ferricyanide also resulted in single modification of each monomer to full conversion (Figure 3.3b).

For both the cpTMV-S65-pAF and cpTMV-S65-3AY constructs, full conversion of each monomer to form one oxidative coupling product (as shown in Figure 3.3b) results in 34 total modifications per double disk assembly. Each of these constructs contain multiple reactive side chains including lysine, tyrosine, and cysteine, along with termini within the cpTMV pore. A construct identical in sequence to cpTMV-S65-pAF and cpTMV-S65-3AY but with serine in place of the ncAAs at position 65 (cpTMV-S65) showed no reactivity when subject to the same oxidative conditions, indicating that only the engineered ncAAs were modified, without cross-reactivity with other amino acids in the cpTMV sequence (Figure 3.3d-e).

Following the confirmation of site selectivity, an oxidative conjugation of the two coupling partners, cpTMV-S65-pAF and cpTMV-S65-3AY, was performed (Figure 3.4a) by mixing the two in a ratio of 1:1 through 10:1 cpTMV-S65-pAF:cpTMV-S65-3AY. A higher stoichiometry of cpTMV-S65-pAF was used to promote the formation of smaller assemblies, rather than large sheets, of asymmetrically conjugated disks. When at a ratio of 10:1 of cpTMV-S65-pAF:cpTMV-S65-3AY, analysis of the mass spectrum resulted in an estimated modification of 31% of monomers of cpTMV-S65-3AY (the limiting coupling partner), or an average of 10.5 modifications per double disk assembly. This does not necessarily demonstrate that the cpTMV-S65-3AY disks were coupled to 10 cpTMV-S65-pAF disks, as each disk contains 34 monomers, and it is possible that there were multiple oxidative coupling linkages between

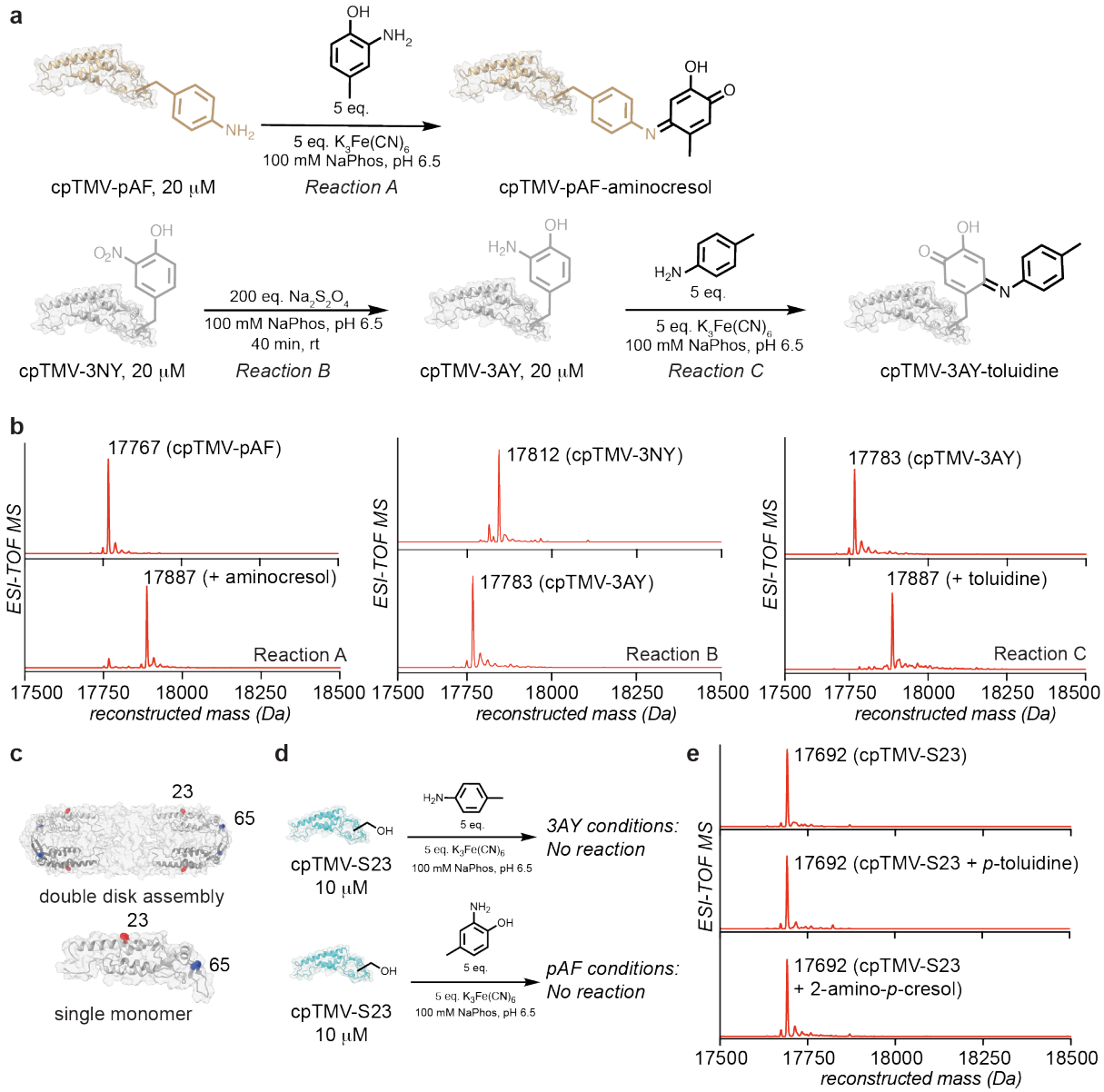


Figure 3.3: Chemical confirmations for accessibility and modification site of ncAA-containing amino acids through small molecule couplings. (a) Conditions are shown for $K_3Fe(CN)_6$-mediated oxidative coupling to cpTMV-S65-pAF (Reaction A) and cpTMV-S65-3AY (Reaction C) following reduction with $Na_2S_2O_4$ (Reaction B). (b) Reconstructed ESI-TOF mass spectra indicate high conversion of each ncAA-containing cpTMV monomer to the expected oxidative coupling product (expected MW: 17887 Da). (c) A cutaway view is provided, showing cpTMV monomers on opposite sides of the disk in gray, sites for protein–protein conjugation in blue, and sites for pigment attachment in red. A close-up view of a single monomer of the individual double disk assembly is also shown. (d) Reaction schemes show control reactions using cpTMV-S23 incubated with *p*-toluidine (3AY conditions) or 2-amino-*p*-cresol (pAF conditions) and $K_3Fe(CN)_6$. (e) No appreciable difference in the mass of cpTMV-S23 (expected MW = 17692) was observed after incubation with small molecules as shown in the reaction schemes in (d).

monomers in a single disk pair. Qualitative analysis using gel electrophoresis also showed formation of conjugates under the oxidative coupling conditions (Figure 3.4a-c). A series of controls showed that cpTMV-S65-pAF, cpTMV-S65-3AY, and $K_3Fe(CN)_6$ are all required to achieve appreciable disk–disk conjugation (Figure 3.4c). This indicates that the oxidative coupling reaction was asymmetric, with cpTMV-S65-pAF disks only reacting with cpTMV-S65-3AY disks and vice-versa, and no coupling of like disks observed.

The effect of the oxidative coupling conditions on the assembly state of cpTMV was next assessed. Decomposition of the self-assembled disks into monomers was not expected from the oxidative coupling conditions used because other virus-like particles have been shown to maintain their assembly state under similar oxidative conditions;[40,41] however, we wanted to ensure that the potential strain placed on the non-covalent interactions between monomers by the close proximity to an adjacent large protein complex did not cause decomposition. The attachment of two intact cpTMV disks would result in a 1.2 MDa, 36 nm-long complex, and even larger complexes may be expected from the conjugation of multiple disks. A size increase from individual to conjugated protein was observed through the appearance of a higher MW species using native gel electrophoresis (Figure 3.4f), by SEC as revealed by a higher-molecular weight shoulder when compared to individual assemblies, and by dynamic light scattering (DLS), which also showed two distinct populations corresponding to individual disks and larger assemblies after coupling, suggesting an increase in size rather than decomposition to monomers upon conjugation. Isolation of the larger-size complexes by SEC and size measurement by SEC, DLS, and native agarose gel electrophoresis showed a successful separation of the larger complexes from individual assemblies (Figure 3.4d-f). A comparison to several controls, including a control with cpTMV-S65-3AY and $K_3Fe(CN)_6$ but lacking cpTMV-S65-pAF, did not show evidence of a size increase and therefore conjugation of cpTMV-S65-3AY to itself, further confirming that this protein–protein coupling reaction occurs asymmetrically.

Disks were conjugated at an equal ratio of cpTMV-S65-pAF:cpTMV-S65-3NY at a concentration of 20 μM for visualization using transmission electron microscopy (TEM). Conjugated disks were observed at dilute concentrations, whereas an uncoupled control did not display the same morphology (Figure 3.5). While many of the TEM images appeared to show the coupled cpTMV disks in a parallel orientation with respect to one another, suggesting that this conformation is preferred in the coupled assemblies, it is possible that there is flexibility in the protein–protein linkage, and some coupled assemblies may exhibit a less parallel orientation. With confirmation that the protein-based model scaffold had been successfully constructed, we moved on to attach synthetic dyes and examine the excited state energy transfer properties of the multi-cpTMV assemblies.

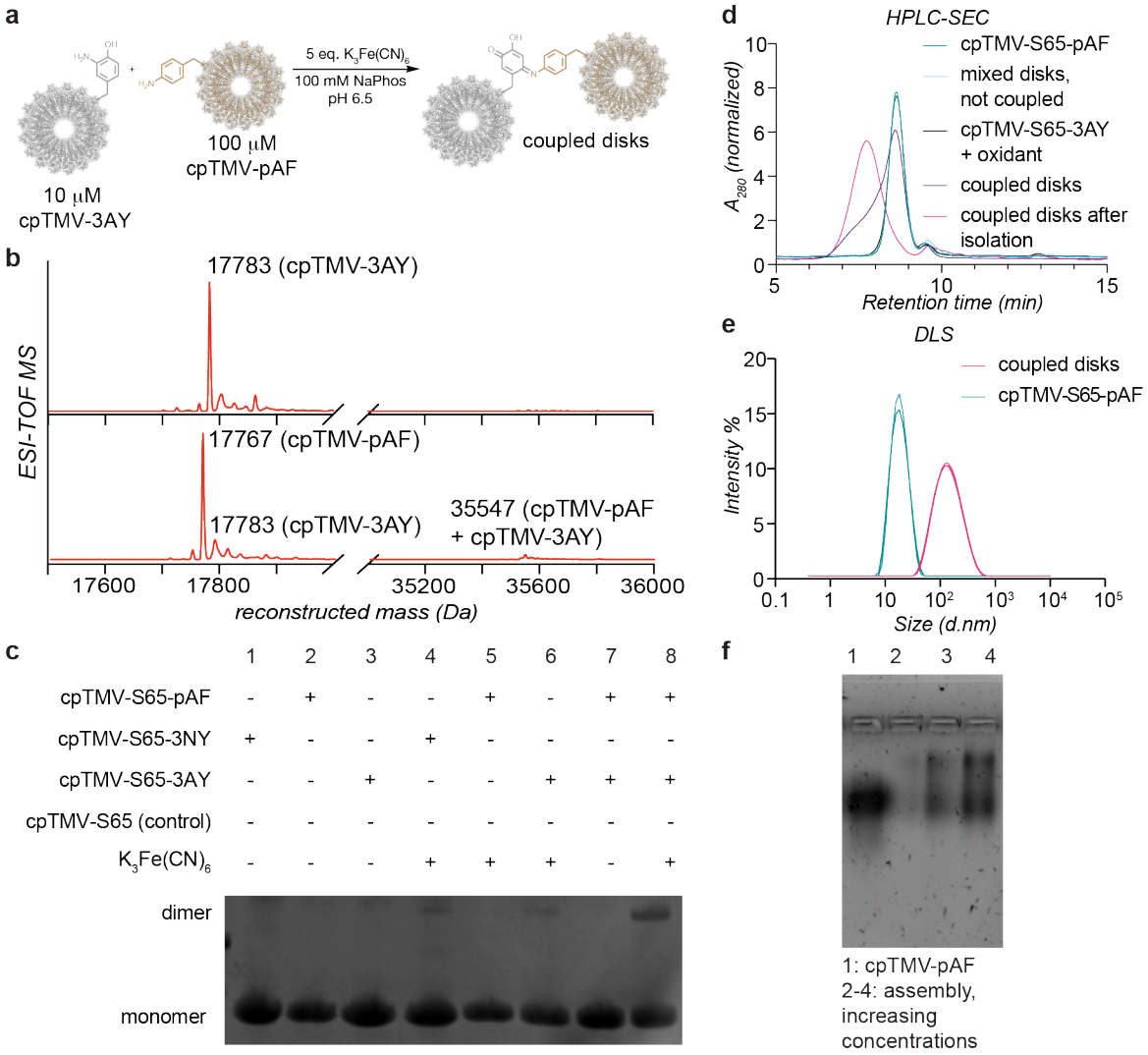


Figure 3.4: Asymmetric coupling of intact cpTMV assemblies. (a) A scheme shows the covalent oxidative coupling of non-canonical amino acids on separate cpTMV disk assemblies. (b) Mass spectrometry shows both reduction of cpTMV-S65-3NY to cpTMV-S65-3AY and coupling of the monomers of the limiting protein, cpTMV-S65-3AY, at an estimated 31% yield (expected MW: 35547 Da). (c) An SDS-PAGE gel shows disk coupling and controls. (d) A size increase after oxidative coupling was verified, and assemblies isolated, using size exclusion chromatography-HPLC. Omitting the oxidant or one of the coupling partners, cpTMV-S65-pAF, prevented the size increase. (e) The increase in size and isolation of assemblies from individual disks was verified by dynamic light scattering. (f) The size increase from individual disks to assemblies was also observable using a native agarose gel: 0.9% agarose in 50 mM NaPhos, pH 7.2, 8 h, 0 °C.

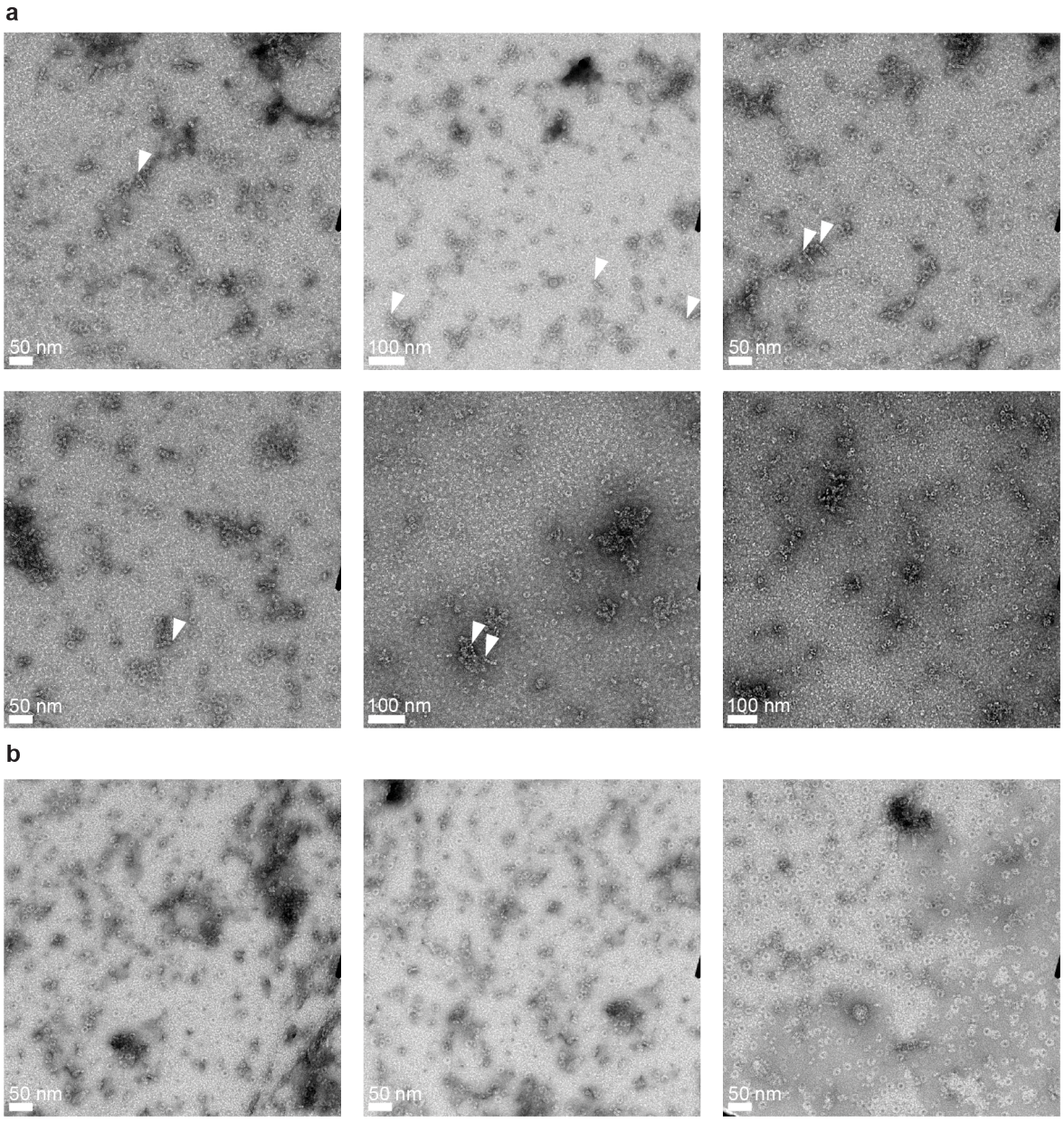


Figure 3.5: TEM images of coupled and mixed cpTMV assemblies. (a) TEM images of cpTMV-S65-pAF coupled to cpTMV-S65-3AY with potassium ferricyanide show multiple examples of disks that appear conjugated at their peripheries. White arrows indicate images of conjugated cpTMV disks from a side view, with disks at or near a 180° angle to one another. (b) TEM images of cpTMV-S65-pAF mixed with cpTMV-S65-3AY in the absence of oxidant show a lower proportion of disks that appear conjugated at their peripheries. Conjugated disks on their sides are not observed.

3.3.3 Measuring energy tranfer between covalently linked donor and acceptor cpTMV complexes

For the installation of light harvesting chromophores, a cysteine was engineered on the exterior surface of each monomer in the cpTMV coat at the S23 position (Figure 3.3c), which has previously been used for chromophore attachment to cpTMV.[28] The pigments selected for this study of energy transfer were Oregon green 488 maleimide (OG488) and Alexa Fluor 594 maleimide (AF594) due to their favorable spectral overlap for Förster resonance energy transfer (FRET), high extinction coefficients, and resistance to photobleaching (Figure 3.6a).[42] These partners have previously been conjugated to TMV,[25] can be selectively excited by tuning the wavelength of light used, and have high spectral overlap between the emission spectrum of the donor OG488 and the absorption spectrum of the acceptor AF594 (Figure 3.6b), enabling energy transfer between the two. The donor chromophores were conjugated to cpTMV-S65-pAF, and acceptor chromophores were conjugated to cpTMV-S65-3NY. Using these maleimide-modified dyes, near-quantitative labeling of all monomers per disk at position S23C was achieved, resulting in circular arrays of 17 dyes attached to the surface of the protein assembly. For a comparison without pigment–pigment interactions, cpTMV-S65-pAF was also labeled at a ratio of one dye per disk (Figure 3.7a-b).

Donor dye-labeled, pAF-containing cpTMV assemblies were then coupled to acceptor dye-labeled, 3AY-containing cpTMV assemblies at a ratio of 2:1 for fluorescence experiments. This coupling led to formation of the desired donor–acceptor model **7** (Figure 3.6e, Figure 3.7c). The constructs used to join the donor and acceptor disks were identical to those used for small molecule labeling experiments (Figure 3.3a-b) with the exception of the S23C mutation for dye attachment. Because the dye attachment proceeded to full conversion and was performed before the disk coupling, all exposed cysteines in these constructs were expected to be protected by their conjugation to the maleimide-containing dyes (Figure 3.7b). Additionally, all other residues on these constructs were shown not to react under the oxidative coupling conditions in small molecule conjugations (Figure 3.3d-e). These factors indicate that the two disks in **7** were joined site-specifically at only positions S65-pAF on the donor disk and S65-3AY on the acceptor disk.

The energy transfer ability of donor pigment-containing cpTMV assemblies to acceptor pigment-containing assemblies (**7**) was then explored (Figure 3.6e). The pigment-labeled, conjugated complexes were compared to several controls in order to understand the contributions of surrounding solvent, protein, and pigments on energy transfer. These controls included the OG488 maleimide dye capped with 2-mercaptoethanesulfonic acid (MESNA) (**1**), cpTMV disks with only sparse OG488 labeling (**2**), cpTMV disks with all monomers labeled with OG488 (**3**), cpTMV disks fully labeled with OG488 and coupled to 4-methylcatechol at the S65 site (**4**), cpTMV disks fully labeled with OG488 and coupled to a cpTMV disk with no chromophore labeling (**5**), and a mixture of cpTMV disks fully labeled with OG488 and cpTMV disks fully labeled with AF594, but not coupled (**6**) (Figure 3.6e, Figure 3.7). For samples **5** and **7**, coupled assemblies were separated from uncoupled disks using SEC, and the fractions containing coupled disks were isolated prior to fluorescence measurements. Spec-

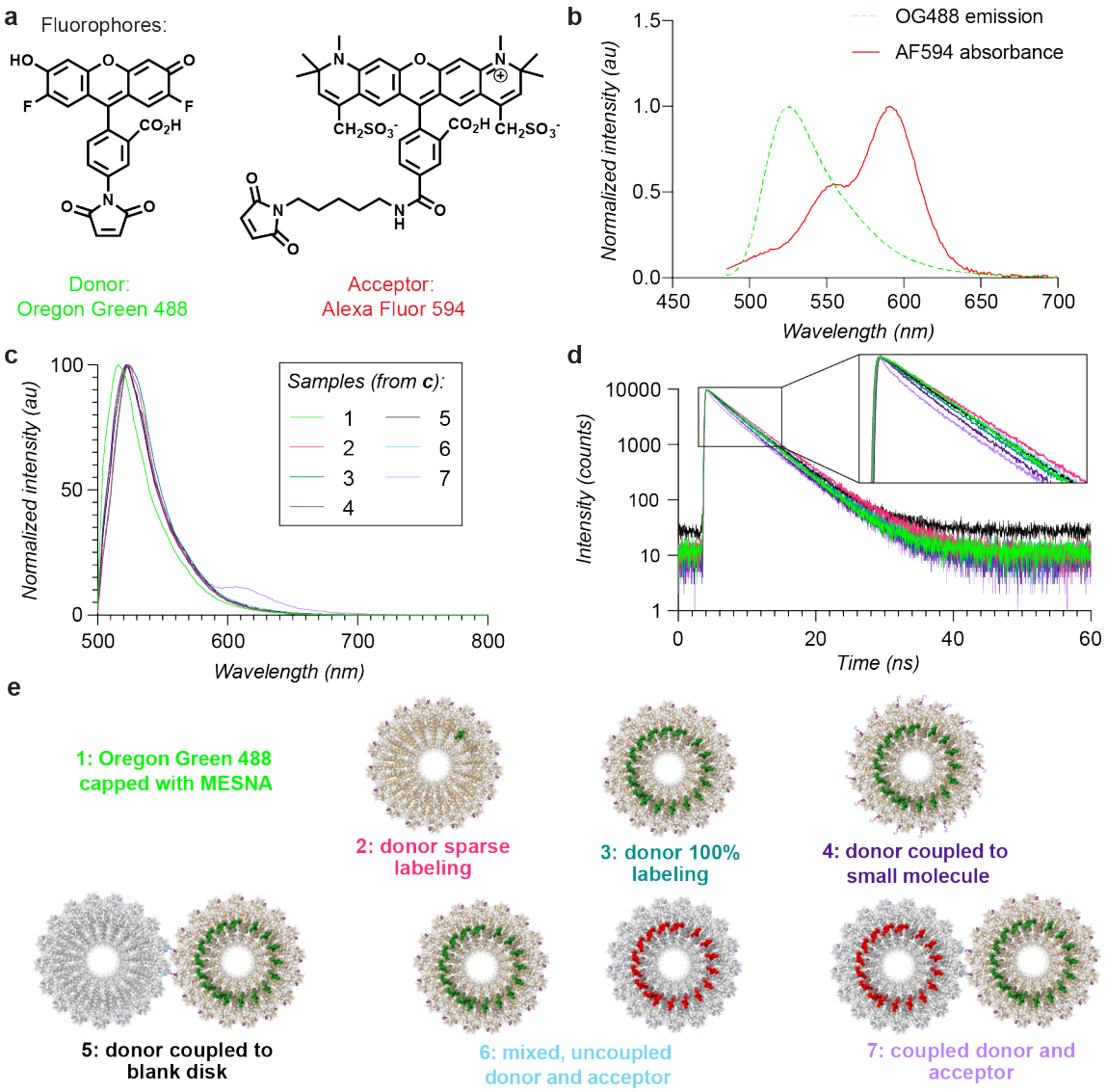


Figure 3.6: Energy transfer between conjugated disk assemblies. (a) The donor and acceptor pair of dyes, Oregon Green 488 and Alexa Fluor 594, are shown with maleimide handles for attachment to the protein surface. (b) The absorbance spectrum of Alexa Fluor 594 and emission spectrum of Oregon Green 488, both conjugated to cpTMV assemblies, show spectral overlap. (c) Emission spectra of coupled assemblies at an excitation wavelength of 465 nm demonstrate that emission only occurs when the donor and acceptor disks are oxidatively coupled and not when they are simply mixed. The legend indicates samples depicted in (e). (d) The fluorescence lifetime of OG488 measured at 524 nm decays more rapidly for the coupled assemblies than for comparable controls; samples are the same as those shown in (e). The inlay shows a magnified version of early timepoints, for better visualization of the shortened lifetime of sample **7** in mauve. (e) A diagram of oxidatively coupled assemblies and a subset of comparison populations are shown, including cpTMV-S65-pAF disks both sparsely (**2**) and completely (**3**) labeled with OG488, a fully labeled cpTMV-S65-pAF disk oxidatively coupled to 4-methylcatechol (**4**), a fully labeled cpTMV-S65-pAF disk oxidatively coupled to a cpTMV-S65-3AY disk bearing no pigments (**5**), a mixture of fully donor-labeled cpTMV-S65-pAF disks and fully acceptor-labeled cpTMV-S65-3AY disks that were not coupled (**6**), and a fully donor-labeled cpTMV-S65-pAF disk oxidatively coupled to a fully acceptor-labeled cpTMV-S65-3AY disk (**7**).

troscopic measurements of each of these samples **1**-**7** were taken at low concentrations (<2 μM dye) to avoid appreciable energy transfer between non-covalently tethered complexes. Emission spectra of each of these populations at an excitation wavelength of 465 nm, which is optimized to excite OG488 without exciting AF594, resulted in emission of the acceptor AF594 only in the coupled donor and acceptor assemblies (**7**) (Figure 3.6c, mauve), but not when uncoupled donor and acceptor assemblies were mixed at the same concentration. A small amount of excitation of AF594 was observed in fully labeled acceptor samples, but direct excitation of the acceptor was not observed in controls containing both donor and acceptor (Figure 3.7e).

