

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Development of Viral Capsid DNA Aptamer Conjugates as Cell-Targeted Delivery Vehicles

Permalink

<https://escholarship.org/uc/item/54x0b673>

Author

Tong, Gary

Publication Date

2012

Peer reviewed|Thesis/dissertation

Development of Viral Capsid DNA Aptamer Conjugates as Cell-Targeted Delivery Vehicles

by

Gary Jen-Wei Tong

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Matthew B. Francis, Chair

Professor Carolyn R. Bertozzi

Professor Douglas S. Clark

Spring 2012

Development of Viral Capsid DNA Aptamer Conjugates
as Cell-Targeted Delivery Vehicles

Copyright © 2012

by Gary Jen-Wei Tong

Abstract

Development of Viral Capsid DNA Aptamer Conjugates as Cell-Targeted Delivery Vehicles

by

Gary Jen-Wei Tong

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

The ability to generate semi-synthetic DNA-protein conjugates has become increasingly important in the fields of chemical biology and nanobiotechnology. As applications in these fields become more complex, there is also an increased need for methods of attaching synthetic DNA to protein substrates in a well-defined manner. This work outlines the development of new methods for site-specific DNA-protein bioconjugation, as well as the development of novel viral capsid DNA aptamer conjugates for cell-targeting purposes.

In order to generate DNA-protein conjugates in a site-specific manner, chemistries orthogonal to native functional groups present on DNA and proteins were exploited. In one method, the attachment of DNA to proteins was achieved *via* oxime formation. This strategy involved the *in situ* deprotection of an allyloxycarbonyl-protected alkoxyamine-bearing DNA in the presence of a protein containing a single ketone group. The utility of this approach was demonstrated in the synthesis of a DNA-GFP conjugate. In addition to the oxime formation route, two oxidative coupling methods were also developed for DNA-protein bioconjugation. The first reaction coupled phenylenediamine-containing DNA to anilines, which had been site-specifically incorporated into proteins, in the presence of NaIO_4 . These reaction conditions were demonstrated on the proteins bacteriophage MS2 and GFP, and were mild enough for the components to retain both protein structure and DNA base-pairing capabilities. The second oxidative coupling reaction conjugated aniline-containing proteins to DNA bearing an *o*-aminophenol moiety. This reaction occurred under similarly mild conditions; however, higher coupling yields were achieved on MS2 at shorter reaction times by using this strategy. In all three of these methods, the generation of a singly-modified product was achieved.

Using one of our oxidative coupling strategies, MS2-DNA aptamer conjugates were synthesized for the development of multivalent cell-targeting delivery vehicles. These agents were generated by selectively functionalizing the interior and exterior surfaces of MS2 with functional molecules and DNA aptamers, respectively, using orthogonal bioconjugation reactions. Interior surface modification was achieved through the incorporation of a uniquely-reactive cysteine residue, while exterior modification occurred *via* the introduction of the non-natural amino acid *p*-aminophenylalanine. MS2 capsids possessing interior fluorophores and exterior DNA aptamers targeted to a Jurkat T cell surface receptor were synthesized using this strategy. In cell-binding experiments, these dual-surface modified capsids were shown to bind target cells in an aptamer-dependent manner. In addition, colocalization experiments using confocal microscopy elucidated their cellular internalization pathway.

Following validation of the cell-targeting capabilities of aptamer-MS2 conjugates, a multivalent photodynamic agent was developed for targeted photodynamic therapy. This agent was synthesized by installing singlet oxygen-generating porphyrins on the interior of MS2 capsids possessing DNA aptamers on the exterior. Upon illumination with 415 nm light, these dual-modified capsids were shown to generate cytotoxic singlet oxygen. In cell experiments, these agents were shown to selectively kill Jurkat cells in a heterogeneous cell mixture.

To my family.

