Title
Development of Scaffolds for Light Harvesting and Photocatalysis from the Coat Protein of Tobacco Mosaic Virus

Author
Dedeo, Michel Toussaint

Publication Date
2012

Peer reviewed|Thesis/dissertation

Development of Scaffolds for Light Harvesting and Photocatalysis from the Coat Protein of Tobacco Mosaic Virus

By

Michel Toussaint Dedeo

A dissertation submitted in partial satisfaction of the

requirements of the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Matt Francis, Chair
Professor Michelle Chang
Professor Seung-Wuk Lee

Spring 2012

Development of Scaffolds for Light Harvesting and Photocatalysis from the Coat Protein of Tobacco Mosaic Virus

by

Michel Toussaint Dedeo

Abstract

Development of Scaffolds for Light Harvesting and Photocatalysis
from the Coat Protein of Tobacco Mosaic Virus

by

Michel Toussaint Dedeo

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Matthew B. Francis, Chair

The utility of a previously developed TMV-based light harvesting system has been dramatically expanded through the introduction of reactive handles for the site-specific modification of the interior and exterior surfaces. Further experiments to reengineer the coat protein have produced structures with unique, unexpected, and useful assembly properties that complement the newly available surface modifications. Energy transfer from chromophores in the RNA channel of self-assembled TMV structures to the exterior was made possible by conjugation of acceptor dyes and porphyrins to the N-terminus. By repositioning the N-terminus to the pore through circular permutation, this process was repeated to create structures that mimic the light harvesting 1 complex of photosynthetic bacteria. To study and improve upon natural photosynthesis, closely packed chromophore arrays and gold nanoparticles were tethered to the pore of stabilized TMV disks through introduction of a uniquely reactive lysine. Finally, a dimeric TMV coat protein was produced to control the distribution and arrangement of synthetic groups with synergistic activity.

This work is dedicated to my parents,

Karen Haas and Vishnu

Table of Contents

Acknowledgements

Matt Francis, for providing boundless enthusiasm and encouragement. Your ambitious directives were balanced with space to explore new ideas and the unexpected findings that can make research so stimulating.

My TMV mentors, Rebekah Miller and Andrew Presley, whose efforts provided a solid foundation for all of the research herein.

Aaron Esser-Kahn, for encouraging me and sharing the molecular biology secrets that have let me transform TMV. Your dedication and enthusiasm towards science was an inspiration.

My collaborators, Rebecca Scheck and Karl Duderstadt, without whose work and help my contributions would be much less.

The Francis group. You have been my teachers, collaborators, and friends for almost six years. I couldn't have asked for a brighter, more stimulating group of people to work with. The willingness of everyone to share their time and expertise has let me accomplish more than I would have thought possible.

The Bertozzi and M. Chang groups, for generously sharing equipment and reagents, and patiently answering the questions which accompany my frequent interruptions of your work.

My friends in the woodshop, for teaching me to surf and facilitating the innumerable lab projects that require drills and saws. My friends on Fuego and at the Cal Sailing Club for helping me spend most of my out-of-lab time in the water.

My family, whose curiosity and open-mindedness has shown me how enjoyable learning and exploring can be.

Chapter 1: Synthetic Templates for Light Harvesting and Photocatalysis

1.1 Abstract

Light harvesting is the process by which photons are absorbed by donor chromophores and transferred to an acceptor, effectively increasing the acceptor's spatial and spectral absorption capability. In order to capture sufficient energy from sunlight, photosynthetic organisms arrange an impressively diverse collection of chromophores and photoprotective molecules around centralized electron transfer groups.[1,2] This process is achieved through the use of protein scaffolds that tune both the crucial spacing and orientation of the components in order to maximize the efficiency of inter-chromophore Förster resonance energy transfer (FRET). These systems can achieve enviably efficient energy conversion but their complexity is currently impossible to recreate synthetically.[3] (Ishizaki 2012) Rather, researchers have used these remarkable natural systems as inspiration to build a wide and growing variety of simpler artificial light harvesting structures with the goal of incorporating them into photovoltaic or photocatalytic devices. Despite this variety, the structures share a common requirement for nanoscale features because of the sensitive dependence of energy transfer on interchromophore distance and orientation.[4] Portions of this chapter have appeared in published form: Dedeo, M. T.; Finley, D. T.; Francis, M. B. Viral Capsids as Self-Assembling Templates for New Materials. *Prog Mol Biol Transl Sci.* **2011**, *103*, 353-92.

1.2 Photosynthesis in Purple Bacteria

Because they are relatively simple, purple bacteria such as *Rb. palustris* have become a model system for the study of photosynthesis (Figure 1-1a). They possess the best characterized photosynthetic apparatus, which consists primarily of two classes of antenna complexes. The first light harvesting 1 (LH 1) complex contains a ring of chromophores surrounding the reaction center, and is sufficient for photosynthesis. The reaction center is capable of a rapid turnover (1000 Hz), though the photons collected by chromophores in the LH1 complex do not approach this capacity.[5](Hu 2004) To increase the energetic and spatial cross-section of energy available to the reaction center, LH2 antenna complexes surround and funnel light into the central LH1 complex. By adjusting the chromophore composition of the LH2 complexes and inter-complex interactions, the organism can optimize the light harvesting based on the level and spectral properties of the ambient light.

The LH2 complex has a ninefold axis of symmetry because it is composed of an inner ring of nine α subunits and an outer ring of nine β subunits. Eighteen overlapping bacteriochlorophyll (BChl) molecules (B850, Figure 1-1c) are sandwiched between the two protein rings with a spacing of 4 Å, which leads to a strong electronic interaction.[6] An additional nine BChls are held by the outer β subunits with a spacing of 8-10 Å and absorb higher energy light (B800, Figure 1-1c) based on their different spacing and environment. Additionally, there are nine carotenoid molecules associated with the phytl tails of the outer BChls.[7]

The LH1 complex shares many structural characteristics with the LH2 complex, but differs in that it must host the reaction center and leave a gap for the diffusion of quinone electron shuttles. Because of this gap, it is not radially symmetric, but it too is composed of αβ-heterodimers. Each of the 16 heterodimer binds two BChl molecules (B875), holding them about 9 Å apart. As shown in Figure 1-1c, these are the last and lowest energy BChls through which energy flows before reaching the reaction center. The reaction center itself is composed of the light, medium, and heavy protein subunits, which arrange two symmetric 'branches' of electron transport cofactors consisting of the P870 dimer, two BChl molecules, bacteriopheophytin molecules, ubiquinone, a carotenoid, and a non-reducing iron atom.[8]

The precise arrangement of these chromophores permits the rapid and efficient transfer of energy from peripheral antenna complexes to the reaction center. In the longest possible path, light absorbed by B800 in the periphery of the LH2 complex from *R. acidophila* is transferred to the B850 ring within 800 fs.[9] The jump from LH2 to LH1 to the reaction center is slower, taking 5-25 ps. Importantly, this timescale is orders of magnitude faster than the lifetime of an isolated LH2 complex, so little absorbed nergy would be dissipated.

1.3 Non-Viral Templates for Light Harvesting and Photocatalysis

Multiporphyrin assemblies were among the first artificial light harvesting systems studied because they offered a simplified model of the chlorophyll antenna found in photosynthetic organisms.[10] Since then, both the understanding of the natural systems and complexity of the models have increased to the point where it has been possible to construct cyclic porphyrin arrays as analogues of LH 2 antenna complexes[11] (Figure 1-2a) and artificial chlorosome light harvesting antenna containing thousands of self-assembling porphyrins.[12] Light harvesting dendrimers were

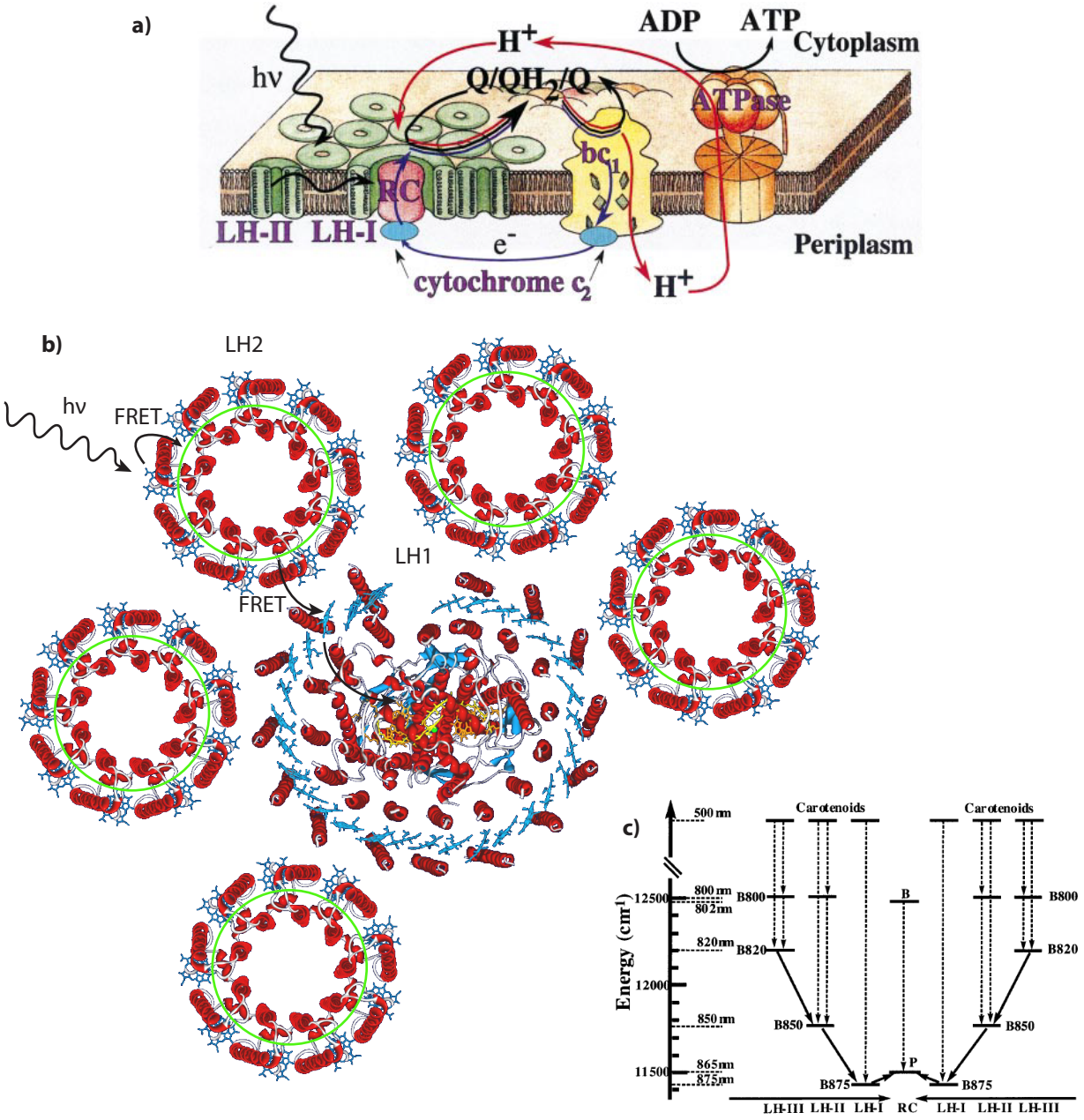

Figure 1-1. Photosynthetic systems in purple bacteria. (a) Energy from absorbed light is used to create a proton gradient across the membrane. The flow of protons down the gradient (red arrows) drives the regeneration of ATP by the ATPase enzyme (orange). Light harvesting 2 (LH 2) complexes absorb light and transfer energy to a central LH 1 complex containing a reaction center (RC) that creates an electron-hole separation. The flow of electrons is shown in blue. They are shuttled from the RC to the ubiquitinone-cytochrome bc complex (yellow) by quinone, and back to the RC by cytochrome c2. (b) A top-down view of light harvesting systems shows the spatial arrangement of chromophores. LH2 complexes (from *Rps. acidophila*) collect and shuttle absorbed energy to a LH1 complex (from *Rb. palustris*), which is then transferred to the cental reaction center. Protein is shown in red, bacteriocholorphyll (BChl) is shown in blue, and the inner ring of BChl molecules in the LH2 complexes is shown as a green circle. (c) Energy diagram shows excitation funneling from high energy chromophores in the periphery to lower energy chromophores and ultimately the reaction center. The B800-B850 chromophores are located in LH2/LH3 complexes, while the B875 are in LH1. Vertical dashed lines indicate intra-complex transfer and diagonal solid lines indicate inter-complex transfer. Adapted from Hu et al., 2004.[5]

developed as an extension of early multiporphyrin assemblies. They are repetitively branched molecules that can form highly symmetric, monodisperse, often spherical shapes (Figure 1-2b). By choosing which chromophores occupy focal points, branches and/or chain ends, spatial control over relative chromophore position can be achieved, directing energy efficiently toward a central acceptor.[13] This central-acceptor arrangement mimics the biological light harvesting LH1 system.[14] Drawbacks of dendritic systems include a labor-intensive synthesis and the potential for contact quenching from adjacent branches.

DNA origami has allowed the computer-aided design of remarkably complex two and three dimensional nanoscale structures.[15] These can be assembled from a long single stranded DNA (e.g. M13 phage genome) molded and stapled together by short designed DNA oligomers. Crucially, the structures can be functionalized at almost any location by incorporating strands that partially hybridize and protrude from the surface to provide a "sticky end" for attachment of complementary DNA, a unique reactive group, or an already-conjugated molecule. Recent applications of this technology to light-harvesting have allowed the study of distance-dependence in competitive FRET processes[16] as well as the design and optimization of a three dye FRET relay (Figure 1-2c).[17] The utility of DNA as a scaffold lie in its unsurpassed control over the number of chromophores in a system, their relative positioning, and the distance between them. However, the relatively high cost of DNA synthesis makes it is difficult to imagine scaling up light harvesting DNA origami past model systems used to study FRET.

A number of scaffolds have been developed recently to try to address the problem of scalability. Like many other light harvesting systems, they rely on the self-assembly of inexpensive, synthetically tractable subunits. Clay nanoparticles feature planar surfaces with periodic anionic features that can bind appropriately functionalized donor and acceptor porphyrins at distances

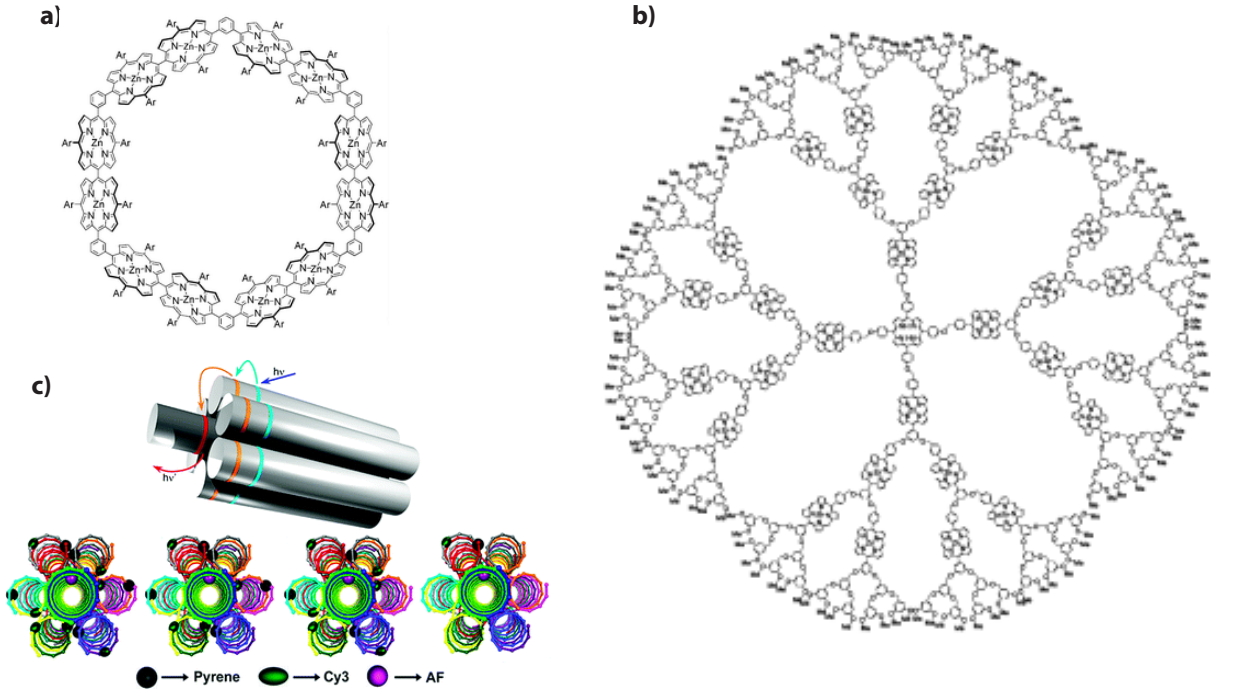

Figure 1-2. Artificial light harvesting systems. (a) cyclic porphyrin arrays can mimic the arrangement of chromophores in LH2 antenna complexes.[11] (b) light harvesting dendrimers can array donor and relay dyes around a central acceptor.[14] (c) DNA origami can provide precice control over relative number and spacing of light harvesting chromophores. (dutta 2011) Figures were adapted from referenced publications.

and orientations that encourage efficient energy transfer (Figure 1-3a).[18] Chromophores covalently attached to the surface of cross-linked micelles have displayed minimal quenching and efficient energy transfer to an electrostatically bound acceptor.[19] Porous metal-organic frameworks (MOFs) were designed to incorporate both "antenna" molecules, which initially absorb most light, and secondary chromophores to which excitons can be delivered from these antennae (Figure 1-3b). Efficient transfer and high quantum yields were reported, rendering these materials potential candidates for conversion of solar to electrical or chemical energy.[20]

Simple light harvesting systems have already been incorporated into excitonic solar cells to broaden their spectral response. Dye sensitized 'Graetzel' cells are photovoltaic devices composed of semiconductor particles sensitized with a chromophore. Most commonly, this is titania sensitized with a ruthenium complex. The sensitizing dye must be capable of being electronically coupled with the semiconductor, which has limited the choices to dyes with incomplete spectral coverage and relatively low extinction coefficients. To overcome this, donor chromophores have been added either to the electrolyte solution[21] (Figure 1-3c) or immobilized on the surface[22] to relay energy to sensitizing dyes via Förster or Dexter mechanisms. In an analogous system using dye-sensitized molecular wires, a donor in solution increased the quantum yields for red photons.[23] Further improvements should be possible by localizing donors close to their acceptors, incorporating a third relay dye, and preventing potentially significant quenching from electrolytes such as I_3^-.

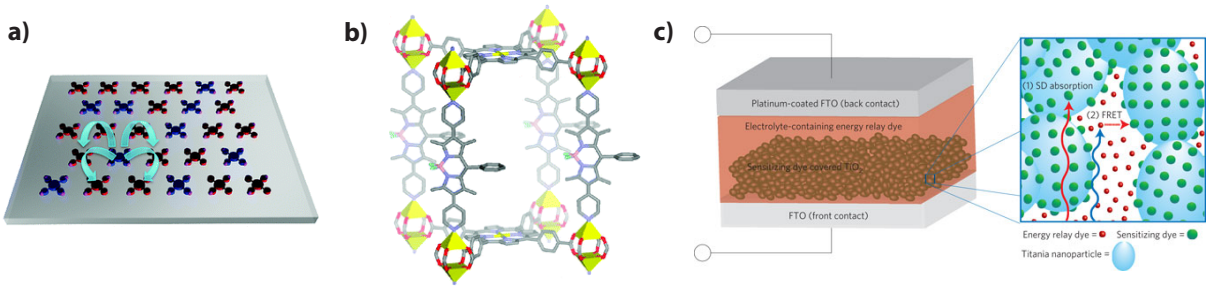

Figure 1-3. Additional artificial light harvesting systems (a) planar clay nanoparticles can posses crystal lattice spacings capable of adsorbing appropriately functionalized porphyrins while allowing energy transfer betwen them.[18] (b) metal organic frameworks provide a new self-assembling structure to array donors and acceptors.[20] (c) The spectral absorbance of Graetzel cells can be expanded by incorporating donor dyes into the electrolyte solution surrounding the titania nanoparticles.[21] Figures were adapted from referenced publications.

1.4 Viral Templates for Light Harvesting and Photocatalysis

1.4.1 Viruses as Nanoscale Building Blocks

Natural photosynthetic systems have inspired a variety of successful artificial light harvesting systems as discussed above. However, these efforts are often hindered by the difficulties associated with the synthesis of organic frameworks containing multiple chromophores and the lack of available scaffolds that are rigid enough to maintain discrete inter-chromophore distances over large length scales. Viral capsids provide promising alternative scaffolds for the positioning of light harvesting molecules, as they readily assemble into regular, highly symmetric, rigid shapes with an ideal periodicity for FRET to occur between attached components (Figure 1-4).

The bulk of the functional materials in nature involve self-assembling proteins, which

typically consist of defined three-dimensional shapes and complex interfaces that come together through large surface interactions. As a result of these properties, protein-based materials can form rigid structures with length scales ranging from nanometers to microns. Theses self-assembly properties are often challenging to achieve using other building blocks, and are important for the

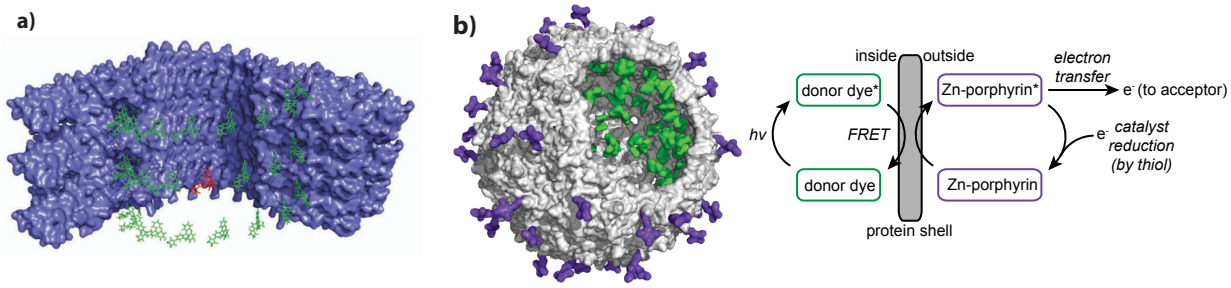

Figure 1-4. Viral capsid-templated chromophore assemblies that mimic some aspects of natural light harvesting systems have been prepared using (a) Tobacco Mosaic Virus (TMV),[54] (b) bacteriophage MS2,[45] and bacteriophage M13.[49]

generation of materials that can propagate nanoscale features into bulk materials with much larger dimensions. Proteins also offer the opportunity for synthetic tailoring, as site-directed mutagenesis can be used to introduce new amino acids chosen from the natural pool or a rapidly growing list of artificial groups.[46] This opens the door to the use of often sophisticated organic synthesis methods to introduce new functional groups, from polymers to catalysts to nanoparticles, on the surfaces of the proteins either before or after the assembly occurs. As an additional advantage, many proteins can be produced on large scale through recombinant expression in bacteria or propagation in plants.

The protein shells of viruses, or "capsids", provide the most well-studied class of self-assembling proteins that have been used for materials science.[24-33] In nature, these structures serve to protect the genomic nucleic acids from degradation and are often active participants in the cell-specific delivery of the genes required for propagation.[34] In general, viral capsids possess a high degree of symmetry and are assembled from many copies of a small number of distinct protein monomers. Many of these features have been resolved through X-ray crystallography cryo-electron microscopy. A large class of viral capsids exists as hollow spherical structures with icosahedral symmetry, providing a high volume-to-surface area ratio for efficient genome encapsulation. Capsids with this geometry exhibit a characteristic "*T*" number, which is related to the total number of protein monomers (*T* x 60) that comprise their structure.[20] As viral capsids with varying *T* numbers are known, it is possible to choose structures with different sizes and complexities to meet the synthetic challenge at hand. A number of viruses also exist as high aspect ratio rods, housing their nucleic acids in extended helices or circular loops. In some cases, purified viral capsid proteins can be used to generate more than one self-assembled state: tobacco mosaic virus (TMV) monomers can assemble into both disks or rods,[35,36] the brome mosaic virus (BMV) can form spherical capsids with either *T*=1 or *T*=3 symmetry,[37] and the length of bacteriophage M13 can be altered by altering its genome size.[38] A structural comparison of several viral capsids used for materials science applications appears in Figure 1-5, and Table 1-1 provides a table listing many of their key properties.

Taken together, viral capsids offer many different architectures for constructing light harvesting and photocatalytic systems. A number of the approaches detailed in the following sections use the intact particles without removing the nucleic acids inside, while some have shown that the genome

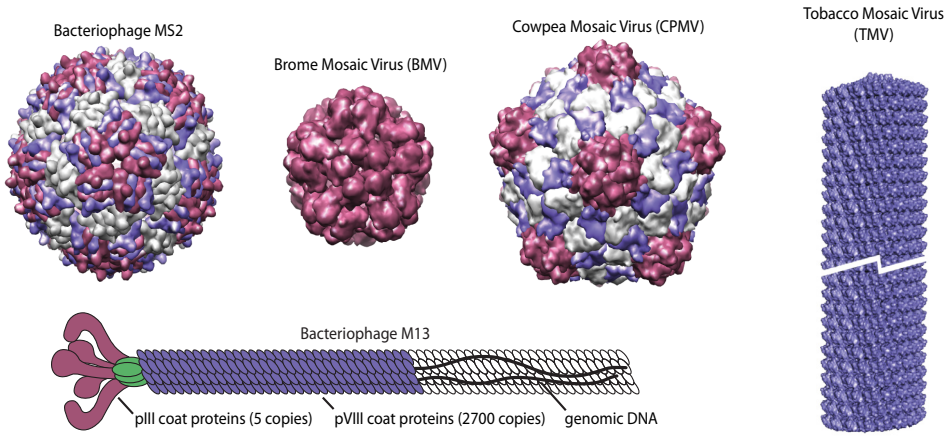

Figure 1-5. Structural comparison of viral capsids used for materials applications. Additional structural details are provided in Table 1.

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

* from *Methanocaldococcus jannaschii*

Table 1-1. Table of key properties for viral capsids commonly used to build new materials.

can be removed to provide new spaces for chemical modification. Site-directed mutagenesis of the coat protein permits the conjugation of groups to precise locations. Multiple groups can be incorporated site-selectively by taking advantage of the rapidly growing set of chemoselective bioconjugation reactions.[39,40] A summary of the chemical strategies that are commonly used for viral capsid modification appears in Figure 1-6. The following examples detail the use of the capsids to integrate multiple chromophores and electron transfer groups, with the goal of creating complex functional assemblies with efficient energy collection and transduction capabilities.

1.4.2 Arraying Chromophores and Catalysts on Cowpea Mosaic Virus (CPMV) and Hepatitis B Virus (HBV)

CPMV was one of the first viral capsids to be covalently modified to template synthetic chromophores.[41] In 2002, the Finn and Johnson groups began their work with virus based materials by exploring the reactivity of the wild-type capsid with a maleimide fluorescein derivative. They observed that roughly 60 dyes could be attached to a capsid comprised of two different monomers. They then installed an additional, much more reactive cysteine in a loop on the exterior surface of one of the monomers, which they found could be alkylated using fewer equivalents and shorter reaction times. This difference in reactivity allowed them to perform sequential reactions, selecting for the introduced cysteine with moderate conditions, followed by more forcing conditions to

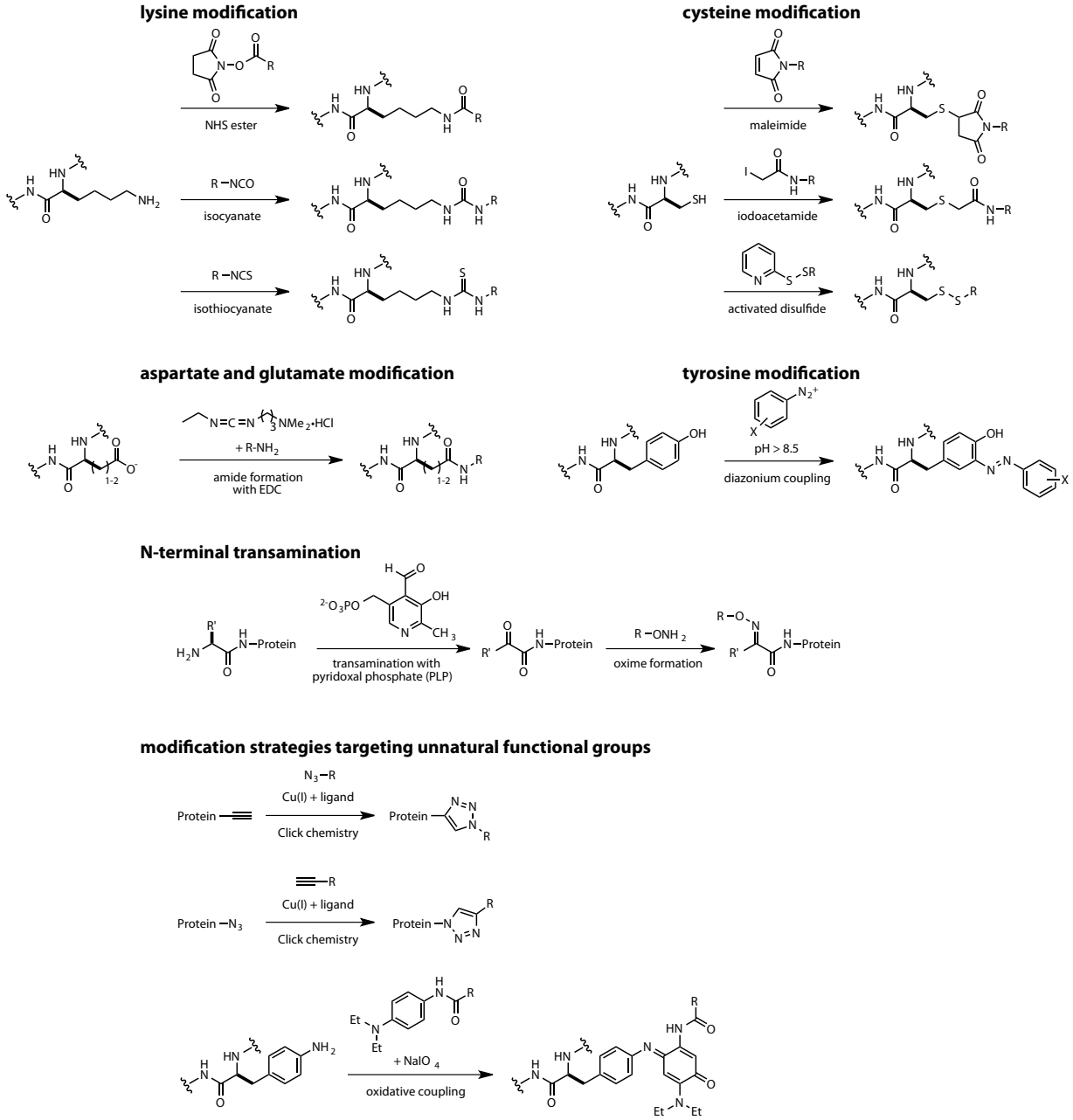

Figure 1-6. Summary of the common chemical strategies that are used to attach new functional groups to viral capsid surfaces.

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

In subsequent papers published the same year, the Finn and Johnson groups explored the reactivity of lysine residues and pushed the envelope for chromophore bioconjugation.[42] They showed that up to 240 chromophores could be installed using NHS-ester chemistry. Lysine bioconjugation was used to introduce a biotin tag, which caused the formation of extended networks upon addition of avidin.[43] They also demonstrated the sequential and independent labeling of

cysteine and lysine residues.

Taking a different approach with another 30 nm icosahedral virus, the Finn group explored non-covalent catalyst attachment to Hepatitis B (HBV). To do this, they introduced a His_6 tag, commonly used for protein purification over immobilized nickel ions, and instead used it to bind iron protoporphyrin IX (or heme) groups from solution.[44] Despite the bulk of both the His_6 tag and the heme, they were able to attach an average of 80 copies to each capsid. The attached hemes were reported to behave as independent redox entities despite their proximity.

1.4.3 Photocatalytic Systems Based on Bacteriophage MS2

Photosynthesis uses the energy of absorbed light to form high-energy electrons, which are ultimately used to generate chemical bonds. Artificial light harvesting systems can also be designed to generate high energy electron carriers, for example reducing the soluble enzymatic redox cofactors $NAD(P)^+$ to NAD(P)H. The 28 nm icosahedral protein shell of bacteriophage MS2 provides two surfaces with reactive handles to array donor chromophores within FRET distance of both each other and a photocatalytic acceptor. In a report by the Francis group, an array of 180 donor chromophores was created by modifying a cysteine introduced to the interior of the empty capsid.[45] A zinc porphyrin was chosen as a suitable acceptor, and conjugated to the unnatural amino acid *p*-aminophenylalanine (*p*aF) installed on the exterior surface through the Schultz *in vivo* method.[46,47] This exterior modification took advantage of a new oxidative coupling reaction also developed in the Francis lab (Figure 1-6). Because the protein shell is only about 1 nm thick, the separation between the interior donors and the exterior acceptors was well within FRET distance (Figure 1-6b). The large number of exterior reactive sites (180) allowed the researchers to screen a variety of catalyst loadings (20-120 zinc porphyrins) to study the energy transfer process.

The flexibility of the system also allowed for a comparison between donors designed to sensitize the porphyrin at either of two wavelengths (350 or 500 nm) where the porphyrin did not strongly absorb. Upon illuminating the donor, the energy transfer to the zinc porphyrin could be measured by the reduction of methyl viologen, which produced a colorimetric response. The catalytic cycle was completed by addition of the sacrificial reductant, β-mercaptoethanol, to reduce the resulting porphyrin radical cations. When coupled to the chromophore array, the porphyrins were capable of photoreduction over a broader set of input wavelengths. Illumination at the peak donor absorption band produced an almost fourfold increase in the reduction of methyl viologen compared to a system with no donor.

1.4.4 Light Harvesting and Photocatalytic Systems Based on Bacteriophage M13

Bacteriophage M13 is another helical virus that has been used to template arrays of dyes and photocatalysts. Its utility for light harvesting was first demonstrated by Scolaro and coworkers, who found that cationic porphyrins would spontaneously adsorb to its surface.[48] These porphyrins served as FRET acceptors for energy absorbed by a surface tryptophan introduced through mutation. However, the light harvesting efficiency was limited both by the low extinction coefficient of tryptophan and the poor overlap between donor emission and acceptor excitation.

The Belcher lab has created light harvesting catalytic systems with M13 as well, taking advantage of their extensive experience interfacing phage with inorganic particles to design complex structures. They began by attaching zinc porphyrins to the 2,700 pVIII coat proteins that run the length of the virus.[49] These were introduced by targeting the N-terminus and a lysine on each monomer, providing distances of 1-2.4 nm between the reactive sites. This was close enough

to allow for significant amounts of electronic coupling. Using transient pump-probe spectroscopy, exciton delocalization was indeed observed in the porphyrin system as quenched fluorescence emission and decreased lifetime. Building on this work in a subsequent report, Belcher and co-workers used phage display to develop an M13 mutant that could also bind iridium oxide (IrO_2) clusters after porphyrin attachment.[50] These clusters could then nucleate the encapsulation of the virus-porphyrin assembly with an iridium oxide shell under mild conditions. The porphyrins then served as photosensitizers, oxidizing the iridium shell, which in turn oxidized water. The integration of porphyrins and catalysts with the viral template was found to improve the quantum yield relative to systems with no porphyrin or no M13 template, highlighting the potential of using self-assembling protein structures to arrange nanoscale components for optimal solar energy collection and conversion.

1.4.5 The Tobacco Mosaic Virus (TMV)

The coat protein of the Tobacco Mosaic Virus (TMV) has appeared in several reports as a means to arrange large arrays of synthetic chromophores. Light harvesting systems have been built from both the intact virus and the partially assembled coat proteins, either extracted from viruses or produced recombinantly. The self-assembly properties of TMV have been thoroughly studied, allowing researchers to access different metastable structures by changing the pH and ionic strength of the solution.[35] The most appealing structures for synthetic manipulation have been double disks, consisting of two stacked rings of 17 monomers each, and helical rods, which contract slightly to a periodicity of 16.33 monomers/turn and can range from tens to thousands of nanometers in length. The subunits in both structures are tightly packed so that attaching a chromophore to the same position on each protein yields a well ordered array with interchromophore spacing well suited for energy transfer via FRET.

TMV has multiple distinct surfaces but few inherently reactive sites for covalent modification, which researchers have overcome in a number of creative ways. In the first example of using covalent chemistry to produce TMV-based materials, the Francis lab targeted native tyrosines on the exterior surface using a newly developed diazonium coupling reaction, and targeted native carboxylic acids in the pore of the rod to insert polymers, chromophores, and chelating groups.[51] [111] Shortly following, two reports were published introducing cystine mutants to facilitate the modification of different surfaces of the TMV rod. The groups of Harris and Culver introduced a cystine on the exterior, which they were able to functionalize with chromophores.[52] They also developed a clever surface immobilization strategy in which the end of the genomic RNA was revealed and hybridized to a complementary strand patterned on a gold surface. To modify the pore of TMV, the Majima group introduced a cysteine at position 99 or 100, which they were able to modify with a pyrene maleimide when the protein was disassembled into monomers.[53] Rather than interfering with the assembly, the pyrene seemed to encourage it by promoting the formation of longer rods than were formed from unmodified protein. Fluorescence measurements of the pyrene-containing rods indicated the formation of excimers, suggesting the increased assembly may be due to favorable π-stacking in the pore.

The Francis group developed a system to collect light across a broad spectrum by creating a closely spaced array of mixed chromophores with spectral properties that allowed for FRET down an energy gradient, analogous to natural photosynthetic systems.[54] Recombinantly derived TMV coat protein lacks genomic RNA, which usually winds through an internal channel in the assembled rods. Taking advantage of this newly available ‘surface’ for covalent modification, the

Francis group introduced a cysteine into position 123 to allow the straightforward modification of monomers with a variety of commercially available maleimide dyes. Subsequent assembly into rods or disks positioned the dyes to within 1.5 nm, which is well within the Förster radius for efficient energy transfer between identical dyes as well as donor-acceptor neighbors (Figure 1-4a).

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

More detailed analyses of the energy transfer rates were performed using time resolved spectroscopic methods in a subsequent collaboration with the Fleming group.[55] Additional studies performed with the Geissler group showed that energy transfer was possible not only to the nearest neighbor, but to neighbors in all directions, providing many paths for energy to travel from the light-absorbing chromophore to the final acceptor.[56] This property is crucial to provide defect tolerance arising from the photobleaching of pigments.