The fluorescence lifetime of the donor chromophore, OG488, was also measured for the coupled donor–acceptor assemblies and controls. The fluorescence lifetime of OG488 was shorter in the coupled assemblies than in each of the controls, particularly at shorter timescales (Figure 3.6d; individual spectra, the instrument response function, and fit residuals are shown in Additional Figures 3.9–3.15). A slight decrease in fluorescence lifetime also occurred for the donor disk when coupled to a small molecule (**4**), but not for the donor disk coupled to an unlabeled acceptor disk (**5**). While the donor disk was 91% coupled to 4-methylcatechol in the small molecule case, resulting in ∼31 oxidative coupling linkages per disk, the inter-disk coupling resulted in only 3–4 linkages per disk. This shows that the oxidative coupling linker may provide an avenue for OG488 energy dissipation, but this does not appear significant for the small number of linkages present in coupled disks, as evidenced by the donor disk coupled to a blank disk case. Notably, the mixture of disks in solution at the same concentration as coupled disks did not demonstrate a decrease in donor fluorescence lifetime when compared to donor disks alone, indicating that the decrease in fluorescence lifetime in the coupled donor and acceptor assemblies (**7**) is primarily due to energy transfer between coupled donor and acceptor complexes rather than nearby complexes in solution.

To deepen our understanding of the inter-disk energy transfer process observed, we deconvolved the fluorescence lifetimes into their constituent components using FluoFit software (Table 3.1). Without conjugation to cpTMV, OG488 maleimide capped with MESNA exhibited decay well-described by a monoexponential fit, with a fluorescence lifetime of 4.1 ns, in agreement with the canonical value of 4.1 ns. When conjugated to cpTMV, the fluorescence lifetime of OG488 was best described by two decay components in all cases. The long component of the fluorescence lifetime remained relatively consistent across samples, with the fluorescence lifetime being slightly increased by comparison to the free dye, indicating that this component describes energy dissipation via dye–solvent interactions. By contrast, the short component of the fluorescence lifetime of the donor dye showed more variation according to its protein and chromophore environment. The short component of the fluorescence lifetime of quantitatively labeled assemblies was always decreased when compared to singly-labeled assemblies. This indicates that the excited state of OG488 dyes in an array on cpTMV was depleted via non-radiative decay from interactions with neighboring dyes (e.g., contact quenching and chromophore aggregation). The most significant decrease in the fluorescence lifetime short component occurred when a donor-labeled disk was oxidatively coupled to an acceptor disk labeled with AF594, shortening the short component of the life-

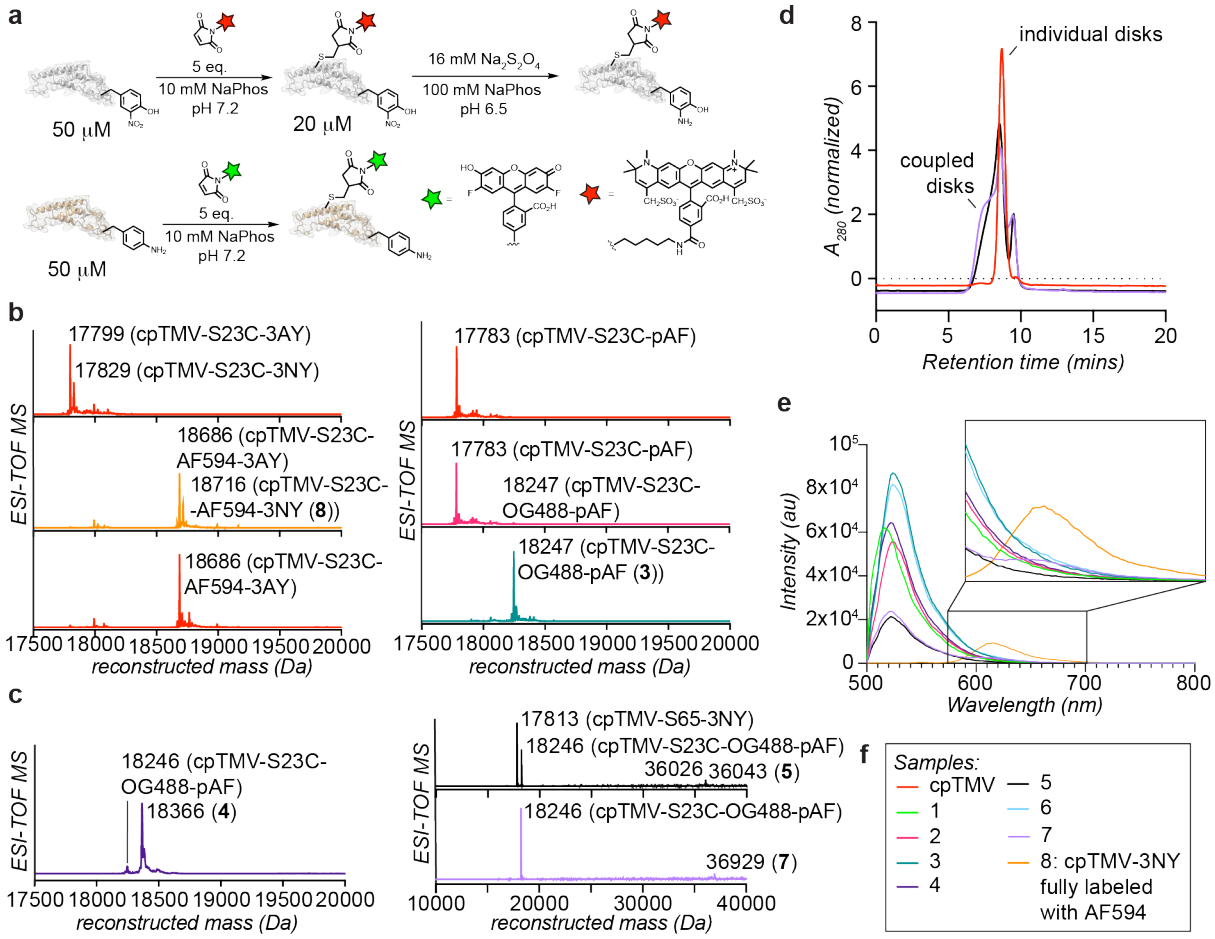


Figure 3.7: Characterization of protein–dye and protein–protein conjugates for energy transfer experiments. (a) Conditions are shown for the conjugation of maleimide dyes to thiol-containing cpTMV and reduction of the non-canonical amino acid 3NY to 3AY. (b) Mass spectra for full labeling of the 3NY-containing cpTMV with AF594 maleimide (expected MW: 18713), reduction of 3NY to 3AY (expected MW: 18683 Da), and single (**2**) and full (**3**) labeling of the pAF-containing cpTMV by OG488 maleimide (expected MW: 18247 Da) are shown. (c) Mass spectrometry data shows cpTMV disks labeled with OG488 and coupled to 4-methylcatechol at the S65 site (**4**) at 91% single modification (expected MW of dye-labeled, coupled disks: 18367 Da), cpTMV disks fully labeled with OG488 and coupled to cpTMV disks with no chromophore labeling (**5**) (expected MW: 36043 Da) and cpTMV disks fully labeled with OG488 coupled to cpTMV disks fully labeled with AF594 (**7**) (expected MW: 36927 Da). (d) Size-exclusion chromatography shows cpTMV disks fully labeled with OG488 and coupled to cpTMV disks with no chromophore labeling (**5**) and cpTMV disks fully labeled with OG488 coupled to cpTMV disks fully labeled with AF594 (**7**). An uncoupled cpTMV sample forming only individual, double-layered disks is shown for comparison. (e) The un-normalized emission spectra of the samples shown in Figure 3.6e in addition to a sample without a donor chromophore show that in the absence of the donor chromophore OG488, excitation and emission of the acceptor chromophore AF594 conjugated to cpTMV-S65-3NY disks can occur. In the presence of a donor chromophore, such as in the case of mixed disks containing donor and acceptor chromophores, this excitation does not occur. (f) A legend for the data presented in this figure, corresponding to the samples shown in Figure 3.6e.

time from 1.5 ± 0.2 ns in mixed assemblies to 0.86 ± 0.05 ns in coupled assemblies. The short component also had the greatest %amplitude contribution to the fluorescence lifetime in the sample containing coupled donor and acceptor disks. This decrease in fluorescence lifetime demonstrates that energy transfer occurred between the asymmetrically coupled assemblies from an array of donor to acceptor chromophores.

Table 3.1: Fluorescence lifetime components of coupled cpTMV assemblies and controls at 465 nm excitation and 524 nm emission

Sample	Long component (ns)	Short component (ns)	Short component (%amplitude)
1	4.11±0.02*	–	–
2	4.79±0.02	2.0±0.2	11.52
3	4.52±0.02	1.79±0.09	21.50
4	4.36±0.02	1.37±0.07	28.69
5	4.52±0.02	1.7±0.1	15.69
6	4.54±0.02	1.5±0.2	9.92
7	4.39±0.02	0.86±0.05	35.82

*values for lifetime components are shown with confidence intervals obtained using support plane error analysis in PicoQuant Fluofit.

The fluorescence decay profile was also used to estimate a timescale for energy transfer from donor to acceptor cpTMV disks in the conjugated system (**7**). Based on structural inspection, we estimate the closest interchromophore distance between a donor and acceptor pair to be 7.5 nm. This distance is based on the distance between residue S23C (the chromophore attachment site) on two closely joined disks in a parallel orientation (red line in Figure 3.8). However, due to the presence of multiple donor–acceptor pairs and distances present in the coupled disk system, there are multiple donor–acceptor pairs that may participate in energy transfer in a single coupled disk system, with 34 chromophores per disk and a distance distribution between ∼7.5 nm for the closest pair and ∼29 nm for the furthest pair. These distances also do not account for the flexibility afforded by the linkers between the chromophore and protein, which allow for some translational and rotational flexibility. Considering only the $1/R^6$ relationship between inter-chromophore distance R and rate of transfer τ_T, there are many donor–acceptor pairs that likely significantly contribute to energy transfer in this model. A subset of the 25 nearest neighbor distances, and the proportion each would be expected to contribute to τ_T based only on the estimated inter-chromophore distance R, are shown in Figure 3.8a-b. This clearly shows that no single pair of donor and acceptor is expected to dominate energy transfer. No single pathway is expected to contribute greater than 12% to the total, and energy transfer is expected to occur via multiple pathways in the absence of information about the chromophores' dipole orientations. Thus, energy transfer calculations should be taken to indicate transfer between disks rather than between specific pairs of donor and acceptor chromophores.[43,44]

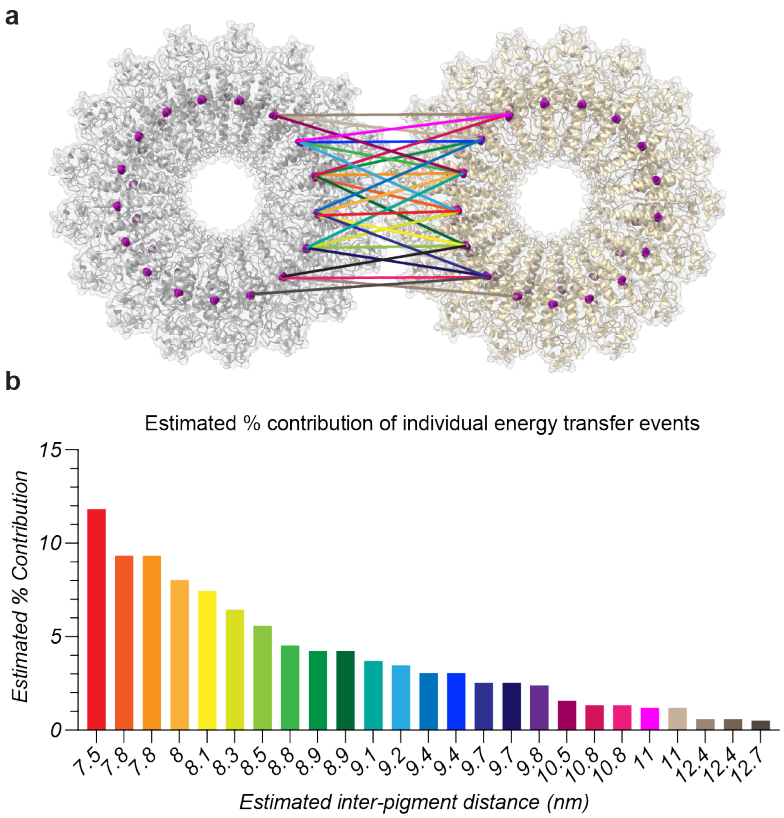


Figure 3.8: An estimate of the contributions of different donor–acceptor chromophore pairs in sample **7** to inter-disk energy transfer. (a) The distance between the chromophore attachment site (S23C) in two cpTMV disks in parallel in close proximity is shown for the closest pair (in red) and 24 additional pairs at increasing distances. This distance was used as a proxy for the inter-chromophore distances. (b) The distances between the chromophore attachment site (S23C) of the pairs shown in (a) and their estimated % contribution to inter-disk energy transfer based on the $1/R^6$ relationship between distance R and energy transfer rate τ_T are shown, indicating that many redundant pathways are likely to significantly contribute to inter-disk energy transfer in this model. This is only an estimate, as the distance calculations are imprecise and orientational flexibility is not accounted for.

Using the experimentally derived fluorescence lifetimes, the timescale of energy transfer between disks τ_T can be estimated at 15 ns, with a corresponding rate of energy transfer k_T of 0.066 ns^{-1} (see Materials and Methods for calculation details). The efficiency E was calculated at 21%, indicating that approximately 21% of the excitonic energy was transferred between the donor and acceptor disks in this system. Despite the large inter-chromophore distances present in this model, the multiple redundant donor–acceptor pathways likely contributed to the observed efficiency.[44] In contrast, energy transfer between LH2 and LH1 in photosynthetic membranes has been found to occur at much faster rates of 3–5 ps.[8] Alongside a lower spectral overlap and extinction coefficient in our donor–acceptor pairs than in the closely coupled chromophores found in bacterial LH2 and LH1, the primary limiting factor in our model for achieving these rapid rates is the large distance between donor and acceptor chromophores, which at ∼7.5 nm is greater than twice the minimum separation expected

between LH2 and LH1 in a photosynthetic membrane.[45] This distance may be tuned in the future to achieve more rapid energy transfer using chromophore modification sites closer to the disk peripheries;[29] for example, given the $1/R^6$ relationship between inter-chromophore distance R and rate of transfer τ_T between FRET chromophore pairs, a decrease in distance from ∼7.5 nm to ∼3 nm could result in greater than a 200-fold decrease in τ_T. Despite the larger inter-chromophore distances, our model may increase energy transfer efficiency through the presence of multiple inter-ring donor–acceptor energy transfer pathways, a feature that also enhances energy transfer efficiency between LH2 and LH1.[46]

In addition, the close association of the photosynthetic pigments with the protein scaffold of LH1 and LH2 restricts the chromophores' relative orientational and translational flexibility and contributes to their efficient energy transfer.[4,5] By contrast, the cpTMV surface modification site used herein positions chromophores on the disks' surface, which is helpful for achieving high modification levels to avoid gaps in the chromophore arrays but allows for orientational flexibility and limits direct protein–chromophore interactions.[28,29] Attaching chromophores to the cpTMV surface using more rigid linkers or a modification site in the cavity region between disks as done previously[28] would allow for the influence of chromophore constraint and chromophore–protein coupling on inter-assembly energy transfer to be investigated.[28,29]

3.4 Conclusion and Outlook

This chapter has demonstrated the production of a protein-based system for simulating energy transfer between photosynthetic light harvesting components. This tailorable platform mimics the photosynthetic environment in pigment orientation and positioning as well as providing the opportunity to study the effect of chromophore–protein interactions on energy transfer between light harvesting complexes. In order to construct this system, a method was developed for asymmetric, site-selective protein–protein conjugation employing non-canonical amino acids. Using this method, protein complexes were site-selectively labeled with a dye at a specific engineered cysteine residue and then coupled to an adjacent protein complex at the engineered non-canonical amino acid residues pAF and 3NY. This model system demonstrated energy transfer from an array of donor chromophores to an array of acceptor chromophores in a well-ordered conformation and solvent environment with an efficiency of 21%. Due to the versatility of the system, this efficiency may be improved by changing the protein modification site, dye linker composition, or dye identities to improve donor–acceptor spectral overlap and decrease donor–acceptor distances.

The installation of ncAAs into proteins via protein engineering and amber codon suppression is becoming increasingly accessible.[47,48] However, predicting which sites will be amenable to ncAA incorporation remains difficult.[49] In this work, of the three sites screened for ncAA incorporation, no incorporation of 3NY was observed at the native Y64 position, despite the structural similarity between tyrosine and 3NY. In contrast, ncAA incorporation did occur at the two adjacent serine positions S63 and S65. Developing a predictive model

for ncAA incorporation, particularly of residues such as pAF and 3NY that can be used as selective handles for protein functionalization, is an interesting target for future work. Previous work in the Francis and Tullman-Ercek groups screened every natural amino acid point mutant of the MS2 virus capsid protein to determine which mutants were able to assemble into icosahedral capsids.[50] Creating a library of every point mutant to ncAAs of interest, such as pAF and 3NY, could be performed with a similar assembly state readout. Further correlating this landscape with parameters derived from protein simulations, such as the solvent-accessible surface area of individual amino acid residues within a protein, may provide guidelines for the incorporation of ncAAs into recombinantly expressed proteins.

The protein–protein conjugation developed herein is the first example of a protein containing an engineered pAF residue being directly, asymmetrically coupled to a protein containing an engineered 3NY residue, adding to the library of protein–protein conjugation strategies. This method may therefore be useful both for designing synthetic light harvesting systems of increasing complexity and other protein-based materials, such as well-controlled antibody-protein conjugates. In Chapter 4, the ability to incorporate the TMV-based light harvesting model into a supported lipid bilayer is demonstrated.[51] The TMV model has also recently been used to identify sources of disorder in biomimetic light harvesting systems and their effect on long-range energy transfer.[29] The ability to couple two TMV assemblies asymmetrically adds another layer of utility to the TMV-based model system by allowing energy transfer between donor and acceptor complexes at controlled distances and orientations to be examined.

3.5 Materials and Methods

General Methods. Unless otherwise noted, all chemicals and solvents were of analytical grade and received from commercial sources. Water (dd-H_2O) used in biological procedures and as reaction solvents was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) and MilliporeSigma (Burlington, MA). The pDule-para-aminoPhe (Addgene #85502),[36] pDule-3-nitroTyrosine (Addgene #85498),[52] and pBad-sfGFP 150TAG (Addgene #85483)[53] plasmids were gifts from Ryan Mehl. Amicon Ultra MWCO centrifugal concentrators were obtained from MilliporeSigma (St. Louis, MO).

General Instrumentation and Sample Analysis. UV-Vis absorption measurements were conducted on a Cary UV-Vis 100 spectrophotometer (Agilent, USA). Protein concentration was determined by UV/Vis analysis on a Nanodrop 1000 instrument (Nanodrop, USA) by monitoring absorbance at 280 nm. Electrospray LC/MS analysis of proteins and their bioconjugates was performed 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. Protein samples were run with a Proswift RP-4H column (Dionex, USA). Protein mass reconstruction was performed

on the charge ladder with Mass Hunter software (Agilent, USA). High Performance Liquid Chromatography (HPLC) was performed on Agilent 1200 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). Size exclusion chromatography (SEC) was performed using a Polysep-GFC-P-5000 column (4.6 x 250 mm) (Phenomenex, USA) at 1.0 mL/min using a mobile phase of 10 mM sodium phosphate buffer, pH 7.2. Dynamic Light Scattering (DLS) was performed on a Zetasizer Nano Series (Malvern Instruments, UK). Measurements were taken in triplicate at protein concentrations of 0.2–1.0 mg/mL in 10 mM sodium phosphate buffer, pH 7.2, at 25 °C.

Gel Analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in a Mini cell tank apparatus (Life Technologies, Carlsbad, CA), using NuPAGE^{TM} Novex^{TM} 4–12% Bis-Tris Protein Gels (Life Technologies). The sample and electrode buffers were prepared according to the suggestions of the manufacturer. All protein electrophoresis samples were heated for 5–10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure the reduction of disulfide bonds. Gels were run for 30 min at 200 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for the assignment of apparent molecular masses. Native agarose gel electrophoresis was performed using 0.9% agarose gels and 50 mM NaPhos buffer, pH 7.2. Samples were mixed with 80% glycerol at a ratio of 1:1 sample:glycerol and allowed to settle in wells for 10 min prior to applying voltage. Native gels were placed on ice and run for 8 h at 25 V to separate the bands. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad, Hercules, CA). Gel imaging was performed on a Gel Doc (Bio-Rad, Hercules, CA).

Transmission Electron Microscopy (TEM). Samples were prepared for TEM analysis using negative staining. Analyte solution (0.2–1.0 mg/mL cpTMV in 10 mM sodium phosphate buffer, pH 7.2) was applied to carbon-coated copper grids for 2 min, followed by rinsing in 4×10 μL droplets of a 1% aqueous solution of uranyl acetate. Grids were left in the final droplet for 1 min. TEM images were obtained at the Berkeley Electron Microscope Lab using an FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage.

Protein expression and purification. The production of *p*-amino-L-phenylalanine (pAF)- and 3-nitro-L-tyrosine (3NY)-containing proteins were performed according to previously published protocols.[38,39] Briefly, the pBAD-cpTMV-S65* vectors (with * designating the stop codon TAG) with either pDule-para-aminoPhe or pDule-3-nitroTyrosine were co-transformed into DH10B *E. coli* cells and plated on LB agar plates containing 50 μg/mL ampicillin and 12.5 μg/mL tetracycline. The resulting colonies were grown overnight in 10 mL LB containing 50 μg/mL ampicillin and 12.5 μg/mL tetracycline at 37 °C and then added to 1 L of arabinose auto-induction media. The 1 L culture was allowed to shake at 37 °C, 220 rpm until it reached an OD_{600} of 0.6–0.8. Then, 10 mL of a 100 mM solution of the non-canonical amino acid (pAF or 3NY) was added to the growth medium to a final concentration of 1 mM. The culture was incubated at 37 °C, 220 rpm for an additional 18 h. Cell

pellets were collected at 8,000 rpm for 30 min, after which the supernatant was discarded and the cell pellets were frozen at -20 °C until purification. After freezing, cell pellets were partially thawed and resuspended in 10 mL lysis buffer, 20 mM triethanolamine (TEA) pH 7.2. Cells were lysed by sonication with a 2 s on, 4 s off cycle for a total of 10 min using a standard disruptor horn at 60% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared at 14,000 rpm for 30 min. The supernatant was treated with 30–40% volume (3–4 mL) of saturated ammonium sulfate and allowed to rotate for 10 min at 4 °C to allow for complete protein precipitation. The precipitated protein was collected at 11,000 rpm for 30 min and resuspended in 10 mL lysis buffer. Complete re-dissolving of the protein and removal of residual ammonium sulfate was accomplished by performing dialysis in 1 L lysis buffer overnight with at least one buffer exchange. The resulting protein solution was treated with 5 μL benzonase (MilliporeSigma, St. Louis, MO) and 4 mg $MgCl_2$ at room temperature for 30 min before the solution was spun down at 10,000 rpm for 10 min. The resulting supernatant was filtered through a 0.22 μm filter and purified using a DEAE column with a 0–180 mM NaCl gradient elution in 20 mM TEA buffer, pH 7.2. The fractions containing cpTMV were further purified using a Sephacryl S-500 column in 10 mM sodium phosphate pH 7.2 elution buffer. Pure fractions were collected and concentrated using 100 kDa MWCO centrifugal concentrators. Purity was confirmed by SDS-PAGE and ESI-TOF MS. Assembly state was confirmed by HPLC-SEC, DLS, and TEM. Protein was flash frozen and stored indefinitely at -80 °C or stored for no longer than 2 weeks at 4 °C without observed decomposition or change in assembly state.