Table of Contents

DEDICATION.....	i
ACKNOWLEDGEMENTS.....	iv
CHAPTER 1: DNA-protein conjugates: novel materials for applications in nanobiotechnology.....	1
1.1 Abstract.....	1
1.2 Introduction.....	2
1.2.1 DNA.....	2
1.2.2 Protein.....	3
1.3 Novel applications for DNA-protein conjugates.....	4
1.3.1 Bioanalytics.....	4
1.3.2 Nanoscale self-assembly.....	7
1.3.3 Artificial enzyme regulation.....	9
1.4 Conclusion.....	11
1.5 References.....	11
CHAPTER 2: Site-specific methods for DNA-protein conjugation.....	14
2.1 Abstract.....	14
2.2 Introduction.....	15
2.2.1 Recent advances in methods for DNA-protein bioconjugation.....	15
2.2.2 Development of new methods for DNA-protein bioconjugation.....	18
2.3 Results and discussion.....	19
2.3.1 Development of an alkoxyamine-bearing DNA for oxime formation.....	19
2.3.2 An oxidative approach for coupling DNA to aniline-containing proteins.....	23
2.3.3 Oxidative coupling of DNA to proteins <i>via o</i> -aminophenols.....	26
2.4 Conclusion.....	27
2.5 Materials and methods.....	28
2.5.1 General procedures and materials.....	28
2.5.2 Instrumentation and sample analysis.....	28
2.5.3 Experimental.....	29
2.6 References.....	33

CHAPTER 3: Viral capsid DNA aptamer conjugates for cell-targeted imaging and drug delivery.....	36
3.1 Abstract.....	36
3.2 Introduction.....	37
3.2.1 Multivalent targeted agents for diagnostic imaging and drug delivery.....	37
3.2.2 Development of a viral capsid-based multivalent cell-targeting vehicle.....	38
3.2.3 Development of a targeted MS2-based agent for photodynamic therapy.....	40
3.3 Results and discussion.....	41
3.3.1 Synthesis of a dual-surface modified viral capsid DNA aptamer conjugate.....	41
3.3.2 Cell-binding and internalization of a viral capsid DNA aptamer conjugate.....	44
3.3.3 Development of a multivalent, cancer-targeted, photodynamic agent.....	45
3.4 Conclusion.....	49
3.5 Materials and methods.....	49
2.5.1 General procedures and materials.....	50
2.5.2 Instrumentation and sample analysis.....	50
2.5.3 Experimental.....	51
3.6 References.....	54

CHAPTER 4: Development of a palladium-catalyzed reaction for the conjugation of pyrene to protein surfaces.....	57
4.1 Abstract.....	57
4.2 Introduction.....	58
4.3 Results and discussion.....	60
4.4 Conclusion.....	65
4.5 Materials and methods.....	66
4.5.1 General procedures and materials.....	66
4.5.2 Instrumentation and sample analysis.....	66
4.5.3 Experimental.....	67
4.6 References.....	71

Acknowledgements

There are many friends and colleagues to thank for all kinds of help over the past six years. I am truly grateful.

The first to thank is Matt Francis. As painful as it was to write this dissertation, it gave me the chance to put everything into perspective. Reading through the chapters, I am truly proud of the work we have accomplished. I have also learned so much in the Francis group, having come to Berkeley with no background in chemical biology. However more than science, I have learned a lot about myself. I am thankful to Matt for his optimism, encouragement, understanding, and support.

Moving on to the rest of the group, there are a number of former members I need to thank. Tilley was a great mentor when I first joined the group. I had never done organic synthesis before Berkeley, and I learned a lot from him. Andrew was (and still is) an awesome friend, and someone with whom I have had many meaningful conversations about research and life (and ANTM). Thanks to Nick for his contribution to Chapter 3. I am proud of our collaboration on both the PDT and DNA origami projects. Dante was an enigma, but a fun one. He also gave a great introduction for my GRS.

Now for my class... Sonny taught me how to do cell culture, and was there for my first scientific triumph in the lab. I will always remember the day we threw some stuff together, shot it onto the flow cytometer, and it worked for the first time. When does that ever happen?! Also, he was a formidable table tennis partner. I am grateful to Michel for giving me his ex-girlfriend's futon. It is a piece of furniture that I have kept all six years. In addition to research, I will miss our conversations about life and love... With all the advances in biomedical research, I am hopeful a cure will be discovered for his sleeve allergy. Kanna has been a great labmate, but also a great friend outside of lab. We have had so many good/ridiculous times it's difficult to recall specific stories. But no matter what anybody says, our Vegas trip was fun. Wesley gave me my first chance to introduce someone for their GRS, which was both nerve-wracking and an honor. Also, he has always been generous with his MS2-T19paF-N87C, which is arguably the most valuable thing in the Francis group. The last person from my year is Zac, who pioneered the work involving the engineering of MS2 capsids, and has always been a strong advocate of learning "the basics".