In addition to the synthetic chromophores discussed above, the empty RNA channel has played host to molecules as large as porphyrins, as reported by the Majima group.[57] By introducing a cysteine at position 127, they created a similar system in which their donor zinc porphyrins and acceptor freebase porphyrins were conjugated to TMV via a maleimide linker. This held them in close proximity to one another upon monomer assembly into short rods. One advantage of using porphyrins is their very high extinction coefficient, though the large gap between excitation and emission prohibited significant donor-to-donor FRET, and therefore long-distance energy transfer.

Having established the possibility of collecting light with a TMV based system, the next challenges were converting the light into useful energy and collecting that energy on a large scale. To address the first challenge, the Francis group redesigned the TMV coat protein to introduce a new reactive site in the pore.[58] The ultimate goal was to place an electron transfer catalyst there to act as a final energy acceptor from the surrounding chromophores in the RNA channel. This project is detailed in chapter 3.

The second challenge involves interfacing TMV-based light harvesting systems with devices to collect the high energy electrons produced, either in the form of chemical bonds or as a photocurrent. For the latter system, an electrically conductive component might be required to connect light harvesting components to an electrode. Carbon nanotubes have a number of properties that would make them ideal in this role, such as high conductivity and a demonstrated ability to interface with a variety of photocatalytic materials.[59,60] Since TMV and nanotubes do not associate on their own, the Francis group developed a method to pre-treat both TMV rods and carbon nanotubes such that they would bind together. In parallel, an exterior tyrosine of TMV was modified to display aldehydes using a diazonium coupling method,[51] while the carbon nanotubes were solubilized with a surfactant-like pyrene-PEG-aminooxy bifunctional linker, which bound the nanotubes with the hydrophobic pyrene.[61] pon mixing, the aminooxy and aldehydes reacted to form oximes, linking the TMV and nanotubes. In this proof of principle work, PEG was chosen to solubilize the nanotubes despite its poor conductivity. In the future, it would need to be replaced with a more conductive polymer to transfer electrons from TMV to nanotubes.

1.5 Conclusion

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

1.6 Literature Cited

1. Nelson, N.; Ben-Shem, A. *Nature reviews. Molecular cell biology* **2004**, *5*, 971–82.
2. Freer, A.; Prince, S.; Sauer, K.; Papiz, M.; Lawless, A. H.; McDermott, G.; Cogdell, R.; Isaacs, N. W. *Structure* **1996**, *4*, 449–462.
3. Ishizaki, A.; Fleming, G. R. *Condensed Matter Physics* **2012**, *3*, 333–361.
4. Albrecht, C. *Principles of fluorescence spectroscopy, 3rd Edition, Joseph R. Lakowicz, editor;* 2008; Vol. 390
5. Hu, X.; Ritz, T.; Damjanovic, A.; Autenrieth, F.; Schulten, K. *Quarterly Reviews of Biophysics* **2002**, *35*, 1–62.
6. Riggsbee, C. W.; Deiters, A. *Trends in Biotechnology* **2010**, *28*, 468-475.
7. Prince, S. M.; Papiz, M. Z.; Freer, A. A.; McDermott, G.; Hawthornthwaite-Lawless, A. M.; Cogdell, R. J.; Isaacs, N. W. *Journal of Molecular Biology* **1997**, *268*, 412-23.
8. Heathcote, P.; Fyfe, P. K.; Jones, M. R.; *Trends in Biochemical Sciences* **2002**, *27* ,79-87.
9. Scholes, G. D.; Fleming, G. R.; Olaya-Castro, A.; van Grondelle, R. *Nature chemistry* **2011**, *3*, 763–74.
10. Milgrom, L. R. *J. Chem. Soc., Perkin Trans.***1983**, *1*, 2535–2539.
11. Nakamura, Y.; Aratani, N.; Osuka, A. *Chemical Society reviews* **2007**, *36*, 831–45.
12. Chappaz-Gillot, C.; Marek, P. L.; Blaive, B. J.; Canard, G.; Bürck, J.; Garab, G.; Hahn, H.; Jávorfi, T.; Kelemen, L.; Krupke, R.; Mössinger, D.; Ormos, P.; Reddy, C. M.; Roussel, C.; Steinbach, G.; Szabó, M.; Ulrich, A. S.; Vanthuyne, N.; Vijayaraghavan, A.; Zupcanova, A.; Balaban, T. S. *J. Am. Chem. Soc.* **2012**, *134*, 944–54.
13. Gilat, S. L.; Adronov, A.; Fréchet, J. M. J. *Angew. Chem. Int. Ed.* **1999**, *38*, 1422–1427.
14. Choi, M.-S.; Yamazaki, T.; Yamazaki, I.; Aida, T. *Angew. Chem. Int. Ed.* **2004**, *43*, 150–8.
15. Nangreave, J.; Han, D.; Liu, Y.; Yan, H. *Current opinion in chemical biology* **2010**, *14*, 608–15.
16. Stein, I. H.; Steinhauer, C.; Tinnefeld, P. *J. Am. Chem. Soc.* **2011**, *133*, 4193–4195.
17. Dutta, P. K.; Varghese, R.; Nangreave, J.; Lin, S.; Yan, H.; Liu, Y. *J. Am. Chem. Soc.* **2011**, *133*, 11985–11993.
18. Ishida, Y.; Shimada, T.; Masui, D.; Tachibana, H.; Inoue, H.; Takagi, S. *J. Am. Chem. Soc.* **2011**, *133*, 14280–6.
19. Peng, H.-Q.; Chen, Y.-Z.; Zhao, Y.; Yang, Q.-Z.; Wu, L.-Z.; Tung, C.-H.; Zhang, L.-P.; Tong, Q.-X. *Angew. Chem. Int. Ed.* **2012**, *51*, 2088–92.
20. Lee, C. Y.; Farha, O. K.; Hong, B. J.; Sarjeant, A. A.; Nguyen, S. T.; Hupp, J. T. *J. Am. Chem. Soc.* **2011**, *133*, 15858–61.
21. Hardin, B. E.; Hoke, E. T.; Armstrong, P. B.; Yum, J.-H.; Comte, P.; Torres, T.; Fréchet, J. M. J.; Nazeeruddin, M. K.; Grätzel, M.; McGehee, M. D. *Nature Photonics* **2009**, *3*, 406–411.

22. Hardin, B. E.; Sellinger, A.; Moehl, T.; Humphry-Baker, R.; Moser, J.-E.; Wang, P.; Zakeeruddin, S. M.; Grätzel, M.; McGehee, M. D. *J. Am. Chem. Soc.* **2011**, *133*, 10662-10667.

23. Shankar, K.; Feng, X.; Grimes, C. A. *ACS Nano* **2009**, *3*, 788–94.

24. Flenniken, M. L.; Uchida, M.; Liepold, L. O.; Kang, S.; Young, M. J.; Douglas, T. *Curr. Top. Microbiol. Immunol.* **2009**, *327*, 71-93.

25. Douglas, T.; Young, M. *Science* **2006**, *312*, 873-875.

26. Evans, D. J. *J. Mater. Chem.* **2008**, *18*, 3746-3754.

27. Fischlechner, M.; Donath, E. *Angew. Chem. Int. Ed.* **2007**, *46*, 3184-3193.

28. Lin, T. *J. Mater. Chem.* **2006**, *16*, 3673-3681.

29. Niemeyer, C. M. *Angew. Chem. Int. Ed.* **2001**, *40*, 4128-4158.

30. Singh, P.; Gonzalez, M. J.; Manchester, M. *Drug Dev. Res.* **2006**, *67*, 23-41.

31. Steinmetz, N. F.; Evans, D. J. *Org. Biomol. Chem.* **2007**, *5*, 2891-2902.

32. Uchida, M.; Klem, M.; Allen, M.; Suci, P.; Flenniken, M.; Gillitzer, E.; Varpness, Z.; Liepold, L.; Young, M.; Douglas, T. *Adv. Mater.* **2007**, *19*, 1025-1042.

33. Young, M.; Debbie, W.; Uchida, M.; Douglas, T. *Annu. Rev. Phytopathol.* **2008**, *46*, 361-384.

34. Harrison, S. C. "Principles of Virus Structure" in Virology, 27-44; ed. B.N. Fields *et al.*; Raven Press, New York; 1985.

35. Klug, A. *Phil. Trans. R. Soc. London, Ser. B* **1999**, *354*, 531-535.

36. Butler, P. J. G. *Phil. Trans. R. Soc. London, Ser. B* **1999**, *354*, 537-550.

37. Sun, J.; DuFort, C.; Daniel, M.; Murali, A.; Chen, C.; Gopinath, K.; Stein, B.; De, M.; Rotello, V. M.; Holzenburg, A.; Kao, C. C.; Dragnea, B. *Proc. Nat. Acad. Sci. USA* **2007**, *104*, 1354-1359.

38. Specthrie, L.; Bullitt, E.; Horiuchi, K.; Model, P.; Russel, M.; Makowski, L. *J. Mol. Biol.* **1992**, *228*, 720-724.

39. Tilley, S. D.; Joshi, N. S.; Francis, M. B. *Proteins: Chemistry and Chemical Reactivity* in the Wiley Encyclopedia of Chemical Biology. 1-16. (2008).

40. Hermanson, G. T. *Bioconjugate Techniques*, 2nd Ed. Academic Press, San Diego, **2008**.

41. Wang, Q.; Lin, T.; Tang, L.; Johnson, J. E.; Finn, M. G. *Angew. Chem. Int. Ed.* **2002**, *41*, 459-462.

42. Wang, Q.; Kaltgrad, E.; T.; Tang, L.; Johnson, J. E.; Finn, M. G. *Chem. Biol.* **2002**, *9*, 805-811.

43. Wang, Q.; Lin, T.; Johnson, J. E.; Finn, M. *Chem. Biol.* **2002**, *9*, 813-819.

44. Prashun, Jr., D.; Kuzelka, J.; Strable, E.; Udit, A. K.; Cho, S.; L, G. C.; er; Quispe, J. D.; Diers, J. R.; Bocian, D. F.; Potter, C.; Carragher, B.; Finn, M. *Chem. Biol.* **2008**, *15*, 513-519.

45. Stephanopoulos, N.; Carrico, Z. M.; Francis, M. B. Nanoscale Integration of Sensitizing Chromophores and Porphyrins with Bacteriophage MS2. *Angew. Chem. Int. Ed.* **2009**, *48*, 9498-9502.

46. Xie, J.; Schultz, P. G. A chemical toolkit for proteins—an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775-782.

47. Carrico, Z. M.; Romanini, D. W.; Mehl, R. A.; Francis, M. B. Oxidative coupling of peptides to a virus capsid containing unnatural amino acids. *Chem. Commun.* **2008**, 1205-1207.

48. Scolaro, L. M.; Castriciano, M. A.; Romeo, A.; Micali, N.; Angelini, N.; Lo Passo, C.; Felici, F. *J. Am. Chem. Soc.* **2006**, *128*, 7446-7447.

49. Nam, Y. S.; Shin, T.; Park, H.; Magyar, A. P.; Choi, K.; Fantner, G.; Nelson, K. A.; Belcher, A. M. Virus-Templated Assembly of Porphyrins into Light-Harvesting Nanoantennae. *J. Am. Chem. Soc.* **2010**, *132*, 1462-1463.

50. Nam, Y. S.; Magyar, A. P.; Lee, D.; Kim, J.; Yun, D. S.; Park, H.; Pollom, T. S.; Weitz, D. A.; Belcher, A. M. Biologically templated photocatalytic nanostructures for sustained light-driven water oxidation. *Nat. Nanotech.* **2010**, *5*, 340-344.

51. Schlick, T. L.; Ding, Z.; Francis, M. B. Dual-surface modification of the Tobacco mosaic virus. *J. Am. Chem. Soc.* **2005**, *127*, 3718-3723.

52. Yi, H.; Nisar, S.; Lee, S.Y.; Powers, M.A.; Bentley, W.E.; Payne, G.F.; Ghodssi, R.; Rubloff, G.W.; Harris, M.T.; Culver, J.N. Patterned Assembly of Genetically Modified Viral Nanotemplates via Nucleic Acid Hybridization. *Nano Lett.* **2005**, *5*, 1931-1936.

53. Endo, M.; Wang, H.; Fujitsuka, M.; Majima, T. Pyrene-Stacked Nanostructures Constructed in the Recombinant Tobacco Mosaic Virus Rod Scaffold. *Chem. Eur. J.* **2006**, *12*, 3735-3740.

54. Miller, R. A.; Presley, A. D.; Francis, M. B. Self-Assembling Light-Harvesting Systems from Synthetically Modified Tobacco Mosaic Virus Coat Proteins. *J. Am. Chem. Soc.* **2007**, *129*, 3104-3109.

55. Ma, Y.; Miller, R. A.; Fleming, G. R; Francis, M. B. *J. Phys. Chem. B.* **2008**, *112*, 6887-6892.

56. Miller, R. A.; Stephanopoulos, N.; McFarland, J. M.; Rosko, A. S.; Geissler, P. L. Francis, M. B. *J. Am. Chem. Soc.* **2010**, *132*, 6068-6074.

57. Endo, M.; Fujitsuka, M.; Majima, T. Porphyrin Light-Harvesting Arrays Constructed in the Recombinant Tobacco Mosaic Virus Scaffold. *Chem. Eur. J.* **2007**, *13*, 8660-8666.

58. Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. Nanoscale Protein Assemblies from a Circular Permutant of the Tobacco Mosaic Virus. *Nano Lett.* **2010**, *10*, 181-186.

59. Guldi, D. M.; Rahman, G. M. A.; Zerbetto, F.; Prato, M. Carbon nanotubes in electron donor-acceptor nanocomposites. *Acc. Chem. Res.* **2005**, *38*, 871-878.

60. Tasis, D.; Tagmatarchis, N.; Bianco, A.; Prato, M. Chemistry of carbon nanotubes. *Chem. Rev.* **2006**, *106*, 1105-1136.

61. Holder, P.; Francis, M. Integration of a self-assembling protein scaffold with water-soluble single-walled carbon nanotubes. *Angew. Chem. Int. Ed.* **2007**, *46*, 4370-4373.

Chapter 2: Modification of the Exterior Surface of TMV Disks and Rods with Photocatalytic Groups

2.1 Abstract

The theoretical feasibility of energy transfer from the RNA channel of TMV to the exterior was established, and three approaches were developed for installing a reactive aldehyde/ketone on the exterior. This energy transfer was confirmed using protein modified in the RNA channel and on the exterior aldehyde with commercially available chromophores. Following the success of this model system, photocatalytic porphyrins were site-specifically conjugated to the exterior of light harvesting rods. Multiple systems with donor chromophores and porphyrins were then constructed and characterized before it became apparent that the porphyrin was incompatible with TMV and would not be able to form well-defined conjugates. Portions of this chapter have appeared in published form: Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008,** *130*, 11762-11770.

2.2 Distance Constraints on Energy Transfer to Acceptors

In order to use the energy collected by a light-harvesting chromophore array inside TMV rods, it must be transferred to a photocatalyst acceptor via FRET, as in natural photosynthetic systems. This process is highly dependent on the distance between donor and acceptor, imposing significant constraints on the architecture of artificial photosynthetic systems. Once excited, these photocatalysts must also either come into contact with a soluble electron shuttle, such as methyl viologen, or be in intimate contact with an electrode to perform the desired redox reactions. The distance that excitation can migrate through a TMV-based light harvesting array before dissipation provides further constraints.

In previous experiments, the intensity of an acceptor's emission would increase with the number of surrounding donors, until a D:A ratio of 101:1.[1] At this point, the antenna effect plateaued, which suggests that energy can reach an acceptor if it is absorbed by a donor in the three turns above or below it (Figure 2-1). The efficiency, on the other hand, dropped off at D:A ratios above 33:1, suggesting that in an ideal case, photocatalyst acceptors should be located at least as frequently as every other turn. Because of this observation, the ends of the rods would be expected to be poor sites for the location of photocatalysts, since this would put them in range of only a small fraction of donor chromophores (<5% for a 300 nm rod).

While the interior channel has been reported to host large planar molecules,[2] it would be exceedingly difficult to interface these with an electrode. The 4 nm pore would also likely be too crowded after conjugation of photocatalysts to allow efficient diffusion of electron shuttle molecules. The exterior of the rod is then the preferred location for photocatalysts because it provides proximity to all chromophores in a rod and ready solvent accessibility.

To estimate the feasibility of this energy transfer from a chromophore attached to S123C in the RNA channel and an acceptor attached to the exterior, theoretical FRET efficiencies were calculated. Building on previous work by Miller *et al*.,[1] the relevant distances were approximated

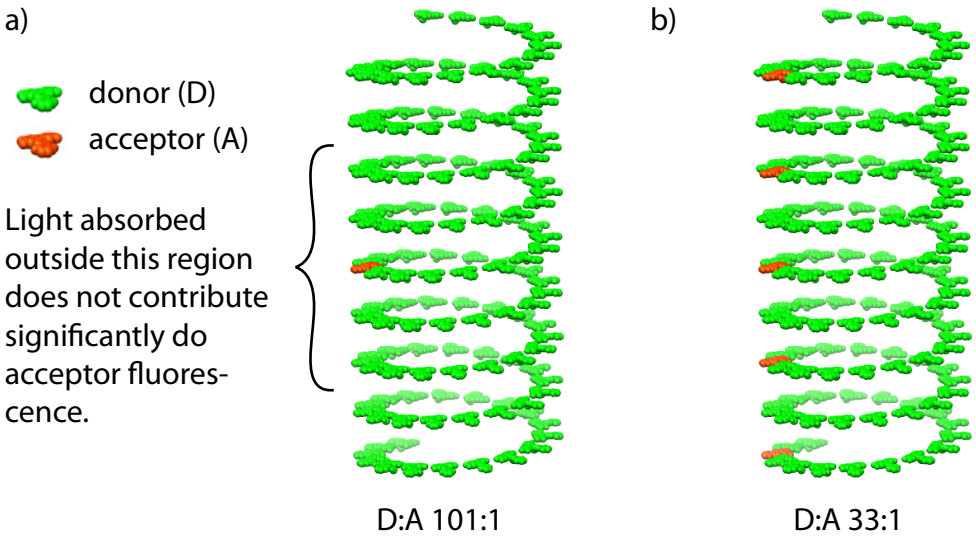

Figure 2-1. Range of energy transfer for excitation in TMV scaffolded OG array. (a) donor contribution to acceptor fluoresence plateaus at a D:A ratio of 101:1 , delineated by bracket. This suggests excitation can travel through about three turns of the helix, or 7 nm, before dissipating. (b) This cartoon illustrates a D:A ratio of 33:1. This is the ratio for which optimal efficiency in rod systems has been observed.

from the fiber diffraction structure of TMV rods (PDB 2tmv) and the highest-resolution crystal structure of disks (PDB 1EI7). The calculations were performed on the experimentally tractable donor:acceptor pair of Oregon Green:AlexaFluor 594.

The Förster radius is a useful benchmark for FRET pairs and can be used to calculate the efficiency of transfer processes.[3] It corresponds to the distance between dyes at which the rate of energy transfer from donor to acceptor is equal to the decay rate of the donor in the absence of an acceptor. At this distance, half of the donor excitations are dissipated by FRET to the acceptor, and the remainder are dissipated via emission and non-radiative decay. The Förster radius can be calculated using the following equation:

$$R_0(\text{Å}) = 0.211(\kappa^2\, \eta^{-4}\, \Phi J)^{1/6} \qquad (1)$$

The orientation factor κ^2 accounts for the relative orientation of their dipole moments, and is approximated as 2/3 for two freely rotating chromophores. This factor can range from 0 for orthogonal dipoles to 4 for parallel dipoles oriented in the same direction. The donor quantum yield Φ represents the ratio of photons emitted to those absorbed in the absence of an acceptor. At rod buffer conditions, Oregon Green was measured to have a quantum yield of 0.17, which is significantly lower than the ~0.7 - 0.8 reported in the product literature (Figure 2-2). The refractive index of the media, η, is 1.33 for aqueous samples, though because proteins offer a more polarizable environment, values up to 1.6 can be used. J is the spectral overlap integral where

$$J = \int_0^\infty F_d(\lambda)\varepsilon_A(\lambda)\, \lambda^4\, d\lambda \qquad (2)$$

in units of $M^{-1}\ cm^{-1}\ nm^4$. $F_d(\lambda)$ is the emission of the donor at λ, with the total fluorescence normalized to one, and is unitless. The acceptor's absorbance $\varepsilon_A(\lambda)$ has units of $M^{-1}\ cm^{-1}$. Despite the fact that FRET is a non-radiative process, the efficiency depends on the fraction of the donor emission which overlaps with the acceptor absorbance, and increases when the acceptor has a high extinction coefficient. The overlap integral was calculated (using a value of 1.33 for η) to be $2.4 \times 10^{15}\ M^{-1}\ cm^{-1}\ nm^4$ for transfer between Oregon Green 488 and AlexaFluor 594, giving a Förster radius of 4.4 nm. For transfer between two Oregon Green molecules, the overlap integral was 1.4×10^{15}, giving a Förster radius of 4.0 nm.

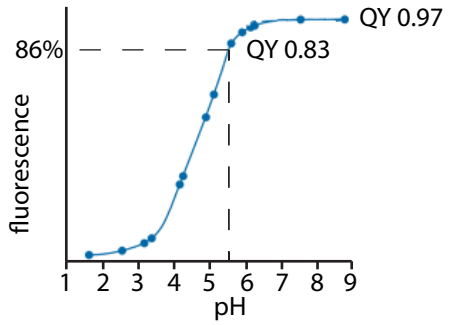

Figure 2-2. pH dependence of fluorescence of Oregon Green 488. The quantum yield of Oregon Green at pH 9 solutions is 0.97.[4] From the manufacturer's product literature, the expected quantum yield is 86% of maximum, or 0.83. Figure adapted from Molecular Probes product literature.

Using the Förster radius, the transfer efficiency can be related to the interchromophore distance using the equation

$$E = R_0^6 / (R_0^6 + r^6) \quad (3)$$

This is the proportion of photons absorbed by the donor which are transferred to the acceptor. This is one of three key parameters to maximize in light harvesting systems, in addition to the fraction of photons absorbed and the efficiency with which the acceptor makes use of the excitation energy. This calculation assumes a single, isolated donor-acceptor pair which is different from a TMV-based chromophore array. The existence of multiple competitive energy transfer pathways renders these efficiency calculations inaccurate in an absolute sense, but valid as relative measures of transfer efficiency.

Due to the close packing of protein monomers in a TMV rod, a donor in the RNA channel is within FRET range of the termini of multiple monomers, creating redundant energy transfer pathways to the exterior of a rod. Calculated efficiencies for the six shortest paths ranged from 37-64% (Figure 2-3). Interestingly, due to the shape and packing of the subunits, the shortest distance is between a donor and acceptor on two different proteins. Despite the moderate estimated transfer efficiencies, the existence of multiple parallel pathways from the RNA channel to the exterior of a TMV rod should compensate for this and permit a significant fraction of the harvested energy to reach photocatalysts on the exterior.

2.3 Chemistry for Exterior Modification

Having established the theoretical feasibility of energy transfer from the RNA channel to the exterior of a TMV rod, we had to decide how to attach donors and acceptors site-specifically. Previous work on intact TMV viruses had shown that tyrosine 139 could be uniquely targeted using diazonium coupling reagents to introduce ketones to the exterior of TMV rods.[5] Unfortunately, this reaction proved less compatible with recombinant TMV, which was prone to denature under the alkaline reaction conditions required. Furthermore, the strongly absorbing azo group that formed during the reaction had the potential to interfere with subsequent light harvesting and energy transfer. In a second approach, a reactive cysteine was introduced to the exterior of viral rods for conjugation of chromophores,[6] but the importance of the cysteine in position 123 for donor immobilization prevented the use of this approach. Instead, a number of new approaches were explored to introduce uniquely reactive handles under mild conditions for the conjugation of photocatalysts.

2.3.1 TMV-Intein Fusion for Expressed Protein Ligation (EPL)

The first reaction we explored targeted the C-terminus, which is located on the exterior of TMV rods. The TMV coat protein was genetically fused to an intein protein, which was capable of cleaving itself off the C-terminus and leaving a thioester in its place. This thioester was uniquely reactive with cysteine analogues and peptides containing an N-terminal cysteine, which would undergo a *S-N* acyl shift to form a stable amide bond. This system had been engineered by Kent and co-workers for the assembly of synthetic peptiedes into full-length proteins[7] and had been adapted by Francis and coworkers for the simpler task of introducing a C-terminal ketone for

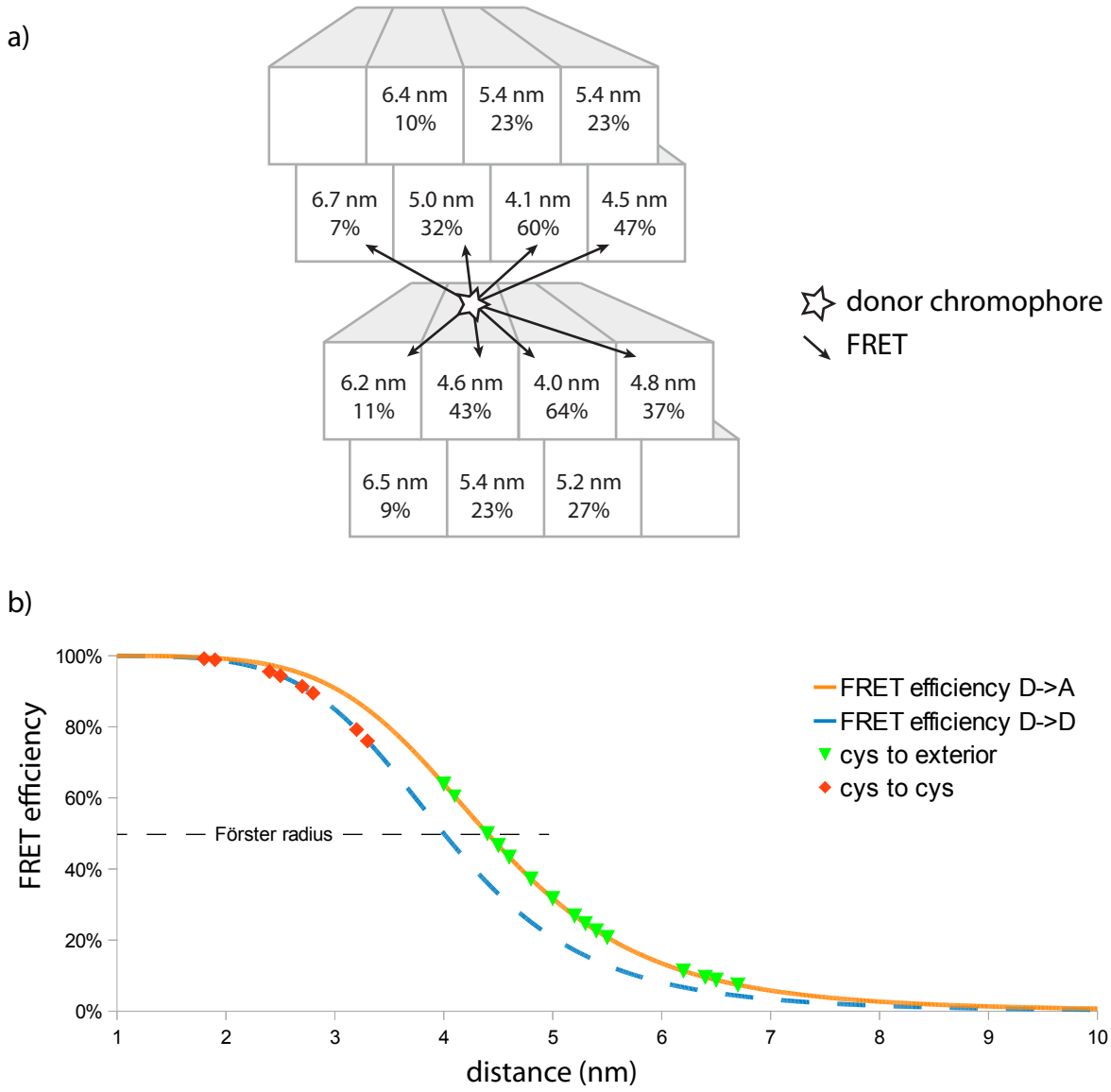

Figure 2-3. FRET efficiency as a function of distance. a) A diagram of protein packing in the TMV rod shows distances and estimated FRET efficiencies for energy transfer from a donor in the RNA channel of a central monomer to an acceptor on the surrounding termini. All four layers are contiguous - the gap between the top and bottom two is only to show the location of the donor dye . b) FRET efficiencies for donor to acceptor (solid orange line) and donor to donor (blue dashed line) are plotted as a function of interchromophore distance. Green triangles show the estimated efficiency of energy transfer of an isolated donor-acceptor system separated by the given distances. For comparison, the analogous efficiency of energy transfer of donor molecules arrayed in the RNA channel is depicted with red diamonds.

further modification with aminooxy compounds (Figure 2-4a).[8] This first step was accomplished by binding the protein fusion to chitin resin and incubating with a mixture of 2-mercaptoethane sulfonate (MESNa) and a cysteine analogue which contained a piperidone ring in place of a carboxylic acid. The MESNa served to catalyze the reaction by attacking the strained peptide bond and forming a semi-stable thioester. This thioester could by hydrolyzed or displaced by the cysteine-piperidone, which could rearrange to form a stable peptide bond. Following elution, the extent of modification was measured by gel shift assay in which a 2K-PEG-ONH_2 is coupled to the

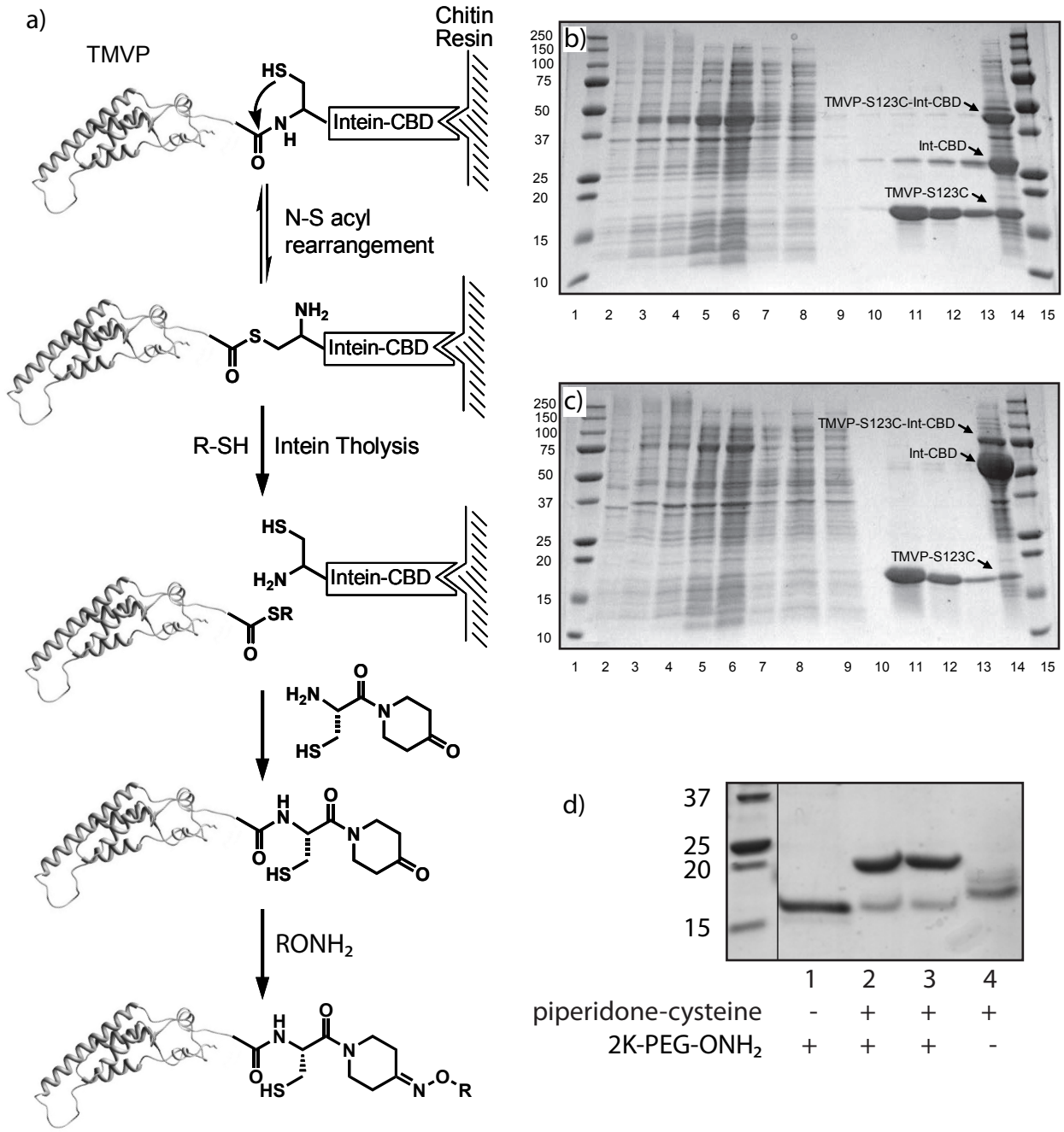

Figure 2-4. Expression of TMV via IMPACT system and introduction and modification of a C-terminal ketone. (s) shows scheme for intein cleavage with a thiol, where R-SH is DTT for simple cleavage or MESNa for native chemical ligation or expressed protein ligation. The introduced piperidone group was reactive with alkoxyamine reagents. SDS-PAGE of TMVP-S123C-Int-CBD shows representative expression and purification from (b) pTXB1 and (c) pTYB1 plasmids. Lanes 1, 15: Molecular weight markers. Lanes 2-6: Crude extract at 0, 3, 6, 9, and 20 hours after induction. Lanes 7-9: Chitin column washes after loading crude lysate. Lane 10: Blank. Lanes 11-13: Fractions eluted after cleavage with DTT for 40 hours, RT. Lane 14: Chitin beads stripped with SDS. (c) shows scheme for intein cleavage with a thiol, where R-SH is DTT for simple cleavage or MESNa for native chemical ligation or expressed protein ligation. The introduced piperidone group was reactive with alkoxyamine reagents. (d) shows an SDS-PAGE gel of TMV cleaved in the absence (lane1) and presence (lane 4) of piperidone-cysteine. The extra bands in lane 4 are attributed to labile "cystone oligomers." Reaction of the protein with 2K-PEG-ONH_2 produced a gel shift quantifiable by densitometry, showing at least 78% of the protein contained a reactive piperidone group.

C-terminal piperidone, creating a discrete band at a higher molecular weight (Figure 2-4d). Using densitometry to analyze gel images indicated at least 78% of the C-termini coupled to PEG. This number represented a lower bound for the extent of C-terminal modification because the oxime formation reaction may not proceed to completion with such a bulky substrate.

Unfortunately, two problems prevented the use of this protein system for the modification of exterior modified disks and rods. The first was a decrease in protein solubility in a number of buffers following the conjugation of the C-terminal piperidone. The more significant problem involved the unavoidable introduction of a second reactive thiol, which made it impossible to modify the protein only at S123C with maleimide reagents. We had hoped the conjugation of a bulky group, such as a porphyrin or PEG chain, to the introduced ketone would block access to this adjacent thiol, but this was not the case. This was established by introducing the piperidone-cystine analogue via EPL, modifying with either AlexaFluor647-ONH_2 or 2K-PEG-ONH_2, then reacting with Oregon Green maleimide (Figure 2-5). Site specific double modification of TMV at S123C and the C-terminus may now be possible using a recently reported ligation reaction involving thioacid/azide amidation.[9] Unlike previous cysteine-free NCL reactions, this reaction can be performed under native protein conditions.

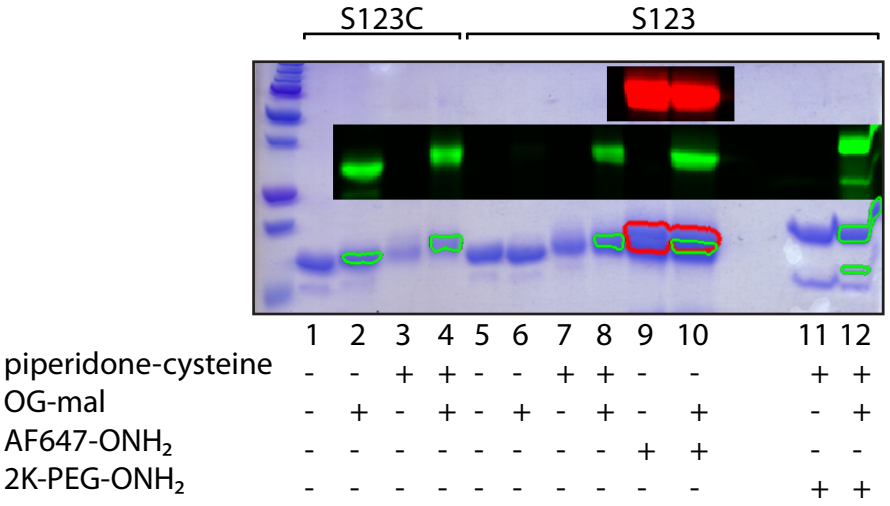

Figure 2-5. SDS-PAGE of TMV showing simultaneous conjugation of alkoxyamine and maleimide groups to a C-terminal piperidone cyteine. Lanes 1-4 are a positive control showing a cysteine introduced at position 123 is reactive with OG-mal. Lanes 5-6 show the bare protein is unreactive to OG-mal. Lanes 7-8 show the thiol of piperidone-cysteine is reactive with OG-mal when the piperidone is not modified. Lanes 9-12 show the thiol of piperidone-cysteine is reactive with OG-mal when the piperidone is first modified with AF647-ONH_2 (9-10) or 2K-PEG-ONH_2(11-12).

2.3.2 Optimization of PLP-Mediated Transamination through Mutagenesis of TMV

As an alternative to using the problematic C-terminal piperidone for exterior modification, we explored the possibility of modifying the N-terminus which is also located on the outside of the rods. The Francis group had recently applied a site-specific biomimetic transamination reaction[10-12] mediated by the biological cofactor pyridoxal 5′-phosphate (PLP, **2**) to the modification of the N-terminus of peptides and proteins.[13,14] The transformation of the N-terminal amino group (**1**) into a ketone or aldehyde (**3**) provides a versatile orthogonal handle for further reaction with alkoxyamines containing functional groups of interest (Figure 2-6a). Despite the abundant display

of lysine side chains on the surface of most proteins, transamination is highly specific to the N-terminus. This is most likely due to the lower pK_a of the α-proton of the N-terminal amino acid relative to the ε-proton of lysine.

From a mechanistic standpoint, PLP-mediated transamination most likely begins with the formation of a Schiff base at the N-terminus (**5**) followed by tautomerization (**5** to **6**) and subsequent hydrolysis, as is the case for many PLP-mediated reactions in biological settings (Figure 2-6).[15,16] In the absence of an enzyme active site, this transformation could be considered to be independent of the identity of the side chain of N-terminal residue. This is an oversimplified view, since our earliest evidence for this reaction resulted from the observation that angiotensin I (N-terminal Asp) underwent decarboxylation along with PLP-mediated transamination.[13] This indicates that the Schiff base intermediate formed between PLP and the N-terminus of a protein or peptide can lead to the formation of alternative products in a fashion that may vary on a case-by-case basis.