Screen of protein expression conditions. The pBAD-cpTMV vectors containing the relevant non-canonical amino acid amber codon (or pBAD-sfGFP N150TAG for GFP controls) were co-transformed with pDule-3-nitroTyrosine into DH10B *E. coli* cells and plated on LB agar plates containing 50 μg/mL ampicillin and 12.5 μg/mL tetracycline. The resulting colonies were grown overnight in 10 mL LB containing 50 μg/mL ampicillin and 12.5 μg/mL tetracycline at 37 °C, and 25 μL of the overnight culture was then added to 5 mL of arabinose auto-induction media[38] with modifications as shown in Figure 3.2a. The cultures were then shaken at 37 °C, 220 rpm. At the times specified in Figure 3.2a, 250 μL of a 20 mM solution of the non-canonical amino acid (3NY) was added to the growth medium to a final concentration of 1 mM. The culture was incubated at 37 °C, 220 rpm for a total of 23 h. 500 μL culture from each expression was then centrifuged at 13,200 rpm for 5 min. The supernatant was discarded, and the pellet was resuspended in 250 μL Bugbuster Protein Extraction Reagent (MilliporeSigma, St. Louis, MO) and allowed to sit at room temperature for 20 min. The solution was then centrifuged for 20 min at 13,200 rpm, the supernatant was removed and retained, and the insoluble portion was resuspended in lysis buffer. The expression level in the supernatant and resuspended insoluble portions were analyzed using gel electrophoresis.

General procedure for labeling cpTMV thiols with maleimide dyes. The following procedure is based on a previously reported procedure with minimal modifications.[28] To 100

μL of cpTMV (100 μM in 10 mM sodium phosphate buffer, pH 7.2) was added 5 equiv. of TCEP to ensure complete thiol reduction prior to chromophore modification. The mixture was briefly vortexed and allowed to sit at room temperature for 20 min. Excess TCEP was removed with a NAP-5 Sephadex G-25 column (GE Healthcare, USA), followed by spin concentration of the eluent in 100 kDa MWCO cutoff filters. To 100 μL of the reduced cpTMV (100 μM in 10 mM sodium phosphate buffer, pH 7.2) was added 0.01 equiv. of maleimide-functionalized chromophore for single modification or 5 equiv. of functionalized chromophore for full modification. The reaction mixture was briefly vortexed and then incubated in 1.5 mL Eppendorf tubes at room temperature with an aluminum foil cover. After 2 h, the crude reactions were purified with a NAP-5 Sephadex G-25 column, followed by spin concentration of the eluent in 100 kDa MWCO concentrators, and an additional NAP-5 Sephadex G-25 column to remove excess chromophores. Further spin concentration was achieved using 100 kDa MWCO concentrators. The protein conjugates were analyzed with MS to gauge extent of modification and HPLC-SEC for assessment of purity and validation of assembly state. Protein conjugates were wrapped in foil and stored at 4 °C for several days or flash frozen and stored at -80 °C for longer periods.

General procedure for labeling pAF-containing cpTMV with aminophenols or catechols. To 96 μL of 20 μM pAF-containing cpTMV in 100 mM sodium phosphate buffer, pH 6.5 was added 5 equiv. of amino-*p*-cresol or 4-methylcatechol (2 μL of 5 mM aminophenol or catechol in 1:10 ACN:100 mM sodium phosphate buffer, pH 6.5) and 5 equiv. of potassium ferricyanide (2 μL of 5 mM potassium ferricyanide in 100 mM sodium phosphate buffer, pH 6.5). The reaction was briefly vortexed and incubated at room temperature. After 2 h, the crude reactions were purified with a NAP-5 Sephadex G-25 column, followed by spin concentration of the eluent in 100 kDa MWCO cutoff filters. The protein conjugates were analyzed with MS to gauge extent of modification.

General procedure for reduction of 3NY-containing cpTMV. To 490 μL of 20 μM 3NY-containing cpTMV in 100 mM sodium phosphate buffer, pH 6.5 was added 200 equiv. of sodium dithionite (10 μL of 200 mM sodium dithionite in 100 mM sodium phosphate buffer, pH 6.5, always prepared immediately prior to use). 800 equiv. of sodium dithionite was used for dye-labeled, 3NY-containing cpTMV. The mixture was briefly vortexed and incubated at room temperature. After 40 min (or 2 h for dye-labeled, 3NY-containing cpTMV), the crude reactions were purified with a NAP-5 Sephadex G-25 column, followed by spin concentration of the eluent in 100 kDa MWCO cutoff filters, and an additional NAP-5 Sephadex G-25 column to remove excess reducing agent. The protein conjugates were analyzed with MS to gauge extent of reduction and HPLC-SEC for assessment of purity and validation of assembly state.

General procedure for labeling 3AY-containing cpTMV with *p*-toluidine. To 96 μL of 20 μM 3AY-containing cpTMV in 100 mM sodium phosphate buffer, pH 6.5 was added 5 equiv. of *p*-toluidine (2 μL of 5 mM *p*-toluidine in 1:10 ACN:100 mM sodium phosphate

buffer, pH 6.5) and 5 equiv. of potassium ferricyanide (2 μL of 5 mM potassium ferricyanide in 100 mM sodium phosphate buffer, pH 6.5). The reaction was briefly vortexed and incubated at room temperature. After 2 h, the crude reactions were purified with a NAP-5 Sephadex G-25 column, followed by spin concentration of the eluent in 100 kDa MWCO cutoff filters. The protein conjugates were analyzed with MS to gauge extent of modification.

Procedure for oxidatively coupling cpTMV-S65-pAF to cpTMV-S65-3AY. To 98 μL of 20 μM pAF-containing cpTMV and 10 μM 3AY-containing cpTMV in 100 mM sodium phosphate buffer, pH 6.5 was added 5 equiv. of potassium ferricyanide (2 μL of 5 mM potassium ferricyanide in 100 mM sodium phosphate buffer, pH 6.5). The reaction was briefly vortexed and incubated at room temperature. After 2 h, the crude reactions were purified with a NAP-5 Sephadex G-25 column, followed by spin concentration of the eluent in 100 kDa MWCO cutoff filters. Separation of the coupled product from uncoupled starting material was performed using a Polysep-GFC-P-5000 column (4.6 × 250 mm). The protein conjugates were analyzed with MS to gauge extent of modification and HPLC-SEC, DLS, and TEM for assessment of assembly state.

Spectroscopic measurements. Samples in buffer were diluted to 1.5 μM Oregon Green 488 using UV-Vis absorption prior to fluorescence measurements. Fluorescence emission and lifetime spectra were collected using a PicoQuant FluoTime FT-300 fluorometer. The samples were transferred to a 1 cm path-length quartz cuvette and excited with a 465 nm PicoQuant pulsed diode laser, with an instrument response function of 150 ps, as measured with a scattering LUDOX sample. Time-resolved emission measurements were performed via time-correlated single photon counting (TCSPC). The lifetime values result from mono- or biexponential reconvolution fitting using PicoQuant FluoFit software version 4.6.6.0, with $X^2 < 1.1$ for all measurements.

FRET and efficiency calculations. The inter-chromophore distances in this system are expected to be over 7 nm for donor–acceptor pairs and ∼1.95 nm for identical chromophores within a single cpTMV disk. Due to these distances, any energy transfer between chromophores in this system would be due to Förster resonance energy transfer (FRET). To calculate the rate of energy transfer between donor and acceptor disks, the expression:

$$\frac{1}{\tau_{DA}} = \frac{1}{\tau_D} + \frac{1}{\tau_T}$$

was used, where τ_{DA} is the fluorescence lifetime of the donor disks when coupled to acceptor disks, τ_D is the fluorescence lifetime of the donor disks in the absence of acceptor disks, and τ_T is the timescale of energy transfer between donors and acceptors.[54] The rate of energy transfer k_T is the inverse of τ_T:

$$k_T = \frac{1}{\tau_T}$$

To calculate the efficiency of energy transfer between donor and acceptor disks, the following expression was used:

$$E = 1 - \frac{\tau_{DA}}{\tau_D}$$

where E indicates energy transfer efficiency.[55]

Inter-chromophore distances were measured using the tape measure tool in ChimeraX, version 1.2.5. To calculate the proportion that each donor–acceptor pair shown in Figure 3.8a would be expected to contribute to energy transfer based solely on inter-chromophore distance, the following expression was used:

$$\text{Estimated \% contribution} = 100 \cdot \frac{1/R_i^6}{\sum 1/R_i^6}$$

where R_i is the distance between a single donor–acceptor pair.

3.6 Additional Figures

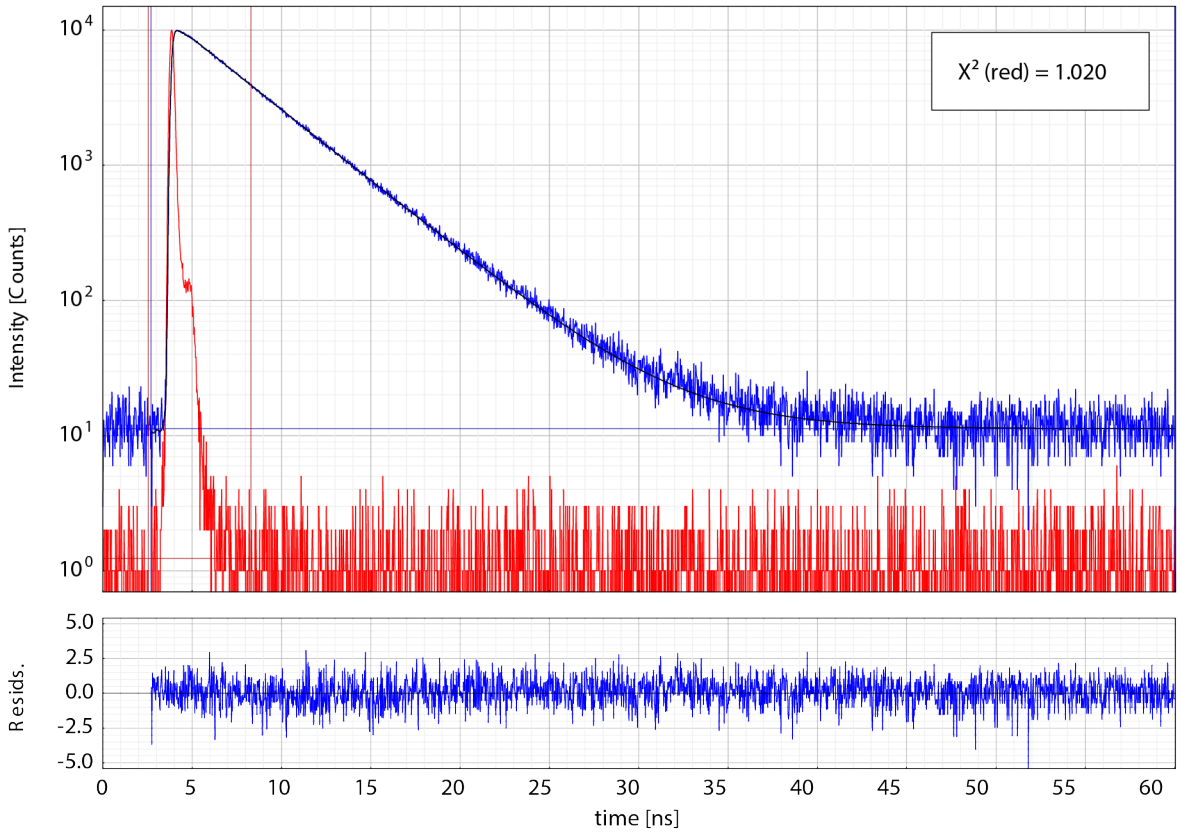


Figure 3.9: Fluorescence emission over time of sample **1**. Emission at 524 nm after excitation at 465 nm of Oregon Green 488 capped with Mesna (**1**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a monoexponential fit as shown by the fit residuals.

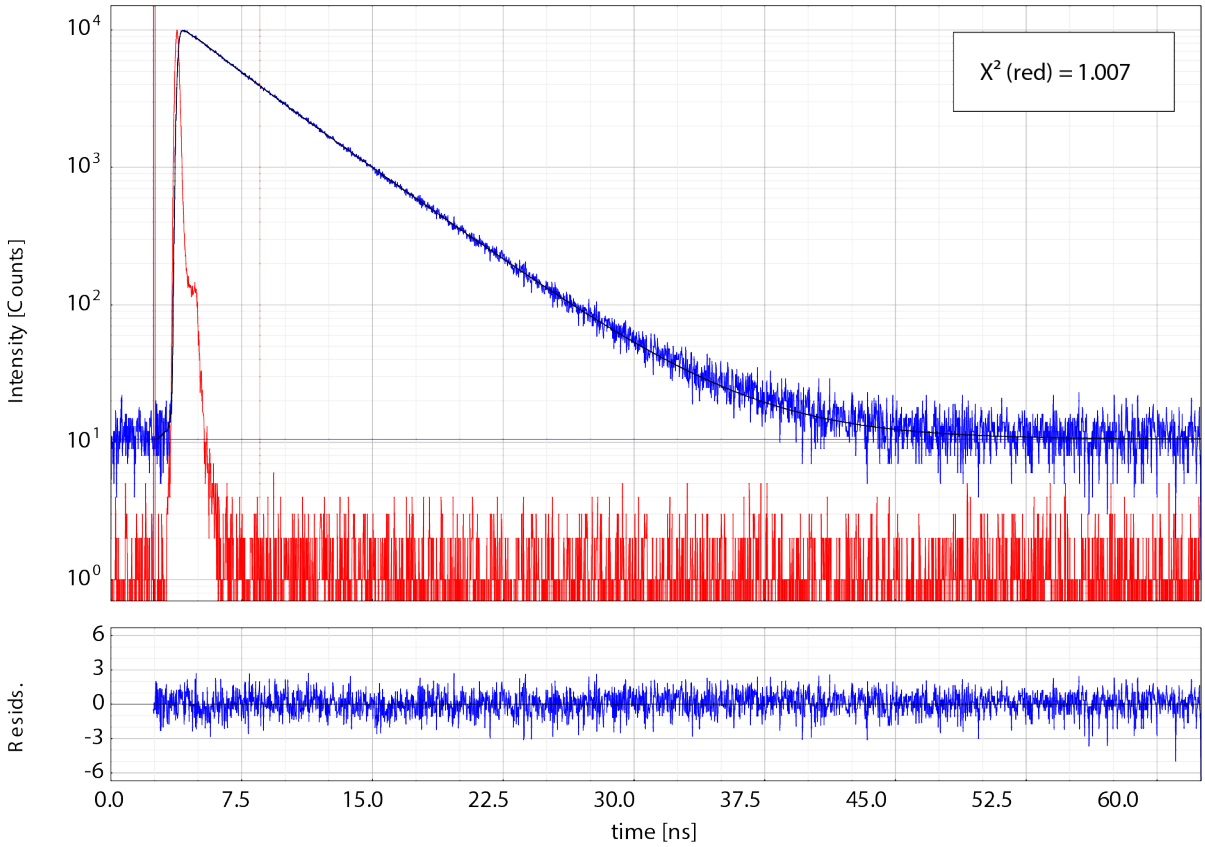


Figure 3.10: Fluorescence emission over time of sample **2**. Emission at 524 nm after excitation at 465 nm of cpTMV-S23C-S65-pAF disk assemblies labeled with Oregon Green 488 at a ratio of one dye per disk (**2**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a biexponential fit as shown by the fit residuals.

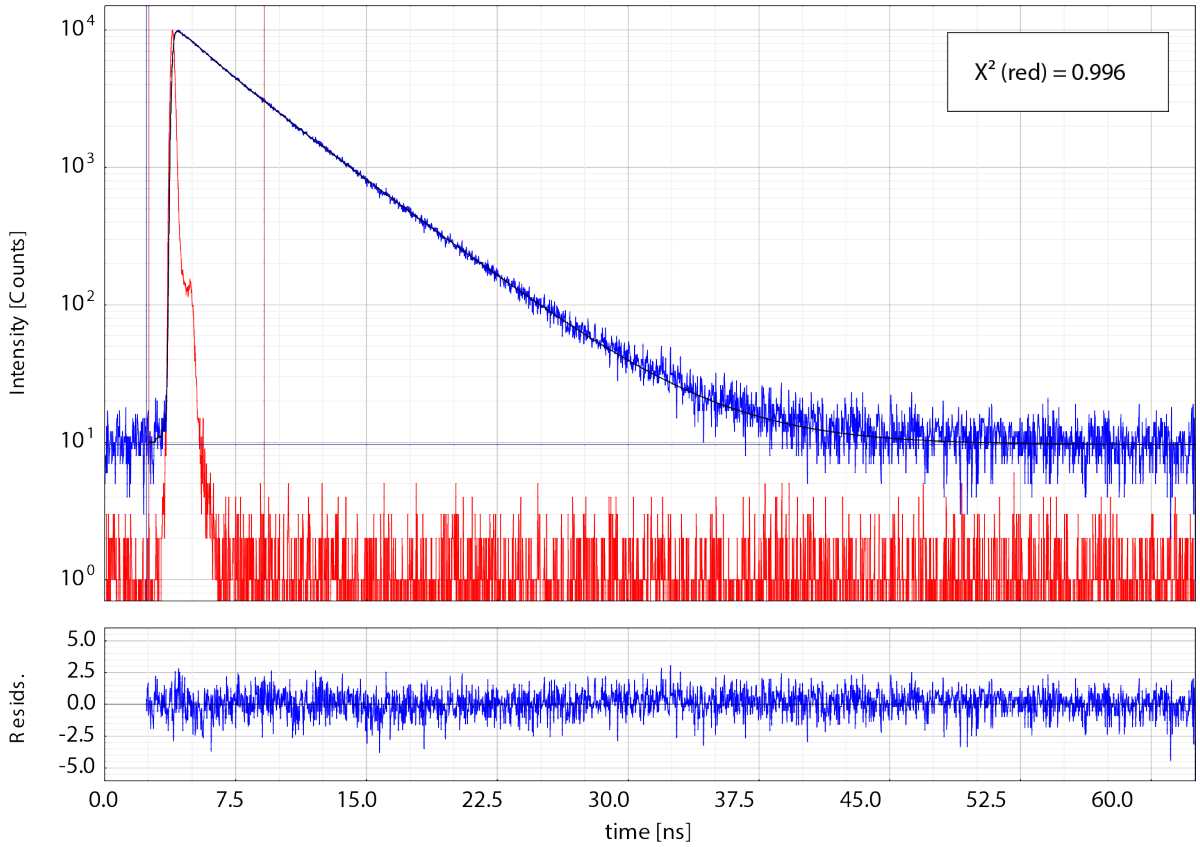


Figure 3.11: Fluorescence emission over time of sample **3**. Emission at 524 nm after excitation at 465 nm of cpTMV-S23C-S65-pAF disk assemblies quantitatively labeled with Oregon Green 488 (**3**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a biexponential fit as shown by the fit residuals.

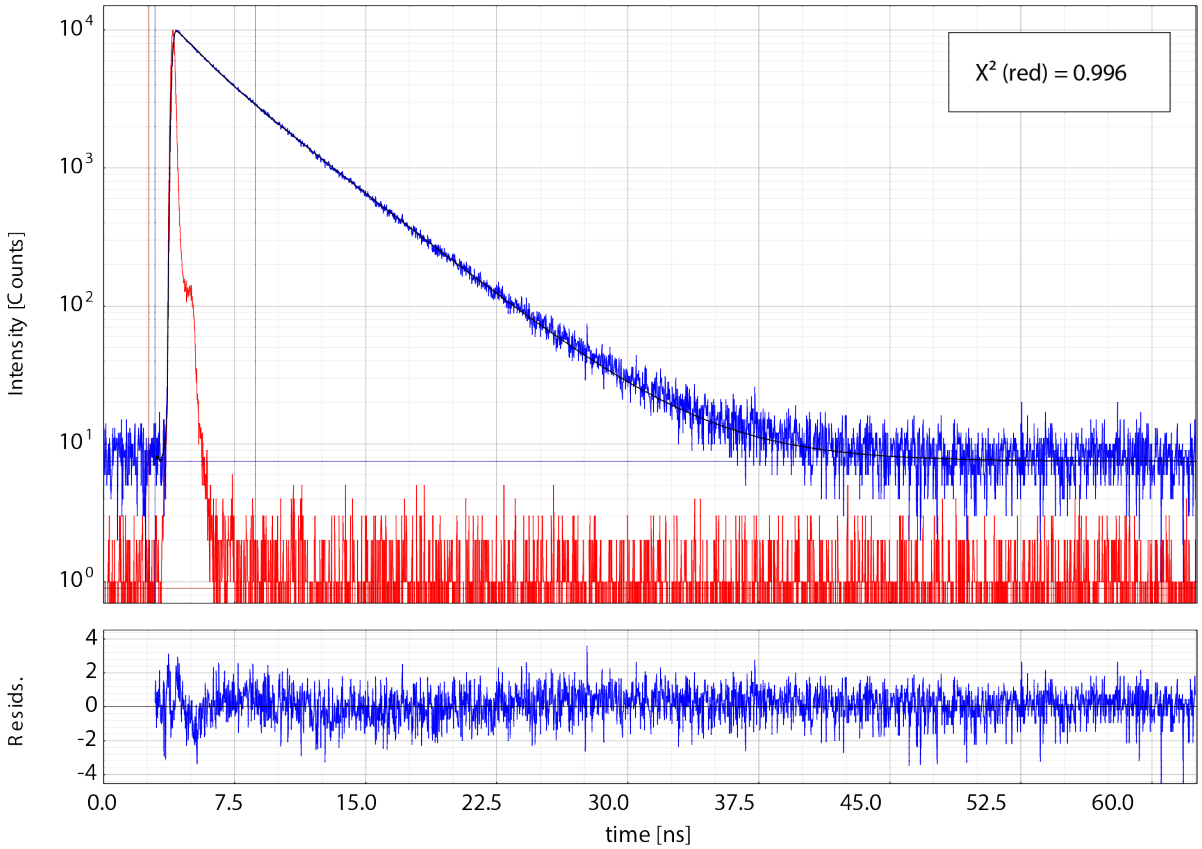


Figure 3.12: Fluorescence emission over time of sample **4**. Emission at 524 nm after excitation at 465 nm of cpTMV-S23C-S65-pAF disk assemblies quantitatively labeled with Oregon Green 488 as well as 4-methylcatechol at the S65-pAF position of each monomer (**4**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a biexponential fit as shown by the fit residuals.