Then came Leah and Chris. Leah has been such a fun person to have in lab, and a great friend to talk to for solid advice. And by "advice", I mean actual advice as well as gossip. She is also the only person in the world who has witnessed me rolling on the floor of Latimer Hall. Chris and I share a similar sense of humor, which made him a great addition to 733. I have enjoyed our numerous lunch outings, our mutual appreciation of yahoo news articles, and our titillating conversation topics. I will always remember the time we went to a bar, got a ride home from random Russians, and then ate all of the pasta out of his parents' pantry.

The next class was a big one. I have always appreciated Allie's scientific insight as well as her wit. In addition, we have shared numerous memorable moments that have involved forehead objects and cheese, among others. Kristen and I will always have our mutual love of Robyn. She also has done great work developing the newest Francis group bioconjugation reaction. Troy is like a walking Wikipedia. If you ever need to know the spawning cycle of sea urchins, he is the one to ask. We also got to fly to Tokyo together on what might have been the best deal I will ever encounter. Then there is Mike, who has been a great person with whom to talk science, and an all-around nice guy. Dan is the co-creator of V-tape®; enough said. And Amy has done really great

stuff modifying microbes. (Also, I will always remember the night we went to Coach Sushi.)

In the next class are Chawita, Stacy, Katherine, Kareem, Jeff, and Farkas (even though you are a postdoc, I think this is when you joined the group). Chawita and I both spent time at the same desk in the sweatshop, and now she is a full-fledged member of 733. She and I have shared many laughs, and she constantly impresses me with her endeavors in both science and business (and the fact that she seems to know every Thai person on Berkeley campus). Stacy has been a great person to work with, and a great staring contest competitor. She is also the only other person in the Francis group working with DNA-MS2, but she has taken it into a new and interesting direction. I will always remember Katherine for her amazing baked goods. I just hope she will remember that Chris and I baked her something special (per request!) when she passed her qualifying exam. Kareem has worked hard to find a good project and I think he has found it. Jeff is from the south. Lastly, Farkas has been a great resource and collaborator in my research. In addition, she has been an amazing sounding board and source of good advice about (especially about jobs) over the last few years.

Then there is Abby. She was the only student to join the group her year, and rounds out the 733 family. She has been a blinding ray of sunshine in 733, she convinced me that socks with boat shoes weren't the worst thing, and she has tolerated all of the conversation topics in 733.

As I finish up, there are a few new members. Jim seems like an appropriate successor to Leah in more ways than one, and I think the "PLP project" is in good hands. Richard has brought both fun and mischief to the lab (e.g. he sent one of the best group emails). For a quiet guy, Tony (a.k.a. Jake) has been quite a surprise. In only a few short months, he has taught me more about body mutilation than I ever wanted to know. Kanwal earned her spot in the lab (i.e. by playing would-you-rather on the way up to Tahoe), and will be a great addition. And I admire the passion that Ioanna brings to both the lab and her MALDIs.

In addition to the Francis group, there are a number of other people I would like to thank from Berkeley. Katie Abole has been an incredible friend and I cannot imagine what my first few years in grad school would have been like without her. I will miss the impromptu conversations in 810 and the foodie moments with Kim Sogi. I also still believe she and I were robbed at the first chem bio bake-off. Thanks to Bryan Dickinson for his expertise and help with the confocal microscopy experiments described in Chapter 3, and for being a great friend. I will miss our outings to the Trappist. Thanks to Jim O'Neil for his help with the Cu-64 radiolabeling experiments described in Chapter 3.

Lastly, I would like to thank all of my family and friends from home. Their unwavering confidence in me, support, and encouragement have been invaluable. In particular, I am grateful to my grandfather, who has been my inspiration throughout graduate school.

CHAPTER 1

DNA-protein conjugates: novel materials for applications in nanobiotechnology

1.1 Abstract

DNA-protein conjugates are the result from the attachment of synthetic nucleic acids to proteins of varying functions, and represent an exciting class of biomaterials with the potential for novel applications. Some examples are provided in this chapter, including applications in bioanalytics, nanoscale self-assembly, and engineered enzyme regulation. This introduction is meant to highlight the types of unique applications made possible through the use of DNA-protein conjugates. Additionally, this chapter serves as a foundation for the work described in subsequent chapters of this dissertation.

1.2 Introduction

In biology, the “central dogma” describes the sequential transfer of genetic information from deoxyribonucleic acid (DNA) to ribonucleic acid (RNA) in a process known as transcription, and from RNA to protein in a process known as translation (Figure 1-1).¹ In this framework, the DNA acts as a “blueprint” for living cells where important genetic information is encoded. And the protein is ultimately the “worker” that is responsible for carrying out tasks specified by the DNA. It is in these roles, that many of us are familiar with DNA and protein – as biomolecules that are critical to survival. However, functionally speaking, DNA and proteins are capable of much more. In order to understand these extrinsic functions, one must look to the structure and some fundamental aspects of these biomolecules.