Mysteriously, recombinant TMV showed no reactivity with PLP (Figure 2-7b). The N-termini

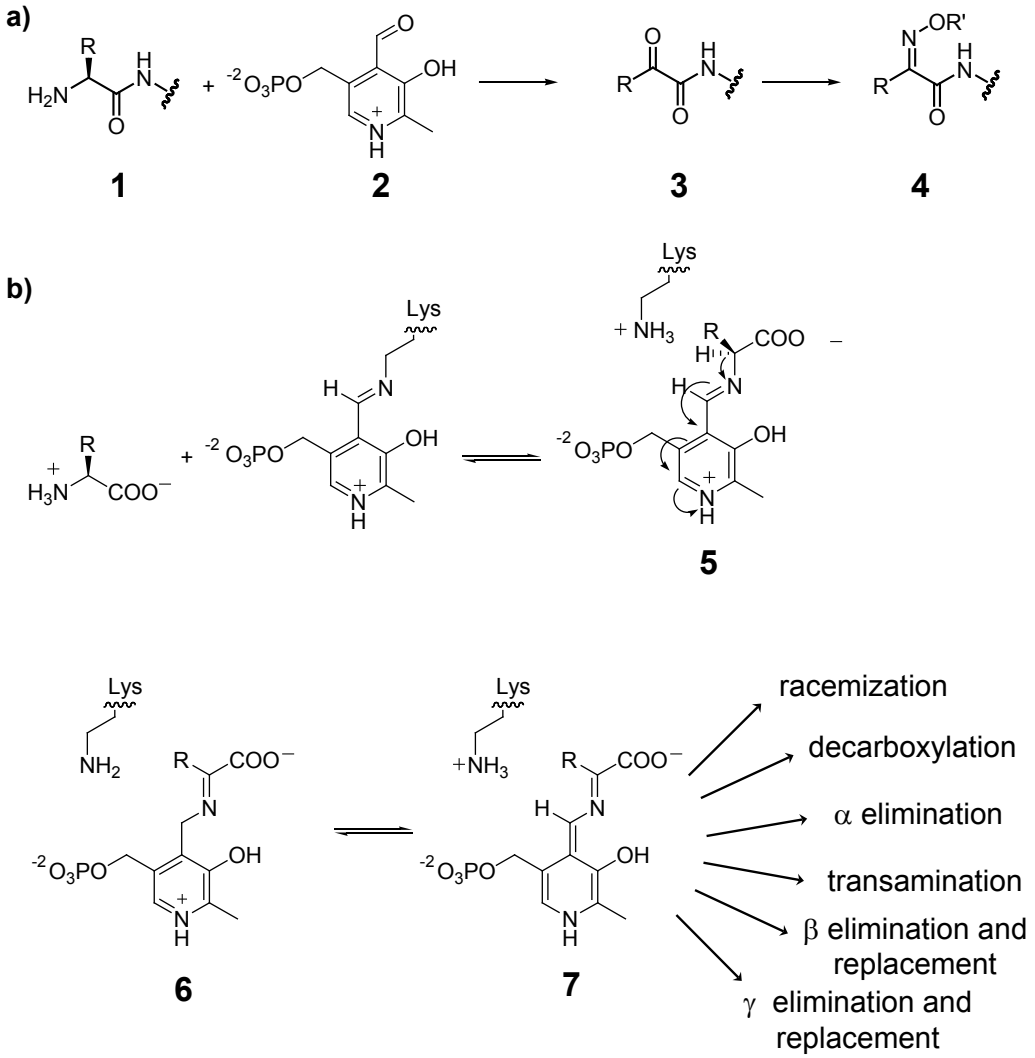

Figure 2-6. Survey of PLP-mediated reaction pathways. a) General scheme for PLP-mediated N-terminal transamination of proteins and peptides. b) The range of reactions mediated by PLP in the biological context.

appeared to be solvent exposed in the crystal structure,[17] and possessed a high 'temperature' score, which suggested that they were unstructured or dynamic and therefore likely to sample sterically permissive conformations for reaction. Equally perplexing was the lack of reactivity observed when oxidizing the N-terminal serines with periodate. Both the chemistry of serine transamination and the structure of the N-terminus were possible culprits, so both variables were tested with a small mutant library.

Previous mutagenesis of TMV in the Francis lab was limited to the introduction of a cysteine at position 123, and unsuccessful attempts to remove the native, buried cysteine at position 27. While it has been possible to generate mutants of the coat protein and express them in plants as part of an infectious virus,[6] mutagenesis and expression in *E. coli* offered a much faster turnaround time.[18] The published procedure to purify recombinant coat protein over anion exchange was faster than extracting viruses from tobacco leaves, but it did not lend itself to purifying multiple proteins in parallel. To facilitate screening many mutants, we took advantage of the affinity tag offered by the intein-CBD of the IMPACT system. This allowed a one-step purification by selective binding of TMV to chitin resin, and elution after incubation with a thiol (Figure 2-4a,b). Once optimized, this system permitted the purification of multiple mutants in parallel with the help of a vacuum manifold.

The available residues to screen were limited by the activity of the methionyl-aminopeptidase enzyme in *E. coli*, which cleaves the start methionine only when the following side chain is small. (Hirel 1989) Using a model protein, Hirel et al. reported >95% methionine cleavage preceding Ala and Gly, >80% preceding Ser, Pro, Thr, Val, and >70% preceding Cys.[19] When wild-type TMV is expressed in E. coli, for instance, mass spectra show no evidence of an N-terminal methionine, suggesting it is completely cleaved to leave an N-terminal serine.

To begin, the N-terminus was extended by inserting a glycine after the start methionine. A temperature screen at 37, 43, and 60 °C showed almost immediate protein precipitation at 43 and 60 °C, and significant precipitate at 37 °C after 15 h. In a subsequent screen, 30 °C was the highest temperature the protein could tolerate for a 48 h period. The next round of mutants was created by replacing the appended N-terminal glycine with the following residues: Ala, Asp, Cys, Ser, Thr, and Pro. The Pro and Thr mutants cleaved *in vivo* or did not express, and mass spectra of the Asp and Cys mutants showed no and partial methionine cleavage, respectively. Preliminary results with N-MetAsp showed almost no modification, so it was not screened further. The remaining four were reacted with PLP for 24 and 44 h, purified via Nap-5, and reacted with 2K-PEG-ONH_2 overnight. The 44 hour reaction gave higher modification, with N-Ala showing the greatest modification, followed closely by N-Gly. A second higher molecular weight band increased with PLP incubation time as well and was likely caused by a second modification by PEG at an unknown site. The hypothesis was put forth that this second band could have arisen from a batch of 2K-PEG-ONH_2 which contained a small fraction of 4K-PEG-ONH_2. This is not the case with all protein reactions with 2K-PEG-ONH_2, since high levels of modification with no second band can be achieved when a ketone is appended to TMV's C-terminus via expressed protein ligation.

In the third screen of N-terminal TMV mutants, an additional residue was inserted to act as a spacer. Histidine was chosen because concurrent peptide studies suggested it might increase reactivity in the second position, and glycine was chosen because the lack of a side chain would provide the most flexibility. Ala, Cys, Gly, and Ser were followed by either His or Gly and transaminated for 24 or 44 hours before purification and PEG modification. Though the AlaHis N-terminus provided the highest reactivity to date, all mutants with a histidine spacer were prone

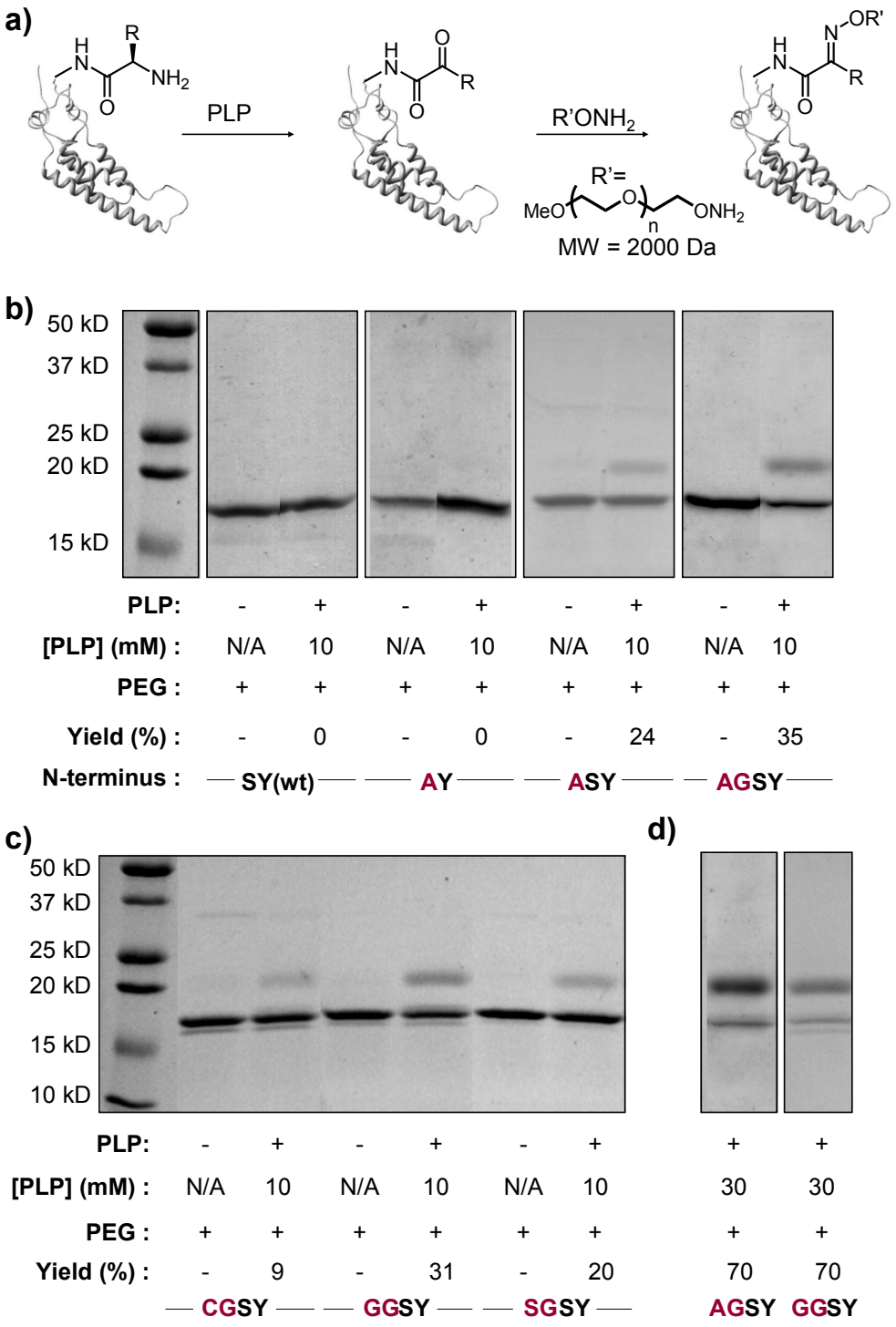

Figure 2-7. Modification of the TMV capsid protein with PLP. (a) Reaction scheme for PLP modification and oxime formation of TMV. (b) SDS-PAGE analysis of TMV following reaction with PLP and treatment with PEG-alkoxyamine. No reactivity was observed with the wt N-terminus (SY) or an S1A mutant (AY) after 44 h of reaction with PLP. Modification was observed after extension of the N-terminus with A or AG, followed by 24 h treatment with PLP. (c) Reactivity was also observed with N-terminally extended mutants containing N-terminal Cys, Gly, and Ser residues. (d) Higher conversion was attained when the PLP concentration was increased to 30 mM. One letter codes labeled in red indicate sites of mutagenesis.

to precipitation, so were not pursued further. The next most reactive mutant was AlaGly, which had no stability issues. After reacting with PLP for 44 h, it appeared to be 66% modified.

Having identified AlaGly as the most reactive mutant with adequate stability, we screened [PLP] to determine if higher modification could be achieved (Figure 2-7c). In a 24 h reaction, 100 mM PLP led to significant protein loss, while 30 mM PLP provided up to 70% modification with little to no precipitation. This compared favorably to the 64% conversion observed with a 44 h reaction in 10 mM PLP. Finally, the CysGly, GlyGly, and SerGly mutants were tested at 30 mM PLP for 24 h to determine if their reactivity would also increase. Only GlyGly appeared significantly more reactive, reaching 70% conversion.

These results generally agreed with solid-phase peptide screens (Figure 2-8). While it would have been possible to create a protein library with all 20 N-terminal residues using a protease-cleavable N-terminal tag , this was unnecessary because both of the best performing residues were available and the majority of recombinant proteins are expressed in *E. coli*, and would therefore be subject to the same N-terminal constraints. The tetrapeptides XKWA containing an N-terminal Ala or Gly had the highest conversion, as did TMV. Peptide screens showed that the identity of the second and third amino acid could have dramatic effects on the transamination but probing these secondary and tertiary amino acid effect one peptide or protein at a time would have been prohibitively time consuming. Subsequent work using a combinatorial peptide library was able to identify optimal N-terminal sequences for transamination by screening a much larger sequence space.[20]

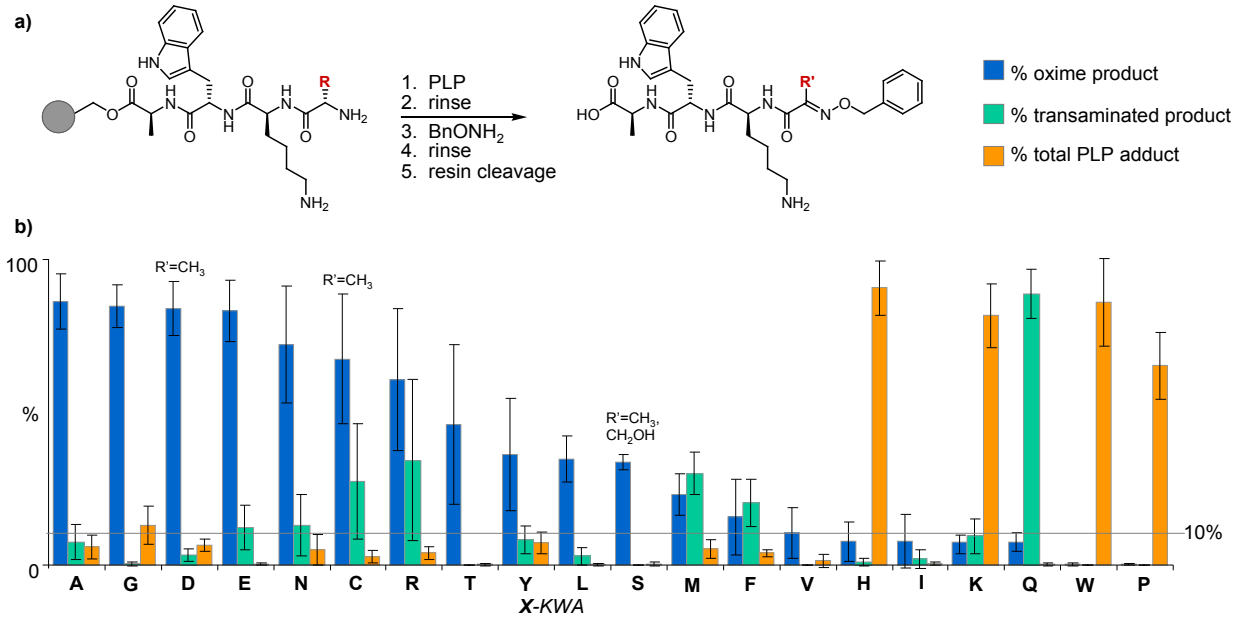

Figure 2-8. Effects of N-terminal amino acid on PLP-mediated transamination. (a) General method for screening a 20 member tetrapeptide library for PLP reactivity. (b) Product distribution for tetrapeptides reacted with PLP. Percentages are based on integration of extracted ion chromatograms. The expected oxime product is in blue, transaminated product that did not form an oxime is in green, and PLP adducts are in orange. The sum of all modified and unmodified (not shown) species is 100%. Error bars represent the standard deviation for three replicate experiments.

2.3.3 Oxidation of Serine

The oxidation of N-terminal serine or threonine to a glyoxylyl group using periodate was first reported in 1939,[21] and was adapted for site-specific protein/peptide modification as early as 1992.[22] The oxidized serine, if quenched in time, is the identical glyoxylyl product of transamination of glycine (Figure 2-9). This group behaves in most reactions like an aldehyde, which is uniquely reactive with nucleophililc alkoxyamines or hydrazines to form site-specific bioconjugates, since proteins lack other electrophilic groups. As discussed above, the side chain of an N-terminal serine is small enough that a preceding methionine is usually cleaved, making it possible to express proteins with an N-terminal serine in *E. coli*.

Historically, serine oxidation for site specific protein modification has not been used widely for three reasons. The N-terminus is suitably located for only a limited number bioconjugation projects, though it can be relocated via circular permutation. This limitation is shared with the PLP-mediated transamination. Periodate is capable of oxidizing many amino acids and sugars (with widely varying rates) which can lead to heterogeneous products, if not a loss of protein folding and function. Cysteine is particularly susceptible as periodate can oxidize it to an unreactive sulfonic acid. Fortunately, double labeling is possible if the cysteine is alkylated prior to the addition of periodate.

As an alternative to the PLP reaction for modifying the N-terminus of TMV, the periodate

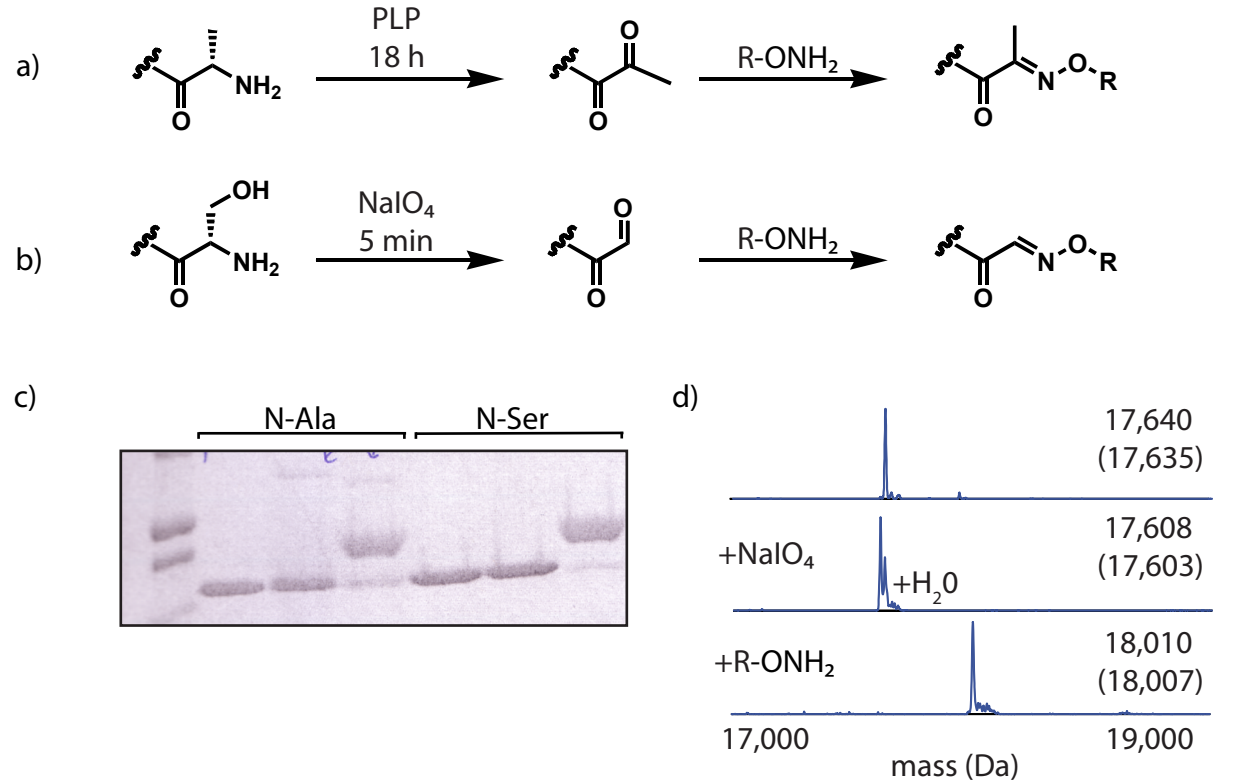

Figure 2-9. Comparison of PLP and NaIO4 for producing ketones or aldehydes at the N-terminus. (a) and (b) show reaction schemes for modifying the N-terminus with PLP and NaIO4, respectively. (c) The SDS-PAGE gel shows TMV with an N-terminal Ala after transamination with PLP and subsequent reaction with 2K-PEG-ONH_2 (lanes 2 and 3). Lanes 5 and 6 show TMV oxidized with NaIO4 and subsequent reaction with 2K-PEG-ONH_2. High conversion is evident in both cases, though the PEG shift appears more complete in the case of N-Ser TMV. (d) shows LCMS spectra of N-Ser TMV, its oxidation with periodate, and reaction with fluorescein thiosemicarbazide. To the resolution of the instrument, both reactions appears to proceed to completion.

oxidation of serine has a number of advantages. The reaction time is 5-10 minutes, compared to 1-24 hours for PLP (TMV requires 18 h). Neither periodate nor any side product contains an aldehyde, permitting the direct addition of aminooxy or hydrazide groups without additional purification steps. The reaction regularly proceeds to completion (Figure 2-9), and produces no side products or adducts on this timescale that are capable of interfering with N-terminal reactivity, unlike PLP.

2.4 Choice of Photocatalyst Accptor

2.4.1 Porphyrins as Photocatalysts in Light Harvesting Systems

Porphyryins have been explored as a final acceptor for artificial light harvesting systems because they are capable of photoinduced electron transfer following absorption of light by the Soret band as well as the Q bands.[23] This absorption profile allows porphyrins to accept energy from donor dyes that absorb in the UV and visible region. Because of their planar, conjugated structure, they tend to aggregate in many solvents including water, which poses a challenge both in their conjugation to biomolecules and the subsequent processing of the hybrid materials. Enzymes avoid this latter problem by binding porphyrins noncovalently in pockets largely isolated from the aqueous environment. Porphyrin chemistry in biology is conducted without the help of light. Instead, photosyntheic organizsms use closely related chlorin molecules for the special pair in photosystem II, which use a photoexcited electron to reduce cofactors and eventually quinone.

In artificial systems, porphyrins have been directly excited[24] or sensitized with a donor chromophore[25] to reduce methyl viologen (MV^{2+}) to its radical cation (MV^{+*}). In addition to offering a spectroscopic readout for the transfer of electrons, this compound has been used as an electron shuttle, e.g. to transfer electrons to platinum to evolve hydrogen gas.[24] It has also been used to reduce quinone to semiquinone in a membrane-bound proton shuttle system.[23] Photoexcited porphyrins have also been reported to sensitize rhodium catalysts capable of regenerating (reducing) $NAD(P)^+$ to NAD(P)H, though in this case it is unclear if the energy is transferred via FRET or as an excited electron.[26] Systems to regenerate NAD(P)H from $NAD(P)^+$ (or vice versa) are of interest because either can be necessary cofactors for many oxidoreductase enzymes capable of producing valuable chiral compounds such as hydroxy acids, amino acids, steroids, and alcohols from prochiral precursors.[27] Due to their high cost and the requirement for stoichiometric equivalents, these soluble cofactors must be produced catalytically *in situ*.

Porphyrins can inject excited electrons into semiconductors as well as small molecules. They have been coated on TiO_2 and studied as alternatives to Ru(bpy) in dye-sensitized solar cells.[28]While they do not require precious Ruthenium and are relatively easy to synthesize, their performance has yet to compete with the more commonly used metal complexes and phthalocyanines. Photoexcited porphyrins have also been reported to transfer electrons into carbon nanotubes, either via non-covalent pyrene electrostatic interactions[29] or covalent modification of oxidized defect sites.[30] This was observed as the quenching of porphyrin fluorescence and shorter lifetimes of the excited states.

2.4.2 Covalent Attachment of Chromophores and Porphyrins to the Exterior of TMV

As a proof of concept, fluorescein thiosemicarbazide was attached to the N-terminus of TMV prepared both by PLP transamination and serine oxidation. The two behaved similarly,

and only the latter will be described here. High levels of modification of disks and rods led to the aggregation and precipitation of the protein, as a significant fraction of the surface area was covered by relatively hydrophobic chromophores. To maintain the solubility of the protein, only a small fraction of the N-termini were modified and the protein was kept in high pH buffers to prevent assembly into disks. By completely modifying the proteins at S123C with an acceptor (AF594) and modifying a subset with donors, we ensured each donor would be adjacent to an acceptor. Energy transfer was then measured by comparing the normalized absorbance spectra with the excitation spectra of the two dye system, monitoring in the tail of the acceptor's emission to minimize bleedthrough of donor fluorescence (Figure 2-10). The ratio of the donor absorbance and its contribution to the acceptor fluorescence gave an efficiency value of 39%, confirming that energy transfer was viable from the RNA channel to the exterior. This value is close to the calculated efficiency value of 43% for transfer within a monomer for the similar Oregon Green:AF594 system. At pH 8, the bulk of the protein is monomeric or in small aggregates, and so the higher efficiency transfer paths calculated in Figure 2-3a are likely unavailable.

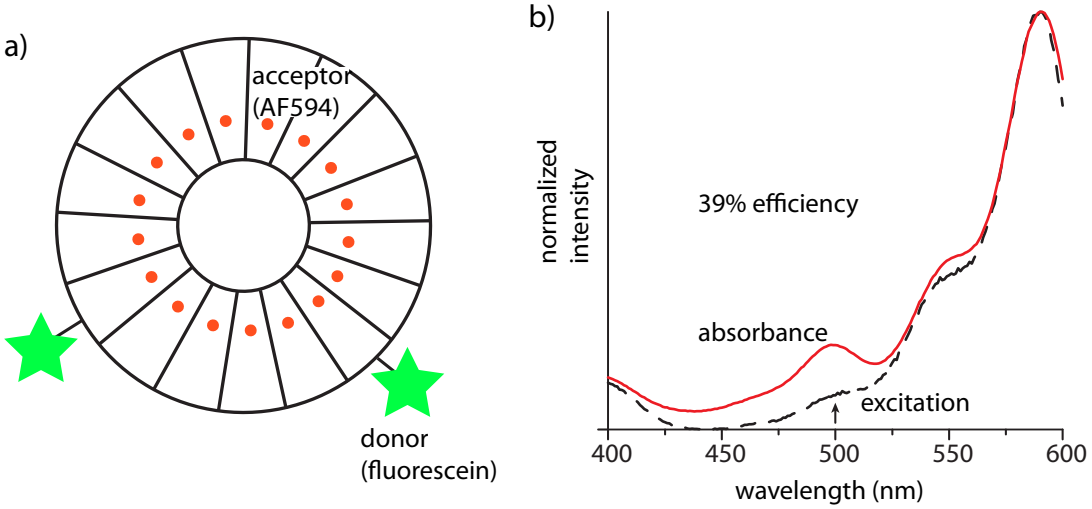

Figure 2-10. Energy transfer from donors on exterior of TMV to acceptors in RNA channel. a) Alexafluor 594 was conjugated to S123C while fluorescein thiosemicarbaizide was reacted with glyoxylyl groups at the N-terminus. b) by comparing the absorbance and excitation spectra of the doubly modified protein, the efficiency of energy can be calculated. The contribution of the donor to the acceptor emission (monitored at 630 nm) relative to the intensity of donor absorbance gives an efficiency value of 39%.

Having established this proof of concept, an oxime bond was used to attach porphyrins to the exterior of TMV by mixing trisulfonate monoalkoxyamine porphyrin with transaminated TMV containing a AlaGly N-terminus. An SDS-PAGE gel comparing the dark precipitate and the lighter supernatant showed both contained slightly gel-shifted fluorescent bands, but the precipitate was significantly more modified (Figure 2-11a). After purification via Nap-5, the yellow solution was analyzed by HPLC, LCMS, and MALDI. The size-exclusion traces suggested a mixture of disks and monomers with porphyrin absorbance and faint porphyrin fluorescence. Only unmodified protein was visible by LCMS, but a small peak at +938 was observed by MALDI (expected +943, (Figure 2-11b).

AlexaFluor 350 (AF350) was first chosen as a donor because its published emission spectra overlapped well with the porphyrin Soret band. The S123C position of TMV was first modified to ~90% with AF350-maleimide before porphyrin-ONH_2 was conjugated in an overnight reaction. A

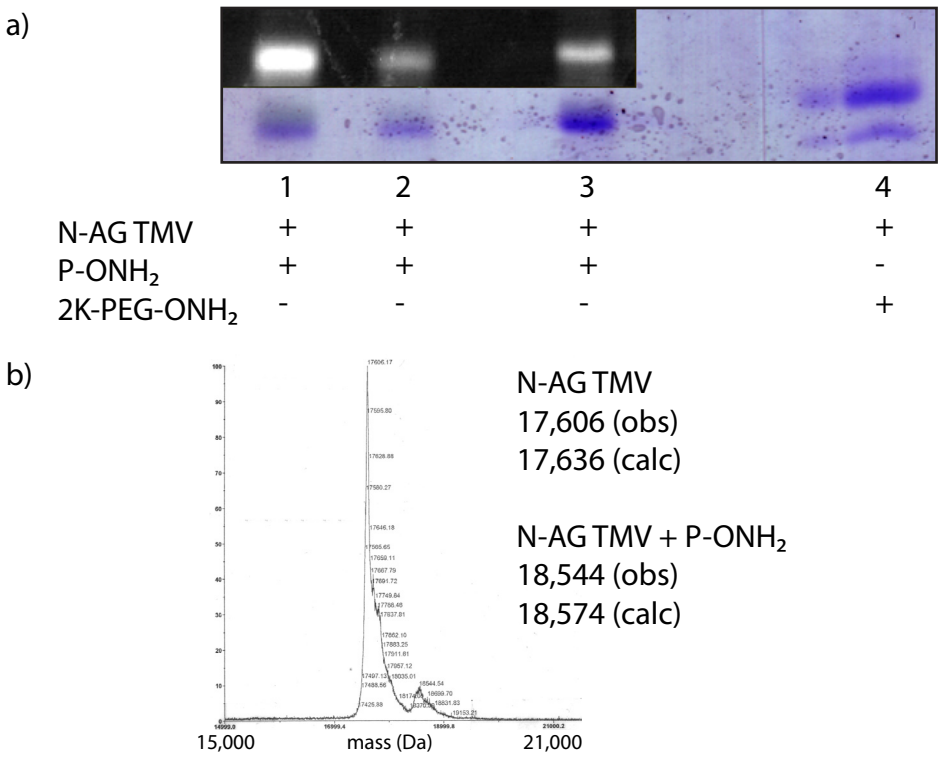

Figure 2-11. Conjugation of porphyin-alkoxyamine (P-ONH_2) to transaminated TMV characterized by SDS-PAGE and MALDI. (a) The protein that precipitated during the reaction (Lane 1) is heavily modified compared to the protein that remained soluble (Lane 2). Lane 3 shows the soluble protein following passage through a Nap-5 column. Lane 4 shows a positive control of TMV reacted with 2K-PEG-ONH_2. (b) A MALDI spectrum of the purified conjugate shows a small peak at +938 (expected +943).

dark band was collected from a NAP-5 gel filtration column and the fluorescence was analyzed the following day. Curiously, the AF350 emission was redshifted from the literature spectra so that it fell squarely between the Soret and Q bands, as measured from an excitation spectra observing the porphyrin emission at 671 nm. Consistent with this poor spectral overlap (and therefore poor FRET efficiency), an overlay of the excitation spectra (monitoring at 671) of the acceptor only and donor+acceptor systems showed no peak corresponding to donor contribution (Figure 2-12).

Conjugates were remade with Oregon Green as the donor chromophore to sensitize the Q bands using both N-AG TMV and permuted TMV (see chapter 3). Despite observing >50% porphyrin modification of permuted protein by SDS-PAGE, neither species showed porphyrin absorbance by size exclusion chromatography on the HPLC. This was attributed to the porphyrin binding to the polysep-5000 column. Furthermore, TEM images showed that neither of the doubly modified species were able to form rods when dialyzed into 100 mM pH 5.5 NaOAc. Taking a different approach, pre-formed rods from TMV which had been transaminated and modified with Oregon Green were reacted with porphyrin alkoxyamine. These samples were purified by passage through a Nap-5 column. The untransaminated TMV rod sample appeared as brown as the transaminated sample, suggesting the porphyrin bound the protein non-specifically. As a control, a solution of the porphyrin without protein was found to bind completely to the top of a Nap-5 column. This indicated that protein, covalently bound or not, was necessary to carry the porphyrin through. This qualitative observation of non-specific binding was confirmed by size exclusion chromatography on the GF-250. The -PLP sample reacted with 1.2 mM porphyrin had increased

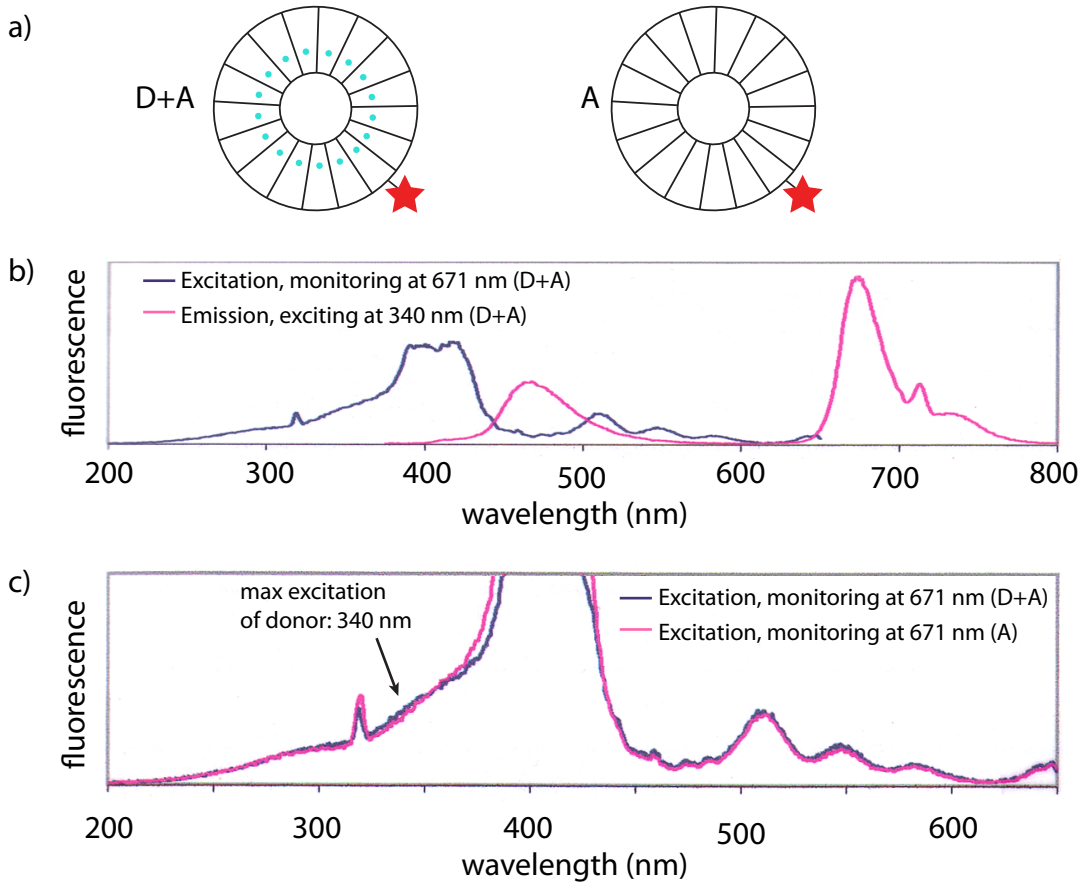

Figure 2-12. Fluorescence spectra of TMV systems containing AF-350 donors and porphyrin acceptors. (a) TMV modified at S123C with AF-350 maleimide and at the N-terminus with a porphyrin alkoxyamine (D+A) was compared with TMV modified only with porphyrin (A). (b) When illuminated at a wavelength that can excite both donor and acceptor, the emission spectra of the (D+A) system (pink line) shows the fluorescence of AF-350 (λmax 465 nm) and the porphyrin at (λmax 671 nm). Notably, the AF-350 emission is red-shifted ~20 nm upon conjugation to TMV compared to the free dye. The excitation spectrum of the (D+A) system (purple line) shows neither the Soret and Q bands overlay with the donor fluoresence. (c) The emission of the porphyrin was monitored with (purple line) and without (pink line) an acceptor present. Because no significant difference is visible between the spectra at the λmax of donor excitation (340 nm), energy transfer from the AF350 to porphyrin is miminal.

absorbance in the Soret region comparable to the +PLP sample reacted with 0.12 mM porphyrin.

Small aggregates were also visible in TEM images containing porphyrin in quantities proportional to the amount of porphyrin added. This result suggested that the porphyrin was also disrupting the structure of the TMV assemblies. The combination of these two problems were evidence that this trisulfonate alkoxyamine porphyrin was incompatible with the TMV light harvesting system. This porphyrin was subsequently attached to the exterior of MS2 and energy transfer was reported from Oregon Green chromophores attached to the inner surface.[25] Work is underway to sandwich photocatalysts between electrodes and TMV-based light harvesting structures.