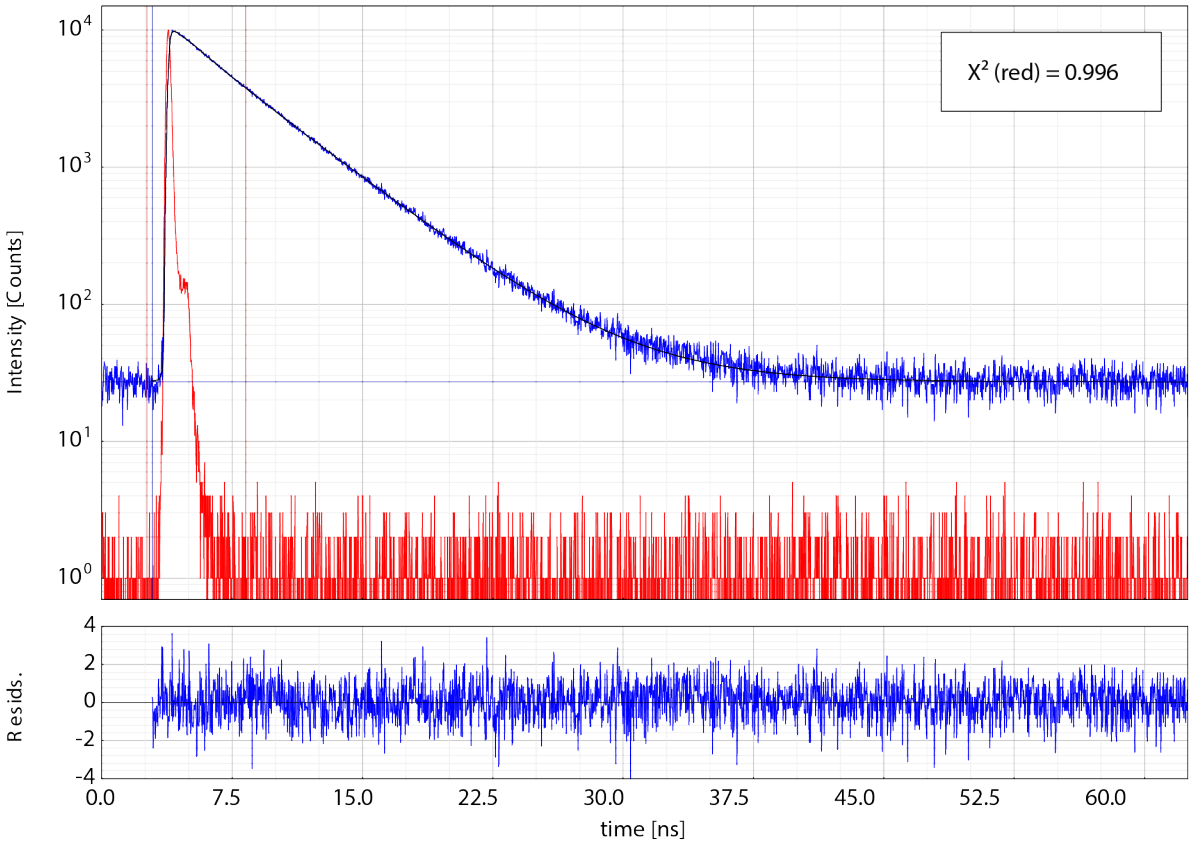


Figure 3.13: Fluorescence emission over time of sample **5**. Emission at 524 nm after excitation at 465 nm of cpTMV-S23C-S65-pAF disk assemblies quantitatively labeled with Oregon Green 488 and coupled to cpTMV-S65-3AY disks without any chromophore labeling (**5**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a biexponential fit as shown by the fit residuals.

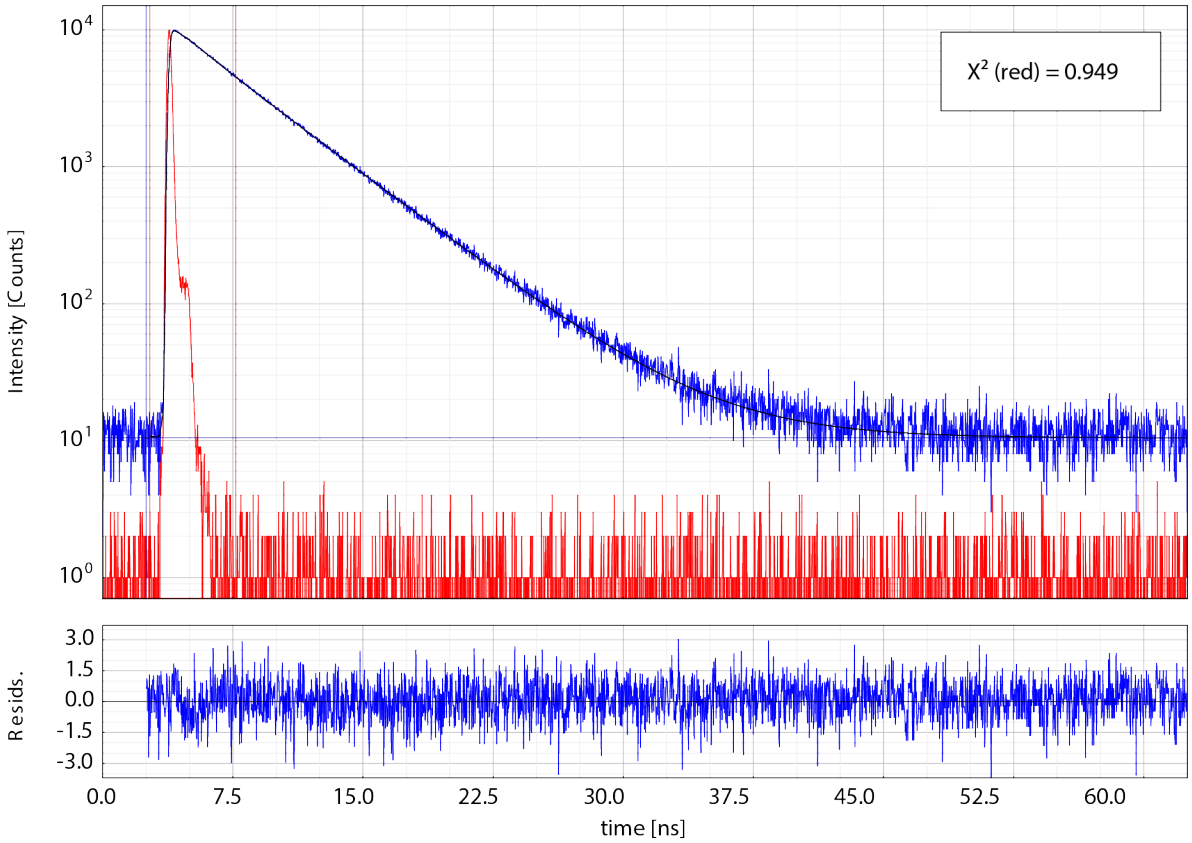


Figure 3.14: Fluorescence emission over time of sample **6**. Emission at 524 nm after excitation at 465 nm of cpTMV-S23C-S65-pAF disk assemblies quantitatively labeled with Oregon Green 488 and mixed, but not conjugated, with cpTMV-S23C-S65-3AY disks quantitatively labeled with Alexa Fluor 594 (**6**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a biexponential fit as shown by the fit residuals.

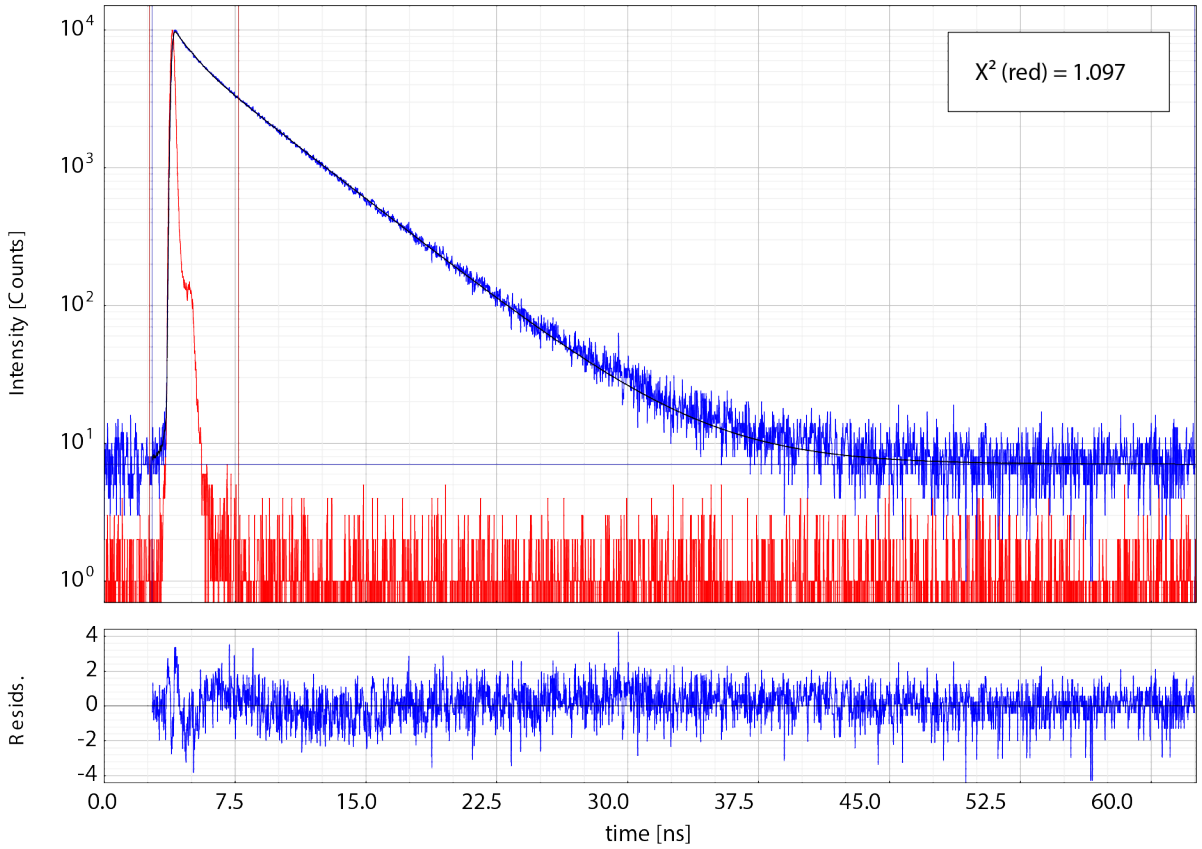


Figure 3.15: Fluorescence emission over time of sample **7**. Emission at 524 nm after excitation at 465 nm of cpTMV-S23C-S65-pAF disk assemblies quantitatively labeled with Oregon Green 488 and coupled to cpTMV-S23C-S65-3AY disks quantitatively labeled with Alexa Fluor 594 (**7**) is shown in blue and the instrument response function in red. The average fluorescence lifetime is well-described by a biexponential fit as shown by the fit residuals.

3.7 Acknowledgements

Thank you to Amanda Li for help with protein production and review of this work. Thank you to Leo Hamerlynck and to Dr. Trevor Roberts for conducting spectroscopic experiments, and input into experimental design and analysis. Thank you to Prof. Naomi Ginsberg for input into experimental design, interpretation of spectroscopic data, and careful review of this work. Thank you to the the staff at the University of California, Berkeley, Electron Microscope Laboratory for their advice and assistance in electron microscopy sample preparation and data collection.

3.8 References

(1) MacGregor-Chatwin, C.; Jackson, P. J.; Sener, M.; Chidgey, J. W.; Hitchcock, A.; Qian, P.; Mayneord, G. E.; Johnson, M. P.; Luthey-Schulten, Z.; Dickman, M. J.; Scanlan, D. J.; Hunter, C. N. *Nature Plants* **2019**, *5*, 879–889.

(2) Dostál, J.; Pšenčík, J.; Zigmantas, D. *Nature Chemistry* **2016**, *8*, 705–710.

(3) Scheuring Simon; Sturgis James N. *Science* **2005**, *309*, 484–487.

(4) Freer, A.; Prince, S.; Sauer, K.; Papiz, M.; Lawless, A. H.; McDermott, G.; Cogdell, R.; Isaacs, N. W. *Structure* **1996**, *4*, 449–462.

(5) Cao, P.; Bracun, L.; Yamagata, A.; Christianson, B. M.; Negami, T.; Zou, B.; Terada, T.; Canniffe, D. P.; Shirouzu, M.; Li, M.; Liu, L.-N. *Nature Communications* **2022**, *13*, 1977.

(6) Qian, P.; Swainsbury, D. J. K.; Croll, T. I.; Castro-Hartmann, P.; Divitini, G.; Sader, K.; Hunter, C. N. *Biochemistry* **2021**, *60*, 3302–3314.

(7) Hashimoto, H.; Uragami, C.; Cogdell, R. J. In *Carotenoids in Nature: Biosynthesis, Regulation and Function*, Stange, C., Ed.; Springer International Publishing: Cham, 2016, pp 111–139.

(8) Sundström, V.; Pullerits, T.; van Grondelle, R. *The Journal of Physical Chemistry B* **1999**, *103*, 2327–2346.

(9) Nagarajan, V.; Parson, W. W. *Biochemistry* **1997**, *36*, 2300–2306.

(10) Hess, S.; Chachisvilis, M.; Timpmann, K.; Jones, M. R.; Fowler, G. J.; Hunter, C. N.; Sundström, V. *Proceedings of the National Academy of Sciences* **1995**, *92*, 12333–12337.

(11) Kramer, T.; Rodriguez, M. *Scientific Reports* **2017**, *7*, 45245.

(12) Kriete, B.; Lüttig, J.; Kunsel, T.; Malý, P.; Jansen, T. L. C.; Knoester, J.; Brixner, T.; Pshenichnikov, M. S. *Nature Communications* **2019**, *10*, 4615.

(13) Xiao, T.; Shen, Y.; Bao, C.; Diao, K.; Ren, D.; Qian, H.; Zhang, L. *RSC Advances* **2021**, *11*, 30041–30045.

(14) Zhang, D.; Yu, W.; Li, S.; Xia, Y.; Li, X.; Li, Y.; Yi, T. *Journal of the American Chemical Society* **2021**, *143*, 1313–1317.

(15) Gong, C.; Sun, S.; Zhang, Y.; Sun, L.; Su, Z.; Wu, A.; Wei, G. *Nanoscale* **2019**, *11*, 4147–4182.

(16) Spillmann, C. M.; Medintz, I. L. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews* **2015**, *23*, 1–24.

(17) Liu, K.; Xing, R.; Li, Y.; Zou, Q.; Möhwald, H.; Yan, X. *Angewandte Chemie International Edition* **2016**, *55*, 12503–12507.

(18) Han, J.; Liu, K.; Chang, R.; Zhao, L.; Yan, X. *Angewandte Chemie International Edition* **2019**, *58*, 2000–2004.

(19) Mikkilä, J.; Anaya-Plaza, E.; Liljeström, V.; Caston, J. R.; Torres, T.; de la Escosura, A.; Kostiainen, M. A. *ACS Nano* **2016**, *10*, 1565–1571.

(20) Petti, M. K.; Lomont, J. P.; Maj, M.; Zanni, M. T. *The Journal of Physical Chemistry B* **2018**, *122*, 1771–1780.

(21) Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. *Nano Letters* **2010**, *10*, 181–186.

(22) Bischoff, A. J.; Harper, C. C.; Williams, E. R.; Francis, M. B. *Journal of the American Chemical Society* **2022**, *144*, 23368–23378.

(23) Endo, M.; Fujitsuka, M.; Majima, T. *Chemistry – A European Journal* **2007**, *13*, 8660–8666.

(24) Dai, J.; Knott, G. J.; Fu, W.; Lin, T. W.; Furst, A. L.; Britt, R. D.; Francis, M. B. *ACS Nano* **2021**, *15*, 8110–8119.

(25) Miller, R. A.; Presley, A. D.; Francis, M. B. *Journal of the American Chemical Society* **2007**, *129*, 3104–3109.

(26) Lee, J.; Lee, D.; Kocherzhenko, A. A.; Greenman, L.; Finley, D. T.; Francis, M. B.; Whaley, K. B. *The Journal of Physical Chemistry B* **2018**, *122*, 12292–12301.

(27) Noriega, R.; Finley, D. T.; Haberstroh, J.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *The Journal of Physical Chemistry B* **2015**, *119*, 6963–6973.

(28) Delor, M.; Dai, J.; Roberts, T. D.; Rogers, J. R.; Hamed, S. M.; Neaton, J. B.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *Journal of the American Chemical Society* **2018**, *140*, 6278–6287.

(29) Hamerlynck, L. M.; Bischoff, A. J.; Rogers, J. R.; Roberts, T. D.; Dai, J.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *The Journal of Physical Chemistry B* **2022**, *126*, 7981–7991.

(30) Carrico, Z. M.; Romanini, D. W.; Mehl, R. A.; Francis, M. B. *Chemical Communications* **2008**, 1205–1207.

(31) Sangsuwan, R.; Obermeyer, A. C.; Tachachartvanich, P.; Palaniappan, K. K.; Francis, M. B. *Chemical Communications* **2016**, *52*, 10036–10039.

(32) ElSohly, A. M.; Netirojjanakul, C.; Aanei, I. L.; Jager, A.; Bendall, S. C.; Farkas, M. E.; Nolan, G. P.; Francis, M. B. *Bioconjugate Chemistry* **2015**, *26*, 1590–1596.

(33) Zhang, J.; Wang, X.; Zhou, K.; Chen, G.; Wang, Q. *ACS Nano* **2018**, *12*, 1673–1679.

(34) Zhang, S.; Zhang, J.; Fang, W.; Zhang, Y.; Wang, Q.; Jin, J. *Nano Letters* **2018**, *18*, 6563–6569.

(35) Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A. *Nature Protocols* **2007**, *2*, 2601–2607.

(36) Mehl, R. A.; Anderson, J. C.; Santoro, S. W.; Wang, L.; Martin, A. B.; King, D. S.; Horn, D. M.; Schultz, P. G. *Journal of the American Chemical Society* **2003**, *125*, 935–939.

(37) Xie, J.; Schultz, P. G. *Nature Reviews Molecular Cell Biology* **2006**, *7*, 775–782.

(38) Ramsey, A. V.; Bischoff, A. J.; Francis, M. B. *Journal of the American Chemical Society* **2021**, *143*, 7342–7350.

(39) Miller, Z. M.; Harper, C. C.; Lee, H.; Bischoff, A. J.; Francis, M. B.; Schaffer, D. V.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **2022**, *33*, 2129–2137.

(40) Seim, K. L.; Obermeyer, A. C.; Francis, M. B. *Journal of the American Chemical Society* **2011**, *133*, 16970–16976.

(41) Behrens, C. R.; Hooker, J. M.; Obermeyer, A. C.; Romanini, D. W.; Katz, E. M.; Francis, M. B. *Journal of the American Chemical Society* **2011**, *133*, 16398–16401.

(42) Panchuk-Voloshina, N.; Haugland, R. P.; Bishop-Stewart, J.; Bhalgat, M. K.; Millard, P. J.; Mao, F.; Leung, W.-Y.; Haugland, R. P. *Journal of Histochemistry & Cytochemistry* **1999**, *47*, 1179–1188.

(43) Bunt, G.; Wouters, F. S. *Biophysical Reviews* **2017**, *9*, 119–129.

(44) Koushik, S. V.; Blank, P. S.; Vogel, S. S. *PLOS ONE* **2009**, *4*, e8031.

(45) Şener, M.; Strümpfer, J.; Timney, J. A.; Freiberg, A.; Hunter, C. N.; Schulten, K. *Biophysical Journal* **2010**, *99*, 67–75.

(46) Ye, J.; Sun, K.; Zhao, Y.; Yu, Y.; Kong Lee, C.; Cao, J. *The Journal of Chemical Physics* **2012**, *136*, 245104.

(47) Saleh, A. M.; Wilding, K. M.; Calve, S.; Bundy, B. C.; Kinzer-Ursem, T. L. *Journal of Biological Engineering* **2019**, *13*, 43.

(48) de la Torre, D.; Chin, J. W. *Nature Reviews Genetics* **2021**, *22*, 169–184.

(49) Hostetler, Z. M.; Ferrie, J. J.; Bornstein, M. R.; Sungwienwong, I.; Petersson, E. J.; Kohli, R. M. *ACS Chemical Biology* **2018**, *13*, 2855–2861.

(50) Hartman, E. C.; Jakobson, C. M.; Favor, A. H.; Lobba, M. J.; Álvarez-Benedicto, E.; Francis, M. B.; Tullman-Ercek, D. *Nature Communications* **2018**, *9*, 1385.

(51) Dai, J.; Wilhelm, K. B.; Bischoff, A. J.; Pereira, J. H.; Dedeo, M. T.; García-Almedina, D. M.; Adams, P. D.; Groves, J. T.; Francis, M. B. *Small* **2023**, *19*, 2207805.

(52) Cooley, R. B.; Feldman, J. L.; Driggers, C. M.; Bundy, T. A.; Stokes, A. L.; Karplus, P. A.; Mehl, R. A. *Biochemistry* **2014**, *53*, 1916–1924.

(53) Miyake-Stoner, S. J.; Refakis, C. A.; Hammill, J. T.; Lusic, H.; Hazen, J. L.; Deiters, A.; Mehl, R. A. *Biochemistry* **2010**, *49*, 1667–1677.

(54) Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer: New York, 2006.

(55) Kyrychenko, A.; Rodnin, M. V.; Ghatak, C.; Ladokhin, A. S. *Analytical Biochemistry* **2017**, *522*, 1–9.

Chapter 4

A membrane-associated light harvesting model is enabled by functionalized assemblies of gene-doubled TMV proteins

4.1 Abstract

Photosynthetic light harvesting requires efficient energy transfer within dynamic networks of light-harvesting complexes embedded within phospholipid membranes. Artificial light-harvesting models are valuable tools for understanding the structural features underpinning energy absorption and transfer within chromophore arrays. Herein a method is developed for attaching the TMV light-harvesting model to a supported lipid bilayer (SLB) to create a model system containing fully synthetic and tunable lipid, protein, chromophore, and solvent components. The protein model consists of tobacco mosaic viral capsid proteins (TMV) that are gene-doubled to create a tandem dimer (dTMV), which breaks the facial symmetry of the double disk to allow for differentiation between the disk faces. A single reactive lysine residue was incorporated into the dTMV assemblies for the site-selective attachment of chromophores for light absorption. On the opposing dTMV face, a cysteine residue was incorporated for the bioconjugation of a peptide containing a polyhistidine tag for association with SLBs. The dual-modified dTMV complexes showed significant association with SLBs and exhibited mobility on the bilayer. The techniques used herein offer a new method for protein–surface attachment and provide a platform for evaluating excited state energy transfer events in a dynamic, fully synthetic artificial light-harvesting system.

This chapter is based on the following publication:

Dai, J.*; Wilhelm, K. B.*; Bischoff, A. J.*; Pereira, J. H.; Dedeo, M. T.; García-Almedina, D. M.; Adams, P. D.; Groves, J. T.; Francis, M. B. A Membrane-Associated Light-Harvesting

Model is Enabled by Functionalized Assemblies of Gene-Doubled TMV Proteins. *Small* **2023**, *19*, 2207805.

This article can be accessed via the following link:
https://onlinelibrary.wiley.com/doi/full/10.1002/smll.202207805

4.2 Introduction

Numerous artificial light-harvesting systems have been developed in pursuit of more efficient photovoltaic and photoelectronic devices and as simplified model systems to probe the underlying structural and chemical features enabling energy transfer in photosynthetic organisms. Substantial effort has gone into creating synthetic light-harvesting mimics based on polymers,[1] dendrimers,[2] nucleic acids,[3] and proteins[4] emulating the pigment arrays, orientations, and environmental interactions present within natural light-harvesting complexes. These modular, parameterized systems provide insight into the role of pigment–protein interactions,[5] the defect tolerance of pigment arrays,[6] and quantum entanglement,[7] among many other features of energy transfer in photosynthesis.[8]

Most of the currently available artificial light-harvesting platforms are solution-phase, stand-alone vehicles lacking a lipid component. However, the lipid outer membrane of photosynthetic bacteria and thylakoid membrane of higher organisms, where most photosynthetic light-harvesting proteins reside, facilitate energy transfer by allowing reorganization of multi-complex assemblies in response to environmental stimuli. For example, the fluidity of bilayers serves a protective function by allowing light-harvesting complexes to aggregate and quench under high light conditions.[9] It is also theorized that the major energetic bottleneck of natural light-harvesting processes occurs during inter-assembly electronic energy transfer,[10] which is contingent upon dynamic, noncovalent associations between light-harvesting complexes within biological membranes.[8] The two-dimensional fluidic movement of light-harvesting complexes within lipid bilayers is a missing component in the vast majority of artificial light-harvesting systems. The only reported example of a membrane-associated artificial light-harvesting model is composed of small molecule chromophores noncovalently embedded in amphiphilic block copolymers,[11,12] and no protein-based systems are available for this purpose. An artificial model composed of membrane-associated, chromophore-bearing protein complexes would closely mimic the fluidity and organization of light-harvesting complexes within native organisms. The modularity of this synthetic model would allow us to probe how two-dimensional movement interacts with pigment composition, constraint, and orientation toward elucidating photosynthetic energy transfer mechanisms and pathways.