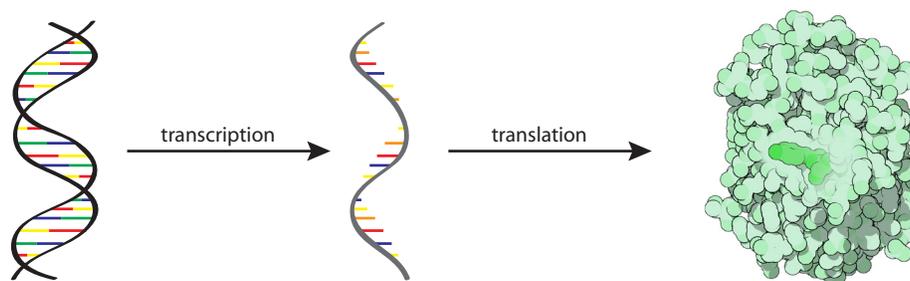


Figure 1-1. The central dogma of molecular biology. Genetic information contained within DNA is first transcribed to messenger RNA. The mRNA transcript may undergo additional processing, at which point the mature mRNA is then translated into protein.

1.2.1 DNA

DNA is a water-soluble biopolymer with alternating deoxyribose and phosphate units forming the backbone of the polymer (Figure 1-2). In addition, there are four bases found in DNA, and the bases are attached at the C1' position on the deoxyribose units. The bases can be divided into two categories: purines and pyrimidines. The purine bases include adenine (A) and guanine (G), and the pyrimidine bases are cytosine (C) and thymine (T). DNA derives its structural diversity from the presence of these four DNA bases, as well as from the order in which they are incorporated. In addition, the structures of the bases allow them to form A/T and G/C base pairs, known as Watson-Crick base pairs, through hydrogen bonding interactions between the bases.² This base-pairing is responsible for the sequence-specificity when two DNA strands hybridize to form a double helix, and is key to many of the applications that utilize DNA. Among these applications are examples involving nano-scale self-assembly,^{3,4} sensing,^{5,6} and cell patterning.^{7,8} DNA base-pairing also allows DNA to form secondary structures, much like RNAs and proteins do, and allows them to be used for molecular recognition. These ssDNAs, called DNA aptamers, can be evolved in an in vitro selection process known as SELEX (Systematic Evolution of Ligands by Exponential Enrichment), and have unique sequences that allow them to specifically bind to target ligands.^{9,10} And these ligands can range from molecules as small as an organic dye, to as large as a macromolecular protein. A notable adaptation of SELEX to living cells, known as cell-SELEX, allows for the identification of DNA aptamers targeted towards whole cells and will be discussed in further detail in chapter 3.¹¹

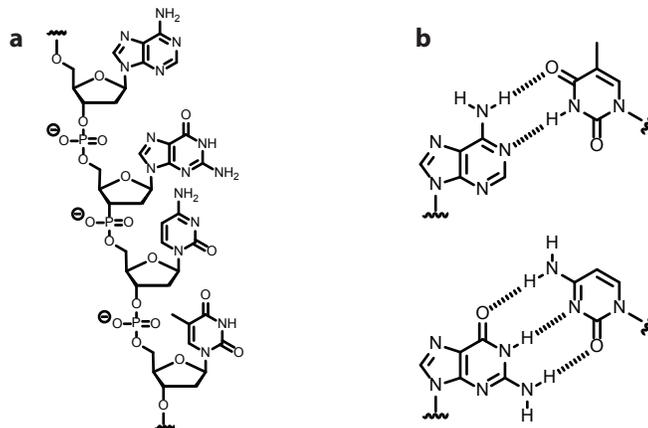


Figure 1-2. DNA structure and base-pairing interaction. **(a)** DNA consists of a backbone of alternating deoxyribose and phosphate units, with the four DNA bases attached to the deoxyribose C1' position. **(b)** DNA base-pairing is largely dictated by hydrogen bond interactions. The base pairs are A/T (top) and G/C (bottom).