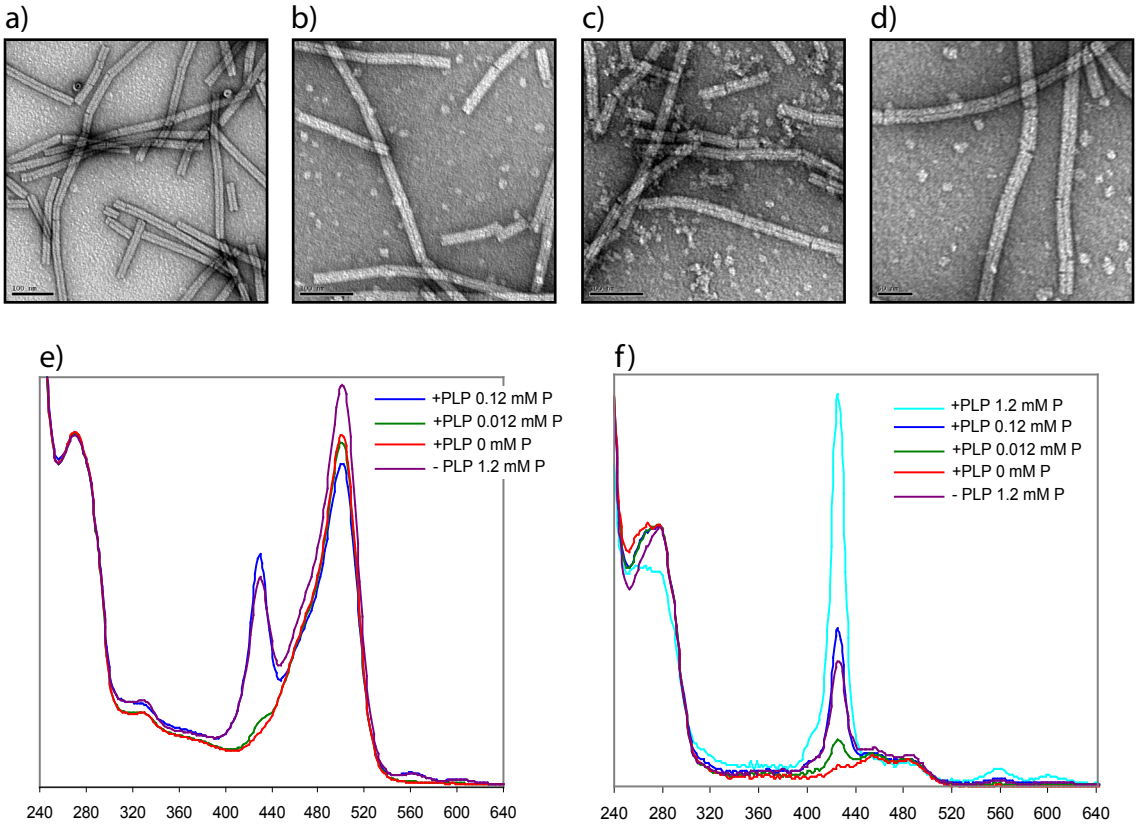

Figure 2-13. TEM images and absorbance spectra of TMV rods conjugated to Oregon Green and aminooxy porphyrin. TEM images show PLP and OG modified TMV reacted with 0.012 mM (a), 0.12 mM (b), and 1.2 mM aminooxy porphyrin (c). (d) shows untransaminated OG rods reacted with 1.2 mM porphyrin. The origin of the debris visible in b-d is unknown, but correlates with the porphyrin absorbance in GF-205 size exclusion (e) and C8 reversed-phase traces (f) of the above samples. The traces were normalized to protein absorbance at 280 nm. The trace of 1.2 mM porphyrin +PLP from size exclusion is unavailable because the sample did no elute from the column.

2.5 Conclusion

To make use of the energy collected by the light harvesting system held in the RNA channel of TMV, we sought to conjugate photocatalysts to the exterior of the rods through the installation of ketone and aldehyde groups. Even after the intein-mediated C-terminal modification proved unworkable, the expression system facilitated the creation of a library of N-terminal mutants of TMV to optimize the PLP-mediated transamination reaction. Using the insights gained from the PLP reaction screening, we were able to optimize the oxidation of an N-terminal periodate as well, which ultimately proved equally useful. While the trisulfonated monoalkoxyamine porphyrin was incompatible with this system, we maintain hope that another photocatalyst will prove more cooperative with the remarkable TMV-based light-harvesting systems.

2.6 Methods and Materials

General Procedures. Unless otherwise noted, all chemicals and solvents were of analytical grade and were used as received from commercial sources. Water (dd-H_2O) used in biological

procedures or as reaction solvents was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Pyridoxal 5'-phosphate monohydrate was obtained from Aldrich (St. Louis, MO). The centrifugations required in spin concentration steps were conducted using an Allegra 64R Tabletop Centrifuge (Beckman Coulter, Inc., Fullerton, CA).

Construction of TMVP Expression Plasmids. A TMV cDNA clone, pTMVP-S123C, was obtained from lab stocks in a PET-20b(+) plasmid. Standard recombinant techniques were used to construct an expression plasmid with pTYB1 vector DNA (NEB, Ipswich, MA) An internal SapI restriction site was removed by introducing a silent mutation (TCT to TCC) using QuikChange mutagenesis (Stratagene, Santa Clara, CA). The primers were obtained from IDT (forward 5´-AGCTCTTTCG AGAGCTCCTC TGGTTTGGTT TGGAC-3´, reverse 5´-GTCCAAACCA AACCAGAGGA GCTCTCGAAA GAGCT-3´). The TMVP-S123C gene was amplified by PCR, using primers obtained from IDT (forward 5´-AGACCATTCA TGTTGTTGCT CAGGTCGC-3´, and reverse 5´-GGTGGTTGCT CTTCCGCAAG TTGCAGGAC-3´). The amplification product was digested sequentially with SapI and NdeI restriction enzymes (NEB) before ligation into pTYB1 with T4 DNA ligase (NEB). The C-terminal residue of TMV was changed from Thr to Gly using QuikChange mutagenesis to facilitate cleavage by the intein. N-terminal mutants were also made using QuikChange mutagenesis.

Expression and Purification of Recombinant TMV Coat Protein (TMVP). Tuner DE3pLysS competent cells (Novagen) were transformed with TMV, and colonies were selected for inoculation of Terrific Broth cultures. When cultures reached mid-log phase as determined by O.D.600, expression was induced by addition of 0.3 mM IPTG (Invitrogen). Cultures were grown for 20 h at 20 °C, harvested by centrifugation, and stored at -80 °C. The following purification steps were performed at 4 °C. Induced cells were thawed and resuspended in 40 mL of Column Buffer (20 mM Tris, pH 8, 500 mM NaCl) and lysed by sonication (Branson Ultrasonics, Danbury, CT.) The cell lysate was cleared by centrifugation at 7,000 rpm for 30 min. The cleared lysate was loaded onto a gravity flow column packed with 5 mL of chitin resin equilibrated with 40 mL of Column Buffer. The column was spun on a rotisserie for 1 h to bind the fusion protein to the chitin resin and then washed with an additional 40 mL Column Buffer. This was followed with 15 mL of Cleavage Buffer (100 mM Tris pH 8, 50 mM DTT). The column was capped with 1 mL of cleavage buffer above the resin and left at 25° C for 20 h. Fractions were analyzed by SDS-PAGE, and fractions containing TMVP were combined and concentrated.

Of the two intein-CBDs available in the commercial IMPACT-CN system from NEB, the larger in pTYB1 was preferred because it bound more tightly to chitin resin. The smaller intein-CBD from the pTXB1 plasmid would disassociate from the chitin and elute with the protein. It was eventually discovered that passage of this eluent over additional chitin resin was shown to provide pure protein.

The identity of the C-terminal residue affects the stability of the amide bond between the protein and intein, as detailed in the product literature. The native threonine residue of TMV led to slow and incomplete *in vitro* intein cleavage. Mutation to glycine, but not serine, improved the cleavage significantly.

General Procedure for the PLP Modification of TMVP. A 1.5 mL Eppendorf tube was charged with a solution of TMV (delivered as 100 µL of a 2 mg ml^{-1} solution in 25 mM potassium phosphate,

pH 6.5; 1 equiv) and a solution of freshly prepared PLP (delivered as 100 μL of a 20 mM solution in 25 mM phosphate buffer, pH adjusted to 6.5 with 1 M NaOH; 175 equiv.) The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 30 °C for 24 h. The PLP was removed from the reaction mixture *via* size exclusion chromatography (Nap-5 desalting column (GE Healthcare) eluting into 25 mM phosphate buffer, pH 6.5.) A 50 uL aliquot of the resulting purified mixture was treated with 50 μL of 0.1 M 2K-PEG-ONH_2 or 25 μl of 25 mM trisulfonate monoalkoxyamine porphyrin in a 1.5 mL Eppendorf tube. The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at rt for 18-20 h. Samples were combined with loading buffer, and analyzed by SDS-PAGE.

General Procedure for the $NaIO_4$ Modification of TMVP. A 1.5 mL Eppendorf tube was charged with a solution of TMV (delivered as 100 μL of a 10 mg ml^{-1} solution in 25 mM potassium phosphate, pH 6.5; 1 equiv) and a solution of $NaIO_4$ (delivered as 3 μL of a 200 mM solution in water; 10.5 equiv.) The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at rt for 5 min. The oxidation was quenched with a solution of Na_2SO_3 (delivered as 3 μl of a 1 M solution in water; 52 equiv.). The lack of small molecule aldehydes permitted addition of alkoxyamine reagents with no further purification.

Gel Analyses. Sodium dodecyl sulfate-poly(acrylamide) gel electrophoresis (SDS-PAGE) was accomplished on a Mini-Protean apparatus from Bio-Rad (Hercules, CA) with 15 % gradient polyacrylamide gels (BioRad, CA), following the protocol of Laemmli.[31] All electrophoresis protein samples were mixed with SDS loading buffer in the presence of dithiothreitol (DTT) and heated to 100 °C for 10 min to ensure reduction of disulfide bonds and complete denaturation unless otherwise noted. Commercially available molecular mass markers (Bio-Rad) were applied to at least one lane of each gel for calculation of the apparent molecular masses. Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA). Protein reaction conversion was estimated from standard optical density measurements of the observed gel bands with Image J™ software (version 1.34s).

Chromophore Attachment to S123C. A thawed aliquot of cpTMVP or AG-TMVP in buffer A was diluted to 100 μl (0.57 mM or 10 mg/ml) with the same buffer. To this solution was added 2.5-3 eq of maleimide functionalized chromophore as a DMSO solution (up to 3% v/v). The reaction mixure was vortexed briefly and was left at rt for 90 min. The bulk of the unreacted chromophore was removed by Nap-5 gel filtration, with elution into 400 μL of the same buffer. When necessary, further purification and/or concentration was accomplished via spin concentration. The extent of protein modification was monitored by LC/ESI-MS (Orbitrap or 150EX) and UV-vis spectroscopy.

Chromophore Attachment to N-terminus of cpTMVP. A 40 μL solution of cpTMVP (previously labeled at C123 with a donor maleimide) in buffer A (0.27 mM or 4.7 mg/mL), was reacted with 3.8 eq of Alexa Fluor 647 C5-aminooxyacetamide (Invitrogen, Carlsbad, CA) from a 25 mM solution in DMSO. The oxime formation was catalyzed by the addition of 47 eq of *p*-anisidine, from a 1 M solution in DMSO, pH adjusted to 6.5 with acetic acid. The reaction mixure was vortexed briefly and was left at rt for 24 h. The unreacted chromophore was removed by spin concentration. Extent of protein modification was monitored by LC/ESI-MS (Orbitrap) and UV-vis spectroscopy (Figure 2-5).

Assembly of cpTMVP Disks and Rods. To form rods, 2 mg/mL (114 μM) solutions of cpTMVP modified with a chromophore at C123 in 100 mM sodium borate, pH 8.0 were dialyzed against 50 mM sodium acetate buffer, pH 5.5 at rt, transferred to 1.5 mL Eppendorf tubes and stored at rt in the dark for 1-3 weeks. Conversion to larger structures was monitored by size-exclusion chromatography HPLC. Rods eluted between 5 and 8 minutes, depending on length. Assemblies were characterized visually by TEM.

Assembly of Hybrid Disks and Rods. AG-TMVP was modified with Oregon Green 488 maleimide (Invitrogen) as described above, then separated from free dye and simultaneously exchanged into 20 mM TEA, pH 8.0 using a Nap-5 column. The same procedure was used to modify cpTMVP with Alexa Fluor 594 (Invitrogen), though exhaustive spin concentration was required to remove the free dye. After overnight equilibration, dissassemly was confirmed by SEC prior to mixing the proteins in a 16:1 or 1:1 AG-TMVP:cpTMVP ratio. A fraction of each mixture was diluted to 0.45-0.5 OD of the chromophore for measurement of the absorbance and fluorescence spectra as described below. The concentrated samples were dialyzed against 400 mM potassium phosphate, pH 6.5 overnight to assemble into disks and rods. Large protein aggregates were removed by centrifugation for 5 min at 18,000 RCF before assembly was confirmed by SEC. Samples were dilluted to 0.45-0.5 OD of the chromophore and absorbance and fluorescence spectra recorded as below.

Size Exclusion Chromatography (SEC). Analytical size exclusion was performed on an Agilent 1100 series HPLC equiped with a PolySep-GFC-P 5000 column (Phenomenex, Torrance, CA). At a flow rate of 1 mL/min, rods eluted between 5-8 minutes, depending on length. Though retention times varied with the buffers used, disks typically eluted from 8-9 minutes, while monomers eluted from 9-10 minutes.

Mass Spectrometry. Prior to analysis, biological samples were desalted and separated from small molecule contaminants using spin concentrators or NAP-5 gel filtration columns (GE Healthcare, Chalfont, UK). Control experiments have indicated that assembled TMVP aggregates eluted in the void volume of these columns, while the bulk of small molecules were retained. Electrospray LC/MS analysis was performed using an LTQ Orbitrap XL hybrid mass spectrometer with an Ion Max electrospray ionization source (Thermo Fisher Scientific, Waltham, MA) connected to a Agilent 1200 series liquid chromatograph (Santa Clara, CA) or an API 150EX system (Applied Biosystems) equipped with a Turbospray source and an Agilent 1100 series LC pump.

Transmission Electron Microscopy (TEM). Samples were prepared for TEM analysis by applying an analyte solution (approximately 0.2 mg/mL in TMVP) to carbon-coated copper grids for 3 min, followed by rinsing with dd-H_2O. The grids were then exposed to a 1.5% aqueous solution of uranyl acetate for 90 s as a negative stain. Images were obtained at the Berkeley Electron Microscope Lab using a FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage. Images were recorded on a Ultra Scan 1000 from Getan (Pleasanton, CA).

Spectroscopic Measurements. The spectroscopic properties of all monomer, disk, and rod solutions were investigated by collecting steady state absorbance and fluorescence measurements. UV-vis

spectroscopic measurements were conducted on a Cary 50 UV-Vis spectrophotometer (Varian, Palo Alto, CA). Samples for fluorescence were diluted to 0.45-0.5 OD and spectra obtained on a Fluoromax-4 spectrofluorometer (SA Instruments, Edison, NJ). For experiments measuring energy transfer to the RNA channel, excitation measurements were collected at 630 nm rather than the emission maximum of the acceptor Alexa Fluor 594 to minimize the direct contributions from donor emission.

2.7 Literature Cited

1. Miller, R. A.; Presley, A. D.; Francis, M. B. *J. Am. Chem. Soc.* **2007**, *129*, 3104-3109.
2. Endo, M.; Wang, H.; Fujitsuka, M.; Majima, T. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2006**, *12*, 3735–40.
3. Albrecht, C. *Principles of fluorescence spectroscopy, 3rd Edition, Joseph R. Lakowicz, editor;* 2008; Vol. 390.
4. Mottram, L. F.; Boonyarattanakalin, S.; Kovel, R. E.; Peterson, B. R. *Organic letters* **2006**, *8*, 581–4.
5. Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B. *J. Am. Chem. Soc.* **2005**, *127*, 3718–23.
6. Yi, H.; Nisar, S.; Lee, S.-Y.; Powers, M. A.; Bentley, W. E.; Payne, G. F.; Ghodssi, R.; Rubloff, G. W.; Harris, M. T.; Culver, J. N. *Nano letters* **2005**, *5*, 1931–6.
7. Dawson, P.; Muir, T.; Clark-Lewis, I.; Kent, S. *Science* **1994**, *266*, 776–779.
8. Esser-Kahn, A. P.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2008**, *130*, 15820–15822.
9. Zhang, X.; Li, F.; Lu, X.-W.; Liu, C.-F. *Bioconjugate chemistry* **2009**, *20*, 197–200.
10. Dixon, H. B. F. *Journal of Protein Chemistry* **1984**, *3*, 99–108.
11. Fields, R.; Dixon, H. B. F. *Methods Enzymol*. **1972,** *25*, 409-419.
12. Wu, P.; Brand, L. *Flourescence Spectroscopy*; Methods in Enzymology; Elsevier, 1997; Vol. 278.
13. Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem. Int. Ed.***2006**, *45*, 5307–11.
14. Scheck, R. A.; Francis, M. B. *ACS Chemical Biology* **2007**, *2*, 247–251.
15. Eliot, A. C.; Kirsch, J. F. *Annu. Rev. Biochem.* **2004**, *73*, 383-415.
16. Snell, E. E. *J. Am. Chem. Soc.* **1945**, *67*, 194-197.
17. Bhyravbhatla, B.; Watowich, S. J.; Caspar, D. L. *Biophysical Journal* **1998**, *74*, 604-15.
18. Shire, S. J.; McKay, P.; Leung, D. W.; Cachianes, G. J.; Jackson, E.; Wood, W. I.; Raghavendra, K.; Khairallah, L.; Schuster, T. M. *Biochemistry* **1990**, *29*, 5119–5126.
19. Hirel, P.-H. *Proceedings of the National Academy of Sciences* **1989**, *86*, 8247–8251.
20. Witus, L. S.; Moore, T.; Thuronyi, B. W.; Esser-Kahn, A. P.; Scheck, R. A.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2010**, *132*, 16812-16817.
21. Nicolet, B. H.; Shinn, L. A. *J. Am. Chem. Soc.* **1939**, *61*, 1615.
22. Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chemistry* **1992**, *3*, 138–146.
23. Gust, D.; Moore, T. A.; Moore, A. L. *Accounts of Chemical Research* **2001**, *34*, 40–48.
24. Itoh, T.; Yano, K.; Inada, Y.; Fukushima, Y. *J. Am. Chem. Soc.* **2002**, *124*, 13437–13441.
25. Stephanopoulos, N.; Carrico, Z. M.; Francis, M. B. *Angew. Chem. Int. Ed.* **2009**, *48*, 9498–

502.

26. Lee, M.; Kim, J. H.; Lee, S. H.; Lee, S. H.; Park, C. B. *ChemSusChem* **2011**, *4*, 581–586.
27. Wichmann, R.; Vasic-Racki, D.; Kragl, U. *Technology Transfer in Biotechnology - Advances in Biochemical Engi*
28. Odobel, F.; Blart, E.; Lagrée, M.; Villieras, M.; Boujtita, H.; El Murr, N.; Caramori, S.; Alberto Bignozzi, C. *Journal of Materials Chemistry* **2003**, *13*, 502–510.
29. Guldi, D. M.; Rahman, G. M. A.; Jux, N.; Tagmatarchis, N.; Prato, M. *Angew. Chem. Int. Ed.* **2004**, *43*, 5526–30.
30. Baskaran, D.; Mays, J. W.; Zhang, X. P.; Bratcher, M. S. *J. Am. Chem. Soc.* **2005**, *127*, 6916–7.
31. Laemmli U. K. *Nature* **1970**, *227*, 680–685.

Chapter 3. Nanoscale Protein Assemblies from a Circular Permutant of the Tobacco Mosaic Virus

3.1 Abstract

The protein coat of the tobacco mosaic virus has been explored extensively for the construction of nanoscale architectures. In previous work, we have reported efficient TMV-based light harvesting systems bearing chromophores in a hollow channel of the assembled protein. We have also reported an N-terminal transamination/oximation method that could be used to attach electrodes and catalytic groups to the exterior surface of the rods. To complement these techniques, we report herein a new circular permutant of the TMV capsid protein that repositions the N- and C-termini to the center of the assemblies. This protein can be produced in very high yield through *E. coli* expression, and self-assembles into light harvesting rods that are much like those assembled from the wild-type protein. However, the disks formed from the permutant structure are stable over a significantly wider pH range, greatly improving the practicality of this assembled form for materials applications. The new position of the N-terminus allows functional groups to be installed in the inner pore of the disks, affording geometries reminiscent of natural photosynthetic systems. The permutant also shows the ability to co-assemble with regular monomers, allowing the future generation of multicomponent rod structures that are modified on the exterior and interior surfaces, as well as in the internal RNA channel. Portions of this chapter have appeared in published form: Dedeo, M. T.; Duderstadt, K. E.; Berger, J. M.; Francis, M. B. Nanoscale Protein Assemblies from a Circular Permutant of the Tobacco Mosaic Virus. *Nano Lett.* **2010**, *10*, 181-6.

3.2 Introduction

The protein capsid of the tobacco mosaic virus is emerging as a versatile building block for the construction of ordered nanoscale materials. The 300 nm viral rods have been used to template the deposition of inorganic oxides,[1] and mutants with altered amino acid composition have shown the ability to control the formation of metal nanoparticles[2] and nickel[3] wires. In terms of organic functionality, ionic interactions and site selective modification reactions have been used to attach polymers,[4,5] carbon nanotubes,[6] and surfaces[7,8] to the capsids, with the overall goal of integrating the protein rods with device components. Through the use of cysteine modification chemistry, chromophores[9,10] and porphyrins[11] have been attached to individual capsid monomers to yield conjugates that self-assemble into efficient light harvesting systems containing thousands of interacting groups. From a practical standpoint, TMV can be produced in large quantities through propagation of the intact virus in plants or expression of the coat protein (TMVP; see Table 3-1 for a list of the protein abbreviations that appear in this chapter) in *E. coli*. Regardless of the origin of the protein, the self-assembly process into rod and disk structures appears to tolerate modest changes in its amino acid composition.[2,3,8-11]

Table 3-1. Protein building blocks used or referenced herein.

Building Block	Origin	Abbreviation
intact viral rods	plant	viral rods
TMV coat protein	plant	pdTMVP
TMV coat protein (S123C)	*E. coli*	TMVP
TMV coat protein (S123C) with Ala Gly appended to N-terminus	*E. coli*	AG-TMVP
circular permutant of TMVP (S123C)	*E. coli*	cpTMVP

These strategies provide the groundwork for the generation of complex TMVP-based materials with multiple integrated components. In the case of light harvesting systems, the obvious next steps involve the introduction of metal complexes that use the collected energy to drive chemical reactions,[12] as well as the addition of "sacrificial" chromophores that can salvage trapped excitons[13] and resist photobleaching.[14] To build these materials, an additional set of chemical modifications is needed to introduce new groups within a few nanometers of the light collecting rings. As one method to do this, we have recently introduced a transamination/oximation strategy that functionalizes a new TMVP mutant (AG-TMVP) at its extended N-terminus, Figure 3-1a.[15] This allows a range of new functional groups to be installed on the external surface of self-assembled TMVP disks and rods. To complement this strategy, we report in this paper a new circular permutant[16,17,18] of TMVP (cpTMVP) that allows the facile construction of internally modified TMVP-based materials. This protein variant relocates the N- and C-termini to a flexible loop that faces the center of disks and rods after assembly, Figure 3-1. This arrangement allows functionalization of the internal pore of the assemblies through the same transamination/oximation process.

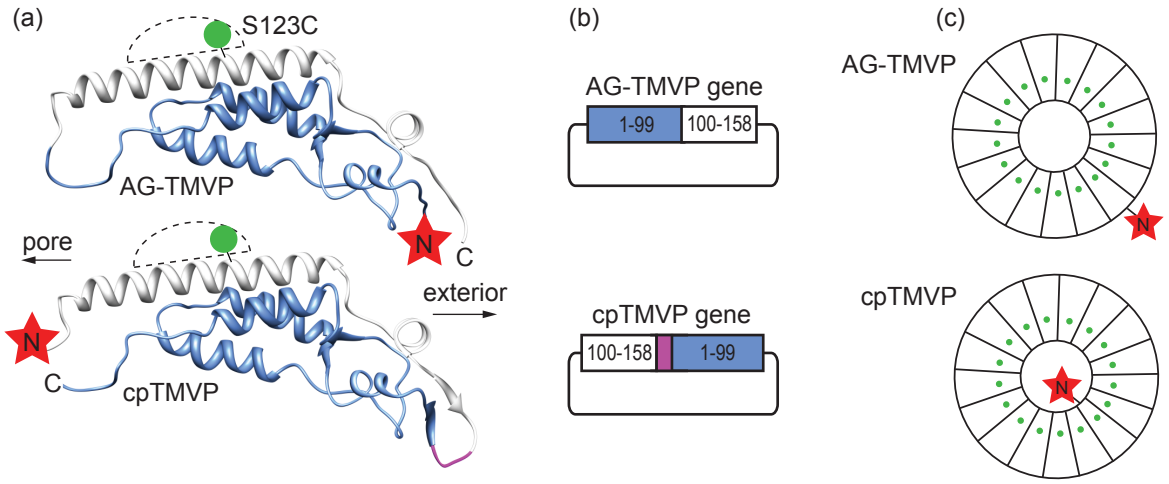

Figure 3-1. Design of a TMVP circular permutant that relocates termini to the pore of disks. (a) The top image shows normal AG-TMVP and the lower image shows the expected structure of a circular permutant, cpTMVP, in which the native termini are connected with a GEG linker (magenta). A flexible loop is interrupted to produce new termini. The location of the reactive cysteine 123 is shown as a green circle inside the RNA channel (dotted line), and the reactive N-terminus as a red star. (b) To create the cpTMVP gene, 5' (blue) and 3' (white) segments of the AG-TMVP gene were switched and connected with a short linker (magenta). (c) Self-assembly into disks and rods positioned the N- and C-termini on the exterior surfaces of AG-TMVP and the pores of cpTMVP.

3.3 Expression and Assembly of Permuted TMV

We began by purchasing a TMVP gene optimized for the codon usage of *E. coli*, in hopes of maximizing protein yield.[19] The construct also possessed the S123C mutation used previously for chromophore attachment. This was rearranged to form the cpTMVP gene using standard molecular biology techniques and cloned into a pET vector. Overexpression of the redesigned gene in *E. coli* produced a soluble protein of the expected mass as seen in Figure 3-2. Purification of cpTMVP proved facile and reliable by precipitation with ammonium sulfate followed by anion exchange, providing highly pure protein in yields up to 100 mg/L culture.

AG-TMVP (and TMVP) assume three major quaternary structures, preferring monomers/small aggregates in basic buffers, disks at neutral pH, and rods at lower pH (though assembly also depends on ionic strength).[20,21] Our first goal was to determine if cpTMVP would behave in a similar fashion. When exchanged into a neutral buffer known to cause disk formation of AG-TMVP,

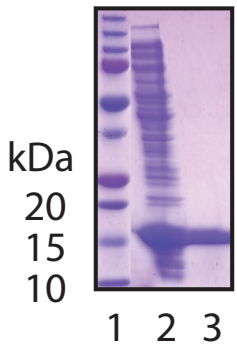

Figure 3-2. Purification of cpTMVP visualized by denaturing polyacrylamide gel electrophoresis (SDS-PAGE). Lane 1: molecular weight marker. Lane 2: clarified lysate from expression in *E. coli*. The overexpression band matched the expected mass of cpTMVP (17707 Da). Lane 3: resuspended pellet from ammonium sulfate precipitation.

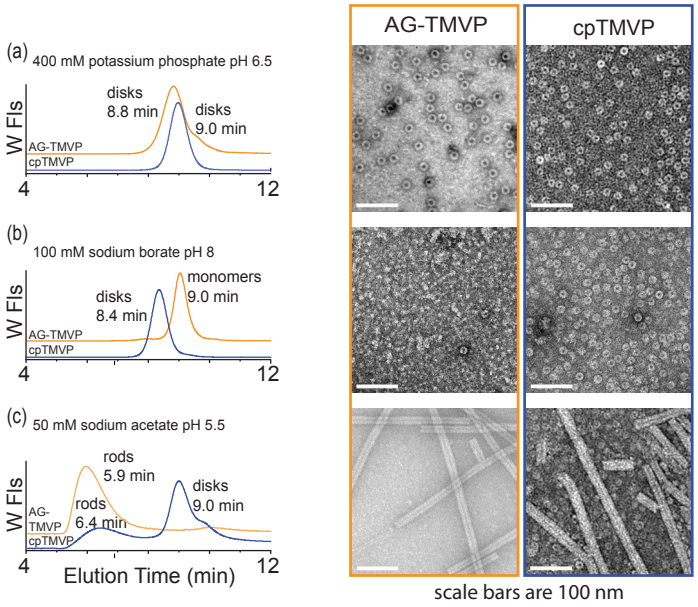

Figure 3-3. Assembly of AG-TMVP and cpTMVP, as measued by size exclusion chromatography (SEC) and transmission electron microscopy (TEM). (a) Disks of AG-TMVP and cpTMVP elute at similar times on SEC and are indistinguishable by TEM. W Fls refers to tryptophan fluorescence (λex = 295 nm, λem = 330 nm). (b) In alkaline buffers (e.g. 100 mM borate, pH 8) cpTMVP is predominantly assembled into disks wheras AG-TMVP is predominantly smaller aggregates, as seen from their retention times on SEC. The left TEM image shows two disks of AG-TMVP surrounded by smaller aggregates while the right shows a dense layer of cpTMVP disks. (c) Rods of cpTMVP and AG-TMVP elute at similar times in a SEC trace and are indistinguishable by TEM, though cpTMVP rods assemble more slowly and are often in equilibrium with disks. Note that SEC traces drift with the elution buffer, and should be taken as relative comparisons.

transmission electron microscopy (TEM) showed cpTMVP as disks with diameters of 16-20 nm and 3-5 nm pores. These structures were indistinguishable from AG-TMVP derived disks (Figure 3-3a). Assembly was further confirmed by size exclusion chromatography (SEC) (Figure 3-3a), as the cpTMVP had a retention time similar to AG-TMVP disks measured in the same mobile phase buffer. AG-TMVP readily dissociated into monomers when the pH was raised to 8, but cpTMVP remained as disks in buffers up to pH 9 for weeks (Figure 3-3b). This increased stability makes

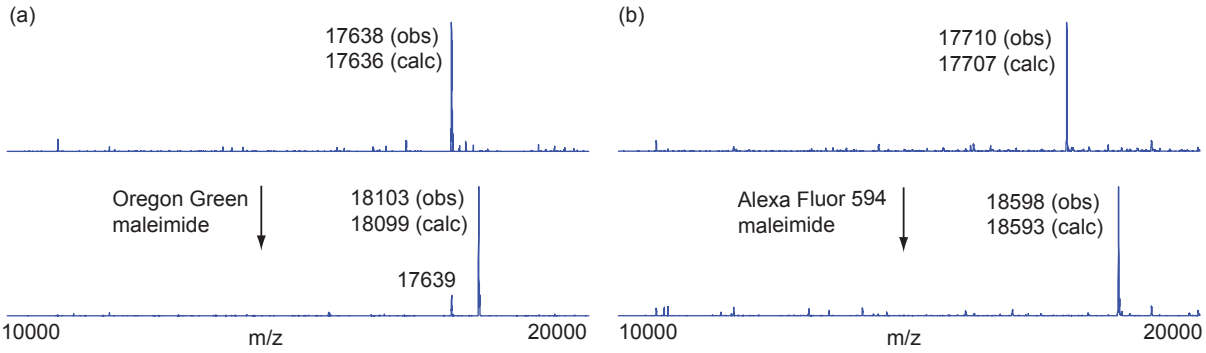

Figure 3-4. Modification of C123S with maleimide dyes. (a) Liquid Chromatography Mass Spectrometry (LCMS) of TMVP before and after conjugation of C123S with Oregon Green maleimide. (b) LCMS of cpTMVP before and after conjugation with Alexa Fluor 594 maleimide.

cpTMVP disks much more practical for the construction of nanoscale rings resembling natural light harvesting systems. Monomers were eventually obtained by screening low ionic strength conditions, such as 20 mM TEA, pH 8 (Figure 3-7b).

To assemble the disks into rods, we first modified S123C with a maleimide dye (Figure 3-4) before dialyzing into a pH 5.5 acetate buffer. This places the dye in or near the RNA binding channel (Figure 3-1a), and has been observed to encourage assembly into rods.[9] Though cpTMVP rods were slower to assemble and often in equilibrium with disks, they were otherwise similar to those formed from AG-TMVP. Both exhibited a dark central pore, an average diameter of 25 nm, and SEC retention times of 5-8 minutes (Figure 3-3c). These assembly data suggest that the circular permutation had only a minor effect on protein's structure and ability to self-assemble, though the permutant's preference for disks in both alkaline and acidic buffers suggests the cpTMVP disk may be more stable than the AG-TMVP disk.

To probe the structural changes responsible for this altered assembly pattern, we crystallized the protein for x-ray diffraction using conditions adapted from work on the plant derived TMVP (pdTMVP) disks.[22] The structure was solved by molecular replacement, using the 'A' disk of a

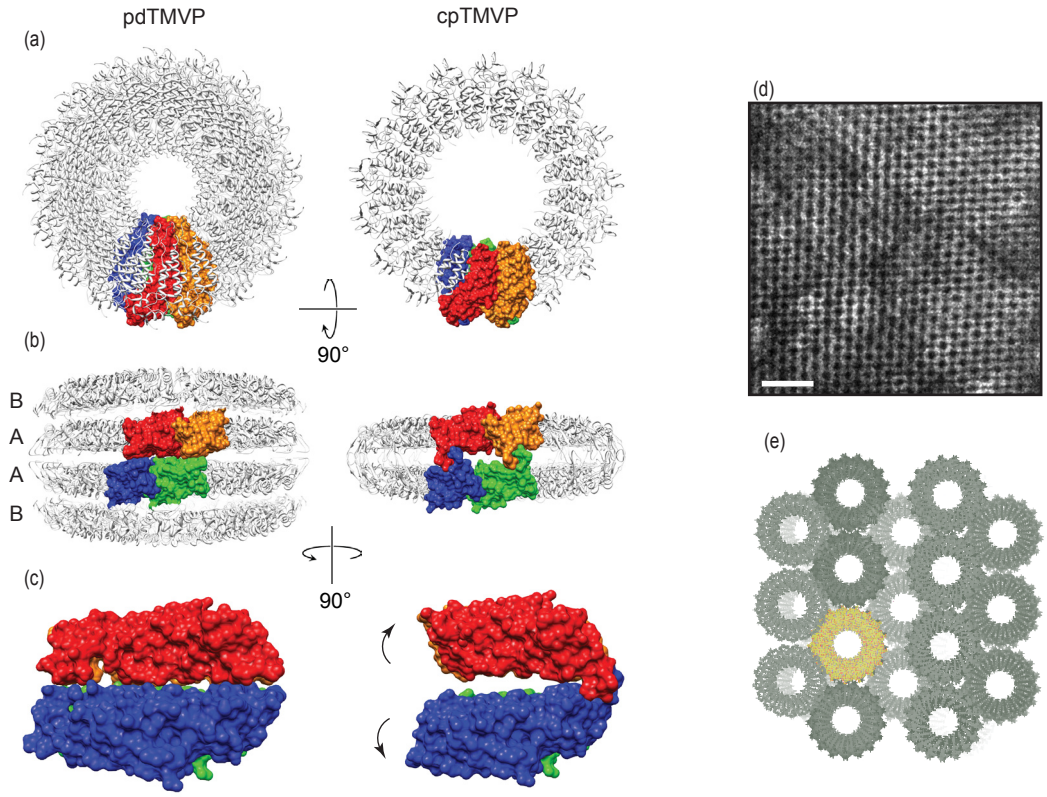

Figure 3-5. Refined atomic models of pdTMVP and cpTMVP. (a) cpTMVP and pdTMVP crystallize as disks containing 17 monomers per layer. Looking down the 17-fold noncrystallographic axis, the size of the pore appears to be exaggerated in cpTMVP relative to pdTMVP because the N- and C-termini are disordered and absent from our model. The actual pore sizes appear to be identical by TEM. (b) The top and bottom disk(s) of cpTMVP and pdTMVP are related by a crystallographic two-fold axis. Compared to the A disk pair of pdTMVP, the disk pair of cpTMVP are rotated with respect to one another about the 17-fold axis to bring their new external loops into contact. (c) The larger gap between disks in cpTMVP relative to the A disk pair of pdTMVP is created by a rigid-body translation, as well as a 2.8° rotation of each monomer (shown by arrows). (d) Highly ordered cpTMVP disks were observed by TEM. The packing resembles that seen in the crystal structure of cpTMVP (e).

2.45 Å structure of pdTMVP. The final model was refined to an R_{free} of 25.24% (full statistics in methods section below).[23] We were reassured to find the permutant disks contained the expected 17 monomers, but were surprised that they crystallizes as a pair of *single* disks with C_2 symmetry (Figure 3-5). This is in contrast to pdTMVP, which crystallized as a pair of A-B *double* disks with C_2 symmetry. Aligning the backbone of cpTMVP and pdTMVP showed a nearly identical fold (RMSD = 0.308 Å, Figure 3-6a). Furthermore, side chain interactions identified as significant for

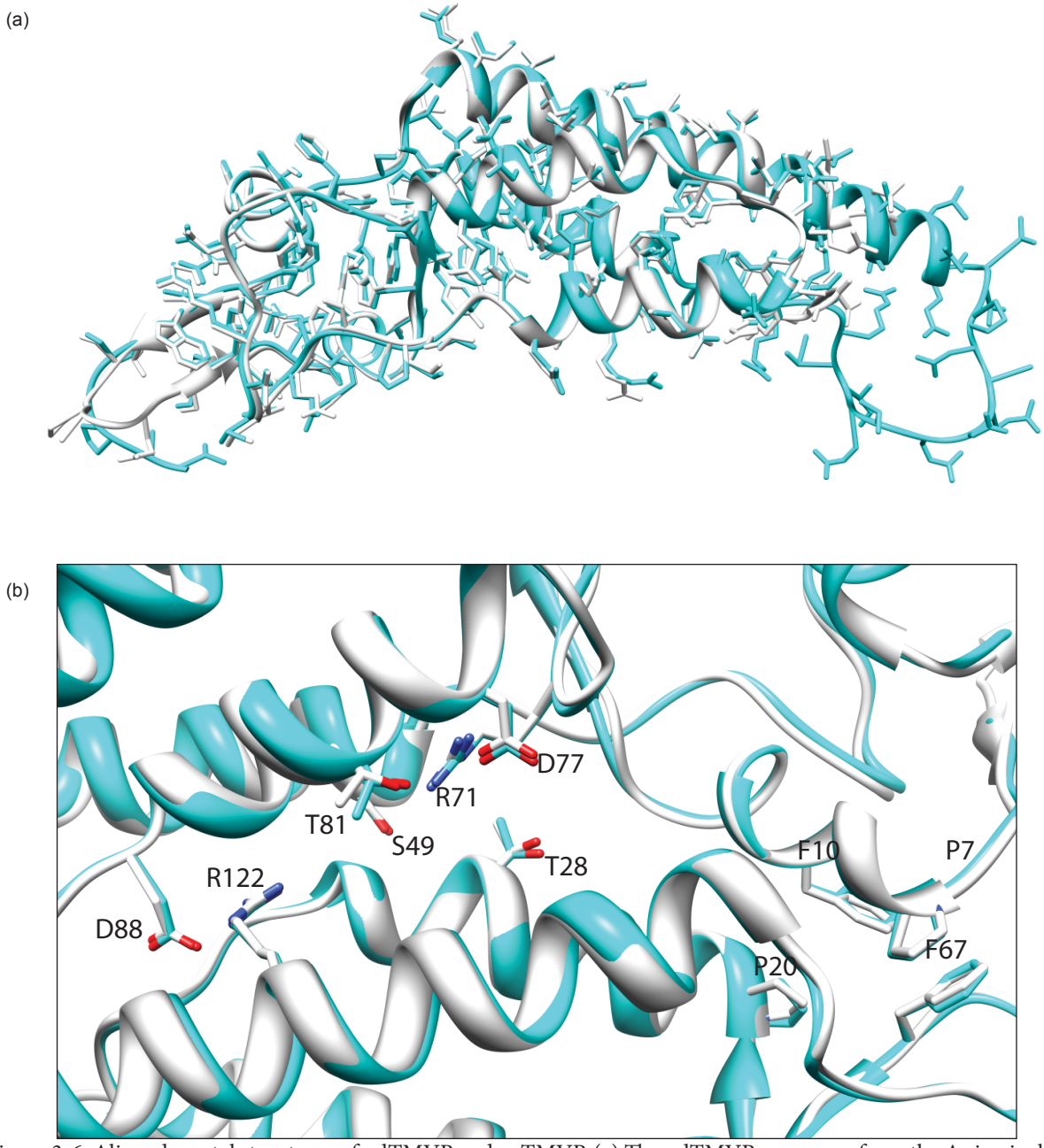

Figure 3-6. Aligned crystal structures of pdTMVP and cpTMVP. (a) The pdTMVP monomer from the A ring is shown in teal and cpTMVP is overlaid in white to show the similarity in protein fold. The RMSD of the alpha carbons was 0.308 Å. (b) Lateral interactions observed in pdTMVP A disks[39] (teal) are maintained in cpTMVP disks (white). Side chains important for ionic, polar, and hydrophilic interactions have similar orientations and spacing in pdTMVP and cpTMVP disks. Residues of pdTMVP are numbered.

the lateral assembly in pdTMVP[23] appear to have been maintained in cpTMVP (Figure 3-6b). As has been observed in an earlier 2.8 Å pdTMVP crystal structure,[24] the central regions of the disks were disordered and so could not be modeled. The same flexible loop region of pdTMVP eluded crystallographers for many years, and when finally modelled,[23] the electron density was insufficient to include side chains.