This chapter utilizes a novel TMV construct with greater control over the surface modification site. Previously engineered TMV complexes self-assemble from monomer units into C_2-symmetric stacks of 2- or 4-disk assemblies, preventing independent functionalization of the top and bottom disks (Figure 4.1a,b).[13,14] Previously unpublished work in the Francis group engineered a tandem dimer of TMV, which joined two copies of the recombinant TMV

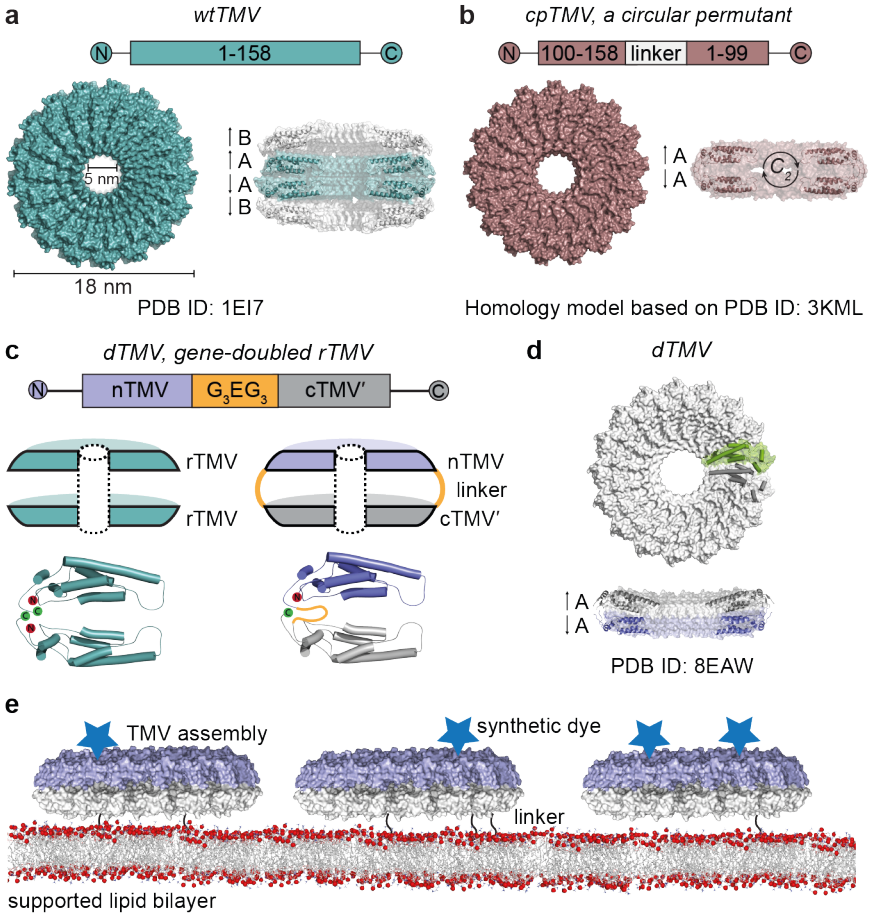


Figure 4.1: Designing synthetic light-harvesting membranes using a novel TMV construct. (a) The disk assembly state of wtTMV (PDB ID: 1EI7) assembles into C_2-symmetric, even-numbered stacks of disks with 17 monomers per disk. (b) cpTMV (PDB ID: 3KML) assembles into a C_2-symmetric, double-layered disk at a broad pH and ionic strength range. (c) A gene-doubled version of recombinant TMV (rTMV) with a 7 amino acid linker has been engineered with the C- and N-termini on the disk periphery. (d) The TMV tandem dimer (dTMV; PDB ID: 8EAW) assembles into a double-layered disk of 17 monomers per disk, which is not C_2-symmetric but has structurally distinct faces. (e) The complex process of photosynthetic energy transfer may be modeled via the covalent attachment of synthetic LHCs based on dTMV to a supported lipid bilayer. The use of synthetic dyes and a synthetic protein complex would allow for stability and tunability of the model system.

(rTMV) sequence with a seven amino acid linker with the sequence GGGEGGG between the C-terminus of one monomer and the N-terminus of another monomer (Figure 4.1c).[15] The two monomers in the tandem dimer, with the N-terminal unit labeled nTMV and the C-terminal unit labeled cTMV′, contained the same amino acid sequence encoded by different nucleotide sequences to ease genetic manipulation. A crystal structure of this tandem dimer, dTMV, was shown to assemble into a double-disk structure with 17 monomers per disk. Unlike wtTMV and cpTMV, assemblies of dTMV break the C_2 symmetry of previous assemblies to allow for differentiation between the top and bottom disk faces, with one con-

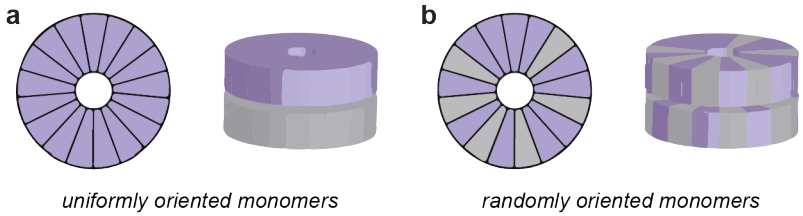


Figure 4.2: Potential orientation of tandem dimers within dTMV assemblies. (a) A schematic shows one possibility of tandem dimers arranged in a uniform orientation within double disk assemblies, with the purple section corresponding to the N-terminal domain of the tandem dimer and the gray section corresponding to the C-terminal domain. (b) A schematic shows a second possibility of tandem dimers arranged in a random orientation within double disk assemblies, with the purple section corresponding to the N-terminal domain and the gray portion corresponding to the C-terminal domain.

cave and one flat dTMV surface (Figure 4.1d).[16] The crystal structure clarified that within the double disk assembly, each tandem dimer of dTMV is positioned with one domain (N- or C-terminal) in the flat and one domain in the concave disk, rather than the complete tandem dimers lying side by side within a single disk. However, there was not sufficient resolution in the linker and N- and C-termini to determine whether the N-terminal domain of dTMV always resides in the same disk (either flat or concave; Figure 4.2a) or whether the N- and C-terminal domains are distributed randomly within the flat and concave disks (Figure 4.2b). The dTMV construct was engineered to install reactive residues on the protein surface, with one bioconjugation handle attached to a chromophore and a second handle for attachment to a fluid supported lipid bilayer (SLB).[17] The SLB provided a two-dimensional surface to which dTMV could bind at high density, while preserving free lateral diffusion on the surface to allow energy exchange between disks (Figure 4.1e).

4.3 Results and Discussion

4.3.1 Asymmetric dual-functionalization of dTMV allows dye-labeled double disks to conjugate to supported lipid bilayers

To begin the construction of a bilayer-associated dTMV model, mutations were selectively introduced into nTMV and cTMV′ to enable two complementary conjugation reactions. S123 was chosen as the bioconjugation site in both nTMV and cTMV′ because it is surface-exposed in both the flat and concave disks and is the site of previous successful chromophore conjugation in other TMV constructs.[18] Two constructs were created: (1) dTMV-S123C–S123′K, in which the cysteine was introduced in the N-terminal nTMV monomer domain and the lysine was introduced in the C-terminal cTMV′ monomer domain; and (2) dTMV-S123K–S123′C, with the opposite arrangement (Figure 4.3a). Aside from these engineered residues, no additional reactive cysteine or lysine residues were expected to be present on the dTMV surface. Both mutants expressed well, producing disks of the expected molecular weight (Figure 4.3b)

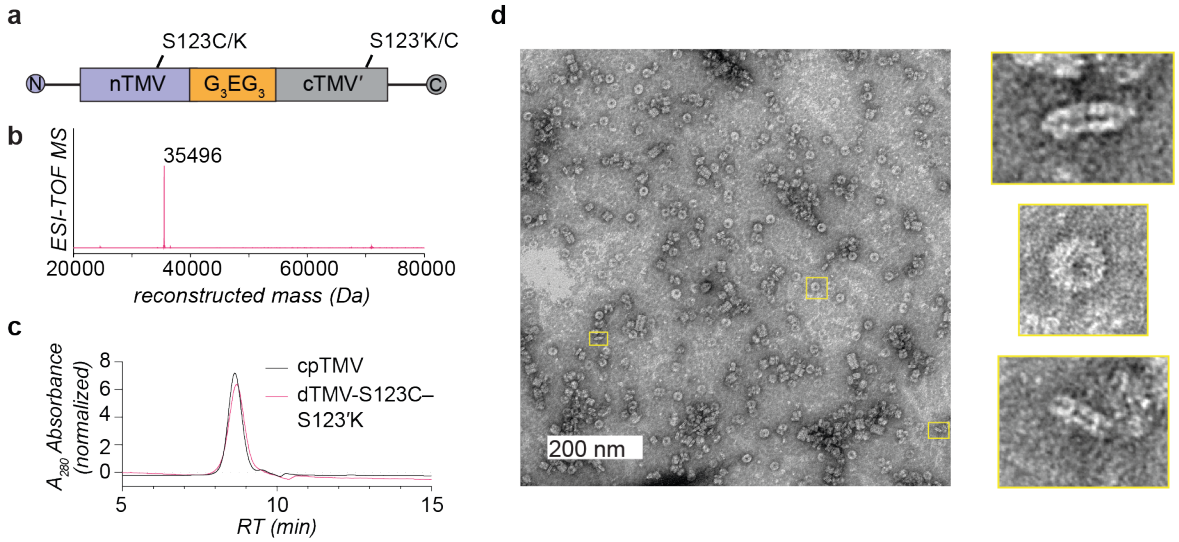


Figure 4.3: Characterization of the assembly state of dTMV. (a) The functionalizable amino acid residues lysine and cysteine were installed on each monomer of the dTMV fused dimer via protein engineering. (b) The mass of a purified dTMV-S123C–S123′K dimer (fused monomers) matches the expected mass of 35496 Da. (c) The assembly state of dTMV-S123C–S123′K was assessed by size exclusion chromatography which showed a size similar to the reported double disk structures of previous TMV constructs with a retention time of 8.5 min. (d) TEM images of dTMV-S123C–S123′K reveal a circular face of assembled capsids and stacks of two or multiples of two disks.

and forming assemblies that were analogous to those of the original dTMV sequence. The double disk morphology was confirmed by size exclusion chromatography (SEC) in comparison to cpTMV, which is known to form double disks[13] (Figure 4.3c), and by transmission electron microscopy (TEM) (Figure 4.3d). The TEM images revealed double disks of the expected diameter (18 nm), and with the characteristic central pore of TMV-based assemblies. Some short, non-helical stacks of disks were also observed in the images, which is consistent with other TMV assemblies.[13]

Disks composed of uniform constructs were incubated with Alexa Fluor 647 (AF647) *N*-hydroxysuccinimide (NHS) ester to achieve fluorescent labeling of the engineered lysine residue with 5–10% of monomers labeled (Figure 4.4a). In previous work, dye labeling of TMV constructs at engineered cysteine residues has been used at levels of 1% or close to 100% labeling to measure both isolated and networked dye excitation properties.[5,19] An intermediate modification level was used in this case to avoid potential off-target modification with the NHS ester dye[20] while ensuring nearly every dTMV disk would contain at least one dye for visualization. A construct without the reactive lysine residue, dTMV-S123–S123′C, showed no modification with AF647 under these conditions (Figure 4.5a). The fluorescently-labeled dTMV disks with an exposed cysteine residue were then incubated with an SLB containing lipids with maleimide head groups. These initial experiments resulted in dTMV sparsely adhering to SLBs, with only marginal increase in density over control bilayers lacking maleimide-functionalized lipids (Figure 4.6). Increasing dTMV incubation concentrations $>$ 5-fold did not significantly increase TMV association to the bilayer. Steric hindrance

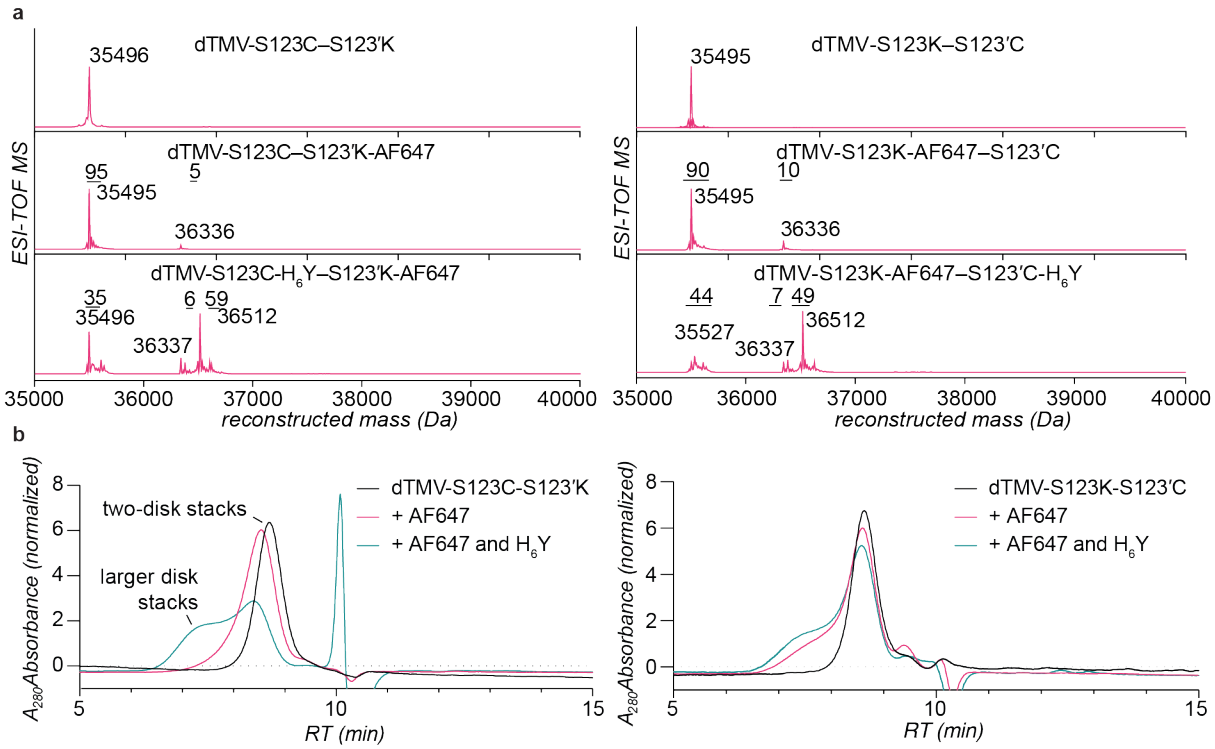


Figure 4.4: Modification of dTMV with Alexa Fluor 647 and H_6Y. (a) Mass data of dTMV-S123K–S123′C and dTMV-S123C–S123′K shows 5–10% modification with Alexa Fluor 647, and 50–60% modification with the H_6Y peptide. (b) Size exclusion chromatography of modified and unmodified dTMV shows a slight size increase upon modification with peptides, but little difference in assembly state between modified and unmodified proteins is indicated. Only fractions in the size range of two disk stacks were collected and used in subsequent experiments involving dTMV assembly association with SLBs.

may have prevented the relatively planar surface of the ∼600 kDa dTMV complex from conjugating to the planar SLB at S123(′)C through the desired linkage.

To address this issue, we sought to incorporate a short and flexible His-tag into the dTMV disks to allow their binding to SLBs containing nickel-chelating lipids. However, expression of TMV with an N-terminal His-tag results in altered assembly states and sensitivities to concentration and buffer,[21,22] requiring a different method for adding the His-tag to dTMV. The termini of dTMV are on the periphery of the disks, so the installation of a terminal His-tag could also result in an edge-on orientation of dTMV on the bilayers, rather than the desired face-on orientation. Covalent modification with a His-tag, as opposed to expression, would allow for control over the surface modification site as well as the number of His-tags per 17-monomer assembly.

To modify the existing dTMV constructs at the S123(′)C position with a His-tag, an oxidative coupling reaction was employed using the enzyme tyrosinase from *Agaricus bisporus* (abTyr) for the conjugation of phenol-containing peptides, small molecules, or proteins to cysteine residues on protein surfaces.[23–25] Short peptide sequences with a C-terminal tyrosine and an N-terminal His-tag were designed to conjugate to the dTMV double disks at S123(′)C

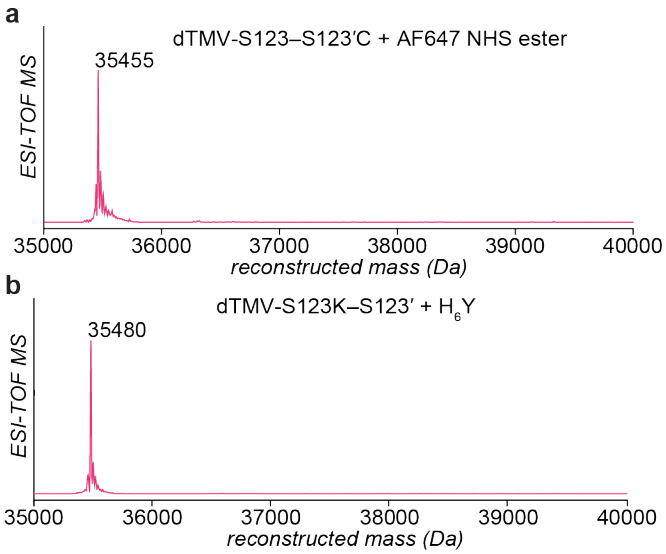


Figure 4.5: Modification site of dTMV with Alexa Fluor 647 and H_6Y. (a) Mass data of dTMV-S123–S123′C (MW: 35455 Da) incubated with Alexa Fluor 647 NHS ester (Expected MW: 36296 Da) shows no modification without a lysine residue engineered at position S123. (b) Mass data of dTMV-S123K–S123′ (MW: 35481 Da) incubated with H_6Y and abTyr (Expected MW: 36499 Da) shows no modification without a cysteine residue engineered at position S123′.

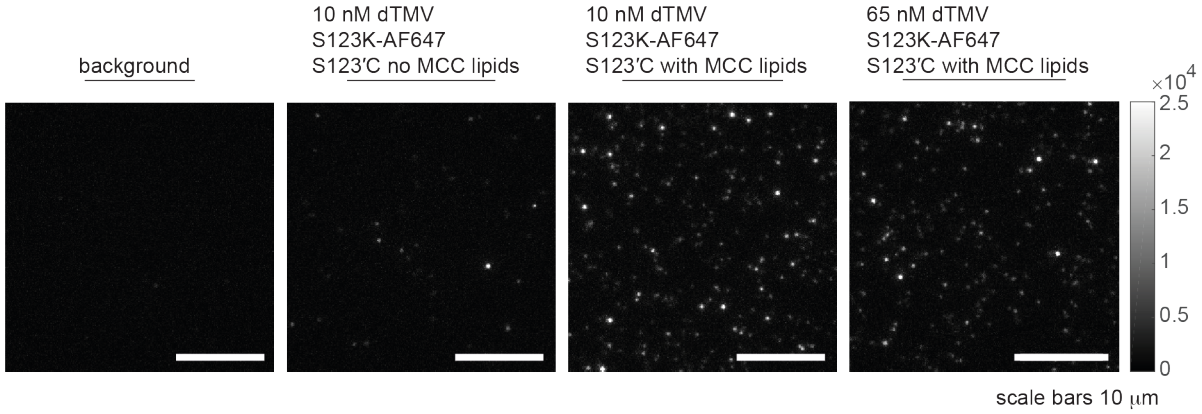


Figure 4.6: Association of dTMV containing an engineered cysteine residue with SLBs. A minor amount of non-specific adherence to the SLB was observed when dTMV-S123K-AF647–S123′C was incubated with an SLB containing no lipids with maleimide head groups. When dTMV-S123K-AF647–S123′C was incubated with maleimide-containing lipids at similar concentrations, the density of mobile complexes on the bilayer was not significantly higher than for the non-maleimide containing SLBs. A higher density was also not achieved by raising the dTMV incubation concentration.

and subsequently coordinate with nickel-nitrilotriacetic acid (Ni^{2+}-NTA) lipids in SLBs. abTyr-catalyzed modification of both dTMV-S123C–S123′K and dTMV-S123K–S123′C with a peptide with the sequence HHHHHHY (H_6Y) resulted in approximately 50% modification, or an average of 8–9 His-tags per double disk assembly, and did not appear to disrupt the assembly state of the double disks significantly (Figure 4.4, Figure 4.7a). Under the same conjugation conditions, no modification with H_6Y of a construct without the reactive engi-

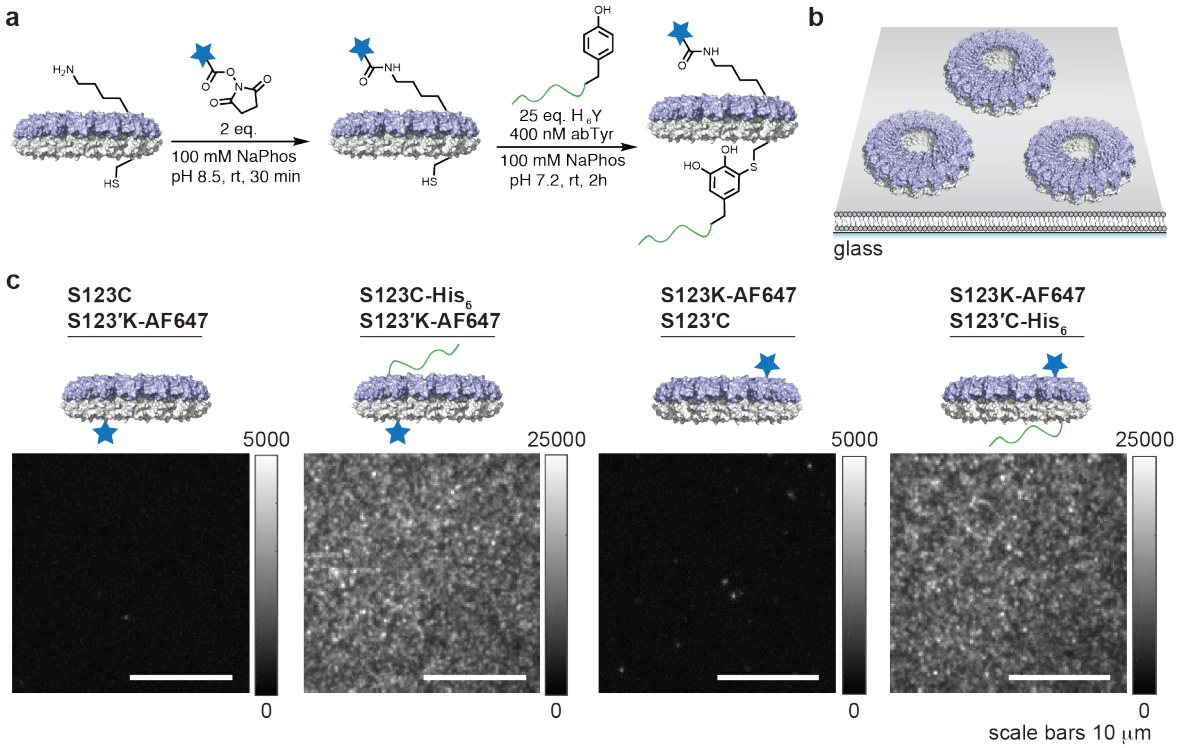


Figure 4.7: Attachment of dTMV to supported lipid bilayers via a non-expressed His-tag. (a) To each engineered dTMV assembly, Alexa Fluor 647 (AF647) was attached to the engineered Lys123 or Lys123′ residue using AF647 NHS ester, and the H_6Y peptide was attached to the Cys123 or Cys123′ residue on the opposite monomer using the enzyme tyrosinase (abTyr). AF647 was used for microscopic visualization, and the H_6Y peptide was used for attachment to nickel-chelating lipids. (b) A schematic shows dTMV assemblies on the surface of a supported lipid bilayer (SLB). (c) Total internal reflection fluorescence (TIRF) microscopy images show that without the H_6Y peptide conjugated to dTMV, neither the S123K-AF647–S123′C or S123C–S123′K-AF647 constructs significantly attached to SLBs containing Ni-NTA lipids. The AF647-labeled constructs functionalized with H_6Y, dTMV-S123C-H_6Y and dTMV-S123′C–H_6Y, attached to SLBs at similar densities when incubated on SLBs with similar concentrations and when modified with similar bioconjugation levels of the H_6Y peptide (59–60% of monomers modified).

neered cysteine residue, dTMV-S123K–S123′, was observed (Figure 4.5b). This modification proceeded similarly both before and after labeling with the AF647 NHS ester dye.

The dTMV-S123(′)K-AF647–S123(′)C-H_6Y modified double disks specifically conjugated to Ni^{2+}-NTA-containing SLBs through multivalent His_6:Ni^{2+}-NTA interactions (Figure 4.7b-c). Densities of 13 and 14 molecules per square micron (μm^{-2}) were reached for dTMV-S123C-H_6Y–S123′K-AF647 and dTMV-S123K-AF647–S123′C-H_6Y, respectively, with a modest 15 nM incubation for 1 h at room temperature in PBS. The maximum possible density for a single layer of dTMV double disk assemblies on an SLB would be ∼3000 μm^{-2}, indicating that the conditions used herein resulted in ∼1/200 of maximum coverage. Thin sheets of TMV disk assemblies linked directly at their peripheries have been synthesized, indicating that electrostatic repulsion does not prevent close packing of TMV assemblies.[26] Going forward, inter-disk energy transfer could be facilitated by increasing the dTMV density on the

SLB using higher incubation concentrations,[27] or by introducing favorable interactions on the edges of the disks.