1.2.2 Protein

Proteins are biopolymers of amino acids that have been polymerized through the formation of peptide bonds. Each of the 20 standard amino acids (and sometimes, non-standard amino acids) can be incorporated into proteins, and the resulting side chains give rise to the chemical and structural diversity of proteins. This diversity has allowed proteins to evolve to have a wide variety of protein functions. In the case of enzymes, proteins are able to fold to create enzyme active sites and catalyze chemical reactions using precisely positioned amino acid residues.^{12,13} In the case of antibodies or binding proteins, proteins are capable of binding specific antigens or protein partners, where the binding specificity arises from the amino acid composition and the protein's resulting tertiary structure. Another functional purpose for proteins that is of particular relevance to this dissertation is their utility as scaffolds for biomaterials. After all, proteins are mono-disperse, nano-scale molecules of precise chemical and structural composition. Furthermore, the various amino acid side chains that protrude from the surface of the protein can be viewed essentially as functional handles for chemical (or enzymatic) modification.

Viral capsid proteins are one class of proteins that has been employed as a scaffold for synthetic materials.¹⁴⁻²⁰ These proteins form the outer shell of a virus, and serve both to protect the viral genome within the capsid and to facilitate the virus's binding and internalization into cells. In general, viruses have evolved to have a number of desirable qualities that make their capsids attractive scaffolds for materials construction. One such quality is that they minimize the number of sequence-unique protein monomers. In most cases, capsids are composed of multiple copies of one, or a small number of, monomers. By doing so, viruses can replicate more efficiently. For materials considerations, this allows one to minimize the number of bioconjugation reactions needed when functionalizing capsids. Another quality is that they self-assemble into higher-ordered structures – the majority of viruses adopt either icosahedral or helical structures (Figure 1-3). Cowpea chlorotic mottle virus, cowpea mosaic virus, and bacteriophage MS2 are examples of viruses that adopt icosahedral structures, while tobacco mosaic virus and bacteriophage M13 are examples of viruses with helical structures. In either case, these structures are highly regular

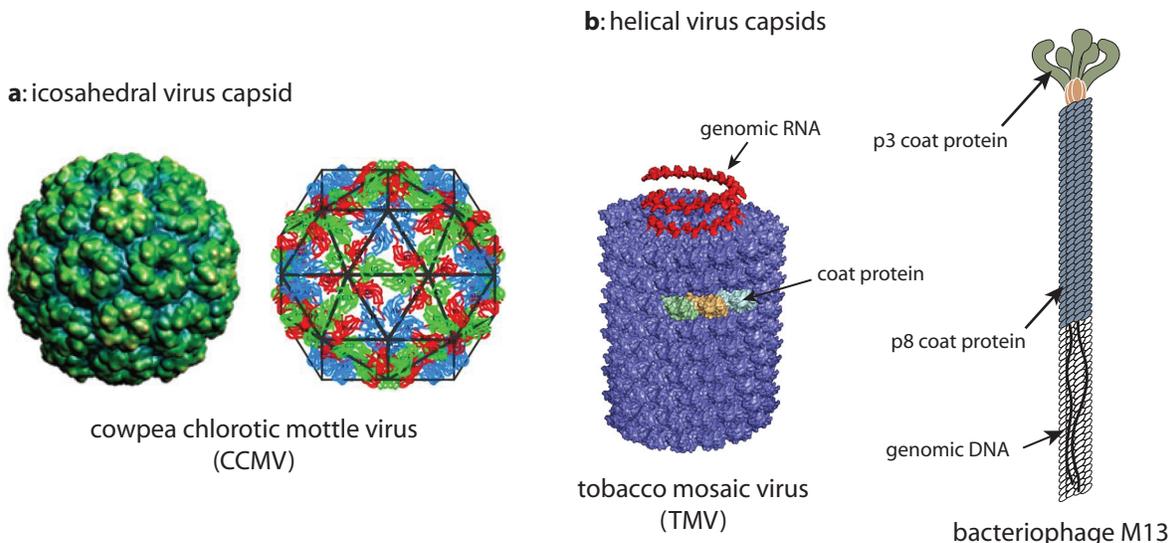


Figure 1-3. The two main virus capsid geometries. **(a)** Icosahedral viral capsids are comprised of 60 protein-containing subunits that self-assemble into a spherical shell around the virus’s genome. In the case of cowpea chlorotic mottle virus (CCMV), each protein subunit contains 3 protein monomers, meaning a total of 180 protein monomers are contained within the CCMV capsid. The schematic on the right shows the arrangement of protein monomers within the 60 triangular subunits that make up the icosahedral capsid. **(b)** Helical capsids assemble around their genomes to form rod-like structures. The tobacco mosaic virus (TMV) and bacteriophage M13 are two examples of viruses that adopt helical structures. Some viruses (TMV) contain multiple copies of a single coat protein, while others (M13) contain multiple copies of different coat proteins. [Stephanopoulos, N. PhD dissertation, UC Berkeley, 2010.]

and symmetric, and are ideal for applications where precise-positioning of chemical modifications may be critical.¹⁸ As a result, viral capsids have been employed as scaffolds for many applications, such as drug delivery, catalysis and light-harvesting.