A fragment of a cpTMVP crystal is shown in a TEM image with the model for comparison (Figure 3-5d and 5e). In addition to using these highly ordered arrangements for structure determination, we are currently exploring their potential as functionalizable protein scaffolds that bridge the nano- and the macroscopic scales.[25]

3.4 Co-assembly of Conventional and Permuted TMV

Given the similarity of the protein folds and lateral interfaces, we wished to test if AG-TMVP and cpTMVP could co-assemble into disks and rods. This property would allow us to produce structures that are modified on both the external and the internal surfaces. Previous work with TMVP[9] showed that donor chromophores conjugated to S123C could undergo fluorescence resonance energy transfer (FRET) to neighboring donor or acceptor chromophores following assembly into rods. We used FRET as a test for co-assembly by labeling S123C of AG-TMVP with a donor chromophore (Oregon Green 488) and cpTMVP with a suitable acceptor (Alexa Fluor 594), Figure 3-7 and 3-4. Monomeric solutions of each labeled protein in 20 mM TEA pH 8 were combined in a 1:1 or 16:1 ratio of donor:acceptor. Before assembly, spectra were collected, and then self-assembly was effected by dialysis against 400 mM potassium phosphate, pH 6.5. After 1-2 d, SEC and TEM indicated that the bulk of the proteins in both solutions had assembled into rods. Insoluble protein aggregates were removed by centrifugation (including long rods and tangles of rods, as observed by TEM). Absorbance measurements of the resulting solutions showed greater loss of donor-labeled AG-TMVP, which shifted the ratio of donor:acceptor from 16:1 to 8:1. Since the exact ratio is not critical for FRET, fluorescence spectra were then recorded of solutions of rods.

Energy transfer, and therefore co-assembly, is clearly shown by both emission and excitation spectra (Figure 3-7e-g). Donor excitation (7f) results in no visible acceptor emission at 612 nm until the monomers are assembled into rods. The excitation spectra collected for the acceptor (3-7g) shows a large increase in contribution from the donor upon assembly. Energy transfer is quantified at 74% by overlaying the absorbance and excitation spectra (3-7g), which is higher than the 39% observed for an 16:1 system using only TMVP.[9] Energy transfer was also observed for analogous systems containing only cpTMVP (data not shown).

3.5 Energy Transfer to the Pore of the Permutant

Having relocated the N-terminus of cpTMVP to the pore of disks and rods, we sought to install centrallized acceptor groups to collect the energy absorbed by the ring of chromophores conjugated to S123C (Figure 3-9a). By examining x-ray structures of TMV rods and disks of TMV coat protein, we were able to establish the feasability of FRET between a donor chromophore in the RNA channel and an acceptor in the central pore. The distances between the side chain of the

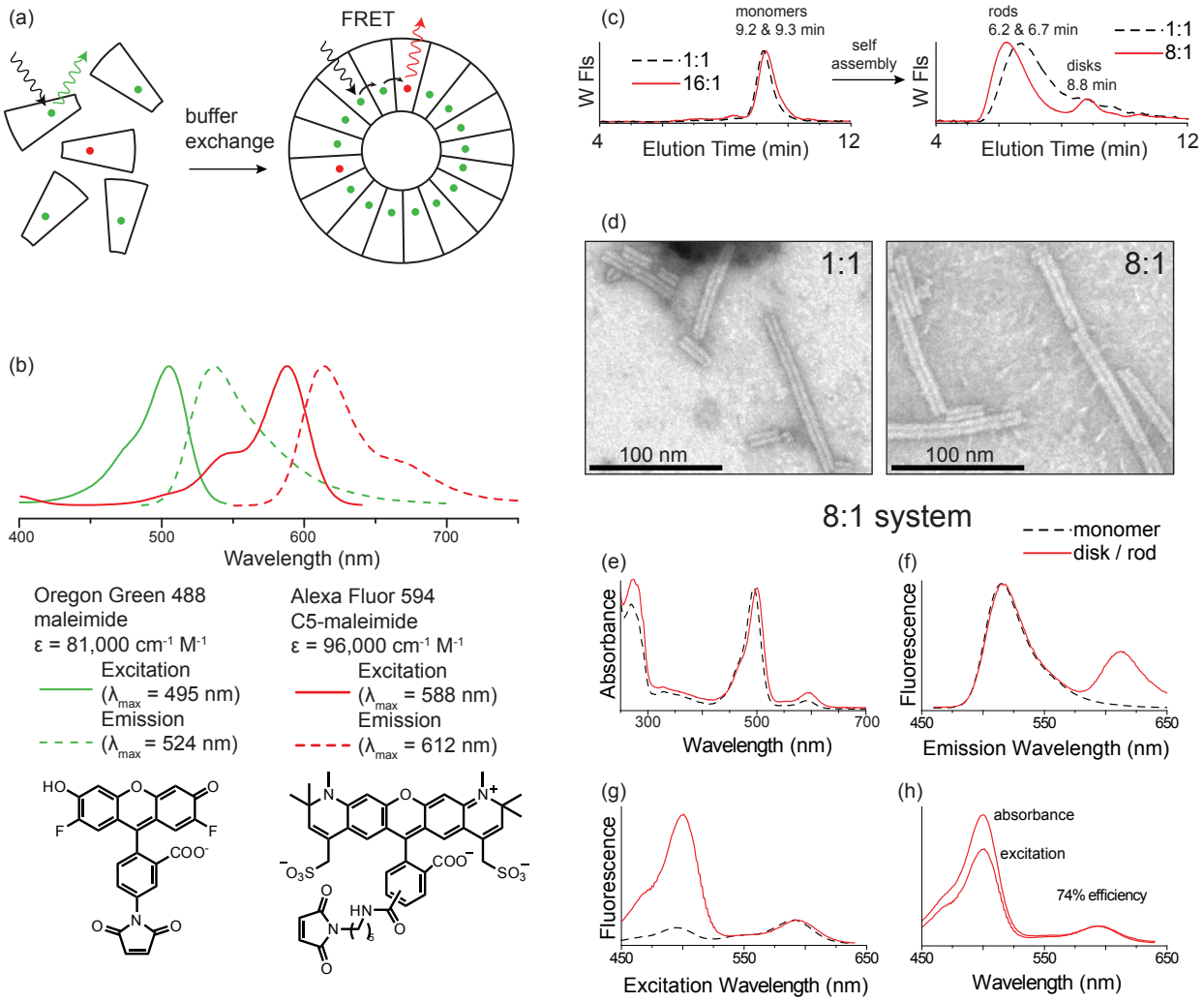

Figure 3-7. Co-assembly of cpTMVP and AG-TMVP measured by SEC, TEM and FRET. (a) A cartoon depicts assembly-dependent FRET. A solution containing monomers of donor-labeled AG-TMVP and acceptor-labeled cpTMVP exhibits little energy transfer. Upon buffer exchange, monomers self-assemble into disks and rods, bringing donor and acceptors within ~1.5 nm to allow FRET. (b) Normalized fluorescence excitation and emission spectra of donor (Oregon Green) and acceptor (Alexa Fluor 594) are plotted to show spectral overlap. (c) Upon buffer exchange, monomers of donor-labeled AG-TMVP and acceptor-labeled cpTMVP self-assemble into rods and disks, as seen by SEC. (d) TEM images indicate most rods are between 100-200 nm. Notably, the width of the 1:1 rods appears to be 80% of the 8:1, pure cpTMVP, or pure AG-TMVP. (e) The absorbance spectra are normalized to the donor. The 16:1 ratio of D:A labeled monomers shifted to 8:1 after self-assembly and removal of large protein aggregates. (f) Donor excitation (450 nm) results in acceptor emission (612 nm) only after self-assembly. The shift in D:A ratios upon assembly potentially exaggerates the difference in acceptor emission. (g) Excitation spectra (em = 650 nm) reveal that the donor chromophore contributes a significant amount of energy towards acceptor emission only after self-assembly. The shift in D:A ratios upon assembly potentially understates the difference in donor contribution. (h) Overlay of the absorbance and excitation spectra exhibits a 74% efficiency in energy transfer.

serine in position 123 and the alanines in position 100 of the surrounding monomers was measured and the FRET efficiency calculated using the methodology presented in chapter 2. The results indicated that there are redundant, highly efficient pathways to transfer energy from the ring of donors to a central acceptor in both disks and rods (Figure 3-8). Such an arrangement is reminiscent of the photosynthetic reaction center, which is surrounded by the light harvesting chromophores of

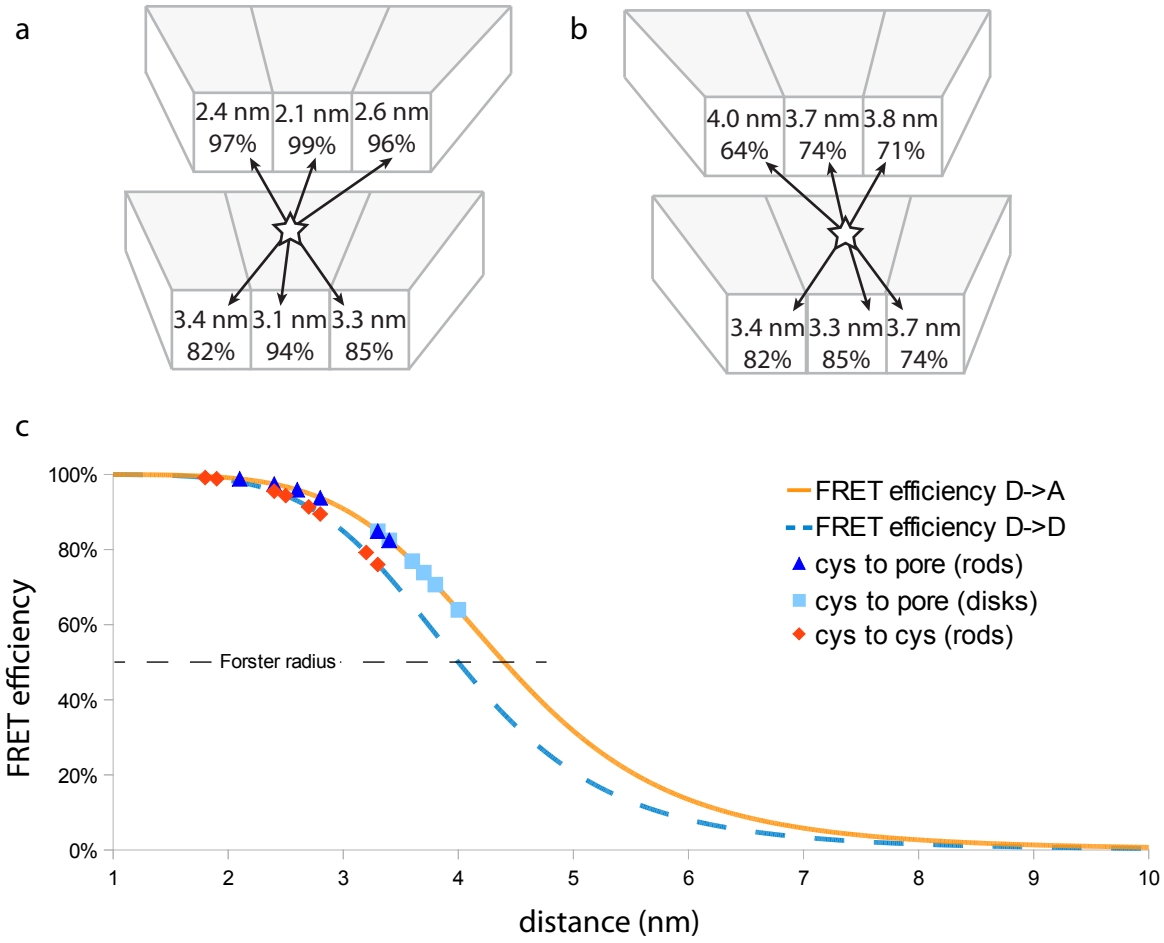

Figure 3-8. FRET efficiency as a function of distance. A diagram of protein packing in the TMV rod (a) and disk (b) shows distances and estimated FRET efficiencies for energy transfer from a donor in the RNA channel to an acceptor in the pore of the surrounding monomers. The layers are contiguous - the gap is only to show the location of the donor dye . In (c) FRET efficiencies for donor to acceptor (solid orange line) and donor to donor (blue dashed line) are plotted as a function of interchromophore distance. Blue triangles and light blue squares show the estimated efficiency of energy transfer of an isolated donor-acceptor system separated by the distances measured for rods and disks, respectively. For comparison, the analogous efficiency of energy transfer between donor molecules arrayed in the RNA channel is depicted with red diamonds.

LH1.[26] Eventually, a photocatalyic acceptor, such as a porphyrin, will be used to drive a chemical reaction with the collected energy, but to test the system we began with another fluorescent dye (Figure 3-8). We chose Alexa Fluor 647 as an acceptor because it has a ketone-reactive aminooxy group, and Texas Red as the donor because its emission spectrum overlaps both its own excitation and that of the acceptor. This feature allows FRET to occur between neighboring donors and from donor to acceptor. The cpTMVP already contained a cysteine for conjugation to a donor, so to prepare it for subsequent conjugation to the acceptor, it was incubated with 10 mM pyridoxal-5'-phosphate (PLP) for 24 hours at 30 ºC to replace over 50% of N-terminal amines with ketones (Figures 10c and 11). Small peaks at +230 and +249 corresponding to covalent N-terminal PLP aldol adducts were also observed, which have not been found to interfere with subsequent experiments (proposed structures appear in Figure 3-10c). The addition of the maleimide-functionalized donor Texas Red was accompanied by a shift in the mass of the protein corresponding to the weight of

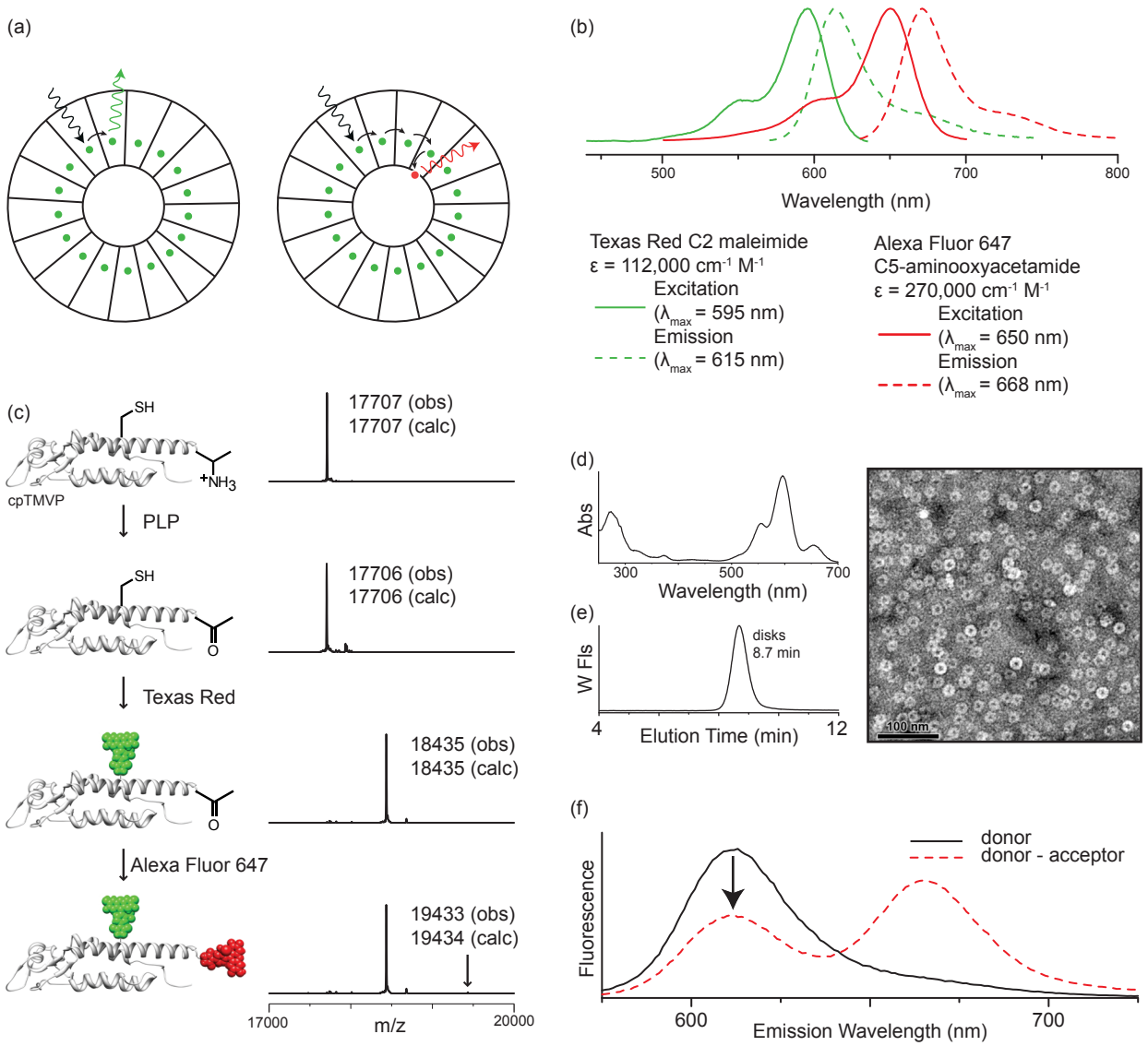

Figure 3-9. Site specific labeling of cpTMVP with donor and acceptor chromophores to measure coassembly. (a) A cartoon representation of a cpTMVP disk shows the approximate locations of donor and acceptor chromophores. (b) Normalized fluorescence excitation and emission spectra of the donor (Texas Red) and acceptor (Alexa Fluor 647) are plotted to show spectral overlap. (c) First, cpTMVP is reacted with PLP to transaminate the N-terminus, introducing a uniquely reactive ketone. Donor chromophore coupling to a cysteine via maleimide chemistry results in complete labeling, as observed by ESI-MS. Subsequent labeling of ~6% of the N-termini with an aminooxy-functionalized acceptor chromophore gives an average of ~1 acceptor per disk of 17 donors, as quantified by UV-vis spectroscopy (d). (e) Double modification does not perturb the structure of disks, as measured by SEC and TEM. (f) Overlay of the emission spectra of donor only and donor - acceptor systems shows energy transfer. Efficiency is defined as the percent decrease in donor fluorescence compared to a system containing no acceptor dyes. Disks containing ~17:1 ratio of donor:acceptor exhibit 45% quenching.

the dye. The oxime-formation between the N-terminal ketone and the acceptor Alexa Fluor 647 aminooxyacetamide proved more challenging, requiring 24 h and addition of *p*-anisidine as a catalyst[27] to install, on average, one acceptor in the pore of each ring of 17 donors. All reactions and measurements were performed in 25 mM potassium phosphate, pH 6.5. Double modification did not appear to perturb the structure of the disks (Figure 3-9e).

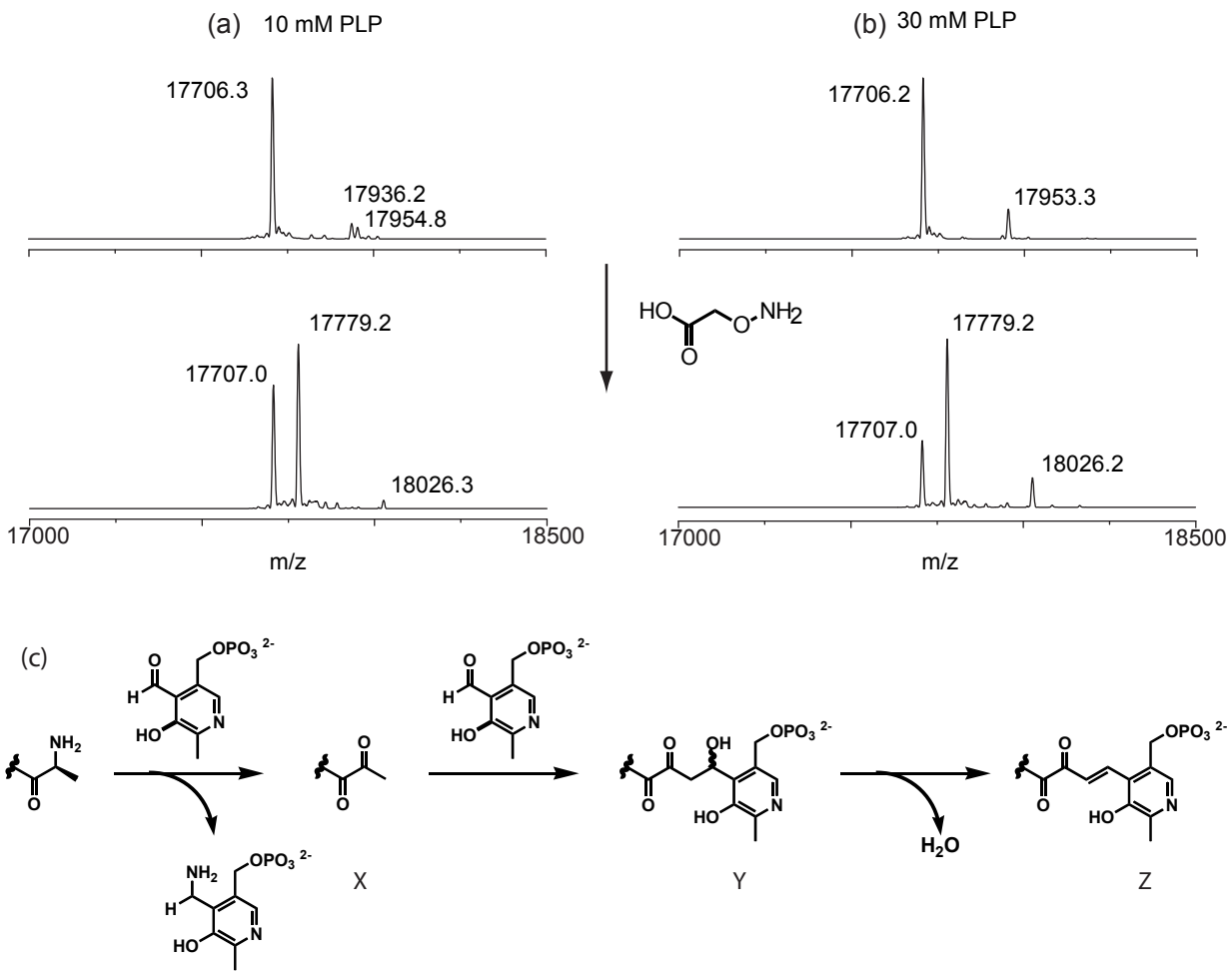

Figure 3-10. Modification of transaminated N-terminus observed by LCMS. To measure the maximum possible modification of the N-terminus, the cpTMVP was incubated with PLP, followed by a large excess of (aminooxy)acetic acid. The degree of conversion is estimated from the relative abundance of signal corresponding to the mass of unmodified (17707) and modified cpTMVP (17779). Small peaks at 17936, and 17953/17955 appear to be the result of aldol condensations and reactions, respectively, between the transaminated N-terminus and PLP. In either case, the resulting ketone is capable of reacting with aminooxy groups, albeit more slowly than the transaminated N-terminus which it replaced. (a) 24 hour incubation with 10 mM PLP leads to >50% transamination and a small fraction of PLP-adduct. (b) Incubation with 30 mM PLP leads to >70% transamination and a larger fraction of PLP-adduct. (c) The N-terminus of cpTMVP is transaminated to give the ketone (X). This can undergo an aldol reaction with PLP to form a stable covalent adduct (Y). Loss of water leads to the formation of the adduct (Z). Both the transaminated N-terminus and the adducts have been observed to participate in oxime formation, though the adducts react more slowly.

Energy transfer from the ring of donors to the central acceptor was measured by comparing the emission spectra (normalized to absorbance) of disks containing only donors to those containing both donors and acceptors. Quenching of the donor is quantified by the decrease in intensity of its emission peak at 615 nm, corresponding to 46% energy transfer (Figure 3-9f).

3.6 Quaternary Structure of Disks in Solution

Despite the clues provided by the crystal structure of the permuted TMV disks, their

assembly state in solution was not definitively established. Because the formation of long rods (Figure 3-3) requires the polar alignment of disks, it was not clear if the apolar double disks observed in the crystal structure were an artifact of the buffer conditions, or one of multiple stable structures. TEM images of disks on their sides showed predominantly double disks, with some stacks of four and no single disks (Figure 3-11a). However, the resolution of the instrument was insufficient to determine whether the double disks were apolar, as in the crystal structure of cpTMVP, or polar, as in the AB double disk of seen in the crystal structure of TMVP. AFM studies to determine if the disks deposited on a poly-L-lysine surface were one or two layers thick were inconclusive. Previous experiments had shown that normally assembled TMV disks with

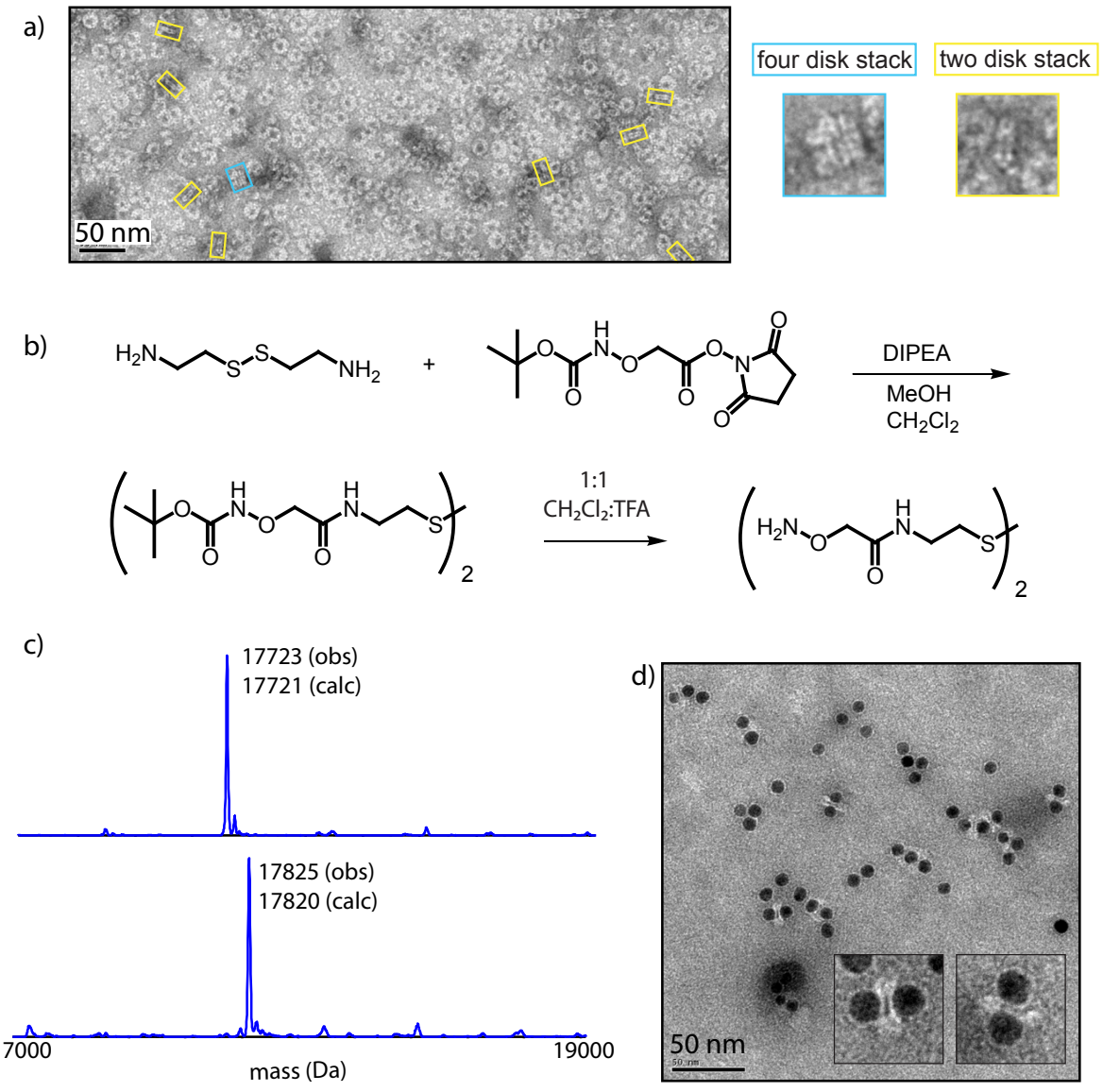

Figure 3-11. Thiolation of the N-terminus and binding of gold nanoparticles as a probe of quaternary structure of permuted TMV. (b) A thiol-alkoxyamine crosslinker was synthesized. (c) cpTMV with an N-terminal serine (top spectrum) was oxidized with periodate and reacted with the alkoxyamine thiol crosslinker (bottom spectrum) to introduce a ring of thiols into the pore. (d) After overnight incubation with 10 nm gold nanoparticles, these thiolated disks bound gold on both sides, as seen by TEM.

thiols introduced in the pore would bind gold nanoparticles on only one face (Chapter 4). We hypothesized that introduction of thiols to the permuted TMV and conjugation to gold could help determine if the disks were polar (binding gold on one side) or apolar (binding gold on both sides).

Because no reaction was observed between gold particles and the cysteine in position 123, a thiol-alkoxyamine crosslinker was synthesized to modify the N-terminus (Figure 3-11b). This two step reaction provided a crosslinker that reacted completely with the glyoxylyl group generated by the oxidation of the N-terminal serine (Figure 3-11c). To avoid precipitation of protein due to extensive crosslinking, the disulfide was reduced by addition of TCEP to the reaction. TEM images of mixtures of thiolated TMV with an excess of gold clearly showed gold binding to both sides, which has not been observed for normally assembled TMV disks under similar (or any) conditions (Figure 3-11d). This strongly suggests that the bulk of the permuted TMV disks are assembled as apolar double disks, as observed in the crystal structure.

3.7 Conclusion

This new protein represents one of the largest departures from the wild-type TMVP sequence to date, and provides many new opportunities for the construction of materials possessing different components in specified locations. As an added benefit, cpTMVP shows significantly increased ability to assemble into disks that are stable across a wide pH range. These assemblies display four anisotropic surfaces that can potentially be functionalized for light collection, catalyst positioning, and surface attachment. Three of these surfaces can now be modified in a site-selective fashion, and methods are currently being developed to modify the final "bottom" surface of the disks. Taken together, this set of chemical modification strategies, combined with the ability of TMV coat proteins to assemble into multiple different structures, provides one of the most versatile protein scaffolds available for the construction of multicomponent materials.

3.8 Materials and Methods

General Procedures. Unless otherwise noted, all chemicals and solvents were of analytical grade and were used as received from commercial sources. Water (dd-H_20) used in biological procedures or as reaction solvents was deionized using a NANOpure purification system (Barnstead, United States). Spin concentration steps were performed using 100,000 molecular weight cutoff spin concentrators from Millipore (Billerica, MA) in an Allegra 64R Tabletop Centrifuge (Beckman Coulter, Fullerton, CA). Large volume dialysis was performed with 3000 molecular weight cutoff SnakeSkin dialysis tubing (Pierce, Rockford, Il). Small volume dialysis was performed with 3500 molecular weight cutoff Slide-A-Lyzer Dialysis Cassettes (Pierce), as indicated below. Protein samples were quantified by the Bradford assay.

Preparation of the DNA Construct for cpTMVP. The starting point for the permutant was a gene for the coat protein of the TMV U1 strain optimized for the codon usage of *E. coli* (Genscript, Piscataway, NJ). This sequence contained the mutation S123C, as well as an Ala Gly addition to the N-terminus to facilitate the PLP mediated transamination reaction. The following procedure was adapted from standard molecular biology techniques:[29,30] The permutant was created by a

two-step polymerase chain reaction (PCR) method in which the N- and C-terminal segments of the TMVP gene were first amplified separately. Both the forward primer for the N-terminal segment (FL) and reverse primer for the C-terminal segment (RL) coded for the flexible linker (GEG), which caused the two segments to overlap and template one another in a second PCR reaction. This linker replaced both the C-terminal Thr and the appended N-terminal Ala Gly of AG-TMVP. The forward primer for the C-terminal segment appended a Gly to the new N-terminus. The gene for the permuted TMV coat protein (cpTMVP) was then cloned into an *E. coli* expression vector (pET 20b, Novagen) using NdeI and EcoRI restriction sites. Protein from this gene was used for crystallography only. Subsequently, the gene was re-cloned using new primers to replace the N-terminal Gly Ala with Ala Gly to facilitate the PLP mediated transamination reaction. This protein was used for the remaining experements, and both are referred to as cpTMVP. The DNA sequences of the primers, optimized AG-TMVP gene, and cpTMVP gene were as follows:

R: GGTGGTTGAATTCTTATTACTGGTTTTCCACTTCAATGATACG
FL: GCGGGCGAAGGCAGCTATAGCATTAC
RL : GCCTTCGCCCGCCGGGCCGC
F: GGTGGTTCATATGGCCGGCAATCCGACCACC

Optimized AG-TMVP DNA sequence:
GCCGGCAGCTATAGCATTACCACCCCGAGCCAGTTTGTGTTTCTGAGCAGCGCCTGGGCGGATCCGATTGAACTGAT-
TAACCTGTGCACCAATGCGCTGGGCAACCAATTTCAGACCCAGCAGGCGCGCACCGTTGTGCAGCGTCAGTTCAGC-
GAAGTTTGGAAACCGAGCCCGCAGGTTACCGTGCGCTTTCCGGATAGCGATTTTAAAGTGTATCGCTATAACGCCGT-
GCTGGATCCGCTGGTGACCGCGCTGCTGGGCGCCTTTGATACCCGTAATCGTATCATTGAAGTGGAAAACCAGGC-
CAATCCGACCACCGCGGAAACCCTGGATGCGACCCGTCGTGTGGATGATGCCACCGTGGCGATTCGCTGTGCCAT-
CAATAACCTGATTGTGGAACTGATTCGTGGCACCGGCAGCTATAACCGTAGCAGCTTTGAAAGCAGCAGCGGCCTG-
GTGTGGACGAGCGGCCCGGCGACC

cpTMVP DNA sequence:
GCCGGCAATCCGACCACCGCGGAAACCCTGGATGCGACCCGTCGTGTGGATGATGCCACCGTGGCGATTCGCTGT-
GCCATCAATAACCTGATTGTGGAACTGATTCGTGGCACCGGCAGCTATAACCGTAGCAGCTTTGAAAGCAGCAGC-
GGCCTGGTGTGGACGAGCGGCCCGGCGGGCGAAGGCAGCTATAGCATTACCACCCCGAGCCAGTTTGTGTTTCT-
GAGCAGCGCCTGGGCGGATCCGATTGAACTGATTAACCTGTGCACCAATGCGCTGGGCAACCAATTTCAGACCCAG-
CAGGCGCGCACCGTTGTGCAGCGTCAGTTCAGCGAAGTTTGGAAACCGAGCCCGCAGGTTACCGTGCGCTTTC-
CGGATAGCGATTTTAAAGTGTATCGCTATAACGCCGTGCTGGATCCGCTGGTGACCGCGCTGCTGGGCGCCTTTGAT-
ACCCGTAATCGTATCATTGAAGTGGAAAACCAG

cpTMVP amino acid sequence with linker shown in red:
AGNPTTAETLDATRRVDDATVAIRCAINNLIVELIRGTGSYNRSSFESSSGLVWTSGPAGEGSYSITTPSQFVFLSSAWAD-
PIELINLCTNALGNQFQTQQARTVVQRQFSEVWKPSPQVTVRFPDSDFKVYRYNAVLDPLVTALLGAFDTRNRIIEVENQ

Buffer A - 25 mM potassium phosphate, pH 6.5

Protein Expression. Tuner DE3pLysS competent cells (Novagen) were transformed with the vector described above, and cultured in Terriffic Broth with 100 µg/L ampicillin. When cultures reached an optical density of 0.6 to 0.8, IPTG was added to a final concentration of 30 µM. Cultures were grown 24 h at 30 °C, harvested by centrifugation, and stored at -80 °C. Cells (from a 2 L expression batch) were thawed, resuspended in 40 ml of 20 mM TEA pH 7.2, and lysed by sonicating with a 2 s on, 6 s off cycle for a total of 20 minutes using a standard disruptor horn at 90% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared by ultracentrifugation for 30 min at 45,000 rpm using a Beckman 45 Ti rotor in an Optima L-80 XP (Beckman Coulter) or 30 min at 10,000 rpm using a using a SLA-600TC rotor in a Sorvall RC5C Plus centrifuge (Waltham, MA). The clarified lysate was decanted, warmed to rt, and

stirred while adding a saturated solution of ammonium sulfate dropwise to a final concentration of 35% (v/v) for AG-TMVP or 15-17.5% for cpTMVP. After 5 min, the white ppt that formed was pelleted by ultracentrifugation (15 min at 45,000 rpm in a Beckman 45 Ti rotor), washed with deionized water, and resuspended in buffer A. This solution was then dialyzed against the same buffer to remove residual ammonium sulfate using SnakeSkin tubing (Pierce). The resulting protein solution was next loaded onto a DEAE column and eluted with a 0 – 300 mM NaCl gradient. Purity was confirmed by SDS-PAGE and HPLC. This preparation provided highly pure cpTMVP in yields up to 100 mg/L culture.