The mobility of dTMV associated with the lipid bilayer was then examined using fluorescence recovery after photobleaching. A section of a bilayer labeled with dTMV-S123C-H_6Y–S123′K-AF647 was bleached, and the fluorescence recovery was monitored and fitted to yield an average diffusion coefficient of 0.13 ± 0.01 $\mu m^2 s^{-1}$ and a mobile fraction of 0.84 ± 0.02 (Figure 4.8a), consistent with the expected movement of a large protein complex on an SLB.[28,29] Single particle tracking also indicated that complexes diffused at varying rates, but that the diffusion rate of a single particle did not vary over time (Figure 4.8b-c). This indicates that the dTMV complexes associated with the Ni in the SLB via a varying number of His-tags, with a greater number of His-tags leading to less mobility on the SLB. The distribution of mobility of single particles is shown in Figure 4.8d, also showing a broad distribution of faster and slower diffusing particles. Analysis of the fluorescence intensity of single particles showed a mean intensity of 2.3 ± 1.5 (SD) chromophores per dTMV assembly (Figure 4.8e). The number of dyes per complex can also be determined by measuring discrete photobleaching steps of single particles, which resulted in an average of 2.1 ± 1.0 (SD) chromophores per dTMV assembly (Figure 4.8f-g). While this is slightly higher than the labeling efficiency calculated from the mass spectrometry data (Figure 4.4), disks containing no chromophores are not captured by measuring particle intensities, so these values are in good agreement. Mobility measurements of the dTMV with the reactive sites on the opposite domains (dTMV-S123K-AF647–S123′C-H_6Y) had similar mobility properties.

To demonstrate the versatility of this strategy, abTyr was also used to modify the surface-exposed cysteine residues of a midsize mammalian protein, bovine serum albumin (BSA), with the H_6Y peptide after labeling the exposed lysine residues of BSA with AF647 NHS ester. A solution concentration of 100 pM dye-labeled, His-tagged BSA yielded a density of 1.4 μm^{-2} on the SLB (Figure 4.9a-c), which is on par with other proteins containing a single His-tag.[27] This demonstrates the potential versatility of the tyrosinase-catalyzed peptide–protein coupling strategy and opens the possibility of attaching His-tags and other peptide sequences to proteins with buried N- and C-termini or for which expression with a peptide tag is not feasible or convenient.

The protein modification strategies used herein have applications extending beyond the design of synthetic light-harvesting models. The site-specific dual functionalization of proteins remains challenging but is extremely important for applications such as probing protein structure and function and developing precision therapeutics and diagnostic materials.[30] The ability to site-selectively conjugate a dye and peptide at specific canonical amino acid residues engineered on a protein surface adds to the toolbox of dual protein functionalization strategies. The covalent attachment of a His-tag to a protein surface site also has situational benefits over its conventional introduction as an N- or C-terminal tag during recombinant protein expression. Alongside protein purification, His-tags are widely used for protein attachment to surfaces.[31] However, genetically engineered His-tag fusion proteins can disrupt protein expression, folding, and enzymatic activity.[32–35] For proteins with buried N- and C-termini or those isolated from native organisms, expression with a His-tag may also be

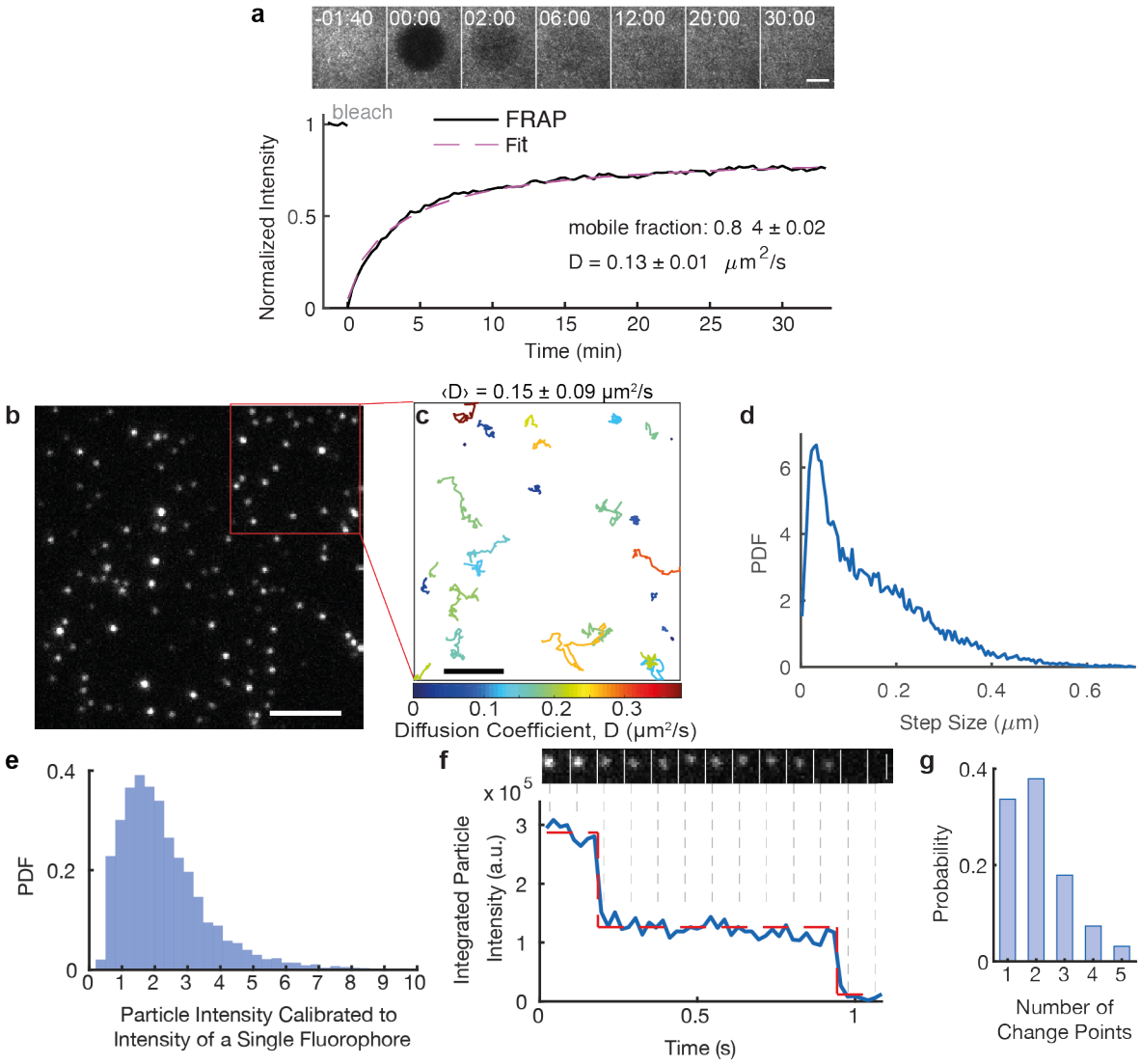


Figure 4.8: dTMV-S123C-H_6Y–S123′K-AF647 is mobile on supported lipid bilayers, as visualized by total internal reflection fluorescence microscopy. (a) Fluorescence recovery after photobleaching of a high density (20 μm^{-2}) of dTMV on the bilayer indicates that 84% of the dTMV disks on the bilayer were mobile, and that the mobile fraction diffused at 0.13 ± 0.01 $\mu m^2 s^{-1}$ on average. Scale bar 10 μm. Error denotes 95% CI. (b) A low density (0.2 μm^{-2}) of disks on the bilayer enabled analysis of single disk particles. Scale bar 5 μm. (c) Tracks of particles depicted in (b) diffusing through time illustrate the varied diffusion rates of particles on the bilayer. Tracks are colored by their diffusion coefficient, D. Scale bar 2 μm. Error denotes standard deviation. (d) The step size distribution of diffusing disks contained contributions from slowly-diffusing particles at smaller step sizes and quickly-diffusing particles at larger step sizes. n = 68,805 steps, n = 2,281 trajectories. (e) The background-subtracted fluorescence intensity distribution of all disks on the bilayer was calibrated to the mean integrated intensity of disks labeled with a single fluorophore. n = 5,934 particles. (f) A representative particle photobleached in two steps, indicating that two fluorophores were conjugated to that particle. Scale bar 1 μm. (g) The step photobleaching traces of particles that bleach completely over the course of the imaging acquisition were analyzed for the number of change points in their fluorescence intensity. n = 96 particles.

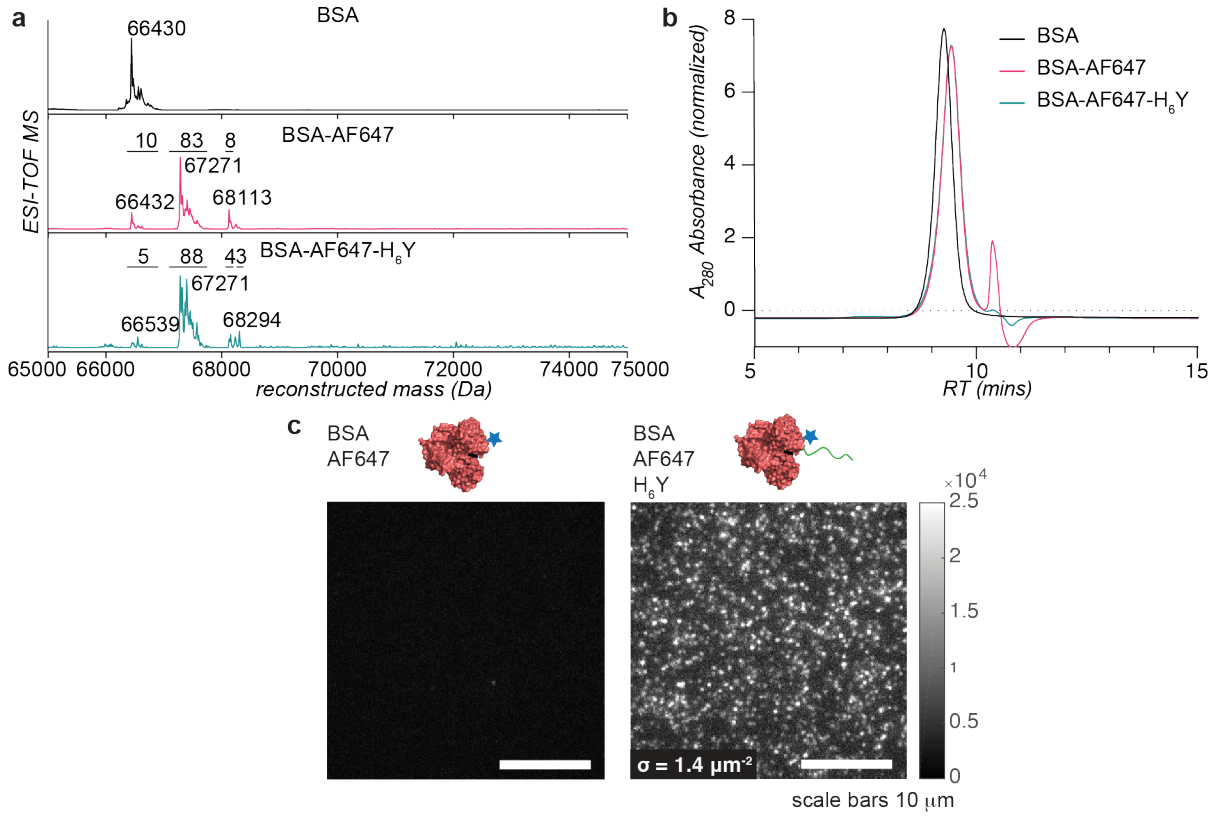


Figure 4.9: Bovine serine albumin (BSA) conjugated to a His-tag and attached to supported lipid bilayers (SLBs). (a) Labeling of BSA with Alexa Fluor 647 (AF647) NHS ester resulted in 83% of proteins being labeled with a single dye and 8% being labeled with two dyes. Addition of the H_6Y peptide to the AF647-labeled BSA resulted in approximately 3% of BSA monomers being labeled with a single H_6Y peptide. (b) Size exclusion chromatography of unmodified and modified BSA shows a slight size decrease upon labeling with both dye and peptide. This may occur because BSA had dimerized pre-labeling, and labeling disrupted interactions at the homodimer interface. (c) Incubation of BSA on a supported lipid bilayer containing lipids with Ni-NTA headgroups resulted in sparse labeling in the sample without a His-tag and denser bilayer labeling in the sample with BSA conjugated to the H_6Y peptide, similar to the dTMV assembly results.

impractical. In contrast, introducing a single cysteine mutation at a convenient point on the protein surface as demonstrated herein or using a native surface-exposed cysteine can allow for post-expressional addition of a His-tag for downstream applications. This expands the library of proteins that can be associated with Ni^{2+}-NTA-containing supported lipid bilayers and other surfaces.

4.3.2 Exploring the orientations of the nTMV and cTMV(′) domains of dTMV

The fluorescence images of dTMV-S123(′)C-H_6Y–S123(′)K-AF647 on SLBs (Figure 4.7) did not provide clarity regarding the relative orientations of tandem dimers within dTMV assem-

blies (Figure 4.2). If similar incubation concentrations of fluorescently labeled dTMV-S123C-H_6Y–S123′K-AF647 and dTMV–S123K-AF647–S123′C-H_6Y on Ni^{2+}-NTA lipid-containing SLBs had resulted in significantly different surface densities, this would have provided evidence that nTMV is always on the flat disk while cTMV′ is always on the concave disk, or vice-versa, due to differing accessibility of the His-tag. However, the observed similarity in surface densities may indicate either that nTMV and cTMV′ are oriented randomly on the flat and concave disks or that the His-tag is equally accessible to the bilayer on either the flat or concave disk despite structural asymmetry.

To attempt to resolve the relative orientations of the tandem dimers within the dTMV double disk, dTMV was labeled at the S123(′)C site with multiple copies of peptides and proteins. The constructs were then imaged by TEM after applying uranyl acetate negative staining. In the case of a uniform monomer orientation, the attached proteins or peptides would be expected to appear on a single side of the disk, while a scrambled orientation would result in modification of both faces. A longer peptide than H_6Y, HHHHHHSGGGGY (H_6SG_4Y), was used to increase visibility of the peptide by TEM, resulting in 70% modification of both dTMV-S123C–S123′K and dTMV-S123K–S123′C (Figure 4.10a). While the results were obscured by the stacking behavior of the dTMV disks, the attached peptide was not clearly resolved in images of either dTMV-S123C-H_6SG_4Y–S123′K or dTMV-S123K–S123′C-H_6SG_4Y (Figure 4.10b-c). When conjugated to an anti-HER2 nanobody containing a C-terminal SGGGGY tag ($nbHER2_{Tyr}$),[36] dTMV-S123K–S123′C appeared to show examples of $nbHER2_{Tyr}$ conjugation to both sides of the double disk (Figure 4.10d-e). The TEM images as well as fluorescence images of modified dTMV are inconclusive, but suggestive of a random orientation of nTMV and cTMV′ on the flat and concave sides of dTMV assemblies.

Whether all dTMV units are oriented with their N-terminal domains on one disk (e.g., the concave disk) and the C-terminal domain on the opposite disk (e.g., the flat disk) or whether they assemble at random, with the flat and concave disks each having a mixture of C- and N-terminal domains, remains an open question. The linker region between each domain was not well resolved in the crystal structure, and imaging of dTMV modified with biomolecules was similarly inconclusive. Because dTMV contains a flat and a concave disk, with the concave disk containing more adjacent monomer interface interactions than the flat disk, engineering at the intra-disk interfaces to remove or to add further residue interactions at only nTMV or cTMV′ may increase the likelihood of dTMV units being uniformly oriented within disk assemblies. Even if nTMV and cTMV′ are randomly oriented within double disk assemblies, there are residues that are more occluded in only the concave or only the flat disk as indicated by differences in their calculated solvent accessible surface areas (SASAs; Table 4.1). As an example, residue Gln36 in the intracavity region between the disks of dTMV is involved in a hydrogen bonding network only present in the concave disk, and its mean SASA is 28.7 ± 0.3 Å^2 lower in the concave disk than in the flat disk. Thus, a careful selection of protein modification site may target only a single side of a dTMV assembly even with a random orientation of dTMV units.

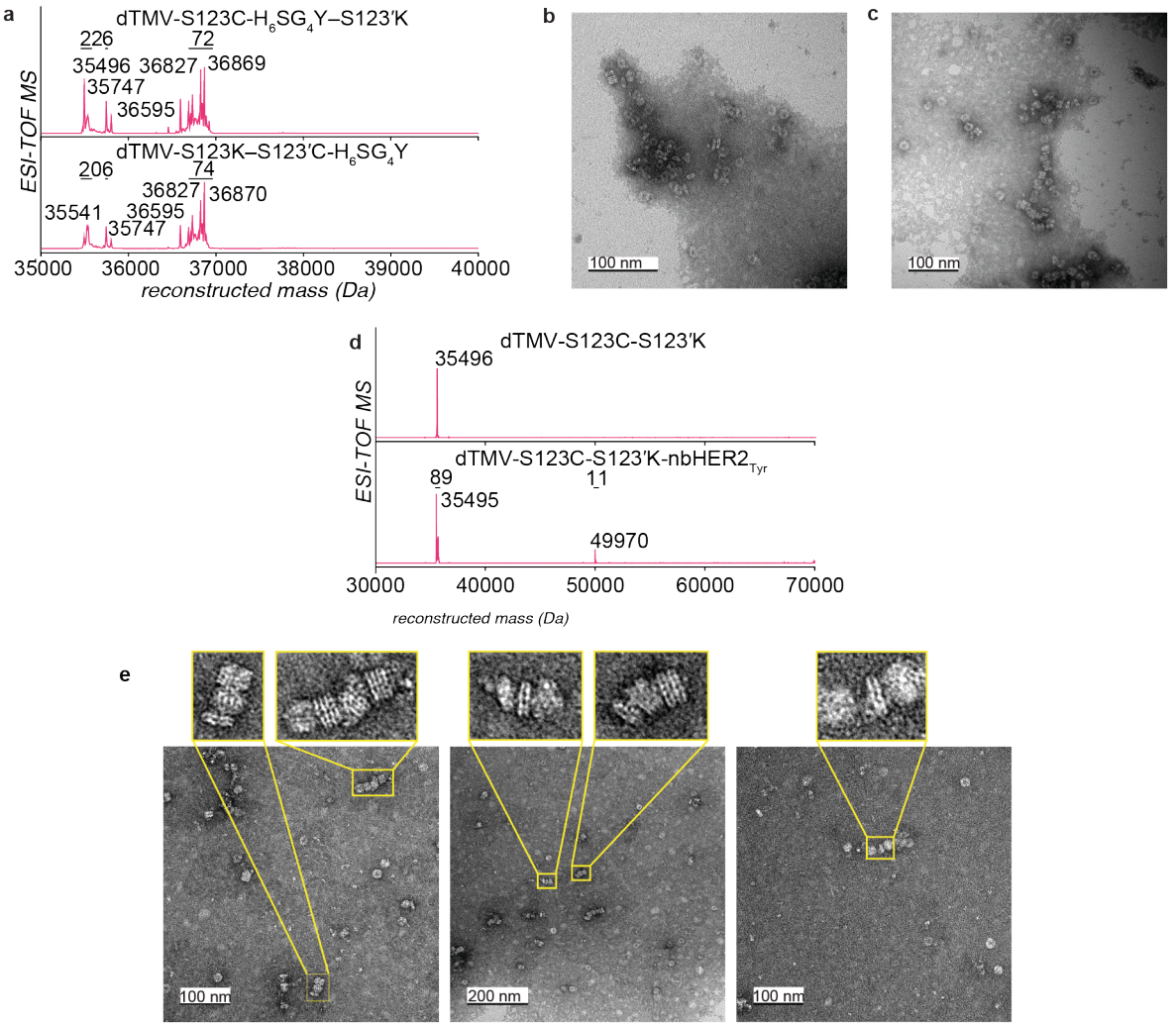


Figure 4.10: TEM images of dTMV conjugated to an anti-HER2 nanobody and the H_6SG_4Y peptide. (a) Mass data of dTMV-S123C–S123′K and dTMV-S123K–S123′C conjugated to the peptide H_6SG_4Y shows 70–75% of monomers were modified with peptide (expected MW of conjugate: 36829 Da). Additional peaks are likely due to peptide impurities. TEM images of dTMV-S123C-H_6SG_4Y–S123′K (b) and dTMV-S123K–S123′C-H_6SG_4Y (c) look similar, both showing examples of interrupted stacks of disks. Few differences appear whether the H_6SG_4Y peptide is attached to the Cys123 position on the nTMV or cTMV′ subunits. (d) Mass data of dTMV-S123C–S123′K conjugated to Human Epidermal Growth Factor Receptor 2 nanobody with a C-terminal SGGGGY tag ($nbHER2_{Tyr}$) shows that 11% of monomers were modified (expected MW of conjugate: 49973 Da), corresponding to an average of 1.9 copies of $nbHER2_{Tyr}$ per disk. (e) TEM images of dTMV-S123C–S123′K conjugated via Cys123 to $nbHER2_{Tyr}$ show some examples of double disks with excess protein on either face.

Determining the orientation of monomers with dTMV may be achieved in the future through several methods. One potential method would be through modification of cysteine residues on one of the constructs reported herein, such as dTMV-S123C–S123′K, with a nanoparticle with high contrast in TEM images. This could be achieved using a method for the modification of gold nanoparticles (AuNPs) by functionalizing the AuNPs with thiols containing a PEG linker and terminal phenol, as previously reported by the Francis lab.[37] Incubating these phenol-modified AuNPs with dTMV containing an exposed thiol on only the N- or C-terminal domain would result in an experiment similar to that with nbHER2$_{\text{Tyr}}$ above. However, due to the high contrast of AuNPs, it would be easier to determine whether AuNPs appeared on a single side or both sides of the dTMV disks in TEM images. Another conclusive method for determining the monomer orientation in dTMV would be through x-ray crystallography structural determination of a mutant containing a point mutation of a small residue to a bulky residue, such as the dTMV-S123K–S123′ mutant. Due to the size difference between lysine and serine, crystallographic data of this mutant would likely be able to distinguish between a uniform or random orientation.

Table 4.1: Solvent-accessible surface area (SASA) of dTMV residues.

nTMV res[a]	cTMV res[a]	nTMV SASA[b]	cTMV SASA[b]	nTMV-cTMV SASA[b]
1	166	38.51 ± 2.57	46.24 ± 0.81	7.73 ± 0.65
2	167	7.85 ± 0.77	8.08 ± 0.58	0.23 ± 0.23
3	168	52.88 ± 1.32	52.15 ± 2.58	-0.73 ± 0.7
4	169	43.38 ± 1.36	40.5 ± 1.61	-2.88 ± 0.51
5	170	110.14 ± 1	101.09 ± 1.77	-9.06 ± 0.49
6	171	63.8 ± 0.58	65.84 ± 1.38	2.04 ± 0.36
7	172	26.97 ± 0.68	23.04 ± 1.87	-3.92 ± 0.48
8	173	29.6 ± 0.99	32.79 ± 1.5	3.2 ± 0.44
9	174	2.66 ± 0.43	5.88 ± 1.23	3.22 ± 0.32
10	175	22.66 ± 1.77	19.23 ± 0.56	-3.43 ± 0.45
11	176	4.32 ± 0.93	1.61 ± 0.48	-2.7 ± 0.25
12	177	29.79 ± 0.88	32.22 ± 1.18	2.43 ± 0.36
13	178	3.55 ± 1.07	3.84 ± 0.49	0.29 ± 0.28
14	179	0.99 ± 0.19	0.64 ± 0.13	-0.35 ± 0.06
15	180	26.24 ± 1.3	27.13 ± 1.44	0.88 ± 0.47
16	181	0 ± 0	0 ± 0	0 ± 0
17	182	1.5 ± 0.3	0.75 ± 0.44	-0.75 ± 0.13
18	183	1.99 ± 0.66	1.48 ± 0.43	-0.51 ± 0.19
19	184	47.3 ± 1.49	48.75 ± 2.06	1.45 ± 0.62
20	185	8.56 ± 1.51	7.53 ± 0.63	-1.03 ± 0.4
21	186	40.85 ± 1.61	42.71 ± 1.77	1.87 ± 0.58
22	187	64.23 ± 8.02	70.96 ± 3.18	6.73 ± 2.09

Continued on next page

Table 4.1 – continued from previous page

nTMV res[a]	cTMV res[a]	nTMV SASA[b]	cTMV SASA[b]	nTMV-cTMV SASA[b]
23	188	0 ± 0	0 ± 0	0 ± 0
24	189	4.15 ± 0.46	3.25 ± 0.32	-0.89 ± 0.14
25	190	37.36 ± 1.43	31.13 ± 1.28	-6.23 ± 0.47
26	191	25.13 ± 3.62	23.98 ± 2.67	-1.15 ± 1.09
27	192	0 ± 0	0 ± 0	0 ± 0
28	193	3.95 ± 0.61	2.13 ± 0.39	-1.82 ± 0.18
29	194	72.36 ± 6.51	77.75 ± 8.26	5.4 ± 2.55
30	195	1.03 ± 0.13	1.57 ± 0.33	0.54 ± 0.09
31	196	0 ± 0	0.06 ± 0.09	0.06 ± 0.02
32	197	4.83 ± 4.23	7.86 ± 0.71	3.03 ± 1.04
33	198	13.28 ± 7.29	48.08 ± 10.52	34.8 ± 3.1
34	199	16.66 ± 2.65	20.71 ± 0.89	4.04 ± 0.68
35	200	0 ± 0	0.12 ± 0.1	0.12 ± 0.02
36	201	1.74 ± 0.31	30.47 ± 1.21	28.72 ± 0.3
37	202	24.87 ± 3.87	79.38 ± 0.5	54.51 ± 0.95
38	203	10.91 ± 1.78	70.22 ± 0.91	59.31 ± 0.49
39	204	46.64 ± 6.75	52.43 ± 3.76	5.79 ± 1.87
40	205	1.49 ± 0.91	17.18 ± 0.67	15.69 ± 0.27
41	206	0 ± 0	3.7 ± 0.51	3.7 ± 0.12
42	207	19.82 ± 6.47	18.74 ± 2.08	-1.08 ± 1.65
43	208	0.44 ± 0.25	10.49 ± 4.33	10.05 ± 1.05
44	209	0 ± 0	3.05 ± 1.27	3.05 ± 0.31
45	210	1.02 ± 0.9	0.49 ± 0.15	-0.54 ± 0.22
46	211	76.99 ± 9.68	42.27 ± 6	-34.73 ± 2.76
47	212	43.62 ± 2.74	34.12 ± 5.74	-9.5 ± 1.54
48	213	0 ± 0	0 ± 0	0 ± 0
49	214	4.76 ± 3.82	8.82 ± 0.71	4.06 ± 0.94
50	215	96.45 ± 9.02	53.4 ± 4.1	-43.04 ± 2.4
51	216	5.98 ± 0.89	4.59 ± 0.6	-1.39 ± 0.26
52	217	6.06 ± 0.71	9.69 ± 0.54	3.63 ± 0.22
53	218	130.44 ± 5.43	96.78 ± 3.97	-33.66 ± 1.63
54	219	55 ± 1.52	79.22 ± 1.74	24.22 ± 0.56
55	220	19.95 ± 1.44	21.01 ± 1.03	1.07 ± 0.43
56	221	0 ± 0	0 ± 0	0 ± 0
57	222	31.58 ± 3.75	45.93 ± 2.28	14.35 ± 1.06
58	223	56.82 ± 2.97	55.32 ± 1.95	-1.5 ± 0.86
59	224	41.11 ± 2.17	50.83 ± 1.24	9.72 ± 0.6
60	225	22.29 ± 1.21	27.83 ± 0.85	5.55 ± 0.36

Continued on next page

Table 4.1 – continued from previous page

nTMV res[a]	cTMV res[a]	nTMV SASA[b]	cTMV SASA[b]	nTMV-cTMV SASA[b]
61	226	83.79 ± 1.09	78.92 ± 1.88	-4.88 ± 0.53
62	227	0 ± 0	0 ± 0	0 ± 0
63	228	34.56 ± 1.14	36.11 ± 0.88	1.55 ± 0.35
64	229	110.16 ± 1.68	102.68 ± 1.83	-7.48 ± 0.6
65	230	98.94 ± 1.78	91.65 ± 1.18	-7.29 ± 0.52
66	231	75.95 ± 1.34	72.26 ± 2.21	-3.69 ± 0.63
67	232	17.83 ± 1.74	20.42 ± 1.87	2.59 ± 0.62
68	233	45.01 ± 1.64	42 ± 1.42	-3.01 ± 0.53
69	234	0.12 ± 0.11	0.26 ± 0.16	0.14 ± 0.05
70	235	6.62 ± 0.49	7.01 ± 0.72	0.39 ± 0.21
71	236	0.14 ± 0.08	0.18 ± 0.12	0.04 ± 0.04
72	237	7.68 ± 0.51	9.84 ± 0.79	2.16 ± 0.23
73	238	21.55 ± 0.87	19.1 ± 1.21	-2.45 ± 0.36
74	239	95.04 ± 0.84	93.86 ± 0.68	-1.19 ± 0.26
75	240	74.1 ± 1.45	73.23 ± 1.57	-0.88 ± 0.52
76	241	0 ± 0	0 ± 0	0 ± 0
77	242	13.51 ± 1.15	24.46 ± 1.17	10.95 ± 0.4
78	243	85.26 ± 0.88	85.11 ± 0.83	-0.16 ± 0.29
79	244	36.34 ± 1.27	38.97 ± 1.28	2.63 ± 0.44
80	245	0 ± 0	0 ± 0	0 ± 0
81	246	14.2 ± 1.35	19.06 ± 1.73	4.87 ± 0.53
82	247	57.71 ± 1.04	59.08 ± 1	1.36 ± 0.35
83	248	0 ± 0	0 ± 0	0 ± 0
84	249	0.67 ± 0.19	0.15 ± 0.14	-0.52 ± 0.06
85	250	19.88 ± 0.69	18.04 ± 0.93	-1.83 ± 0.28
86	251	17.19 ± 0.51	15.66 ± 0.83	-1.52 ± 0.24
87	252	0 ± 0	0 ± 0	0 ± 0
88	253	7.24 ± 0.52	1.52 ± 0.41	-5.71 ± 0.16
89	254	22.63 ± 0.72	41.23 ± 0.91	18.6 ± 0.28
90	255	46.99 ± 0.87	176.34 ± 2.45	129.35 ± 0.63
91	256	23.43 ± 1.85		
92	257	70.7 ± 3		
93	258	10.39 ± 0.68		
94	259	12.69 ± 1.68		
95	260	189.51 ± 3		
96	261			
97	262			
98	263			

Continued on next page

Table 4.1 – continued from previous page

nTMV res[a]	cTMV res[a]	nTMV SASA[b]	cTMV SASA[b]	nTMV-cTMV SASA[b]
99	264			
100	265			
101	266	129.03 ± 2.4		
102	267	72.69 ± 4.56		
103	268	53.5 ± 2.35		
104	269	81.09 ± 2.52		
105	270	46.45 ± 1.45		
106	271	10.54 ± 1.57		
107	272	5.75 ± 0.95		
108	273	67.26 ± 1.63		
109	274	16.61 ± 1.51	128.76 ± 1.53	112.15 ± 0.52
110	275	2.66 ± 0.19	94.54 ± 0.81	91.88 ± 0.2
111	276	11.34 ± 0.48	35.16 ± 2.53	23.82 ± 0.62
112	277	148.35 ± 2.99	171.24 ± 1.27	22.89 ± 0.79
113	278	16.2 ± 0.95	128.7 ± 1.58	112.5 ± 0.45
114	279	0 ± 0	35.29 ± 1.72	35.29 ± 0.42
115	280	39.8 ± 1.34	37.33 ± 1.36	-2.47 ± 0.46
116	281	77.62 ± 1.02	65.16 ± 1.32	-12.47 ± 0.4
117	282	6.74 ± 0.43	16.3 ± 1.1	9.56 ± 0.29
118	283	6.86 ± 0.56	3.65 ± 0.34	-3.21 ± 0.16
119	284	96.77 ± 0.76	95.01 ± 3.81	-1.76 ± 0.94
120	285	45.18 ± 0.83	49.19 ± 1.07	4.01 ± 0.33
121	286	0 ± 0	0.03 ± 0.03	0.03 ± 0.01
122	287	18.41 ± 0.93	14.31 ± 0.58	-4.11 ± 0.27
123	288	73.26 ± 1.58	70.26 ± 1.05	-3 ± 0.46
124	289	6.69 ± 0.46	8.99 ± 1.03	2.3 ± 0.27
125	290	0 ± 0	0 ± 0	0 ± 0
126	291	34.9 ± 3.04	42.52 ± 1.39	7.62 ± 0.81
127	292	76.4 ± 1.21	73.44 ± 3.09	-2.96 ± 0.8
128	293	0 ± 0	0 ± 0	0 ± 0
129	294	0.51 ± 0.66	4.66 ± 0.53	4.15 ± 0.21
130	295	76.34 ± 1.66	73.45 ± 5.85	-2.89 ± 1.48
131	296	28.19 ± 1.24	31.58 ± 7.7	3.39 ± 1.89
132	297	12.57 ± 1.89	9.6 ± 0.75	-2.97 ± 0.49
133	298	39.83 ± 0.85	43.65 ± 3.34	3.82 ± 0.84
134	299	177.74 ± 1.17	169.32 ± 8.86	-8.41 ± 2.17
135	300	5.69 ± 0.36	7.11 ± 0.48	1.42 ± 0.14
136	301	34.98 ± 1.32	33.26 ± 1.82	-1.71 ± 0.55

Continued on next page

Table 4.1 – continued from previous page

nTMV res[a]	cTMV res[a]	nTMV SASA[b]	cTMV SASA[b]	nTMV-cTMV SASA[b]
137	302	9 ± 0.6	9.14 ± 0.36	0.14 ± 0.17
138	303	16.04 ± 0.48	15.37 ± 0.59	-0.68 ± 0.19
139	304	27.96 ± 0.89	27.52 ± 1.66	-0.44 ± 0.46
140	305	42.8 ± 1.34	40.95 ± 1.03	-1.85 ± 0.41
141	306	62.9 ± 2	69.51 ± 1.34	6.61 ± 0.59
142	307	85.21 ± 0.88	84.66 ± 0.91	-0.55 ± 0.31
143	308	42.38 ± 0.81	42.46 ± 1.16	0.08 ± 0.34
144	309	0 ± 0	0.05 ± 0.08	0.05 ± 0.02
145	310	31.8 ± 1.33	33.4 ± 0.98	1.59 ± 0.4
146	311	96.17 ± 0.87	94.94 ± 0.88	-1.23 ± 0.3
147	312	66.85 ± 1.68	68.48 ± 2.16	1.63 ± 0.66
148	313	9.58 ± 0.43	8.84 ± 1.15	-0.73 ± 0.3
149	314	44.8 ± 1.01	44.05 ± 1.43	-0.75 ± 0.42
150	315	0.46 ± 0.29	0.21 ± 0.23	-0.25 ± 0.09
151	316	100.88 ± 1.43	95.97 ± 1.25	-4.91 ± 0.46
152	317	56.21 ± 2.54	38.61 ± 1.73	-17.6 ± 0.75
153	318	66.27 ± 1.76	82.95 ± 2.13	16.68 ± 0.67
154	319	117.04 ± 2.54	82.44 ± 3.32	-34.6 ± 1.01
155	320	102.37 ± 3.66	118.17 ± 1.3	15.81 ± 0.94
163		80.69 ± 3.26		
164		17.69 ± 1.14		
165		18.94 ± 1.25		

[a]'res' indicates residue number in dTMV sequence [b]values shown are averages of all 17 monomers ± standard error

4.4 Conclusion and Outlook

Few artificial light-harvesting systems have incorporated the two-dimensional fluidity afforded to natural light-harvesting complexes by their embedment within a lipid membrane. Herein, a well-characterized light-harvesting model constructed from modified virus-like particles was adapted to facilitate their attachment to a planar lipid bilayer. This model was constructed from a newly reported dual-functional double-disk protein assembly consisting of 17 tandem dimer TMV subunits. By sequentially attaching a synthetic pigment and a His-tag-containing peptide to the dTMV surface, we have constructed a fully synthetic biomimetic model composed of discoidal protein arrays bound to chromophores and then associated with an SLB containing nickel-chelating lipids. Total internal reflection fluorescence (TIRF) microscopy measurements coupled with quantitative single molecule analysis enabled

detailed characterization of the fluorescently labeled, bilayer-associated dTMV double disks and demonstrated their mobility on the SLB.

The asymmetry of dTMV assemblies will allow for a more thorough interrogation of the role of the protein environment on chromophore excited state lifetimes and energy transfer in protein-based model light-harvesting systems. Unlike C_2-symmetric disk assemblies, the intracavity region of dTMV is narrower and can be selectively mutated on either the flat or concave surface to construct an asymmetric chromophore environment. In chromophore-labeled TMV complexes, these features of dTMV may allow for significant protein–chromophore coupling, a feature contributing to lengthened excited state lifetimes and high-efficiency energy transfer.[8] The solvent-exposed surface of dTMV is also easily accessible for functionalization, manifested by the successful step-wise conjugation of dyes and peptides described herein, leading to possibilities of coupling single fluorescently labeled dTMV assemblies to form donor–acceptor pairs (as in Chapter 3) or coupling dTMV assemblies to other light-harvesting proteins. The asymmetry of the complex lends itself to the directional attachment to a fluid lipid bilayer and could also be used to quantitate excited state energy transfer between mixed populations of dTMV disks that have been separately decorated with donor and acceptor chromophores and attached to an SLB. Such organization is reminiscent of the arrangement of LH1 and LH2 in photosynthetic membranes.

This synthetic system, with the SLB associated with a glass coverslip, facilitates analysis via a variety of fluorescence microscopy[38,39] and spectroscopic methods[28,40,41] to measure the diffusion and energy transfer behavior of the dye-labeled dTMV assemblies on the bilayer. Additionally, the underlying substrate of the SLB may be synthesized from or lithographically patterned with electronically active materials, allowing for energy transfer between or energy harvesting from the chromophore-labeled dTMV complexes.[42–45]

The ability to attach dTMV to an SLB, together with the ability to join disks at their periphery demonstrated in Chapter 3, open the possibility of creating a functional photosynthetic membrane using entirely synthetic components. The greatest density of dTMV complexes on the SLB in this work was 14 molecules per μm^{-2}, or approximately one disk per 71 nm^2 on average. This density may hamper the ability of disks to transfer energy to one another if a population of dTMV disks labeled with donor dyes and dTMV disks labeled with acceptor dyes were incubated together on a SLB. Higher incubation concentrations may lead to higher densities and favorable conditions for energy transfer. Alternatively, combining a strategy to conjugate disks together as in Chapter 3 with the ability to attach disks to an SLB may allow for energy transfer to be observed between dTMV disks attached to an SLB.

A platform for attaching dTMV light-harvesting complexes to an SLB, followed by joining disks at their periphery, would require at least three independent functionalizations. While lysine was used here for chromophore attachment to dTMV, the reaction does not proceed to full conversion to produce a complete array of chromophores even under optimal conditions.[16] Using the engineered cysteine residue for the attachment of maleimide-functionalized chromophores, as in Chapter 3, would likely lead to more complete chromophore arrays. Following this, an engineered aniline residue at one of the S123 positions could be used for

His-tag attachment using similar conditions to those reported herein, with the H_6Y peptide and abTyr. Anilines on protein surfaces have been shown to react with phenols in the presence of abTyr in previous work.[37] Finally, the SLB-associated, dye-labeled dTMV complexes could be linked at their peripheries through the modification of engineered peripheral sites. For symmetric protein-protein coupling, lysine could be engineered on the periphery of each protein complex, and a bifunctional NHS ester linker could be used to covalently link the disks. This would not distinguish between forming donor–acceptor pairs versus donor–donor or acceptor–acceptor pairs; however, controlling the stoichiometry of donors and acceptors may still lead to acceptors being primarily surrounded by donor disks. For asymmetric protein–protein coupling, the N-terminus of the donor disk could be joined to an engineered lysine residue on the acceptor disk via orthogonal or sequential modifications. This could potentially be achieved using a 2-pyridinecarboxaldehyde derivative for modification of the N-terminus with a reactive functionality[46] or pyridoxal 5′-phosphate to install a ketone at the N-terminus.[47] A heterobifunctional linker could subsequently be used to asymmetrically couple donor and acceptor disks. Furthermore, engineering histidine residues within the cavity region of dTMV into this platform for the ligation of hemes as reported previously for cpTMV[48] could introduce the ability to transfer energy from surface chromophores to interior hemes. This would introduce the ability to convert light energy to charge separation ability in the SLB-associated dTMV light harvesting model.

4.5 Materials and Methods

General methods and materials. Unless otherwise noted, all chemicals and solvents were of analytical grade and were received from commercial sources. The phospholipids 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC), 1,2-dioleoyl-*sn*-glycero-3-[(*N*-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] nickel salt (Ni-NTA-DOGS), and 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[4-(*p*-maleimidomethyl)cyclohexane-carboxamide] (MCC-DOPE) sodium salt for preparation of supported membranes were purchased from Avanti Polar Lipids (Alabaster, AL) as chloroform solutions. Alexa Fluor 647 NHS ester was purchased from Sigma-Aldrich (St. Louis, MO). Water (dd-H_2O) used as reaction solvent was deionized using a Barnstead NANOpure purification system (Thermo Fisher, Waltham, MA). All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Protected amino acids and resins for solid phase peptide synthesis were obtained from Novabiochem (Merck KGaA, Darmstadt, Germany). Spin concentration was performed using 100,000 or 30,000 molecular weight cutoff spin concentrators from Millipore (Burlington, MA). Dialysis was performed with either Slide-A-Lyzer Dialysis Cassettes (Pierce, Rockford, Il) for small volume samples or dialysis tubing (Fisherbrand, Pittsburgh, PA) for large volume samples.

Mass spectrometry. Protein and small molecules were analyzed using liquid chromatography (1200 series, Agilent Technologies, USA) that was connected in line with an Agilent 6224

Time-of-Flight (TOF) mass spectral system equipped with a Turbospray ion source. Protein samples were run with a Proswift RP-4H column (Dionex, USA). Protein mass reconstruction was performed on the charge ladder with Mass Hunter software (Agilent Technologies, USA).

High Performance Liquid Chromatography. HPLC was performed on Agilent 1260 Infinity Series HPLC Systems (Agilent, USA). Sample analysis for all HPLC experiments was achieved with an in-line diode array detector (DAD) and an in-line fluorescence detector (FLD). Size exclusion chromatography (SEC) was performed using a Polysep-GFC-P-5000 column (4.6 × 250 mm) (Phenomenex, USA) at 1.0 mL/min using a mobile phase of 10 mM sodium phosphate buffer, pH 7.2.

Gel Analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in a Mini cell tank apparatus (Life Technologies, Carlsbad, CA), using NuPAGE^{TM} Novex^{TM} 4–12% Bis-Tris Protein Gels (Life Technologies). The sample and electrode buffers were prepared according to the suggestions of the manufacturer. All protein electrophoresis samples were heated for 5–10 min at 95 °C in the presence of 1,4-dithiothreitol (DTT) to ensure the reduction of disulfide bonds. Gels were run for 30 min at 200 V to separate the bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for the assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad, Hercules, CA). Gel imaging was performed on a Gel Doc (Bio-Rad, Hercules, CA).

Transmission Electron Microscopy. TEM analysis of dTMV and its conjugates with peptide and nanobody was carried out at the Berkeley Electron Microscope Lab with an FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage. Samples were prepared for analysis by applying analyte solution (5 μL of 10 μM protein in 10 mM NaPhos pH 7.2) to carbon-coated copper grids for 2 min. The sample application was followed by rinsing in 1% uranyl acetate solution (4 × 10 μL), leaving the sample in the last droplet for 1 min before wicking away the excess solution.

Protein expression and purification. DH10B competent cells were transformed with the plasmids containing the dTMV variants described above. Colonies were selected for inoculation in Terrific Broth with 100 μg/L ampicillin at 37 °C. When cultures reached an optical density of 0.6 to 0.8, 0.01% arabinose was added. After growing for 20 h at 20 °C, the cells were harvested by centrifugation (8000 rpm, 15 min) and the cell pellet was stored at -20 °C. Cells were resuspended in 20 mL lysis buffer (20 mM triethanolamine [TEA] pH 7.2) supplemented with benzonase and 2 mM $MgCl_2$. Cells were lysed by sonication with a 2 s on, 4 s off cycle for a total of 10 min using a standard disruptor horn at 60% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared at 14,000 rpm for 30 min. A saturated solution of ammonium sulfate was added to the supernatant to reach a final concentration of 15%. The mixture was rotated for 10 min at 4 °C to allow the complete protein precipitation. The precipitated protein was then collected at 10,000 rpm

for 30 min and then resuspended in low salt buffer (20 mM TEA, pH 7.2). The resulting protein solution was dialyzed against the buffer to remove the residual ammonium sulfate before loading onto a DEAE column and purifying with a 0–300 mM NaCl gradient elution in buffer (20 mM TEA, pH 7.2). Purity was confirmed by SDS-PAGE and ESI-TOF MS. Pure fractions were pooled, and fractions containing desired dTMV in addition to impurities were further purified using a HiPrep^{TM} 26/60 Sephacryl® S-500 HR column (GE Healthcare, USA).

General procedure for peptide synthesis. Solid-phase peptide synthesis was performed following an established literature procedure with minimal modifications.[49] Side chain protecting groups used were: His (Trt), and Ser (*t*Bu). The resin linker used was benzyloxybenzyl alcohol (Wang) polystyrene. Synthesis was accomplished manually, using 5 equiv. of amino acids in dimethylformamide (DMF) with *O*-(1*H*-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU, Novabiochem, Merck KGaA, Darmstadt, Germany) as the coupling reagent with 10 equiv. of *N*,*N*-diisopropylethylamine (DIPEA, Sigma-Aldrich, St. Louis, MO) as an additive. Once all amino acids had been coupled to the peptide on the resin, the N-terminal Fmoc group was removed using 20% piperidine (Sigma-Aldrich, St. Louis, MO) in DMF. Peptides were cleaved from resin using a cocktail of 95% trifluoroacetic acid (TFA, EMD Millipore, Billerica, MA), 2.5% water, and 2.5% triisopropylsilane (TIPS, Sigma-Aldrich, St. Louis, MO). Crude peptides were precipitated in cold diethyl ether, analyzed, and further purified if needed by reversed-phase HPLC, and lyophilized before use. Data for each of the peptides is found in Table 4.2.

Table 4.2: Peptide Data Table

Name	Sequence	Calculated MW [Da]	Found MW [Da]
H_6Y	HHHHHHY	1004.0	1003.4
H_6SG_4Y	HHHHHHSGGGGY	1319.3	1318.5

General procedure for labeling of dTMV with AF647 NHS ester. The proteins were first exchanged into the reaction buffer (100 mM sodium phosphate, pH 8.5). To 100 μL of dTMV (100 μM) was added 2 equiv. of AF647 NHS ester. The reaction mixture was briefly agitated and then incubated in 1.5 mL Eppendorf tubes at room temperature with an aluminum foil cover. After 15 min, the reactions were quenched by 1 mM hydroxylamine. The crude reaction was purified with a NAP-5 Sephadex G-25 column (GE Healthcare, USA) followed by purification using a Polysep-GFC-P-5000 column (4.6 × 250 mm) (Phenomenex, USA) to remove the unreacted chromophores. The fractions that showed absorption at 280 nm were combined and subjected to spin concentration using 100 kDa MWCO filters. The protein conjugates were analyzed with LC-MS and HPLC-SEC for assessment of conjugation level, purity, and validation of assembly state.

General procedure for labeling of dTMV with H_6Y or H_6SG_4Y peptide. The proteins were first exchanged into the reaction buffer (10 mM sodium phosphate, pH 7.2).

To 100 μL of peptide (500 μM) was added 400 nM abTyr, followed by 25 μM protein. The reaction mixture was briefly agitated and then incubated in 1.5 mL Eppendorf tubes at room temperature with an aluminum foil cover. After 2 h, excess peptide and abTyr were removed via repeated centrifugation through 100 kDa molecular weight cutoff filters and used directly for TEM imaging or further purified for fluorescence imaging using a Polysep-GFC-P-5000 column (4.6 × 250 mm). The fractions in the size range containing TMV disk stacks were combined and subjected to spin concentration using 100 kDa MWCO filters. The protein conjugates were analyzed with MS and HPLC-SEC for assessment of conjugation level, purity, and validation of assembly state.

Procedure for labeling of Bovine Serum Albumin (BSA) with AF647 NHS ester. BSA lyophilized powder (Sigma-Aldrich, St. Louis, MO) was dissolved just before use into the reaction buffer (100 mM sodium phosphate, pH 8.5) at a concentration of 100 μM. To 50 μL of BSA (100 μM) was added 2 equiv. of AF647. The reaction mixture was briefly agitated and then incubated in 1.5 mL Eppendorf tubes at room temperature with an aluminum foil cover. After 1 h, the reactions were quenched by 1 mM hydroxylamine. The crude reaction was purified with a NAP-5 Sephadex G-25 column followed by purification using a Polysep-GFC-P-5000 column (4.6 × 250 mm) to remove the unreacted chromophores. The fractions that showed absorption at 280 nm were combined and subjected to spin concentration using 30 kDa MWCO filters. The protein conjugates were analyzed with LC-MS and HPLC-SEC for assessment of conjugation level and purity.

Procedure for labeling the BSA-AF647 conjugate with H_6Y peptide. BSA-AF647 was first exchanged into the reaction buffer (10 mM sodium phosphate, pH 7.2). To 100 μL of H_6Y peptide (200 μM) was added 400 nM tyrosinase from *Agaricus bisporus* (abTyr), followed by 100 μM protein. The reaction mixture was briefly agitated and then incubated in 1.5 mL Eppendorf tubes at room temperature with an aluminum foil cover. After 2.5 h, excess peptide and abTyr were removed via repeated centrifugation through 30 kDa molecular weight cutoff filters followed by purification using a Polysep-GFC-P-5000 column (4.6 × 250 mm). The fractions that showed absorption at 280 nm were combined and subjected to spin concentration using 30 kDa MWCO filters. The protein conjugates were analyzed with MS and HPLC-SEC for assessment of conjugation level and purity.