Protein Crystallization. Purified cpTMVP in unbuffered water was spin concentrated, centrifuged to remove any precipitated protein, and determined to be 25 mg/mL using UV analysis. Diffraction-quality crystals were formed by mixing equivalent volumes of concentrated cpTMVP and crystallization buffer (300 mM ammonium sulfate, 100 mM HNO_3, pH 8.0) in a hanging drop suspended over crystallization buffer to allow for vapor diffusion at 18 °C. Crystals formed over 1-3 weeks. They were harvested in 25% (v/v) glycerol cryoprotectant, where the glycerol only replaced the water in the mother liquor, and immediately frozen in liquid nitrogen.

X-ray diffraction Data Collection, Solution and Refinement. Single-wavelength native data were collected at the Advanced Light Source in Lawrence Berkeley National Lab, Beamline 8.3.1.[31] The crystals diffracted to 3.0 Å in the space group C2 (Data Table S1) with one full ring of 17 monomers per asymmetric unit. The diffraction data were indexed and scaled in HKL2000.[32] Molecular replacement was performed with Phenix.automr,[33] using the full 'A' ring of pdTMVP generated using the pdb 1EI7 as a search model. The positions of individual chains were rotationally and translationally refined with rigid-body restraints, using Phenix.refine.[34] The final model was produced with iterative rounds of model building in COOT[35] and refinement (individual xyz, grouped ADPs and with seventeen fold noncrystallographic symmetry) using Phenix.refine. MOLPROBITY was used to validate the structure during model building.[36] Molecular graphics renderings were computed using CHIMERA[37] except for Figure 3e, which was produced using COOT.

Table S1. Crystallography Statistics

Data Collection	Native
Resolution	50-3.00
Wavelength, Å	1.1159
Space group	C2
Unit cell dimensions, Å	220.8, 177.2, 102.8
I/σ^a	7.5 (1.6)
% $R_{sym}{}^a$	11.6 (44.7)
% Completeness[a]	94.8 (81.8)
Redundancy[a]	3.0 (2.3)
No. of reflections (unique)	221932 (73255)

Refinement statistics (49-3.01 A)	
No. of reflections (working/test set)	58611 / 1629
% R_{work} / % R_{free}	23.91 / 25.30
Number of atoms	
Protein	18,700
rms deviation	
Bond lengths, Å	0.004
Bond angles, °	0.935
Ramachandran (molprobity) favored / allowed %	97.84 / 100

[a] Numbers in parentheses refer to highest resolution shell.

Protein Data Bank Accession Code. Coordinates and structure factors for cpTMVP were deposited in the Protein Data Bank (PDB) under access code 3KML.

N-terminal Transamination of cpTMVP with PLP. A 2 mg/mL solution of cpTMVP in buffer A was combined with an equal volume of 20 or 60 mM PLP in the same buffer. The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 30 °C for 24 h. The PLP was removed from the reaction mixture via spin concentration before the protein was quantified, aliquoted, and snap frozen in dry ice/ethanol for storage in buffer A at -20 °C.

General Procedure for $NaIO_4$ Modification of TMVP. A 1.5 mL Eppendorf tube was charged with a solution of TMV (delivered as 100 µL of a 10 mg ml^{-1} solution in 25 mM potassium phosphate, pH 6.5; 1 equiv) and a solution of $NaIO_4$ (delivered as 3 µL of a 200 mM solution in water; 10.5 equiv.) The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at rt for 5 min. The oxidation was quenched with a solution of Na_2SO_3 (delivered as 3 µl of a 1 M solution in water; 52 equiv.). The lack of small molecule aldehydes permitted addition of alkoxyamine reagents with no further purification.

Measuring Extent of N-terminal Transamination. Two thawed aliquots of cpTMVP that had been reacted with either 10 or 30 mM PLP were each diluted to 100 µL (0.57 mM or 10 mg/mL) with buffer A. To half of each solution was added 46 eq of aqueous (aminooxy)acetic acid (TCI, Portland, OR) from a 131 mM solution in water, pH adjusted to 6.5 with NaOH. The oxime formation was catalyzed by the addition of 35 eq of *p*-anisidine, from a 1 M solution in DMSO, pH adjusted to 6.5 with acetic acid. The reaction mixure was vortexed briefly and was left at rt for 24 h. The small molecules were removed by spin concentration. The extent of protein modification was monitored by LC/ESI-MS (Orbitrap) (Figure S4).

Chromophore Attachment to S123C. A thawed aliquot of cpTMVP or AG-TMVP in buffer A was diluted to 100 µl (0.57 mM or 10 mg/ml) with the same buffer. To this solution was added 2.5-3 eq of maleimide functionalized chromophore as a DMSO solution (up to 3% v/v). The reaction mixure was vortexed briefly and was left at rt for 90 min. The bulk of the unreacted chromophore

was removed by Nap-5 gel filtration, with elution into 400 μL of the same buffer. When necessary, further purification and/or concentration was accomplished via spin concentration. The extent of protein modification was monitored by LC/ESI-MS (Orbitrap or 150EX) and UV-vis spectroscopy (Main text, Figure 5 and Figure S2).

Chromophore Attachment to N-terminus of cpTMVP. A 40 μL solution of cpTMVP (previously labeled at C123 with a donor maleimide) in buffer A (0.27 mM or 4.7 mg/mL), was reacted with 3.8 eq of Alexa Fluor 647 C5-aminooxyacetamide (Invitrogen, Carlsbad, CA) from a 25 mM solution in DMSO. The oxime formation was catalyzed by the addition of 47 eq of *p*-anisidine, from a 1 M solution in DMSO, pH adjusted to 6.5 with acetic acid. The reaction mixure was vortexed briefly and was left at rt for 24 h. The unreacted chromophore was removed by spin concentration. Extent of protein modification was monitored by LC/ESI-MS (Orbitrap) and UV-vis spectroscopy (Figure 5).

Assembly of cpTMVP Disks and Rods. To form rods, 2 mg/mL (114 μM) solutions of cpTMVP modified with a chromophore at C123 in 100 mM sodium borate, pH 8.0 were dialyzed against 50 mM sodium acetate buffer, pH 5.5 at rt, transferred to 1.5 mL Eppendorf tubes and stored at rt in the dark for 1-3 weeks. Conversion to larger structures was monitored by size-exclusion chromatography HPLC. Rods eluted between 5 and 8 minutes, depending on length. Assemblies were characterized visually by TEM.

Assembly of Hybrid Disks and Rods. AG-TMVP was modified with Oregon Green 488 maleimide (Invitrogen) as described above, then separated from free dye and simultaneously exchanged into 20 mM TEA, pH 8.0 using a Nap-5 column. The same procedure was used to modify cpTMVP with Alexa Fluor 594 (Invitrogen), though exhaustive spin concentration was required to remove the free dye. After overnight equilibration, dissassemly was confirmed by SEC prior to mixing the proteins in a 16:1 or 1:1 AG-TMVP:cpTMVP ratio. A fraction of each mixture was diluted to 0.45-0.5 OD of the chromophore for measurement of the absorbance and fluorescence spectra as described below. The concentrated samples were dialyzed against 400 mM potassium phosphate, pH 6.5 overnight to assemble into disks and rods. Large protein aggregates were removed by centrifugation for 5 min at 18,000 RCF before assembly was confirmed by SEC. Samples were dilluted to 0.45-0.5 OD of the chromophore and absorbance and fluorescence spectra recorded as below.

Size Exclusion Chromatography (SEC). Analytical size exclusion was performed on an Agilent 1100 series HPLC equiped with a PolySep-GFC-P 5000 column (Phenomenex, Torrance, CA). At a flow rate of 1 mL/min, rods eluted between 5-8 minutes, depending on length. Though retention times varied with the buffers used, disks typically eluted from 8-9 minutes, while monomers eluted from 9-10 minutes.

Mass Spectrometry. Prior to analysis, biological samples were desalted and separated from small molecule contaminants using spin concentrators or NAP-5 gel filtration columns (GE Healthcare, Chalfont, UK). Control experiments have indicated that assembled TMVP aggregates eluted in the void volume of these columns, while the bulk of small molecules were retained. Electrospray LC/MS analysis was performed using an LTQ Orbitrap XL hybrid mass spectrometer with an

Ion Max electrospray ionization source (Thermo Fisher Scientific, Waltham, MA) connected to a Agilent 1200 series liquid chromatograph (Santa Clara, CA) or an API 150EX system (Applied Biosystems) equipped with a Turbospray source and an Agilent 1100 series LC pump.

Transmission Electron Microscopy (TEM). Samples were prepared for TEM analysis by applying an analyte solution (approximately 0.2 mg/mL in TMVP or cpTMVP) to carbon-coated copper grids for 3 min, followed by rinsing with dd-H_2O. The grids were then exposed to a 1.5% aqueous solution of uranyl acetate for 90 s as a negative stain. Images were obtained at the Berkeley Electron Microscope Lab using a FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage. Images were recorded on a Ultra Scan 1000 from Getan (Pleasanton, CA).

Spectroscopic Measurements. The spectroscopic properties of all monomer, disk, and rod solutions were investigated by collecting steady state absorbance and fluorescence measurements. UV-vis spectroscopic measurements were conducted on a Cary 50 UV-Vis spectrophotometer (Varian, Palo Alto, CA). Samples for fluorescence were diluted to 0.45-0.5 OD and spectra obtained on a Fluoromax-2 spectrofluorometer (SA Instruments, Edison, NJ). For experiments measuring energy transfer to the pore, excitation measurements were collected at 680 nm rather than the emission maximum of the acceptor Alexa Fluor 647 (665 nm) to minimize the direct contributions from donor emission. For experiments measuring energy transfer between co-assembled AG-TMVP and cpTMVP, excitation measurements of were collected at 650 nm rather than the emission maxiumum of the acceptor Alexa Fluor 594 (612 nm) to minimize the direct contribution from donor emission.

AFM. Disks of cpTMVP and AG-TMVP were formed by exchange into 400 mM potassium phosphate, pH 6.5, and assembly confirmed by SEC. Solutions of 2 mg/mL protein were spotted on freshly cleaved, poly-lysine coated mica and rinsed with ddH_2O before drying under a stream of N_2. Measurements were performed on a Molecular Imaging PicoSPM II with a PicoScan 2500 controller (RHK, Troy, MI) in tapping mode. Cantilevers (TESP, from Veeco, Camarillo, CA) were used with resonant frequencies between 305-341 kHz. Images were taken with 1μm size, 1024 pixels/line, 1024 lines. The scanning speed was 1 line/s.

Polyacrylamide Gel Electrophoresis (PAGE). For protein analysis, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)[38] was accomplished on a Mini-Protean apparatus (Bio-Rad, USA). Commercially available markers (Bio-Rad, USA) were applied to one lane of each gel for calculation of apparent molecular weights. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Fisher, USA). Gel imaging was performed on a CanoScan LiDe 30 scanner (Canon, Japan).

3.9 Literature Cited

1. Shenton, W.; Douglas, T.; Young, M.; Stubbs, G.; Mann, S. *Adv. Mater.* **1999**, *11*, 253-256.
2. Dujardin, E.; Peet, C.; Stubbs, G.; Culver, J. N.; Mann, S. *Nano Lett.* **2003**, *3*, 413-417.
3. Knez, M.; Bittner, A. M.; Boes, F.; Wege, C.; Jeske, H.; Mai, E; Kern, K. *Nano Lett.* **2003**, *3*, 1079-1082.
4. Schlick, T. L.; Ding, Z.; Francis, M. B. *J. Am. Chem. Soc.* **2005**, *127*, 3718-3723.
5. Niu, Z.; Bruckman, M. A.; Li, S.; Lee, L. A.; Lee, B.; Pingali, S. V.; Thiyagarajan, P.; Wang, Q. *Langmuir* **2007**, *23*, 6719-6724.
6. Holder, P. G.; Francis, M. B. *Angew. Chem. Int. Ed.* **2007**, *46*, 4370-4373.
7. Vega, R. A.; Maspoch, D.; Salaita, K.; Mirkin, C. A. *Angew. Chem. Int. Ed.* **2005**, *44*, 6013.
8. Yi, H.; Nisar, S.; Lee, S.; Powers, M. A.; Bentley, W. E.; Payne, G. F.; Ghodssi, R.; Rubloff, G. W.; Harris, M. T.; Culver, J. N. *Nano Lett.* **2005**, *5*, 1931-1936.
9. Miller, R. A.; Presley, A. D.; Francis, M. B. *J. Am. Chem. Soc.* **2007**, *129*, 3104-3109.
10. Ma, Y. Z.; Miller, R. A.; Fleming, G. R.; Francis, M. B. *J. Phys. Chem. B* **2008**, *112*, 6887-6892.
11. Endo, M.; Fujitsuka, M.; Majima, T. *Chem. Eur. J.* **2007**, *13*, 8660.
12. Stephanopoulos, N.; Carrico, Z. M.; Francis, M. B. *Angew. Chem. Int. Ed.* **2009***, 48*, *in press*.
13. Bopp, M. A.; Jia, Y.; Li, L.; Cogdell, R. J.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10630.
14. Niyogi, K. *Curr. Opin. Plant Biol.* **2000**, *3*, 455.
15. Scheck, R. A.; Dedeo, M. T.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2008**, *130*, 11762-11770.
16. Goldenberg, D. P.; Creighton, T. E. *J. Mol. Biol.* **1983**, *165*, 407-413.
17. Paavola, C. D.; Chan, S. L.; Li, Y.; Mazzarella, K. M.; McMillan, R. A.; Trent, J. D. *Nanotechology* **2006**, *17*, 1171–1176.
18. Regan L. *Curr. Opin. Struct. Biol.* **1999**, *9*, 494-499.
19. Gouy, M.; Gautier, C. *Nucl. Acids Res.* **1982**, *10*, 7055-7074.
20. Klug, A. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1999**, *354*, 531-535.
21. Butler, P. J. G. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **1999**, *354*, 537-550.
22. Leberman, R.; Finch, J. T.; Gilbert, P. F. C.; Witz, J.; Klug, A. *J. Mol. Biol.* **1974**, *86*, 179-180.
23. Bhyravbhatla, B.; Watowich, S. J.; Caspar, D. L. *Biophys. J.* **1998**, *74*, 604-615.
24. Bloomer, A.; Champness, J.; Bricogne, R.; Klug, A. *Nature* **1978**, *276*, 362-368.
25. Zheng, J.; Birktoft, J. J.; Chen, Y.; Wang, T.; Sha, T.; Constantinou, P. E.; Ginell, S. L.; Mao, C.; Seeman, N. C. *Nature* **2009**, *461*, 74-77.

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

27. Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem. Int. Ed.* **2006**, *45*, 7581-7584.

28. Bradford, M. *Anal. Biochem.* **1976**, *72*, 248.

29. Ausubel, F. M., Brent, R., Kingsont, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K., Eds. Short Protocols in Molecular Biology; Fourth Edition; Wiley, NY, 1999; pp 8-20.

30. Sambrook, J., Russel, D. W., Eds. *Molecular Cloning*; Cold Spring Harbor Laboratory Press, NY, 2001; Section 13.37.

31. MacDowell, A. A.; Celestre, R. S.; Howells, M.; McKinney, W.; Krupnick, J.; Cambie, D.; Domning, E. E.; Duarte, R. M.; Kelez, N.; Plate, D. W.; Cork, C. W.; Earnest, T. N.; Dickert, J.; Meigs, G.; Ralston, C.; Holton, J. M.; Alber, T.; Berger, J. M.; Agard, D. A.; Padmore, H. A. *Synchrotron Rad.* **2004**, *11*, 447-455.

32. Otwinowski, Z.; Minor, W. In *Macromolecular Crystallography Part A*; Academic Press, 1997; Vol. Volume 276, pp. 307-326.

33. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *Acta Cryst.* **2002,** *D58*, 1948-1954.

34. Adams, P.; Gopal, K.; Grosse-Kunstleve, R.; Hung, L.; Ioerger, T.; McCoy, A.; Moriarty, N.; Pai, R.; Read, R.; Romo, T.; Sacchettini, J.; Sauter, N.; Storoni, L.; Terwilliger, T. *Journal of Synchrotron Radiation* **2004**, *11*, 53-55.

35. Emsley, P.; Cowtan, K. *Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.

36. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucl. Acids Res.* **2007**, *35*, W375-383.

37. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem*. **2004**, *25*, 1605–1612.

38. Laemmli, U. K. *Nature* **1970**, *227*, 680–685.

39. Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R., Klug, A. *Nature* **1978**, *267*, 362-368.

Chapter 4: A TMV Scaffold for Pore, Exterior, and Channel Modification

4.1 Abstract

A mutant of TMV was generated with no lysines or other NHS-ester reactive groups, which allowed us to reintroduce a new lysine residue for site-specific modification. Seven different sites were tested, the most successful being T104K in the pore of disks and rods. This reintroduced lysine was targeted simultaneously with the N-terminus on the exterior and the cysteine on the top face, providing a third reactive handle. This approach was used to create closely packed chromophore structures in the pore of TMV reminiscent of bacterial chlorosomes, and to bind gold nanoparticles to the pore of the assembled disks.

4.2 Background

4.2.1 Lysine Conjugation

Conjugation to the amine sidechain of lysine is a common strategy to functionalize proteins. A variety of electrophiles are available that react to form covalent bonds with varying specificity for lysine over other nucleophilic groups present (Figure 1-6).[1] Multiple lysines are found on the surface of most proteins, which provides the opportunity for high loading of one or multiple synthetic groups, at the expense of site-specificity. Another consideration when using modifying lysines is that many of the reaction of the amine to produce an amide bond removes a positive charge, affecting the isoelectric point and possibly the solubility of the protein.

4.2.1 Metal Nanoparticles in Nanotechnology and Light Harvesting

Gold and silver nanoparticles have been incorporated into nanomaterials for a multitude of purposes. Commercially available gold particles range from 1.8 to at least 100 nm and silver particles range from 10 to at least 150 nm. Their density makes them excellent markers for transmission electron microscopy (TEM) and light microscopy samples, and when conjugated to antibodies, they can be used to label samples with high resolution.[2,3] Many applications of gold and silver nanoparticles take advantage of their surface plasmon resonance, which can lead to a local enhancement in the electromagnetic field. This property makes the absorption spectrum of the particles sensitive to aggregation. These properties have been exploited for numerous highly sensitive sensor applications.[4,5] Our interest lies in the influence of the nanoparticles on the absorption, transmission, and emission properties of nearby chromophores. We hope that by gaining a better understanding of the fundamental chromophore-gold interactions we can develop hybrid light harvesting structures with greater absorption and more efficient energy transfer. We are encouraged by reports of metal-nanoparticle-enhanced light harvesting and fluorescence of LH1,[6] LH2,[7] and photosystem I complexes.[8] These properties and others have motivated the incorporation of metal nanoparticles into both silicon-based[9] and organic[10] photovoltaic devices.

The effect of a nanoparticle on the local electromagnetic field is strongly dependent on the distance from the metal surface as well as the presence of any neighboring nanoparticles.[11] The exact relationship between the metal-chromophore distance and the effect on fluorescence is not clear, and different groups has reported conflicting results (Figure 4-1). One reason for this is likely due to the variety of systems that have been developed to hold nanoparticles and fluorophores at specific distances, including single[12] and double-stranded DNA,[13] DNA origami,[14] and scanning probe microscopy.[15] The size of the particles and identities of the dyes also confound the ability to compare results directly. The rigidity of a protein scaffold such as TMV may provide more definitive data on the distance-dependence of fluorescence quenching and enhancement.

4.3 Introduction of a Third Reactive Site to TMV

Previous work demonstrated that it was possible to modify TMV simultaneously at a cysteine in the RNA channel (Chapter 1) and at the N-terminus on the exterior (Chapter 2). The next goal was to create a TMV scaffold with these two reactive sites and a third uniquely reactive site in the pore. In addition to the pore modifications detailed in Chapter 3, the Majima group provided a precedent for modification of the pore by introducing a cysteine into positions 99 and

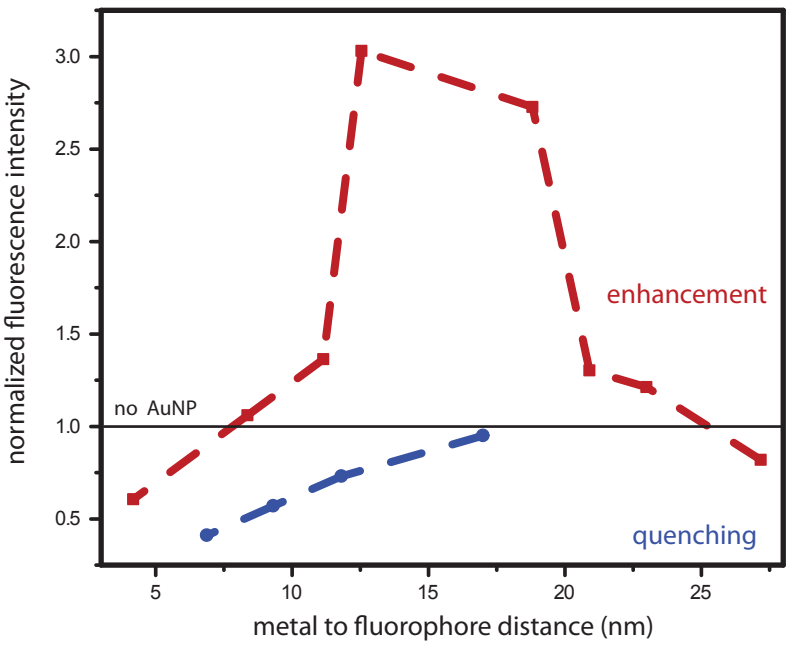

Figure 4-1. Summary of literature reports on the proximity effect of gold nanoparticles on fluorescence of organic chromophores. Reports of fluorescence quenching and enhancement by gold nanoparticles vary widely. References 16-26 report fluorescence enhancement, while 27-32 report quenching.

100 and alkylating with a maleimide-bearing pyrene.[33] We instead sought to create a uniquely reactive lysine in this region. To begin, we eliminated the solvent exposed nucleophiles on the surface of TMV that could also react with NHS esters.

After reaction with NHS esters the LCMS spectra of TMV showed up to 5 additions, but it was not obvious which residues were involved in these reactions. The two lysines and the N-terminus should form stable conjugates, while cysteine 123 and the numerous serine and threonine residues could form transient (thio)esters. Consistent with this prediction, I observed the loss of some adducts after overnight incubation with 50 mM 2-mercaptoethanesulfonate (MESNa). In an attempt to elucidate the residues involved, I mutated the lysines at positions 53 and 68 to arginines. The positively charged lysines in wild-type TMV form salt bridges that mediate the assembly of TMV into disks and rods, and arginine was thus chosen to maintain this positive charge.

Surprisingly, these K53R K68R (RR) mutants reacted with only one NHS-ester, indicating that most of the remaining residues were no longer reactive. Upon further examination, we observed that these RR TMV mutants formed stable disk structures in buffers ranging from pH 6-9, where normal TMV would have dissociated into monomers and small aggregates. We attributed the limited reactivity to this stable assembly, as two of the three remaining reactive residues were now likely buried in interfaces. Subsequent experiments indicated that the K53R mutation was sufficient to stabilize the disk assembly. Though cross reactivity between lysines and maleimides is more commonly observed than between cysteines and NHS-esters,[1] cysteine 123 was the next nucleophile tested for reactivity. Neither mutants with the original serine 123 nor cysteine 123 alkylated with a maleimide showed reactivity with NHS esters, identifying this cysteine as the last reactive residue. With a protein unreactive towards NHS esters, we proceeded to screen sites for the introduction of a uniquely reactive lysine.

The following residues of RR TMV were mutated to lysine: T5, T6, N101, T103, T104,

S154, and T158. The N-terminus of these new mutants was extended with AlaGly. The protein used for preliminary reactivity studies discussed above had the wild-type N-terminus. Max Fishman, an undergraduate under my supervision, performed the bulk of the characterization of these mutants.

4.3.1 T158K

This outside lysine mutant purified very well. There were little to no contaminants in the purified protein according to gel electrophoresis. However, the protein readily formed monomers in solutions above pH 7.0. The disks that it formed at pH 7.0 and below had retention times on the HPLC of around 8.5-8.85 minutes. TEM pictures showed that T158K made disks, partial disks, and a few stacked disks in solution. The protein precipitated at high pH.

There were even more problems with this protein when we tried to modify the protein with maleimides and NHS-esters. The protein would precipitate when a high concentration of maleimide (above 2.5 mM) was added to the solution.

4.3.2 T6K

This outside lysine mutant proved to be easier to work with than T158K. It would not precipitate as readily when maleimide or NHS-esters were added and it remained soluble at high pH, though it tended to be a mix of monomers and disks. The protein would modify to completion with maleimide while at pH 6.5 and would doubly modify with NHS-esters.

4.3.3 T5K

This outside lysine mutant was very similar to T6K; however it had a better yield and the protein was highly pure after putting it on the DEAE twice. This protein also tended to be a mixture of disks and monomers or partial disks and monomers above pH 6.5.

This protein would react to completion with a maleimide. However, reacting with NHS esters produced a heterogeneous mixture of unmodified, +1, and +2 additions. It was discovered that if we reacted the protein with PLP first to transaminate the N-terminus, the amount of double modification by NHS-esters decreased dramatically. This suggested that the N-terminus was modified with NHS-esters.

4.3.4 S154K

This outside lysine mutant reacted well with maleimides and it also reacted with NHS-esters. However it did show double modification upon NHS-ester addition like the other proteins. This outside lysine mutant seemed to form a homogenous mixture of disks according to SEC and showed signs of forming stacked disks. This is a unique characteristic among the outside lysine mutants and makes it worth investigating in future studies.

4.3.5 T104K

Based on the highest resolution crystal structure of the disk (PDB 1ei7),[34] we hypothesized that the pore region spanned residues 98 to 104. It should be noted that this region is largely disordered in the crystal and should be taken as an approximate guide. The tight packing of the adjacent loops (alpha carbons as close as 8 Å) led to some concern that the conjugation of bulky groups might be self-limiting due to steric crowding. To maximize solvent accessibility and minimize potential crowding, residues 101, 103, and 104 were chosen on the edge of the pore, closest to the introduced cysteine 123 (Figure 4-2). These RR TMV lysine mutants were largely

indistinguishable in expression, assembly, and modification, though T104K appeared slightly more reactive with NHS esters. This mutant was chosen for further study.

This mutant seems to remain as disks or as a mixture of disks and stacked disks at pH values ranging from 6-9. It is much more stable and less prone to precipitation behaved than any of the outside lysine mutants. It can be site specifically modified in three different regions.

T104K can be modified with maleimide small molecules to completion and a large percentage of it can be modified by malimide dyes. It was doubly modified by NHS-esters, but to a lower degree than the outside lysine mutants. No FRET experiments were performed on this protein, but it was modified with OG-maleimide and TAMRA succinimidyl ester or OG-maleimide and AF594 succinimidyl ester. These experiments showed no double modification by T104K with the succinimidyl ester dyes.

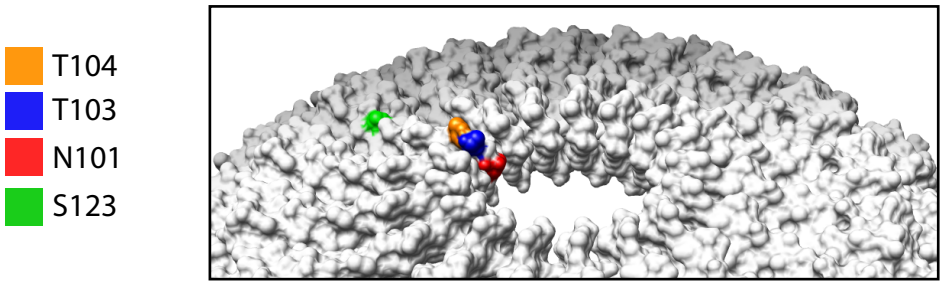

Figure 4-2. Positions screened in pore of TMV for introduction of reactive lysine. The crystal structure of the top ring of the double disk (1ei7) shows the close packing of monomers to form a central pore. The residues T104, T103, and N101 selected for mutation to lysine are highlighted. The position of S123 is shown for reference.

4.4 Triple Modification of TMV

In the process of trying to modify the RR TMV protein with alkoxyamines at the transaminated N-terminus, maleimides at S123C, and NHS esters at T104K, it became apparent that incomplete PLP-mediated transamination left N-terminal amines capable of reacting with NHS esters. This reactivity was unobserved until this point because the native N-terminus used in earlier experiments was not solvent accessible to NHS esters. In an attempt to transaminate the N-termini completely, reactions were screened with up to100 mM PLP for up to 48 hours. However, this approach was abandoned because the products often had more than one PLP adduct and the subsequent oxime formation was incomplete. Instead, a serine was appended to the N-terminus, with and without a glycine spacer, for periodate oxidation to a glyoxylyl group.[1] Both mutants appeared to oxidize completely within five minutes using 5 mM freshly prepared $NaIO_4$. Encouragingly, the subsequent oxime formation also appeared to proceed to completion (Figure 2-9d).

Using this new RR TMV mutant, we developed a strategy to modify each monomer on the exterior, RNA channel, and pore surfaces. Because periodate can oxidize cysteines to unreactive sulfonic acids, S123C was first alkylated with maleimide. Sodium periodate was added to the unpurified mixture, quenched with sodium sulfite, and the subjected to the NHS-ester. This mixture was purified on a Nap-5 gel filtration column and reacted with an alkoxyamine. Each reaction appeared to proceed to completion as determined by LCMS (Figure 4-3). The additional +18

and +33 masses could correspond to the oxidation of the newly formed thioethers to sulfoxides or sulfones, respectively. To our knowledge, this is the first demonstration of three site-specific reactions applied to a single protein.

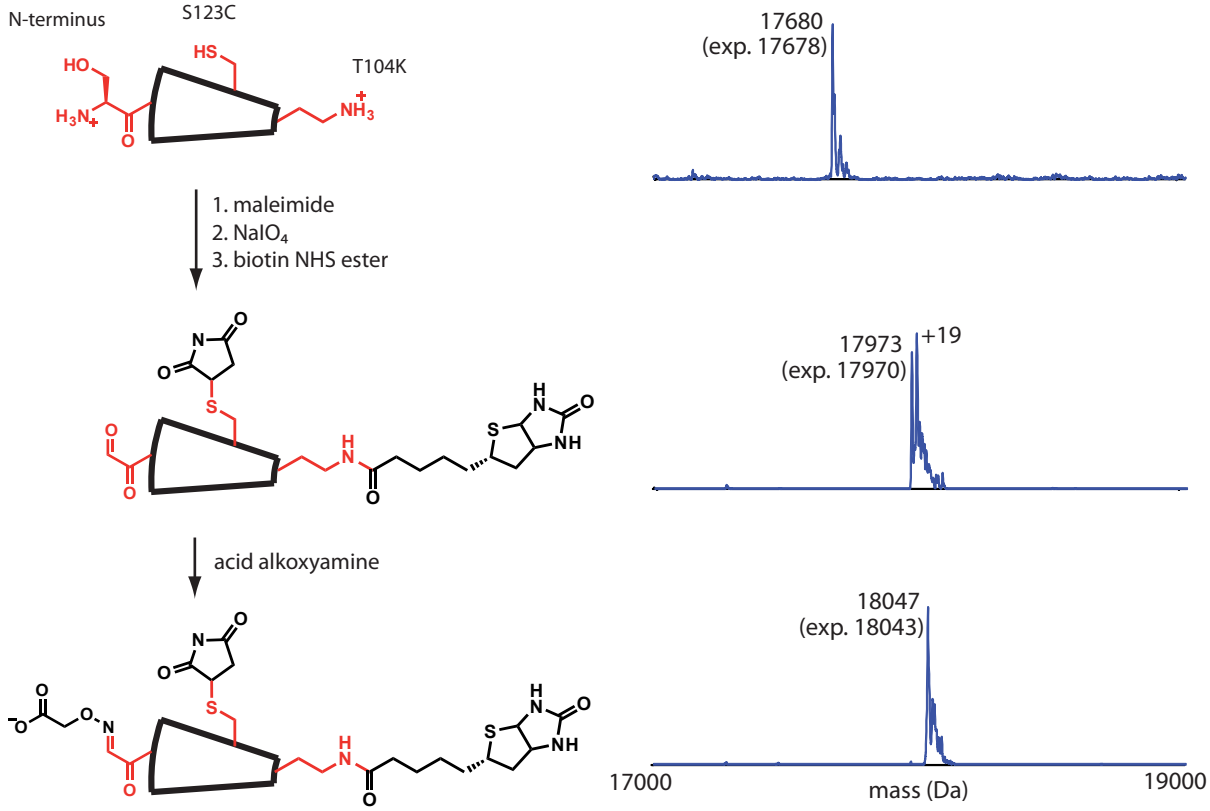

Figure 4-3. Site specific modification of three surfaces of TMV. The high solubility of each reagent permitted quantitative labeling in each case.

4.5 Immobilization of Gold Nanoparticles in the Pore

Gold nanoparticles can be attached to proteins through covalent bonds to the surface of the gold particle,[35] reactive handles on passivating ligands,[36] or through non-specific electrostatic interactions.[37] We sought to attach TMV directly to the surface of the gold with a thiol to avoid non-specific electrostatic interactions, maintain control over the location of conjugation, and minimize the heterogeneity of protein-nanoparticle samples. Gold nanoparticles were prepared for attachment by exchanging their citrate ligands for bis(*p*-sulfonatophenyl)phenylphosphine (BSPP).[38] This ligand has been shown to decrease non-specific interactions with biomolecules and facilitate conjugate analysis by agarose gel electrophoresis relative to citrate. We observed no gold binding to TMV containing a solvent exposed cysteine 123 in an alpha helix, which we attributed to the inability of the short side chain to penetrate the ligand shell. Instead, we sought to modify TMV to display thiols with a longer alkyl spacer.

Three crosslinkers were reacted with T104K and screened for their ability to bind gold to the pore of RR TMV. 2-Iminothiolane (Traut's reagent) is a commercially available crosslinker

that reacts with primary amines, cleaving a cyclic thioether to form a thiol.[1] Upon incubation with T104K RR TMV, however, a mass spectrum showed only two peaks that were 15 and 23 amu less than the expected +102 amu. Furthermore, this conjugate showed no reactivity towards maleimides. This result was surprising but not unprecedented, as Mokotoff *et al.* reported the re-formation of a cyclic thioether above neutral pH in which ammonia was lost (Figure 4-4a).[39] Though this thioether is unable to form a covalent bond with gold, the reported product contains an iminium group which could attract the negatively charged ligands surrounding the gold particles through electrostatic interactions.

The other two crosslinkers were chemically similar but varied in their length. *N*-succinimidyl 3-(2-pyridyldithio)-propionate (SPDP) and sulfosuccinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate (Sulfo-LC-SPDP) contained a pyridyl disulfide, which upon reduction or disulfide exchange, expelled the non-reactive pyridine-2-thione (Figure 4-4b,c). The absorbance of this molecule at 343 nm allowed the extent of the disulfide reduction to be monitored by absorbance. While both NHS-esters were capable of reacting fully with the protein, higher concentrations of the longer Sulfo-LC-SPDP were required, presumably due to steric hinderance. After conjugation to T104K and reduction with TCEP, a positive control showed both SPDP variants reacted completely with a bulky maleimide dye (Figure 4-5). The close spacing of dyes conjugated in this manner was of potential interest to studies of natural photosynthetic systems composed of closely spaced chromophores. Further research is being pursued in collaboration with the Whaley group.

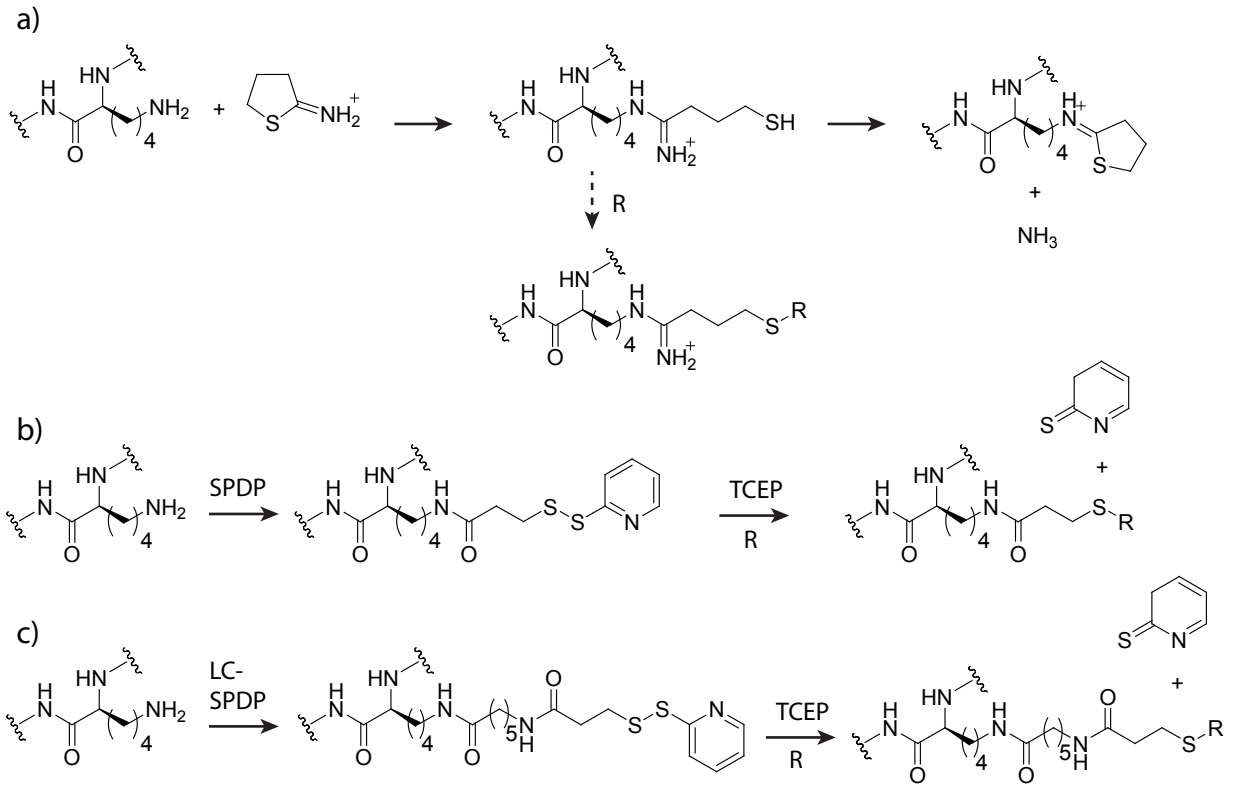

Figure 4-4. Crosslinkers for thioloating T104K. (a) Traut's reagent rapidly rearranged to form an unreactive thioether before the thiol could be alkylated. (b) and (c) SPDP and LC-SPDP reacted completely with T104K and after reduction. The resulting thiol was reactive with both maleimides and gold.

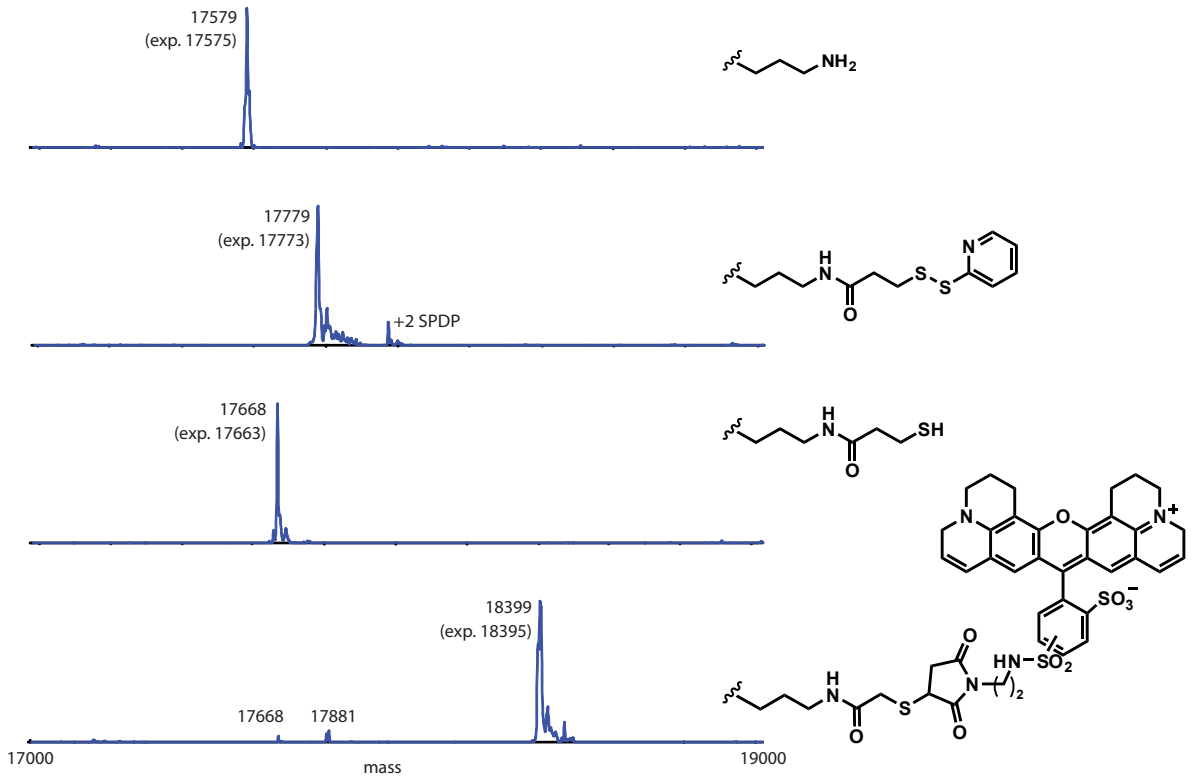

Figure 4-5. LCMS spectra show the steps involved in thiolating T104K of TMV. Despite its size, the Texas Red maleimide chromophore modified almost all of the TMV monomers.

Gold nanoparticles were attached to the pore of RR TMV by incubating the thiolated protein with BSPP-stabilized gold overnight in the presence of the reducing agent *tris*(2-carboxyethyl) phosphine (TCEP). Both TCEP and BSPP contain phosphines capable of binding gold, but neither can bind as strongly as a thiol.[40] It was uncertain how the competition of TCEP with BSPP affected the ligand shell of the nanoparticles but there was precedent for use of up to 10 mM TCEP to reduce disulfides when attaching thiolated biomolecules to gold particles.[41]

All three lysine conjugates were observed to bind gold particles by TEM and native polyacrylamide gel analysis (Figure 4-6). Curiously, the Traut's modified protein appeared more often in chains of alternating disk(-gold-disk)$_n$ (Figure 4-6d). We first tried the unmodified T104K mutant as a negative control, but found that it also bound gold to the pore, presumably through electrostatic interactions. Consistent with this hypothesis, RR TMV without a lysine in the pore showed no affinity for gold. Nanoparticles with diameters of 5, 10, and 15 nm were all successfully attached to the protein in this manner.

4.6 Purification of Gold-Protein Conjugates

The overnight incubation of gold and thiolated RR TMV disks produced a heterogeneous mixture of free gold and disks with 0-2 gold particles attached to one face. Because our first goal was to study a single disk bound to a single gold particle, we attempted to remove the other species. The first approach was to take advantage of the fact that high concentrations of salt could cause gold particles to precipitate by shielding the negative charges that held them apart.

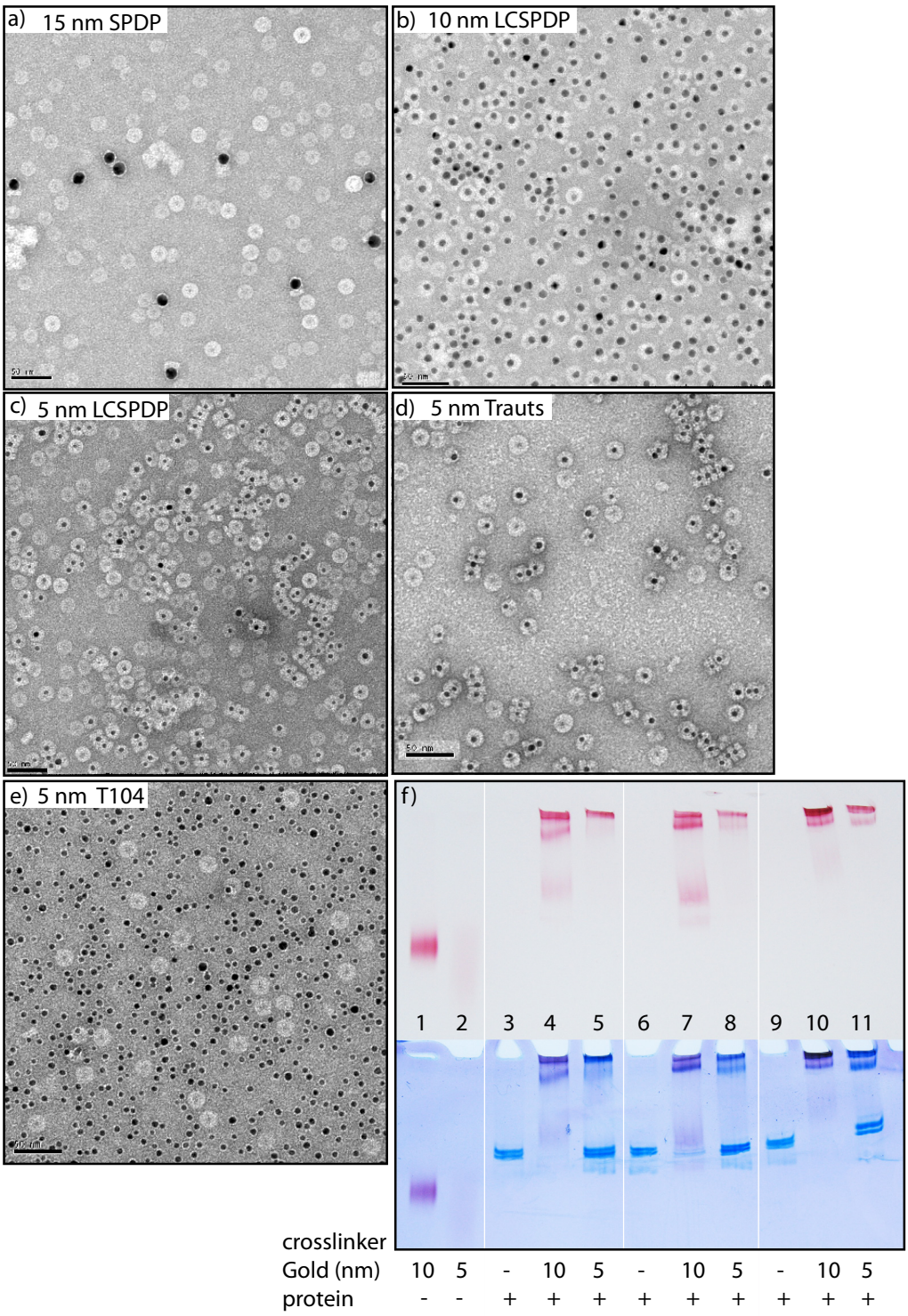

Figure 4-6. TEM and SDS-PAGE of TMV-gold conjugates. (a) 15 nm gold bound to TMV with SPDP. (b) 10 nm gold bound to TMV with LC-SPDP. (c) 5nm gold bound to TMV with LC-SPDP. (d) 5 nm gold bound to TMV with Traut's reagent. (e) 5nm gold with T104 TMV. (f) The native PAGE gel was imaged before and after staining with Coomasie. It showed migration of gold (red) in absence of protein (1,2). For each crosslinker, the pattern was the same: in the absence of gold (3,6,9) the TMV disks migrated as two bands. With 10 nm gold (lanes 4,7,10) almost no protein but some gold migrates into the gel. With 5 nm gold (lanes 5,8,11) no gold, but some protein migrates into the gel.

Although the precipitation worked quite well in the absence of protein, the presence of conjugated or unconjugated protein prevented it. The second technique was to centrifuge the gold, with or without a sucrose pad or gradient, relying on the drastically different densities of gold, TMV, and gold-TMV conjugates. All of these processes were greatly facilitated by the high extinction coefficients of the gold particles, which rendered them visible without staining.

When gold was mixed with an excess of protein, the major products were bare disks and disks with a single gold particle. These two populations could be visualized by native polyacrylamide gel. In Figure 4-7e, lane 4 shows a mixture of fast-moving free protein and slow moving gold-bound protein. Following two rounds of centrifugation at 21.1 rcf, the pellet containing the gold-protein species (lane 5) was separated from the supernatant containing the unconjugated protein (lane 6). In the absence of a thiol on the disks, no slow-moving bands were observed (lanes 1-3), and the gold migrated separately from the protein. TEM images corroborated these observations.

It was more challenging to separate mixtures of free gold, protein-gold conjugates, and free protein. Following ultracentrifugation through a 40% sucrose solution above a 70% sucrose pad, some protein-dye conjugates could be visualized by fluorescence at the top of the tube, some red color (presumably gold-protein conjugate) appeared to migrate to the interface between different sucrose layers, and a small amount of gold (presumably unconjugated) formed a pellet. This appeared very promising and the interemediate band was removed with a pipette and dialyzed to remove the sucrose. TEM images showed gold attached to small, irregular white blotches, suggesting that the disks did not survive the centrifugation and that the conjugated protein was probably denatured. This approach was not pursued further.

The third technique used to separate gold, protein-gold conjugates, and free protein was native agarose gel electrophoresis. Although this method lacks the resolution of polyacrylamide gel electrophoresis, it offers the advantage of recovering the separated species. Bare protein would run slowly, bare gold would run quickly, and mixtures of the two would run at intermediate rates, regardless of whether they were covalently joined. This co-migration prevented the separation of free gold from the conjugates. Furthermore, little to no separation was observed between a mixture of gold-conjugated disks and free disks. Because of these challenges, purification and further optimization were abandoned.

4.7 Conclusion

Through site-directed mutagenesis, we developed a promising new protein scaffold based on the Tobacco Mosaic Virus. The K53R mutation produced a dramatic change in assembly properties and provided disks stable over a wide pH range, making these disks more useful as material building blocks. The introduction of a uniquely reactive lysine into the pore further expands the protein's utility by providing a rigid scaffold for the study of tightly packed chromophores. The thiolation of the pore lysines, conjugation to gold, and purification through centrifugation allows the formation of homogenous, monodisperse gold-protein conjugates. Future work will make use of the additional reactive handles for incorporating chromophores and assembling higher-order structures.

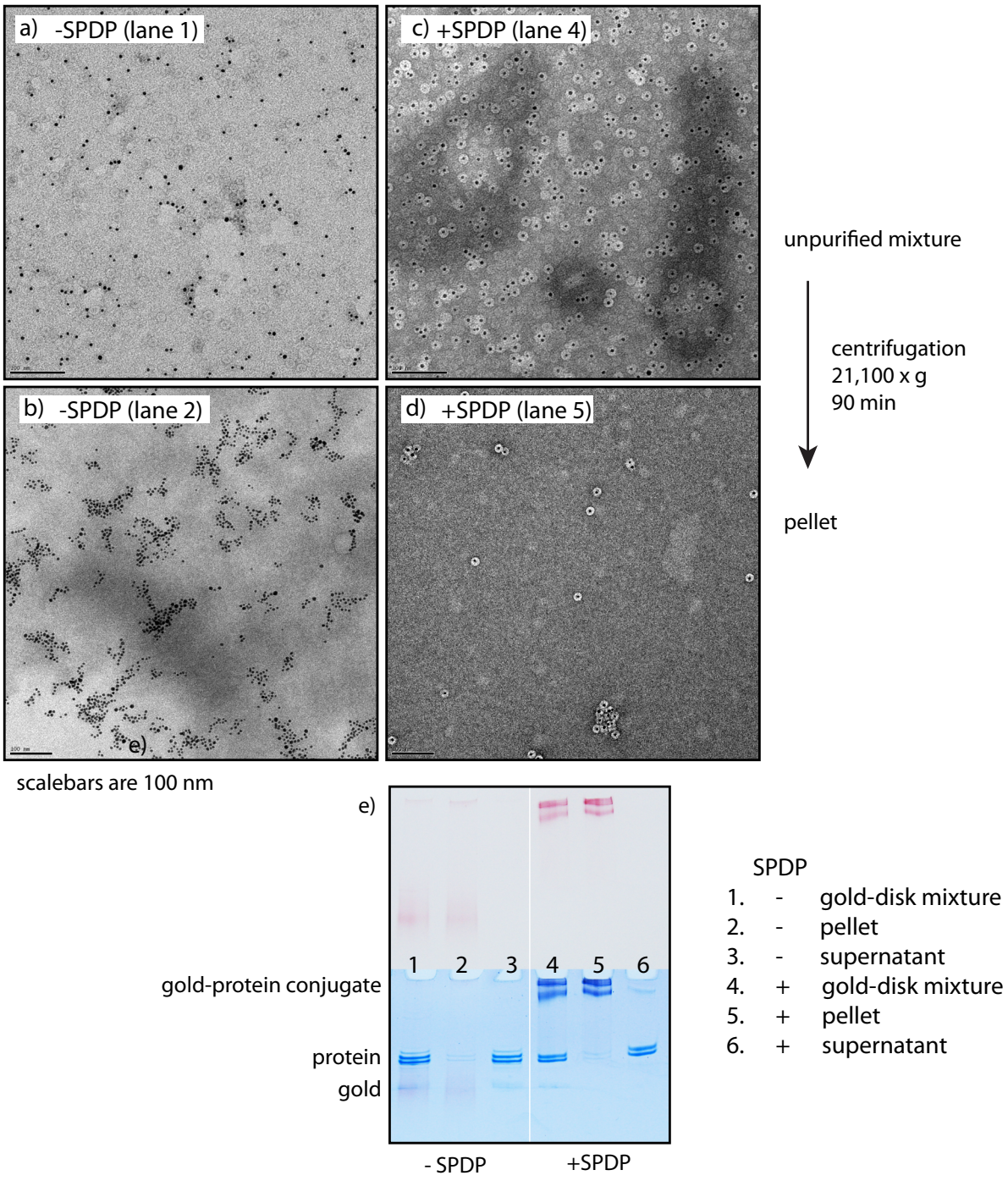

Figure 4-7. Removal of unbound TMV by centrifugation. (a) and (b) are a negative control showing purification depends on covalent attachment. TEM images show 5 nm gold nanoparticles and TMV disks lacking the SPDP crosslinker before (a) and after (b) centrifugation. Without a covalent attachment to the gold, the disks remain in the supernatant. (c) shows disks with SPDP crosslinker bound to gold nanoparticles before centrifugation. The resuspended pellet (d) shows protein disks bound to gold and no free disks. (e) The lack of quickly moving protein in lane 5 confirms that free protein was retained in the supernatant (lane 6).

4.8 Materials and Methods

General Procedures. Unless otherwise noted, all chemicals and solvents were of analytical grade and were used as received from commercial sources. Water (dd-H_20) used in biological procedures or as reaction solvents was deionized using a NANOpure purification system (Barnstead, United States). Spin concentration steps were performed using 100,000 molecular weight cutoff spin concentrators from Millipore (Billerica, MA) in an Allegra 64R Tabletop Centrifuge (Beckman Coulter, Fullerton, CA). Protein samples were quantified by the Bradford assay.

Preparation of the DNA Construct for cpTMVP. The starting point for the RR TMV mutants was a gene for the coat protein of the TMV U1 strain optimized for the codon usage of *E. coli* (Genscript, Piscataway, NJ). This sequence contained the mutation S123C, as well as an Ala Gly addition to the N-terminus to facilitate the PLP mediated transamination reaction. Site-directed mutagenesis was performed using QuikChange mutagenesis (Stratagene, Santa Clara, CA).

Buffer A - 25 mM potassium phosphate, pH 6.5

Protein Expression. Tuner DE3pLysS competent cells (Novagen) were transformed with the vector described above, and cultured in Terriffic Broth with 100 µg/L ampicillin. When cultures reached an optical density of 0.6 to 0.8, IPTG was added to a final concentration of 30 µM. Cultures were grown 24 h at 30 °C, harvested by centrifugation, and stored at -80 °C. Cells (from a 2 L expression batch) were thawed, resuspended in 40 mL of 20 mM TEA pH 7.2, and lysed by sonicating with a 2 s on, 6 s off cycle for a total of 20 minutes using a standard disruptor horn at 90% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared by ultracentrifugation for 30 min at 45,000 rpm using a Beckman 45 Ti rotor in an Optima L-80 XP (Beckman Coulter) or 30 min at 10,000 rpm using a using a SLA-600TC rotor in a Sorvall RC5C Plus centrifuge (Waltham, MA). The clarified lysate was decanted, warmed to rt, and stirred while adding a saturated solution of ammonium sulfate dropwise to a final concentration of 35% (v/v). After 5 min, the white ppt that formed was pelleted by ultracentrifugation (15 min at 45,000 rpm in a Beckman 45 Ti rotor), washed with deionized water, and resuspended in buffer A. The resulting protein solution was next loaded onto a DEAE column and eluted with a 0 – 300 mM NaCl gradient. Purity was confirmed by SDS-PAGE and HPLC. This preparation provided pure RR TMV in yields up to 30 mg/L culture.

Assembly of RR TMVP Disks. Disks were the only species observed between pH 6-9 by size exclusion chromatography (SEC) when working with RR TMV and RR T104K TMV. Curving stacks of disks were obsered by TEM, but may be a artifact of drying since larger structures were rarely observed by SEC.

General Procedure for PLP Modification of TMVP. A 1.5 mL Eppendorf tube was charged with a solution of TMV (delivered as 100 µLof a 2 mg ml^{-1} solution in 25 mM potassium phosphate, pH 6.5; 1 equiv) and a solution of freshly prepared PLP (delivered as 100 µLof a 20 mM solution in 25 mM phosphate buffer, pH adjusted to 6.5 with 1 M NaOH; 175 equiv.) The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 30 °C for 24 h. The PLP was removed from the reaction mixture *via* size exclusion chromatography (Nap-5 desalting

column (GE Healthcare) eluting into 25 mM phosphate buffer, pH 6.5.) A 50 μL aliquot of the resulting purified mixture was treated with 50 μLof 0.1 M 2K-PEG-ONH_2 or 25 μLof 25 mM trisulfonate monoalkoxyamine porphyrin in a 1.5 mL Eppendorf tube. The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 25 °C for 18-20 h. Samples were combined with loading buffer, and analyzed by SDS-PAGE.

General Procedure for $NaIO_4$ Modification of TMVP. A 1.5 mL Eppendorf tube was charged with a solution of TMV (delivered as 100 μLof a 10 mg ml^{-1} solution in 25 mM potassium phosphate, pH 6.5; 1 equiv) and a solution of $NaIO_4$ (delivered as 3 μLof a 200 mM solution in water; 10.5 equiv.) The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 25 °C for 5 min. The oxidation was quenched with a solution of Na_2SO_3 (delivered as 3 μLof a 1 M solution in water; 52 equiv.). The lack of small molecule aldehydes permitted addition of alkoxyamine reagents with no further purification.

General Procedure for the Thiolation of T104K. A 1.5 mL Eppendorf tube was charged with a solution of RR T104K TMV (delivered as 100 μLof a 10 mg ml^{-1} solution in 25 mM potassium phosphate, pH 6.5; 1 equiv). To increase the reactivity of the lysines, the pH was raised by the addition of 50 μL of 100 mM TEA, pH 8. To this solution was addded 4 μLof a 100 mM solution of SPDP in DMSO (7 eq.) The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 25 °C for 2 h. To this solution was added 8 μLof a 200 mM solution of TCEP in water. The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 25 °C for 20 min before reaction with maleimides or gold.

Chromophore Attachment to S123C. A thawed aliquot of RR TMV in buffer A was diluted to 100 μl (0.57 mM or 10 mg/ml) with the same buffer. To this solution was added 2.5-3 eq of maleimide functionalized chromophore as a DMSO solution (up to 3% v/v). The reaction mixure was vortexed briefly and was left at 25 °C for 90 min. The bulk of the unreacted chromophore was removed by Nap-5 gel filtration, with elution into 400 μL of the same buffer. When necessary, further purification and/or concentration was accomplished via spin concentration. The extent of protein modification was monitored by LC/ESI-MS (Orbitrap or 150EX) and UV-vis spectroscopy.

Triple Modification of RR T104K N-SS TMV. A thawed aliquot of RR T104K TMV with a serine appended to the N-terminus (N-SS) in buffer A was diluted to 25 μl (0.57 mM or 10 mg/ml) with the same buffer. To this solution was added 0.5 μL of 500 mM maleimide (17 eq.) The reaction mixure was vortexed briefly and was left at 25 °C for 50 min. To this solution was added 2 μL of 200 mM $NaIO_4$ (14 eq.) The reaction mixure was vortexed briefly and was left at 25 °C for 5 min. before quenching with a 2 μLof a 1 M solution of Na_2SO_3 in water. The mixture was dilluted with an equal volume of 100 mM TEA, pH 8, to increase the pH and the reactivity of the lysines. To this solution was added 4 μL of 100 mM biotin-NHS (28 eq.) in DMF. The reaction mixure was vortexed briefly and was left at 25 °C for 2 h before purification via Nap-5 gel filtration, eluting into 200 μL of buffer A. To 15 μL of this solution was added 5 μL of 100 mM aminooxy acetic acid in water and 1 μL of a 1 M anisidine solution in DMSO, pH adjusted to 6.5 with acetic acid. The reaction mixure was vortexed briefly and was left at 25 °C for 18 h before analysis by LCMS.

Gold Nanoparticle Ligand Exchange for Conjugation to Proteins. A 10 mL aqueous solution of

5, 10, or 15 nm citrate-stabilized Au nanoparticles (Ted Pella, Redding, CA) was stirred with 5 mg of bis(*p*-sulfonatophenyl)phenylphosphine (BSPP) for 18 h at 25 °C. Solid NaCl was added until the color changed from burgundy to blue. The best yields were obtained when the following steps were performed as quickly as possible. The solution was transferred to a 15 mL centrifuge tube and centrifuged at 1000 x g for five minutes to pellet the nanoparticles. The yellow supernatant was removed and the precipitate was dissolved in 200 μL of 0.5 mM BSPP in water and transferred to a 1.5 mL Eppendof tube. To this solution was added 1 mL of MeOH, changing the color to purple. This was centrifuged at 13,000 x g for 3 min to pellet the nanoparticles again. The supernatant was removed and the pellet was resuspended in 100 μLof 0.5 mM BSPP solution. A sample of this solution was dilluted 100-fold for quantification by UV/vis, using extinction coefficients obtained from the manufacturer of $9.7x10^6$ for 5 nm gold, $9.55x10^7$ for 10 nm gold, and $2.4x10^8$ for 15 nm gold.

General Method for the Conjugation of TMV Disks to Gold Nanoparticles and their Purification. Concentrated solutions of BSPP-coated gold nanoparticles and thiolated TMV disks were combined in a 1:1 molar ratio of disk:gold in the presence of 5 mM TCEP to prevent disulfide formation. These were mixed briefly and allowed to couple over 1-2 days. For the samples shown in Figure 4-7, the scale of the reactions was 360 pmol in 116 μLto give [Au] = [disks] = 3.1 μM. The presence of free protein in lane 4 of Figure 4-7e suggested that the disks were actually in excess. The purification of this excess protein was accomplished by two 45 min centrifugations at 21,100 x g, removing the buffer and washing the pellet after each. The pellet was resuspended in buffer containing 0.5 mM BSPP.

Gel Analyses. Denaturing sodium dodecyl sulfate-poly(acrylamide) gel electrophoresis (SDS-PAGE) was accomplished on a Mini-Protean apparatus from Bio-Rad (Hercules, CA) with 15 % gradient polyacrylamide gels (BioRad, CA), following the protocol of Laemmli.[42] All electrophoresis protein samples were mixed with SDS loading buffer in the presence of dithiothreitol (DTT) and heated to 100 °C for 10 min to ensure reduction of disulfide bonds and complete denaturation unless otherwise noted. Commercially available molecular mass markers (Bio-Rad) were applied to at least one lane of each gel for calculation of the apparent molecular masses. Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA).

Native poly(acrylamide) gel electrophoresis was accomplished on a Mini-Protean apparatus from Bio-Rad (Hercules, CA) with 10-20 % gradient polyacrylamide gels (BioRad, CA). In general, samples were applied directly to the wells of the gel without addition of denaturing or reducing agents or heat. Gels were run for 30-40 minutes at 30 mA, imaged with a color digital camera, and stained with Coomassie before imaging again.

Size Exclusion Chromatography (SEC). Analytical size exclusion was performed on an Agilent 1100 series HPLC equiped with a PolySep-GFC-P 5000 column (Phenomenex, Torrance, CA). At a flow rate of 1 mL/min, rods eluted between 5-8 minutes, depending on length. Though retention times varied with the buffers used, disks typically eluted from 8-9 minutes, while monomers eluted from 9-10 minutes.

Mass Spectrometry. Prior to analysis, biological samples were desalted and separated from small molecule contaminants using spin concentrators or NAP-5 gel filtration columns (GE Healthcare,

Chalfont, UK). Electrospray LC/MS analysis was performed using an LTQ Orbitrap XL hybrid mass spectrometer with an Ion Max electrospray ionization source (Thermo Fisher Scientific, Waltham, MA) connected to a Agilent 1200 series liquid chromatograph (Santa Clara, CA) or an API 150EX system (Applied Biosystems) equipped with a Turbospray source and an Agilent 1100 series LC pump.

Transmission Electron Microscopy (TEM). Samples were prepared for TEM analysis by applying an analyte solution (approximately 0.2 mg/mL in TMVP) to carbon-coated copper grids for 3 min, followed by rinsing with dd-H_2O. The grids were then exposed to a 1.5% aqueous solution of uranyl acetate for 90 s as a negative stain. Images were obtained at the Berkeley Electron Microscope Lab using a FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage. Images were recorded on a Ultra Scan 1000 from Getan (Pleasanton, CA).

4.9 Literature Cited

1. Hermanson, G. T. *Bioconjugate Techniques*, 2nd Ed. Academic Press, San Diego, **2008**.
2. Baschong, W.; Wrigley, N. G. *Journal of Electron Microscopy Technique* **1990**, *14*, 313–323.
3. Sokolov, K.; Follen, M.; Aaron, J.; Pavlova, I.; Malpica, A.; Lotan, R.; Richards-Kortum, R. *Cancer Res.* **2003**, *63*, 1999–2004.
4. Mayer, K. M.; Hafner, J. H. *Chemical Reviews* **2011**, *111*, 3828–57.
5. Rosi, N. L.; Mirkin, C. A. *Chemical Reviews* **2005**, *105*, 1547–62.
6. Mackowski, S.; Wörmke, S.; Maier, A. J.; Brotosudarmo, T. H. P.; Harutyunyan, H.; Hartschuh, A.; Govorov, A. O.; Scheer, H.; Bräuchle, C. *Nano Lett.* **2008**, *8*, 558–64.
7. Beyer, S. R.; Ullrich, S.; Kudera, S.; Gardiner, A. T.; Cogdell, R. J.; Köhler, J. *Nano Lett.* **2011**, *11*, 4897–4901.
8. Kim, I.; Bender, S.; Hranisavljevic, J.; Utschig, L.; Huang, L.; Wiederrecht, G.; Tiede, D. *Nano Lett.* **2011**, *11*, 3091-3098.
9. Atwater, H. A.; Polman, A. *Nature materials* **2010**, *9*, 205–13.
10. Chen, F.-C.; Wu, J.-L.; Lee, C.-L.; Hong, Y.; Kuo, C.-H.; Huang, M. H. *Applied Physics Letters* **2009**, *95*, 013305.
11. Lakowicz, J. R. *Analytical Biochemistry* **2005**, *337*, 171-94.
12. Dulkeith, E.; Ringler, M.; Klar, T. A.; Feldmann, J.; Muñoz Javier, A.; Parak, W. J. *Nano Letters* **2005**, *5*, 585-589.
13. Singh, M. P.; Strouse, G. F. *J. Am. Chem. Soc.***2010**, *132*, 9383-9391.
14. Acuna, G. P.; Bucher, M.; Stein, I. H.; Steinhauer, C.; Kuzyk, A.; Holzmeister, P.; Schreiber, R.; Moroz, A.; Stefani, F. D.; Liedl, T.; Simmel, F. C.; Tinnefeld, P. *ACS Nano* **2012**, *6*, 3189-95.
15. Bharadwaj, P.; Anger, P.; Novotny, L. *Nanotechnology* **2007**, *18*, 044017.
16. Chhabra, R.; Sharma, J.; Wang, H.; Zou, S.; Lin, S.; Yan, H.; Lindsay, S.; Liu, Y. *Nanotechnology* **2009**, *20*, 485201.
17. Kulakovich, O.; Strekal, N.; Yaroshevich, A.; Maskevich, S.; Gaponenko, S.; Nabiev, I.; Woggon, U.; Artemyev, M. *Nano Lett.* **2002**, *2*, 1449–1452.
18. Tovmachenko, O. G.; Graf, C.; van den Heuvel, D. J.; van Blaaderen, A.; Gerritsen, H. C. *Advanced Materials* **2006**, *18*, 91–95.
19. Dulkeith, E.; Morteani, A. C.; Niedereichholz, T.; Klar, T. A.; Feldmann, J.; Levi, S. A.; van Geddes, C. D.; Cao, H.; Gryczynski, I.; Gryczynski, Z.; Fang, J.; Lakowicz, J. R. *J. Phys. Chem. A* **2003**, *107*, 3443–3449.
20. Aslan, K.; Gryczynski, I.; Malicka, J.; Matveeva, E.; Lakowicz, J. R.; Geddes, C. D. *Current Opinion in Biotechnology* **2005**, *16*, 55–62.

21. Cady, N. C.; Strickland, A. D.; Batt, C. A. *Molecular and Cellular Probes* **2007**, *21*, 116–124.
22. Fu, Y.; Zhang, J.; Lakowicz, J. R. *J. Am. Chem. Soc.* **2010**, *132*, 5540–5541.
23. Lakowicz, J. R. *Analytical Biochemistry* **2001**, *298*, 1–24.
24. Shimizu, K. T.; Woo, W. K.; Fisher, B. R.; Eisler, H. J.; Bawendi, M. G. *Phys. Rev. Lett.* **2002**, *89*, 117401.
25. Zin, M. T.; Leong, K.; Wong, N.-Y.; Ma, H.; Sarikaya, M.; Jen, A. K.-Y. *Nanotechnology* **2009**, *20*, 015305.
26. Eustis, S.; El-Sayed, M. A. *Chem. Soc. Rev.* **2006**, *35*, 209–217.
27. Gueroui, Z.; Libchaber, A. *Phys. Rev. Lett.* **2004**, *93*, 166108.
28. Singh, M. P.; Strouse, G. F. *J. Am. Chem. Soc.* **2010**, *132*, 9383–9391.
29. Veggel, F. C. J. M.; Reinhoudt, D. N.; Möller, M.; Gittins, D. I. *Phys. Rev. Lett.* **2002**, *89*, 203002.
30. Jennings, T. L.; Singh, M. P.; Strouse, G. F. *J. Am. Chem. Soc.* **2006**, *128*, 5462–5467.
31. Dulkeith, E.; Ringler, M.; Klar, T. A.; Feldmann, J.; Muñoz Javier, A.; Parak, W. J. *Nano Lett.* **2005**, *5*, 585–589.
32. Schneider, G.; Decher, G.; Nerambourg, N.; Praho, R.; Werts, M. H. V.; Blanchard-Desce, M. *Nano Lett.* **2006**, *6*, 530–536.
33. Endo, M.; Wang, H.; Fujitsuka, M.; Majima, T. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2006**, *12*, 3735–40.
34. Bhyravbhatla, B.; Watowich, S. J.; Caspar, D. L. *Biophysical Journal* **1998**, *74*, 604-15.
35. Ackerson, C. J.; Jadzinsky, P. D.; Jensen, G. J.; Kornberg, R. D. *J. Am. Chem. Soc.* **2006**, *128*, 2635–2640.
36. Malecki, M.; Hsu, A.; Truong, L.; Sanchez, S. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 213-218.
37. Zahr, O. K.; Blum, A. S. *Nano Lett.* **2012**, *12*, 629-33.
38. Loweth, C. J.; Caldwell, W. B.; Peng, X.; Alivisatos, A. P.; Schultz, P. G. *Angewandte Chemie International Edition* **1999**, *38*, 1808–1812.
39. Mokotoff, M.; Mocarski, Y. M.; Gentsch, B. L.; Miller, M.; Zhou, J. -H; Chen, J.; Ball, E. D. *The Journal of Peptide Research* **2001**, *57*, 383–389.
40. Warner, M. G.; Reed, S. M.; Hutchison, J. E. *Chemistry of Materials* **2000**, *12*, 3316–3320.
41. Wijaya, A.; Schaffer, S. B.; Pallares, I. G.; Hamad-Schifferli, K. *ACS nano* **2009**, *3*, 80–6.
42. Laemmli U. K. *Nature* **1970**, *227*, 680–685.

Chapter 5: The Structure and Reactivity of a TMV Coat Protein Dimer

5.1 Abstract

A genetic fusion of two TMV coat protein monomers was created to provide an addressable biomaterial scaffold with alternating units. Two TMV genes with identical codons but distinct nucleotide sequences were connected with a seven residue linker, permitting facile genetic manipulation for the introduction of uniquely reactive residues to each domain. The assembly of these structures was studied to determine the packing pattern.

5.2 Background

Homodimerization and homooligomerization of proteins is a commonly used by Nature to help create complex structures while minimizing genome size.[1] Among the many thousands of non-covalent homodimer proteins known, notable examples include restriction enzymes that recognize palindromic DNA sequences[2] and transmembrane receptors.[3] While many of these protein homodimers associate transiently and therefore rely exclusively on non-covalent interactions, others are joined covalently by cysteine disulfides to provide additional stability.[4,5,6] Notable examples of proteins that naturally form covalent homodimers are IgG antibodies[7] and the coat protein of bacteriophage Qβ.[8]

Joining two proteins through genetic fusion is a commonly used technique to combine their desirable properties. This has gained widespread use to facilitate protein expression and purification,[9] track proteins *in vivo*,[10] and increase half-life of therapeutic proteins.[11] The engineering of covalent homodimers is far less common, and the reports found start from non-covalent homodimers. Single chain estrogen receptors have been made to study signaling,[12] the bacteriophage PP7 coat protein was dimerized to increase stability,[13] I-DmoI was dimerized to create a novel site specific DNA endonuclease,[14] the MS2 coat protein was dimerized to compensate for destabilizing insertions,[15] and cytochrome bc_1 was dimerized to study its function.[16] In the last two examples, the dimerization served the purpose to break the symmetry of the dimer, allowing the creation of systems in which the two halves (A and B) could be modified independently. This method of producing covalent 'heterodimers' from non-covalent homodimers avoids the stochastic distribution of products (AA, AB, BA, BB) that would be produced by the random assembly two different species.

5.3 Production of the TMV Coat Protein Dimer

We created a dimer of TMV in hopes that we could uniquely modify each monomer, providing control over the spatial organization of two components (A and B). We hypothesized that upon assembly, these groups would arrange in an alternating ABAB pattern (Figure 5-1). This would provide greater flexibility in the study of interactions between donors, acceptors, photoprotective groups, and photocatalysts on the surfaces of TMV assemblies.

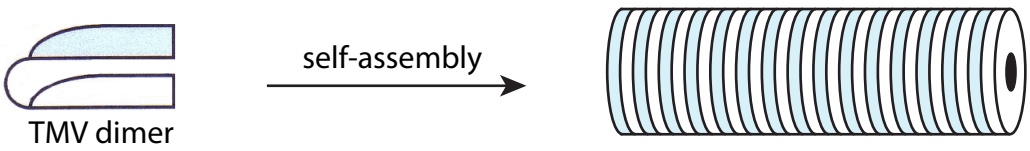

Figure 5-1. Hypothesized assembly of the TMV dimer. A covalent linkage between the two domains allows different synthetic components to be addressed to each.

The TMV coat protein dimer was first constructed by cloning a codon-optimized TMV gene upstream of a wild-type TMV gene in a pTYB1 plasmid (discussed in Chapter 2). The codons used in each TMV gene were distinct enough that primers designed to hybridize to one gene did not bind the other. This allowed the selective mutation of residues in either of the fused monomers. Quikchange mutagenesis was first used to replace the restriction site between the domains with a flexible linker of 1, 3, 5, or 7 glycine residues. In an expression screen, the dimers connected with

1 and 5 glycine residues underwent significant *in vivo* cleavage of their fused intein affinity handle, preventing their purification. Similar *in vivo* cleavage has been observed with many other TMV-intein constructs, and is a well-documented problem in the IMPACT manual. Its cause is unclear, and it may not reflect the ultimate stability or workability of the purified protein. Of the remaining two constructs, the 7 glycine linker gave a higher yield, and so was chosen for further work (Figure 5-2a). With an N-terminal AlaGly, PLP-mediated transamination and reaction with 2K-PEG-ONH_2 reached 53% conversion (Figure 5-2b). Because short rods were observed to aggregate laterally, a glutamate was inserted into the linker to increase hydrophilicity and provide electrostatic repulsion (Figure 5-2c). While the intein based expression and purification system facilitated the generation of small mutant libraries due to the ease of parallel purification (detailed in chapter 2), there were two significant drawbacks. Expression and purification of the dimer as an intein-fusion provided at most 2 mg/L of expression, which was not enough for structural studies such as crystallography. Furthermore, contaminating proteins would sometimes co-purify, complicating the subsequent analysis.