Procedure for constructing dTMV-nbHER2$_{\mathbf{Tyr}}$ conjugate. The proteins were first exchanged into the reaction buffer (10 mM sodium phosphate, pH 7.2). To 100 μL of nbHER2$_{Tyr}$ (50 μM) was added 400 nM abTyr, followed by 25 μM dTMV-S123C–S123′K. The reaction mixture was briefly agitated and then incubated in 1.5 mL Eppendorf tubes at room temperature with an aluminum foil cover. After 2 h, excess nbHER2$_{Tyr}$ and abTyr were removed via repeated centrifugation through 100 kDa MWCO filters. The protein conjugates were analyzed with LC-MS and HPLC-SEC for assessment of conjugation level, purity, and validation of assembly state.

Supported membrane preparation. Supported membranes were formed on glass coverslips as described elsewhere.[50] Briefly, 25 mm # 1.5 thickness round glass coverslips (Thomas Scientific, Swedesboro, NJ) were ultrasonicated for 30 min in 50:50 isopropyl alcohol/water, rinsed thoroughly in Milli-Q water (MilliporeSigma, Billerica, MA), etched for 5 min in freshly prepared piranha solution (3:1 sulfuric acid/hydrogen peroxide), and again rinsed thoroughly in Milli-Q water. Coverslips were then assembled in Attofluor chambers (Invitrogen, Waltham, MA) that had been cleaned in 50:50 isopropyl alcohol/water, rinsed thoroughly in Milli-Q water, and dried. Bilayers were formed by rupturing small unilamellar vesicles (SUVs) on the cleaned glass substrate. SUVs for bilayer formation were prepared by mixing a molar ratio of 98% DOPC and 2% Ni-NTA-DOGS in chloroform in a round bottom flask, drying the lipid mixture with a rotovap, and resuspending the mixture to 0.5 mg/mL in Milli-Q water. For bilayers including maleimide-headgroup lipids, the lipid mixture contained molar ratios of 95% DOPC, 2% Ni-NTA-DOGS, and 3% MCC-DOPE. The suspension of lipids in water was then sonicated with a probe sonicator (Analis Scientific Instruments, Namur, Belgium) for a total of 1 min using pulses of 15 s at 32% amplitude with 10 s pauses while sitting in an ice bath to prevent the lipid suspension from warming. Sonicated solutions were then centrifuged at 21,000 $\times$ g for 20 min at 4 °C to remove titanium particles from the sonicator and lipid aggregates. SUV solutions in water were mixed 1:1 with PBS and 300 µL of this solution was added to each assembled Attofluor chamber and incubated for 35 min. The membrane was washed with PBS and then incubated in imaging buffer (20 mM HEPES, 137 mM NaCl, 1 mM $CaCl_2$, 2 mM $MgCl_2$, 5 mM KCl, 0.7 mM Na_2HPO_4, 6 mM D-glucose, and 0.2% w/v BSA) for 30 min to block bilayer defects and prevent nonspecific interactions of fluorescent proteins of interest with the surface. dTMV was diluted in an imaging buffer to 4$\times$ the desired working concentration, then diluted 1:4 into the Attofluor chambers and incubated for 30–35 min. The solution incubation concentrations of dTMV proteins were 15–40 nM for experiments requiring high protein density on the bilayer and and 100–500 pM for single particle tracking experiments.

TIRF microscopy. TIRF experiments were performed on a motorized inverted microscope (Nikon Eclipse Ti-E; Technical Instruments, Burlingame, CA) equipped with a motorized Epi/TIRF illuminator (Nikon), Lumen Dynamics X-Cite® 120LED Fluorescence Illumination System (Excelitas Technologies, Waltham, MA, USA), Perfect Focus System (Nikon), and a motorized stage (Applied Scientific Instrumentation MS-2000, Eugene, OR). A laser launch with 488, 561, and 640 nm (Coherent OBIS, Santa Clara, CA) diode lasers was controlled by an OBIS Scientific Remote (Coherent) and aligned into a fiber launch custom built by Solamere Technology Group (Salt Lake City, UT). Only the 640 laser line was used in this study. A dichroic beamsplitter (ZT640rdc with ET660LP emission; Chroma Technology, Bellows Falls, VT) reflected the laser light through the oil emersion objective lens (Nikon; 1.49, numerical aperture, 100x, TIRF; Technical Instruments). All emissions were collected through an emission filter (ET700/75M, Chroma Technology) and recorded using an electron multiplying (EM)-CCD (iXon 897DU; Andor, South Windsor, CT) after passing through a laserblocking filter (Z488/647M; Chroma Technology). Exposure times, multidimensional

acquisitions, and time-lapse periods for all experiments were set using Micro-Manager.[51] A transistor-transistor logic signal from the appropriate laser triggered the camera exposure. The laser intensities were measured at the sample for each experiment day so that a constant laser intensity is used for each type of imaging across different days.

AF647-labeled dTMV was imaged using the 640 nm laser line. dTMV density on bilayers was imaged using 20 ms exposure time, 0.7–1.2 mW power at the sample, and a camera setting of 500 gain. Imaging of AF647-labeled dTMV diffusion on the supported membrane was performed with a streaming acquisition of 20 ms exposure time at 5.2 mW power at the sample and 1000 gain for 50 frames. Oxygen scavengers were not used to minimize photobleaching in order to analyze the number of discrete step photobleaching events along particle trajectories. Images for the dTMV single particle intensity distribution were taken with the same acquisition settings as for the diffusion data. For FRAP measurements, TIRF images before and after bleaching were collected every 20 s at 0.2 mW power and 20 ms exposure time. The bleaching was performed in TIRF, closing the front aperture to define the bleached spot, with 11 mW power continuously for 1.5 s. All imaging was done at 24 °C.

Image Analysis. Image analysis was performed in Matlab (MathWorks, Natick MA) using custom scripts, available upon request.

Quantitation of particle density on SLBs. An intensity vs. density calibration curve was built from at least 20 micrographs each of 4 bilayers at multiple dTMV densities for which particles were countable, using the same imaging conditions (20 ms, 1.2 mW at sample, 500 gain) for each sample. Particles in a 28 μm × 28 μm area of even illumination were counted using TrackMate, with the particle diameter and intensity threshold set by eye then applied uniformly across the data sets. The mean intensity of the same area was measured for each bilayer. The slope and y-intercept of this linear intensity vs. density calibration curve was used to calculate the density of dTMV on high-density bilayers. Images of high-density bilayers were taken with decreased illumination intensity, but as fluorescence intensity scales linearly with illumination intensity, this change in imaging conditions was easily accounted for to calculate bilayer density.

FRAP analysis. FRAP data were normalized according to the workflow described in Carnell, M et al.[52] Briefly, a circular FRAP region fully contained in the bleached area was defined, and at each timepoint t, the mean intensity in this area, FRAP(t), was subtracted by intensity in that area immediately after bleaching ($\mathrm{FRAP}_{\mathrm{bleach}}$). This quantity was normalized to the pre-bleach intensity of the FRAP area ($\mathrm{FRAP}_{\mathrm{pre\text{–}bleach}}$) and the intensity of a reference area at time t ($\mathrm{Ref}_{\mathrm{pre\text{–}bleach}}/\mathrm{ref}(t)$). The fully normalized FRAP recovery trace was calculated as,

$$\mathrm{Fullnorm}(t) = \frac{\mathrm{Ref}_{\mathrm{pre\text{–}bleach}}}{\mathrm{ref}(t)} \cdot \frac{\mathrm{FRAP}(t) - \mathrm{FRAP}_{\mathrm{bleach}}}{\mathrm{FRAP}_{\mathrm{pre\text{–}bleach}} - \mathrm{FRAP}_{\mathrm{bleach}}}$$

The resulting recovery trace was fit to a diffusion-limited recovery model,

$$\text{Fullnorm}(t) = \text{A} \cdot \text{e}^{\frac{2\tau_\text{D}}{t}} \left(I_0 \left(\frac{2\tau_\text{D}}{t} \right) + I_1 \left(\frac{2\tau_\text{D}}{t} \right) \right)$$

where

$$\tau_\text{D} = \frac{\text{w}^2}{4\text{D}}$$

and where A is the plateau intensity, w is the bleach area radius, D is the diffusion coefficient, and $I_0()$ and $I_1()$ are modified Bessel functions of the first kind. This model assumes bleaching occurs instantaneously, bleaching has a step-function profile, recovery is diffusion limited, diffusion is lateral and equal in all directions, and there is a single diffusing population with a single diffusion coefficient. Our experiment generally meets all assumptions except for the last, but a fit can allow us to extract a mean diffusion coefficient from the highly heterogeneous sample.

Single particle diffusion analysis. Particles in a 28 μm × 28 μm area of even illumination were tracked using TrackMate using the Difference of Gaussians spot detector, a particle diameter of 0.4 μm, threshold of 300, and using the median filter and sub-pixel localization. The simple LAP tracker was used to link particle trajectories, using a maximum linking distance of 0.4 μm and not allowing gaps in tracks. The data from TrackMate was then exported to Matlab for analysis and visualization. The diffusion coefficient for individual tracks was calculated from the mean square displacement between adjacent frames along the track. The step size distribution of 2,281 particles was compiled from particle localizations from every third frame along particle trajectories longer than 10 frames. This 60 ms time interval minimizes the effect of particle localization errors, which can introduce artifacts with shorter time intervals.[53] Due to the highly heterogeneous diffusion within the sample, we chose not to fit the step size distribution to a model with a specific number of diffusing species.

Quantitation of single particle intensities. TMV particles presented on bilayers at low density (0.05 μm^{-2}) were localized in 28 μm × 28 μm areas of even illumination using TrackMate using parameters stated above. Total particle intensity was determined by integrating a 7 pixel × 7 pixel (0.75 μm × 0.75 μm) area centered around the particle and subtracting the intensity of an average background area of the same size. These intensities were then calibrated to the intensity of a single fluorophore, determined by integrating a 7 pixel × 7 pixel area around single emitters,[54] to obtain a distribution of the number of fluorophores bound to TMV particles.

Step photobleaching analysis. Tracks obtained for single particle diffusion analysis were analyzed for the number of step photobleaching events. Only tracks of particles that were reliably tracked throughout the acquisition and fully bleached within the acquisition were

analyzed to ensure that the number of fluorophores attached to each disk was measured accurately. The particle intensity in each frame was determined as stated above. Upon particle bleaching, an additional 10 frames (or until the end of the acquisition) were analyzed for total background-subtracted intensity at the last identified particle location and appended to the particle intensity trace. The intensity trace was analyzed using a Bayesian change point algorithm[55] to determine the number of change points, or fluorophore bleaching events, in the particle trajectory. All intensity traces and identified change points were verified manually.

Correlating particle intensity and mobility. The single-particle diffusion coefficients of dTMV assemblies were plotted against their integrated intensities in the first frame of imaging. Particles not present in the first frame (diffused into the frame later in the movie) were excluded from analysis, as they could have already undergone photobleaching events.

Procedure for generating cpTMV homology model. For visualization purposes, a homology model of cpTMV containing all amino acid residues was generated based on the crystal structure of cpTMV (PDB code: 3KML)[13] and the crystal structure of wtTMV (PDB code: 1EI7),[14] both obtained from the Protein Data Bank. Initial structural preparation was conducted using The Pymol Molecular Graphics System, Version 2.4.2. First, a monomer of wtTMV was superimposed on a monomer of cpTMV, and the residues resolved in the crystal structure of wtTMV but not cpTMV (residues 92–110 of wtTMV and 1–12, 154–161 of cpTMV) were fused to the unresolved N- and C-termini of cpTMV. The bond between residues 99 and 100 of wtTMV was then cleaved *in silico* and an N-terminal glycine was added to produce the N- and C-termini of cpTMV. Following this, cpTMV was symmetry expanded to create the double disk assembled structure consisting of two C_2-symmetric disks, each containing 17 monomers. The Schrödinger Maestro package (version 2022-1)[56] was used for subsequent structural preparation and molecular dynamics simulations. The Desmond system builder was used to solvate the double disk structure in an orthorhombic box with periodic boundaries at 10 Å from the protein of water molecules described using the TIP3P model,[57,58] neutralized with sodium ions, with the addition of 150 mM NaCl in an OPLS4 force field.[59] A molecular dynamics simulation was performed with an NPT ensemble of T = 300 K, P = 1 bar, a Coulombic cutoff radius of 9.0 Å, and a 100 ps simulation time with sampling time of 5 ps. This short simulation time was chosen to relax sidechain and solvent interactions for the protein representations shown herein, but not alter the overall quaternary structure of the assembly. The pressure control was applied using the Martyna-Tobias-Klein barostat method[60] with a 2.0 ps relaxation time, and the temperature control was applied using the Nosé-Hoover thermostat method with a 1.0 ps relaxation time.[61] The trajectory was analyzed, and the lowest energy frame was used as the homology model.

Procedure for obtaining solvent-accessible surface area (SASA) of dTMV residues. The SASA of each amino acid residue of dTMV was obtained using the crystal structure of dTMV reported herein as an input structure. The Schrödinger Bioluminate package[62] was

used to perform a Residue Analysis of all resolved residues present on all 17 tandem dimers of the dTMV crystal structure to obtain the SASA of each residue.

Statistical analysis. All fluorescence imaging data are representative of three independent experiments. Statistical analyses for imaging data are described above. For the SASA data gathered from each dTMV monomer in the crystal structure (n = 17), the SASA values of residues with incomplete neighboring residues were removed, and the remaining residues of nTMV and cTMV′ were aligned. The corresponding residues of each tandem dimer were averaged, and the averaged value of each residue of nTMV and the corresponding cTMV′ residue were compared by taking the difference between them. The error in the difference was evaluated by taking the standard error of the difference between means.[63,64] The resulting values are shown in Table 4.1. Statistical analyses were performed in Microsoft Excel for Mac, Version 16.62.

4.6 Acknowledgements

Thank you to Dr. Jing Dai for developing the idea of attaching TMV to a bilayer and for engineering and obtaining the crystal structure of dTMV. Thank you to Dr. Kiera Wilhelm for synthesizing the bilayers, attaching constructs to the bilayer, and conducting fluorescence imaging experiments and data analysis. Thank you to both Dr. Jing Dai and Dr. Kiera Wilhelm for excellent scientific discussions and help with explaining our work. Thank you to Dr. Michel Dedeo for creating the dTMV construct. Thank you to Derek García-Almedina for providing nbHER2$_{\mathrm{Tyr}}$. Thank you to Dr. Jose Pereira and Prof. Paul Adams for assistance obtaining the dTMV crystal structure. Thank you to Prof. Jay Groves for discussions about dTMV attachment to the SLB, and for input into structuring and presenting this work. Thank you to the the staff at the University of California, Berkeley, Electron Microscope Laboratory for their advice and assistance in electron microscopy sample preparation and data collection. Thank you to Dr. Kathleen Durkin and the Berkeley Molecular Graphics and Computation Facility for help with protein modeling.

4.7 References

(1) Jiang, Y.; McNeill, J. *Chemical Reviews* **2017**, *117*, 838–859.

(2) Adronov, A.; Fréchet, J. M. J. *Chemical Communications* **2000**, 1701–1710.

(3) Dutta, P. K.; Varghese, R.; Nangreave, J.; Lin, S.; Yan, H.; Liu, Y. *Journal of the American Chemical Society* **2011**, *133*, 11985–11993.

(4) Park, H. et al. *Nature Materials* **2016**, *15*, 211–216.

(5) Delor, M.; Dai, J.; Roberts, T. D.; Rogers, J. R.; Hamed, S. M.; Neaton, J. B.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *Journal of the American Chemical Society* **2018**, *140*, 6278–6287.

(6) Miller, R. A.; Stephanopoulos, N.; McFarland, J. M.; Rosko, A. S.; Geissler, P. L.; Francis, M. B. *Journal of the American Chemical Society* **2010**, *132*, 6068–6074.

(7) Whaley, K. B.; Sarovar, M.; Ishizaki, A. *Procedia Chemistry* **2011**, *3*, 152–164.

(8) Scholes, G. D.; Fleming, G. R. In *Adventures in Chemical Physics*; Advances in Chemical Physics, 2005, pp 57–129.

(9) Akhtar, P.; Görföl, F.; Garab, G.; Lambrev, P. H. *Chemical Physics* **2019**, *522*, 242–248.

(10) Yang, M.; Fleming, G. R. *The Journal of Chemical Physics* **2003**, *119*, 5614–5622.

(11) Adams, P. G.; Collins, A. M.; Sahin, T.; Subramanian, V.; Urban, V. S.; Vairaprakash, P.; Tian, Y.; Evans, D. G.; Shreve, A. P.; Montaño, G. A. *Nano Letters* **2015**, *15*, 2422–2428.

(12) Calver, C. F.; Schanze, K. S.; Cosa, G. *ACS Nano* **2016**, *10*, 10598–10605.

(13) Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. *Nano Letters* **2010**, *10*, 181–186.

(14) Bhyravbhatla, B.; Watowich, S. J.; Caspar, D. L. *Biophysical Journal* **1998**, *74*, 604–615.

(15) Dedeo, M. T. Development of Scaffolds for Light Harvesting and Photocatalysis from the Coat Protein of Tobacco Mosaic Virus, Ph.D. United States – California: University of California, Berkeley, 2012.

(16) Dai, J. Artificial Bioenergetic Functional Models Templated by Viral Capsid Proteins, Ph.D. United States – California: University of California, Berkeley, 2020.

(17) Groves, J. T.; Boxer, S. G. *Accounts of Chemical Research* **2002**, *35*, 149–157.

(18) Miller, R. A.; Presley, A. D.; Francis, M. B. *Journal of the American Chemical Society* **2007**, *129*, 3104–3109.

(19) Hamerlynck, L. M.; Bischoff, A. J.; Rogers, J. R.; Roberts, T. D.; Dai, J.; Geissler, P. L.; Francis, M. B.; Ginsberg, N. S. *The Journal of Physical Chemistry B* **2022**, *126*, 7981–7991.

(20) Boll, L. B.; Raines, R. T. *ChemBioChem* **2022**, *23*, e202200258.

(21) Li, X.; Song, B.; Chen, X.; Wang, Z.; Zeng, M.; Yu, D.; Hu, D.; Chen, Z.; Jin, L.; Yang, S.; Yang, C.; Chen, B. *PLOS ONE* **2013**, *8*, e77717.

(22) Bruckman, M. A.; Soto, C. M.; McDowell, H.; Liu, J. L.; Ratna, B. R.; Korpany, K. V.; Zahr, O. K.; Blum, A. S. *ACS Nano* **2011**, *5*, 1606–1616.

(23) Mogilevsky, C. S.; Lobba, M. J.; Brauer, D. D.; Marmelstein, A. M.; Maza, J. C.; Gleason, J. M.; Doudna, J. A.; Francis, M. B. *Journal of the American Chemical Society* **2021**, *143*, 13538–13547.

(24) Marmelstein, A. M.; Lobba, M. J.; Mogilevsky, C. S.; Maza, J. C.; Brauer, D. D.; Francis, M. B. *Journal of the American Chemical Society* **2020**, *142*, 5078–5086.

(25) Maza, J. C.; Bader, D. L. V.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Krska, S. W.; Parish, C. A.; Francis, M. B. *Journal of the American Chemical Society* **2019**, *141*, 3885–3892.

(26) Zhang, S.; Zhang, J.; Fang, W.; Zhang, Y.; Wang, Q.; Jin, J. *Nano Letters* **2018**, *18*, 6563–6569.

(27) Nye, J. A.; Groves, J. T. *Langmuir* **2008**, *24*, 4145–4149.

(28) Chung, J. K.; Lee, Y. K.; Denson, J.-P.; Gillette, W. K.; Alvarez, S.; Stephen, A. G.; Groves, J. T. *Biophysical Journal* **2018**, *114*, 137–145.

(29) Kaizuka, Y.; Groves, J. T. *Biophysical Journal* **2004**, *86*, 905–912.

(30) Xu, L.; Kuan, S. L.; Weil, T. *Angewandte Chemie International Edition* **2021**, *60*, 13757–13777.

(31) Wieneke, R.; Tampé, R. *Angewandte Chemie International Edition* **2019**, *58*, 8278–8290.

(32) Zhao, D.; Huang, Z. *Bioinorganic Chemistry and Applications* **2016**, *2016*, ed. by Brabec, V., 8206854.

(33) Meng, L.; Liu, Y.; Yin, X.; Zhou, H.; Wu, J.; Wu, M.; Yang, L. *Applied Biochemistry and Biotechnology* **2020**, *190*, 880–895.

(34) Woestenenk, E. A.; Hammarström, M.; van den Berg, S.; Härd, T.; Berglund, H. *Journal of Structural and Functional Genomics* **2004**, *5*, 217–229.

(35) Zhu, S.; Gong, C.; Ren, L.; Li, X.; Song, D.; Zheng, G. *Applied Microbiology and Biotechnology* **2013**, *97*, 837–845.

(36) Maza, J. C.; García-Almedina, D. M.; Boike, L. E.; Hamlish, N. X.; Nomura, D. K.; Francis, M. B. *ACS Central Science* **2022**, *8*, 955–962.

(37) Ramsey, A. V.; Bischoff, A. J.; Francis, M. B. *Journal of the American Chemical Society* **2021**, *143*, 7342–7350.

(38) Kondo, T.; Chen, W. J.; Schlau-Cohen, G. S. *Chemical Reviews* **2017**, *117*, 860–898.

(39) Huang, W. Y.; Chiang, H.-K.; Groves, J. T. *Biophysical Journal* **2017**, *113*, 1807–1813.

(40) Lin, W.-C.; Iversen, L.; Tu, H.-L.; Rhodes, C.; Christensen, S. M.; Iwig, J. S.; Hansen, S. D.; Huang, W. Y. C.; Groves, J. T. *Proceedings of the National Academy of Sciences* **2014**, *111*, 2996–3001.

(41) Penwell, S. B.; Ginsberg, L. D. S.; Noriega, R.; Ginsberg, N. S. *Nature Materials* **2017**, *16*, 1136–1141.

(42) Groves, J. T.; Ulman, N.; Boxer, S. G. *Science* **1997**, *275*, 651–653.

(43) Lohmüller, T.; Iversen, L.; Schmidt, M.; Rhodes, C.; Tu, H.-L.; Lin, W.-C.; Groves, J. T. *Nano Letters* **2012**, *12*, 1717–1721.

(44) Iversen, L. et al. *Science* **2014**, *345*, 50–54.

(45) Yudovich, S.; Marzouqe, A.; Kantorovitsch, J.; Teblum, E.; Chen, T.; Enderlein, J.; Miller, E. W.; Weiss, S. *Biophysical Journal* **2022**, *121*, 2624–2637.

(46) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nature Chemical Biology* **2015**, *11*, 326–331.

(47) Crochet, A. P.; Kabir, M. M.; Francis, M. B.; Paavola, C. D. *Biosensors and Bioelectronics* **2010**, *26*, 55–61.

(48) Dai, J.; Knott, G. J.; Fu, W.; Lin, T. W.; Furst, A. L.; Britt, R. D.; Francis, M. B. *ACS Nano* **2021**, *15*, 8110–8119.

(49) Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. *Molecular Biotechnology* **2006**, *33*, 239–254.

(50) Lin, W.-C.; Yu, C.-H.; Triffo, S.; Groves, J. T. *Current Protocols in Chemical Biology* **2010**, *2*, 235–269.

(51) Edelstein, A. D.; Tsuchida, M. A.; Amodaj, N.; Pinkard, H.; Vale, R. D.; Stuurman, N. *Journal of Biological Methods* **2014**, *1*, e10.

(52) Carnell, M.; Macmillan, A.; Whan, R. In *Methods in Membrane Lipids*, Owen, D. M., Ed.; Springer New York: New York, NY, 2015, pp 255–271.

(53) Lin, J. J.; O'Donoghue, G. P.; Wilhelm, K. B.; Coyle, M. P.; Low-Nam, S. T.; Fay, N. C.; Alfieri, K. N.; Groves, J. T. *Biophysical Journal* **2020**, *118*, 2879–2893.

(54) Axmann, M.; Schütz, G. J.; Huppa, J. B. *JoVE* **2015**, e53158.

(55) Ensign, D. L.; Pande, V. S. *The Journal of Physical Chemistry B* **2010**, *114*, 280–292.

(56) Schrödinger Release 2021-4: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021.

(57) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(58) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, Pullman, B., Ed.; Springer Netherlands: Dordrecht, 1981, pp 331–342.

(59) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. *Journal of Chemical Theory and Computation* **2021**, *17*, 4291–4300.

(60) Martyna, G. J.; Tobias, D. J.; Klein, M. L. *The Journal of Chemical Physics* **1994**, *101*, 4177–4189.

(61) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *The Journal of Chemical Physics* **1992**, *97*, 2635–2643.

(62) Schrödinger Release 2022-1: BioLuminate, Schrödinger, LLC, New York, NY, 2021.

(63) Altman, D. G., *Practical statistics for medical research*; Chapman and Hall: London, 1991.

(64) Kirkwood, B. R.; Sterne, J. A. C., *Essential Medical Statistics, 2nd Edition*; Wiley-Blackwell: Oxford, 2003.