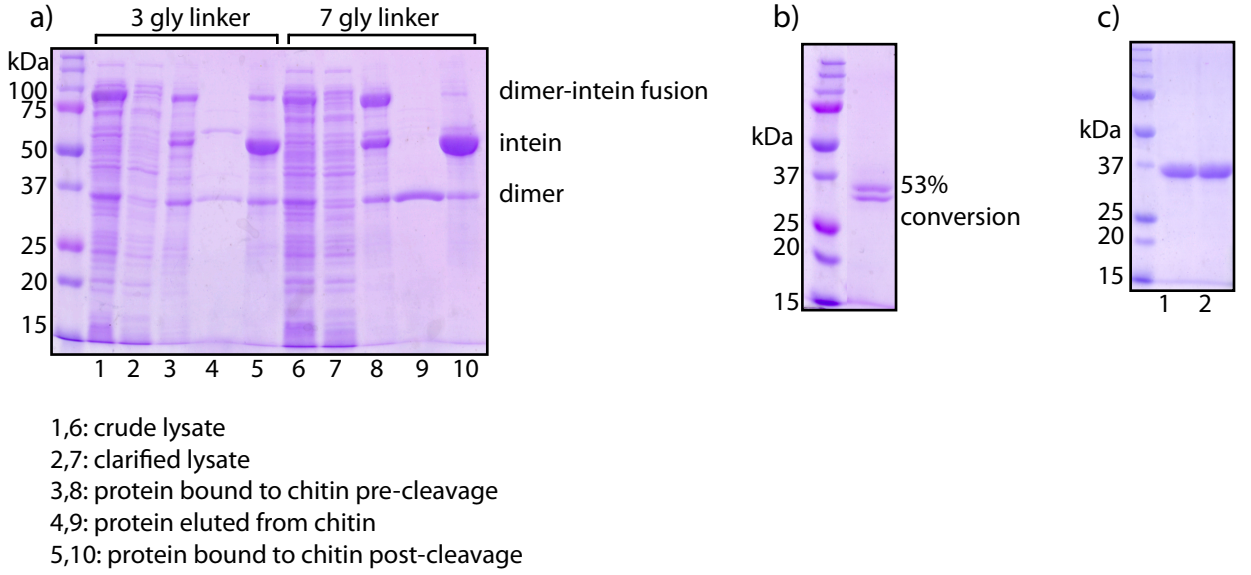

Figure 5-2. SDS-PAGE gels of the expression, purification, and modification of TMV coat protein dimer from a pTYB1 vector. (a) The best yield of 2 mg / L was obtained for dimer containing a linker of 7 glycines between the TMV coat proteins (lane 9). (b) Following PLP transamination and oximation with 2K-PEG-ONH2, 53% coversion was observed by densitometry. (c) Expression and purification of TMV dimer was not affected by the introduction of a glutamate residue in place of the fourth (lane 1) or fifth (lane 2) residue of the linker.

To obtain TMV coat protein dimer protein in higher yields and purity, we tested expression from a number of other vectors. The pET20b plasmid used for most TMV expression produced large quantities of insoluble protein, presumably because the folding machinery of the *E. coli* could not keep up with high expression levels dictated by the T7 promoter. The increased sensitivity of this construct relative to normal TMV can be rationalized by imagining the extra non-productive protein folding pathways, such as intra- and inter-protein domain swapping.[17] Decreasing the temperature from 30 to 20 °C did not improve the protein solubility, so the gene was cloned into vectors with weaker promoters. The pBAD plasmid contained a relatively weak P_{BAD} promoter, which was reported in the product literature to aid the soluble expression of protein.[18] Unfortunately,

the restriction site adjacent to the start codon is NcoI, which restricts the first nucleotide to guanine and therefore the N-terminal amino acids to A, G, E, D, and V. The more commonly used NdeI restriction site provides no such limitations, but its presence elsewhere in the plasmid prevented its addition to the multiple cloning site. Expression of the dimer from the pBAD plasmid produced 10-30 mg/L of TMV coat protein dimer protein (Figure 5-3).

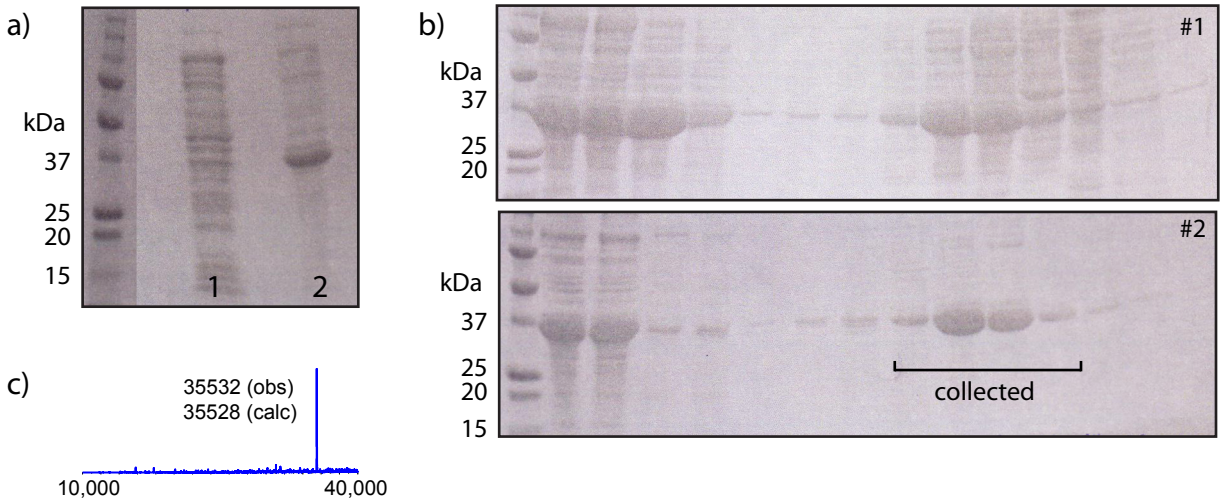

Figure 5-3. Expression and purification of a TMV coat protein dimer from the pBAD vector. (a) The clarified lysate was treated with NH4SO4 and centrifuged to separate contaminating proteins in the superatant (lane 1) from the TMV dimer in the pellet (lane 2). (b) The resuspended pellet was purified over an anion exchange colunm twice to provide pure protein. (c) The observed mass matched the calulated mass, as determined by ESI-MS.

5.4 Assembly of TMV Coat Protein Dimer

Much like the RR TMV mutants (chapter 4), the TMV coat protein dimer forms stable disks at pH 8 (Figure 5-6c) as well as rods of stacked disks. The assembly pattern of the dimer is important for its rational use as a scaffold for nanofabrication. We hypothesized that six non-redundant disk assemblies were possible (Figure 5-4). To distinguish between the structures, a S123C mutation

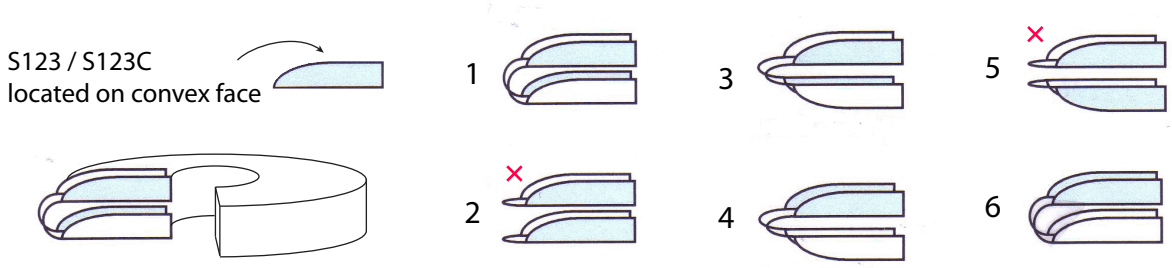

Figure 5-4. Hypothesized assembly structures for disks of TMV coat protein dimers. The cartoons depict the possible ways in which two adjacent TMV dimers can arrange themselves within a disk. Blue and white are used to distinguish the N- and C-terminal domains, respectively. Redundant structures were omitted for clarity. Since the disks formed by the TMV coat protein dimer (Figure 5-6) appear to be the same size as normal TMV disks, we assume that both contain 17 monomers or domains per layer. This would eliminate model structures 2 and 5, since the side-by-side arrangement would lead to an even number of domains per layer (16 or 18).

was introduced into the first, second, or both domains. We hoped to distinguish between possible structures based on the relative reactivity of these cysteines. For instance, structure **6** in Figure 5-4 might show greater reactivity when the cysteine is on the N-terminal (blue) domain compared to the C-terminal (white) domain. However, no difference in reactivity was observed when titrating in a malemide-functionalized dye or 2K-PEG-maleimide (data not shown).

A bulkier maleimide was generated by reacting the lysines of hen egg white lysozyme with a NHS ester-maleimide crosslinker (Figure 5-5a). Lysozyme is well-suited for this due

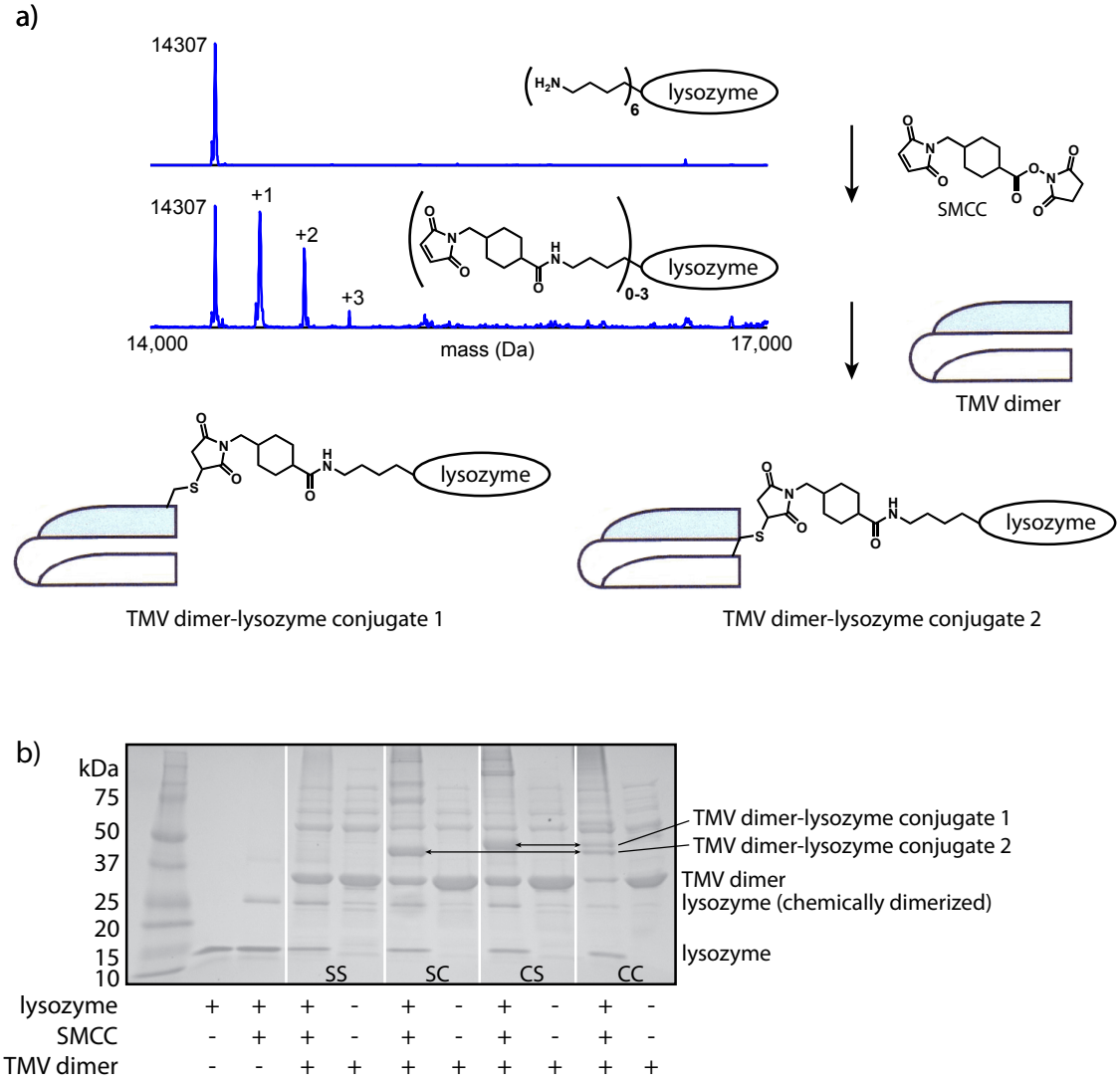

Figure 5-5. Probing the assembly of the TMV coat protein dimer by conjugation to maleimide-bearing lysozyme. (a) To create a maleimide with a sterically bulky substituent, the crosslinker succinimidyl-4-(n-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) was reacted with lysozyme, which contains six lysines and no cysteines. Mass spectra indicated a distribution of 0-3 additions of SMCC to lysozyme, and that the crosslinkers retained their reactive maleimide groups. These were incubated with disks of TMV dimer having undergone S123C mutations on neither (SS), the C-terminal (SC), N-terminal (CS), or both (CC) of the domains. (b) The protein coupling reactions were analyzed by SDS-PAGE, which showed that lysozyme conjugates to the N- and C-terminal domains of the TMV dimer travelled at slightly different rates. This difference made it possible to visualize bands for the modification of both N- and C-terminal domains of the TMV dimer in the same lane of the 'CC' protein modified with lysozyme. The observation that both domains are equally reactive ruled out the possibility that the dimer assembled into disk structure number 6 in Figure 5-4.

to its small size, commercial availability, and lack of free thiols (the eight cysteines form four disulfide bonds). While the six lysines and the N-terminus can be modified with NHS esters under forcing conditions, three of the lysines are particularly reactive.[19] A heterogeneous mixture of 0-3 conjugates was produced and mixed with TMV coat protein dimer disks. A SDS-PAGE gel showed that cysteines on both domains were equally reactive (Figure 5-5b), and thus suggested that the TMV coat protein dimers did not form structure **6** (Figure 5-4). All proteins remained in disk form following the reaction, ruling out the possibility of disassembly (and therefore spurious results) due to lysozyme conjugation. The possible assembly states could have been narrowed further if it were possible to observe a TMV coat protein dimer with a cysteine in both domains reacting with two lysozyme proteins to make a ~64.6 kDa complex. It is unfortunately difficult to distinguish larger molecular weight bands containing TMV from contaminating proteins.

Preparations were made to perform an analogous experiment to study the accessibility of surface thiols with gold nanoparticles in the place of maleimide-bearing lysozyme (Figure 5-6). Because the S123C mutant is unable to bind gold, a lysine was introduced onto the edge of the pore

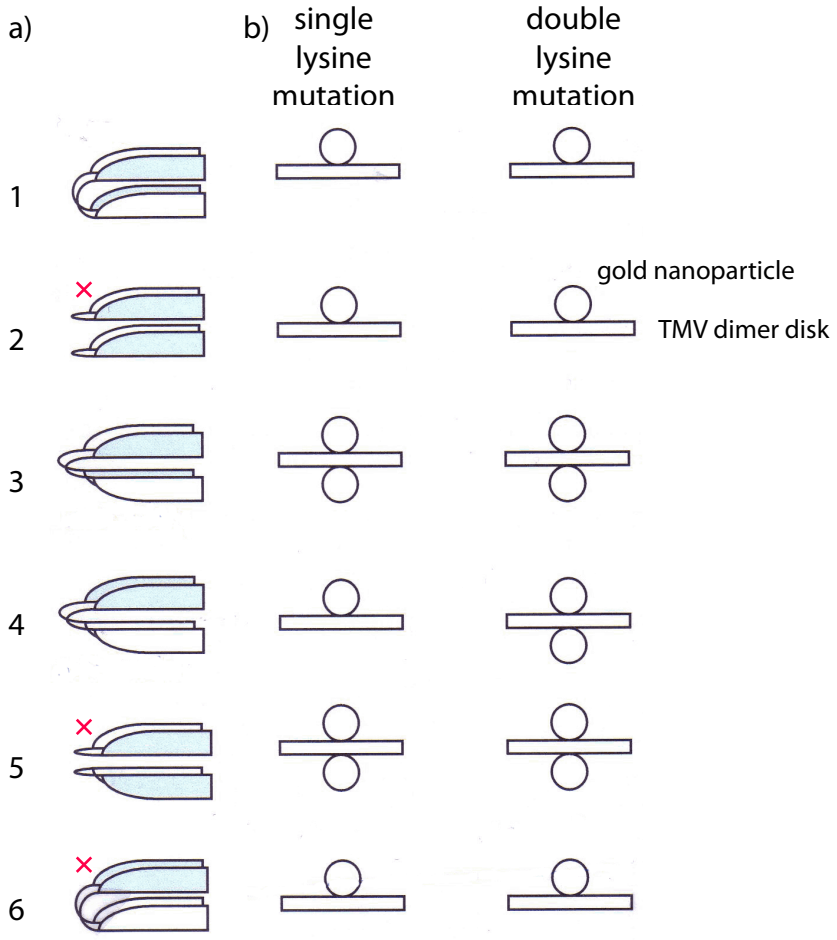

Figure 5-6. Probing the assembly of the TMV coat protein dimer by conjugation to a gold nanoparticle. This proposed experiment may determine if the TMV coat protein dimer disks are apolar, similar to the permutant (Figure 3-11d), or polar, similar to the RR mutants (Figure 4-6). (a) Previous experiments have eliminated three of the six hypothesized assembly structures. A lysine can be introduced on the edge of the pore on one or both of the domains. After conjugation to a thiol-bearing crosslinker, this lysine residue should bind gold nanoparticles, as discussed in Chapters 3 and 4. The reactivity of dimers in each assembly state is predicted in (b) for single and double lysine mutants.

of the N-terminal domain for thiolation, as discussed in chapter 3 and 4. The cysteines were capped with maleimide and biotin-NHS was added to measure the reactivity of the protein. As shown by **Figure 5-7a**, two additions of NHS-biotin were observed, which is consistent with the presence of the extended N-terminus and T104K mutation. The lack of reactivity of K53 and K68 is attributed to the stable assembly of the TMV coat protein dimer into disks, as measured by SEC (Figure 5-6b). TEM images of the TMV coat protein dimer solution after purification show a mixture of disks and rods composed of stacked disks. The cause of the discrepancy between SEC and TEM is unknown. Experiments to conjugate gold to the disks will begin shortly.

There are two hypotheses as to why the stacked disk rods formed by the TMV coat protein dimer do not 'anneal,' or transition from a 17-fold radially symmetric assembly to the helical assembly that characterizes the intact virus and structures of normal TMV at low pH. The first is that the transition of normal TMV stacked disks to rods requires a 'tightening' of the rod from 17 monomers per turn to 16 1/3, and this might stretch the linker of every protein unfavorably. The second hypothesis is that if the TMV coat protein dimer assembles as an apolar disk (structures **3-5** in Figure 5-6a and the permuted TMV, Figure 3-5), assembly into an annealed rod would require significant reorganization to orient the domains of both rings in the same direction. This appears possible for the permuted TMV, judging by the formation of long rods (Figure 3-3), but the addition of a covalent linkage between the rings may prove an insurmountable barrier.

Finally, crystallography screens based on the conditions used for cpTMVP did not produce any crystals.

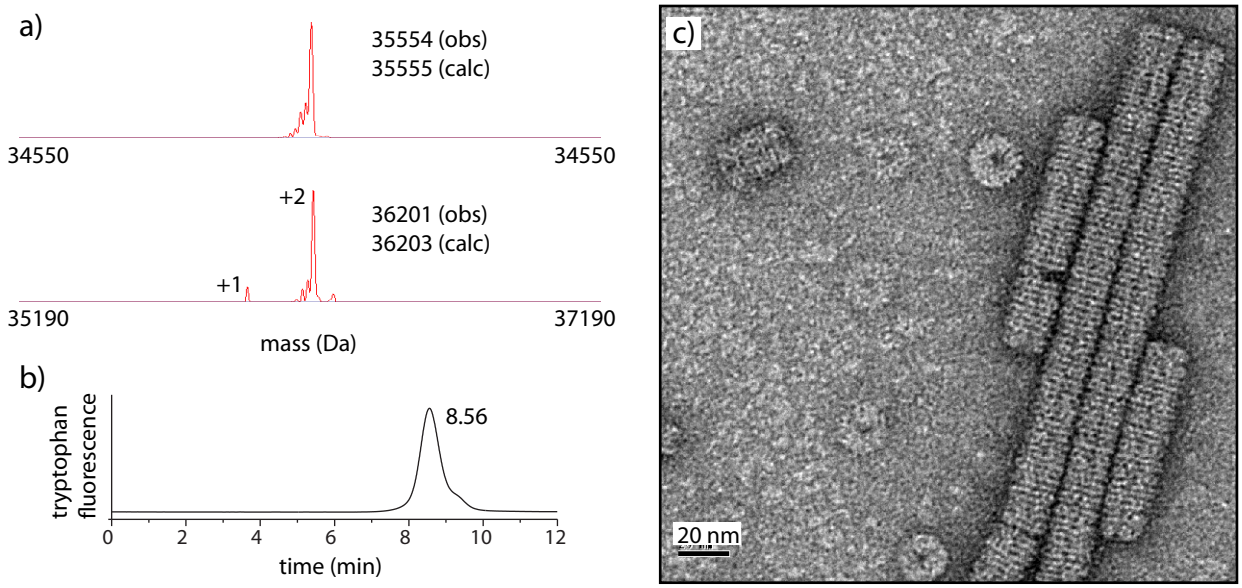

Figure 5-7. Modification and assembly of a TMV coat protien dimer with a lysine in the pore. A TMV dimer mutant was created with an N-terminus extended with a serine, S123C mutations in both domains, and a T104K mutation in the N-terminal domain. The cysteines were capped with maleimide and the lysine reactivity measured by the addition of biotin-NHS. (a) Mass spectra showed two additions, consistent with modification of the N-terminus and T104K. SEC traces of the disks in 100 mM TEA, pH 8, suggested the assembly state was disks before (not shown) and after (b) the reactions. (c) TEM images recorded after the reactions showed both disks and rods of stacked disks.

5.5 Conclusion

A new scaffold was produced by joining two TMV coat protein monomers with a flexible peptide linker. After expression, these proteins provided stable disks and stacked-disk rods in which each 'monomer' has a known neighbor, so the pair can be modified using synthetic groups that complement the function of the other. The possibilities include chromophore pairs for studying FRET, chromophores paired with photoprotective anthracine derivatives, or cooperative molecular catalysts. Studies on the exact nature of the assembly are underway, and will expand the use of TMV-based proteins as nanoscale building blocks.

5.6 Methods and Materials

General Procedures. Unless otherwise noted, all chemicals and solvents were of analytical grade and were used as received from commercial sources. Water (dd-H_20) used in biological procedures or as reaction solvents was deionized using a NANOpure purification system (Barnstead, United States). Spin concentration steps were performed using 100,000 molecular weight cutoff spin concentrators from Millipore (Billerica, MA) in an Allegra 64R Tabletop Centrifuge (Beckman Coulter, Fullerton, CA). Protein samples were quantified by the Bradford assay.

Preparation of the DNA Construct for the TMV Coat Protein Dimer. The starting point for the TMV coat protein dimer constructs was the S123C mutant of the coat protein of the TMV U1 strain, cloned into a pTYB1 vector (NEB, Ipswich, MA). A spacer and EcoRI restriction site were introduced between the NdeI (start codon) and the beginning of the gene to provide 14 a nucleotide spacer between the NdeI site and new EcoRI site. This was performed using QuikChange mutagenesis (Stratagene, Santa Clara, CA). A TMV gene containing the S123C mutation and optimized for the codon usage of *E. coli* (Genscript, Piscataway, NJ) was PCR amplified with an NdeI restriction site on the 5' end and an EcoRI restriction site on the 3' end. Both vector and insert were doubly digested before ligation with T4 DNA ligase (NEB). QuikChange mutagenesis was used to replace the EcoRI restriction site with a series of 1-7 amino acid linkers.

For expression from the pBAD vector (Life Technologies), the gene for the TMV coat protein dimer was PCR amplified from the pTYB1 vector with primers encoding an NcoI restriction site on the 5' end and a HindIII restriction site on the 3' end. Both pBAD vector and insert were doubly digested before ligation with T4 DNA ligase (NEB). QuikChange mutagenesis was used to change the N-terminus and introduce a T104K mutation into the first gene.

```
>oTMV-TMV both domains have S123C, linker GGGEGGG
AGCTATAGCATTACCACCCCGAGCCAGTTTGTGTTTCTGAGCAGCGCCTGGGCGGATCCG
ATTGAACTGATTAACCTGTGCACCAATGCGCTGGGCAACCAATTTCAGACCCAGCAGGCG
CGCACCGTTGTGCAGCGTCAGTTCAGCGAAGTTTGGAAACCGAGCCCGCAGGTTACCGTG
CGCTTTCCGGATAGCGATTTTAAAGTGTATCGCTATAACGCCGTGCTGGATCCGCTGGTG
ACCGCGCTGCTGGGCGCCTTTGATACCCGTAATCGTATCATTGAAGTGGAAAACCAGGCC
AATCCGACCACCGCGGAAACCCTGGATGCGACCCGTCGTGTGGATGATGCCACCGTGGCG
ATTCGCTGTGCCATCAATAACCTGATTGTGGAACTGATTCGTGGCACCGGCAGCTATAAC
CGTAGCAGCTTTGAAAGCAGCAGCGGCCTGGTGTGGACGAGCGGCCCGGCGACCGGCGGC
GGTGAAGGCGGTGGTTCTTACAGTATCACTACTCCATCTCAGTTCGTGTTCTTGTCATCA
```

GCGTGGGCCGACCCAATAGAGTTAATTAATTTATGTACTAATGCCTTAGGAAATCAGTTT
CAAACACAACAAGCTCGAACTGTCGTTCAAAGACAATTCAGTGAGGTGTGGAAACCTTCA
CCACAAGTAACTGTTAGGTTCCCTGACAGTGACTTTAAGGTGTACAGGTACAATGCGGTA
TTAGACCCGCTAGTCACAGCACTGTTAGGTGCATTCGACACTAGAAATAGAATAATAGAA
GTTGAAAATCAGGCGAACCCCACGACTGCCGAGACGTTAGATGCTACTCGTAGAGTAGAC
GACGCAACGGTGGCCATAAGGTGCGCGATAAATAATTTAATAGTAGAATTGATCAGAGGA
ACCGGATCTTATAATCGGAGCTCTTTCGAGAGCTCTTCTGGTTTGGTTTGGACCTCTGGT
CCTGCAACT

>oTMV-TMV both domains have S123C, linker GGGEGGG
SYSITTPSQFVFLSSAWADPIELINLCTNALGNQFQTQQARTVVQRQFSEVWKPSPQVTV
RFPDSDFKVYRYNAVLDPLVTALLGAFDTRNRIIEVENQANPTTAETLDATRRVDDATVA
IRCAINNLIVELIRGTGSYNRSSFESSSGLVWTSGPATGGGEGGGSYSITTPSQFVFLSS
AWADPIELINLCTNALGNQFQTQQARTVVQRQFSEVWKPSPQVTVRFPDSDFKVYRYNAV
LDPLVTALLGAFDTRNRIIEVENQANPTTAETLDATRRVDDATVAIRCAINNLIVELIRG
TGSYNRSSFESSSGLVWTSGPAT

Expression and Purification of TMV Coat Protein Dimer from pTYB1. Tuner DE3pLysS competent cells (Novagen) were transformed with a plasmid encoding the TMV coat protein dimer, and colonies were selected for inoculation of 1L Terrific Broth cultures. When cultures reached mid-log phase as determined by O.D.600, expression was induced by addition of 0.3 mM IPTG (Invitrogen). Cultures were grown for 20 h at 20 °C, harvested by centrifugation, and stored at -80 °C. The following purification steps were performed at 4 °C. Induced cells from 1 L of media were thawed and resuspended in 40 mL of Column Buffer (20 mM Tris, pH 8, 500 mM NaCl) and lysed by sonication (Branson Ultrasonics, Danbury, CT.) The cell lysate was cleared by centrifugation at 7,000 rpm for 30 min. The cleared lysate was loaded onto a gravity flow column packed with 5 mL of chitin resin equilibrated with 40 mL of Column Buffer. The column was spun on a rotisserie for 1 h to bind the fusion protein to the chitin resin and then washed with an additional 40 mL Column Buffer. This was followed with 15 mL of Cleavage Buffer (100 mM Tris pH 8, 50 mM DTT). The column was capped with 1 mL of cleavage buffer above the resin and left at 25° C for 20 h. Fractions were analyzed by SDS-PAGE, and fractions containing TMV coat protein dimer were combined and concentrated.

Of the two intein-CBDs available in the commercial IMPACT-CN system from NEB, the larger in pTYB1 was preferred because it bound more tightly to chitin resin. The smaller intein-CBD from the pTXB1 plasmid would disassociate from the chitin and elute with the protein. It was eventually discovered that passage of this eluent over additional chitin resin was shown to provide pure protein.

The identity of the C-terminal residue affects the stability of the amide bond between the protein and intein. Consistent with observations tabulated in the IMPACT-CN manual and detailed in (ref Johnson 2006), the native threonine residue of TMV led to slow and incomplete *in vitro* intein cleavage. Mutation to glycine, but not serine, improved the cleavage significantly.

Expression and Purification of TMV Coat Protein Dimer from pBAD. DH10B competent cells (Life Technologies) were transformed with the pBAD vector encoding the TMV coat protein dimer gene and cultured in Terriffic Broth with 100 µg/L ampicillin. When cultures reached an optical density of 0.6 to 0.8, arabinose was added to a final concentration of 0.01%. Cultures were grown 24 h at 20 °C, harvested by centrifugation, and stored at -80 °C. Cells (from a 1 L expression batch) were thawed, resuspended in 40 mL of 20 mM TEA pH 7.2, and lysed by sonicating with

a 2 s on, 6 s off cycle for a total of 20 minutes using a standard disruptor horn at 90% amplitude (Branson Ultrasonics, Danbury, CT). The resulting lysate was cleared by ultracentrifugation for 30 min at 45,000 rpm using a Beckman 45 Ti rotor in an Optima L-80 XP (Beckman Coulter) or 30 min at 10,000 rpm using a using a SLA-600TC rotor in a Sorvall RC5C Plus centrifuge (Waltham, MA). The clarified lysate was decanted, warmed to rt, and stirred while adding a saturated solution of ammonium sulfate dropwise to a final concentration of 35% (v/v). After 5 min, the white ppt that formed was pelleted by ultracentrifugation (15 min at 45,000 rpm in a Beckman 45 Ti rotor), washed with deionized water, and resuspended in buffer A. The resulting protein solution was next loaded onto a DEAE column and eluted with a 0 – 300 mM NaCl gradient. Purity was confirmed by SDS-PAGE and HPLC. This preparation provided pure TMV coat protein dimer in yields up to 20 mg/L culture.

Buffer A - 25 mM potassium phosphate, pH 6.5

General Procedure for the PLP Modification of TMV Coat Protein Dimer. A 1.5 mL Eppendorf tube was charged with a solution of TMV coat protein dimer (delivered as 100 μL of a 2 mg ml^{-1} solution in 25 mM potassium phosphate, pH 6.5; 1 eq) and a solution of freshly prepared PLP (delivered as 100 μL of a 20 mM solution in 25 mM phosphate buffer, pH adjusted to 6.5 with 1 M NaOH; 175 eq. The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at 30 °C for 24 h. The PLP was removed from the reaction mixture *via* size exclusion chromatography (Nap-5 desalting column (GE Healthcare) eluting into 25 mM phosphate buffer, pH 6.5.) A 50 uL aliquot of the resulting purified mixture was treated with 50 μL of 0.1 M 2K-PEG-ONH_2. The mixture was briefly agitated to ensure proper mixing and was incubated without further agitation at rt for 18-20 h. Samples were combined with loading buffer, and analyzed by SDS-PAGE.

Lysozyme-TMV Coat Protein Dimer Crosslinking with SMCC. A 200 μL solution of 0.8 mM hen egg white lysozyme in buffer A was prepared. To this solution was added 5 eq of succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC, 16 μL of a 50 mM DMSO solution). The reaction mixture was vortexed briefly and was left at 25 °C for 30 min. The bulk of the unreacted crosslinker was removed by Nap-5 gel filtration, with elution into 400 μL of the same buffer. The extent of protein modification was monitored by LC/ESI-MS (Orbitrap or 150EX) and yield was monitored by UV-vis spectroscopy. To 10 μL of a 0.28 mM solution of SMCC-lysozyme was added 15 μL of 0.12 mM TMV coat protein dimer for a lysozyme:TMV coat protein dimer ratio of 1:0.6. These reacted for 1 h before analysis by SDS-PAGE.

Maleimide Attachment to S123C and Biotin-NHS Attachment to T104K. A 50 μL aliquot of TMV coat protein dimer in buffer A (0.29 mM or 10 mg/mL) was combined with 36 eq of maleimide (1 μL of a 500 mM solution in DMSO). The reaction mixure was vortexed briefly and was left at 25 °C for 30 min. To this solution was added 7 eq of biotin-NHS (2 μL of a 100 mM solution in DMF). The bulk of the unreacted maleimide and biotin-NHS was removed by Nap-5 gel filtration, with elution into 200 μL of 100 mM TEA, pH8. The extent of protein modification was monitored by LC/ESI-MS (Orbitrap or 150EX).

Gel Analyses. Sodium dodecyl sulfate-poly(acrylamide) gel electrophoresis (SDS-PAGE) was

accomplished on a Mini-Protean apparatus from Bio-Rad (Hercules, CA) with 15 % gradient polyacrylamide gels (BioRad, CA), following the protocol of Laemmli.[20] All electrophoresis protein samples were mixed with SDS loading buffer in the presence of dithiothreitol (DTT) and heated to 100 °C for 10 min to ensure reduction of disulfide bonds and complete denaturation unless otherwise noted. Commercially available molecular mass markers (Bio-Rad) were applied to at least one lane of each gel for calculation of the apparent molecular masses. Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA). Protein reaction conversion was estimated from standard optical density measurements of the observed gel bands with Image J™ software (version 1.34s).

Size Exclusion Chromatography (SEC). Analytical size exclusion was performed on an Agilent 1100 series HPLC equiped with a PolySep-GFC-P 5000 column (Phenomenex, Torrance, CA). At a flow rate of 1 mL/min, rods eluted between 5-8 minutes, depending on length. Though retention times varied with the buffers used, disks typically eluted from 8-9 minutes, while monomers eluted from 9-10 minutes.

Mass Spectrometry. Prior to analysis, biological samples were desalted and separated from small molecule contaminants using spin concentrators or NAP-5 gel filtration columns (GE Healthcare, Chalfont, UK). Electrospray LC/MS analysis was performed using an LTQ Orbitrap XL hybrid mass spectrometer with an Ion Max electrospray ionization source (Thermo Fisher Scientific, Waltham, MA) connected to a Agilent 1200 series liquid chromatograph (Santa Clara, CA) or an API 150EX system (Applied Biosystems) equipped with a Turbospray source and an Agilent 1100 series LC pump.

Transmission Electron Microscopy (TEM). Samples were prepared for TEM analysis by applying an analyte solution (approximately 0.2 mg/mL in TMV coat protein dimer) to carbon-coated copper grids for 3 min, followed by rinsing with dd-H_2O. The grids were then exposed to a 1.5% aqueous solution of uranyl acetate for 90 s as a negative stain. Images were obtained at the Berkeley Electron Microscope Lab using a FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage. Images were recorded on an Ultra Scan 1000 from Getan (Pleasanton, CA).

5.7 Literature Cited

1. Marianayagam, N. J.; Sunde, M.; Matthews, J. M. *Trends in biochemical sciences* **2004**, *29*, 618–25.
2. Pingoud, A.; Jeltsch, A. *Nucleic Acids Research* **2001**, *29*, 3705–3727.
3. Kumar, V.; Chambon, P. *Cell* **1988**, *55*, 145–156.
4. Yasui, N.; Kitago, Y.; Beppu, A.; Kohno, T.; Morishita, S.; Gomi, H.; Nagae, M.; Hattori, M.; Takagi, J. *The Journal of Biological Chemistry* **2011**, *286*, 35247–56.
5. Smith, P. J.; Brandt, W. F.; Stickells, B. J.; Von Holt, C. *Comparative Biochemistry And Physiology B Comparative Biochemistry* **1992**, *103*, 975–980.
6. Linsley, P. S.; Nadler, S. G.; Bajorath, J.; Peach, R.; Leung, H. T.; Rogers, J.; Bradshaw, J.; Stebbins, M.; Leytze, G.; Brady, W. *The Journal of Biological Chemistry* **1995**, *270*, 15417–15424.
7. Chatron, P.; Pontet, F. *Annales de Biologie Clinique* **1992**, *50*, 565-575.
8. Golmohammadi, R.; Fridborg, K.; Bundule, M.; Valegaard, K.; Liljas, L. *Structure* **1996**, *4*, 543–554.
9. Riggs, P. In *Current Protocols in Molecular Biology*; Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
10. Kahana, J. A.; Silver, P. A. In *Current Protocols in Molecular Biology*; Ausubel, F. M.; Brent, R.; Kingston, R. E.; Moore, D. D.; Seidman, J. G.; Smith, J. A.; Struhl, K., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2001.
11. Huang, C. *Current Opinion in Biotechnology* **2009**, *20*, 692–9.
12. Li, X.; Huang, J.; Yi, P.; Bambara, R. A.; Hilf, R.; Muyan, M. *Mol. Cell. Biol.* **2004**, *24*, 7681–7694.
13. Caldeira, J. C.; Peabody, D. S. *Journal of Nanobiotechnology* **2007**, *5*, 10.
14. Silva, G. H.; Belfort, M.; Wende, W.; Pingoud, A. *Journal of Molecular Biology* **2006**, *361*, 744–54.
15. Peabody, D. S. *Archives of Biochemistry and Biophysics* **1997**, *347*, 85–92.
16. Czapla, M.; Borek, A.; Sarewicz, M.; Osyczka, A. *Protein Engineering, Design and Selection 25*, 11.
17. Rousseau, F.; Schymkowitz, J.; Itzhaki, L. *Implications of 3D Domain Swapping for Protein Folding, Misfolding and Function*; Landes Bioscience, 2000.
18. Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. *J Bacteriol*. **1992**, *177*, 4121-4130.
19. Teske, C. A.; Simon, R.; Niebisch, A.; Hubbuch, J. *Biotechnology and Bioengineering* **2007**, *98*, 193–200.
20. Laemmli U. K. *Nature* **1970**, *227*, 680–685.