1.3 Novel applications for DNA-protein conjugates

In light of these functionalities of both DNA and protein, researchers have sought ways to synthesize hybrid DNA-protein conjugates with the goal of developing novel applications.^{21–29} And in fact, there have been a number of reports in the literature where these conjugates have been synthesized and applied in three general areas: bioanalytics, nano-scale self-assembly, and enzyme regulation.

1.3.1 Bioanalytics

Immuno-Polymerase Chain Reaction

One application of DNA-protein conjugates is an antigen-detection system based on the enzyme-linked immunosorbent assay (ELISA). This system couples the antigen specificity of antibodies with the ability for polymerase chain reaction (PCR) amplification of DNA. In ELISAs, antibodies conjugated with an enzyme are used to detect the presence of antigens. If the antigen is present, the antibody-enzyme conjugate will remain bound. Upon addition of substrates for the enzyme, a colorimetric readout will then indicate the presence and relative concentration of the

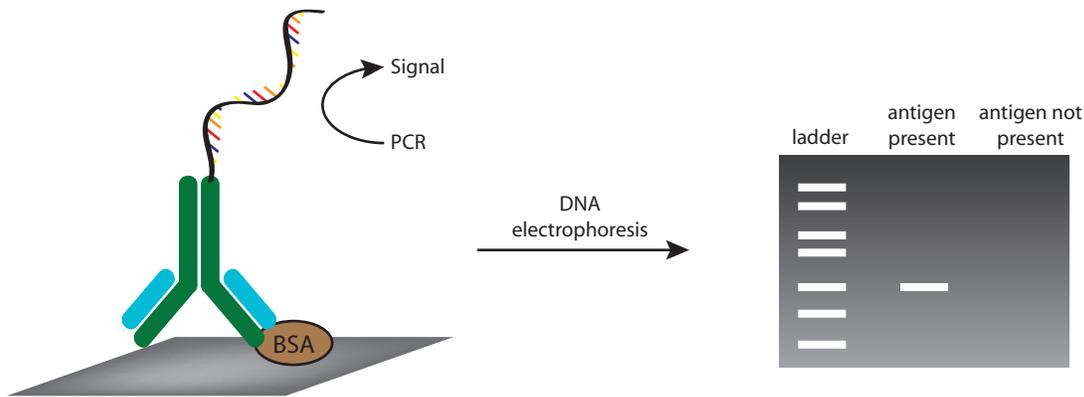


Figure 1-4. Immuno-polymerase chain reaction schematic. In a typical example, an antibody-DNA conjugate is bound to an antigen. The presence of the antigen-antibody-DNA complex is then detected by the presence of a specific PCR product after DNA gel electrophoresis.

antigen in the sample. Similar in concept, immuno-PCR (IPCR) uses an antibody conjugated to DNA to detect the presence of an antigen. Instead of a linear colorimetric readout, however, the antigen is detected by the presence of a PCR product that has been exponentially amplified through PCR (Figure 1-4).

In their initial report, Sano *et al.* used bovine serum albumin (BSA) as a test antigen.³⁰ They incubated α -BSA antibodies that were conjugated to linear plasmid DNA with varying amounts of BSA, and then performed PCR using primers that resulted in a 260-bp PCR product. Using agarose gel electrophoresis, a 260-bp PCR product could be visualized when antigen was present. Upon quantification, the antibody-DNA conjugate showed a clear difference between positive signal vs. background, and the researchers estimated the lower detection limit of this technique to be only 580 antigen molecules. When they compared their IPCR results to an ELISA using the same antibody conjugated to alkaline phosphatase, they determined that IPCR was three to five orders of magnitude higher in sensitivity than ELISA.

Biochips

DNA microarrays are often used for genotyping, studying gene expression profiles, or detecting single nucleotide polymorphisms, and have become an established tool in biomedical research. This is in large part due to microarray fabrication technology that is made possible due to DNA's chemical robustness. Analogous microarray chips for proteins and antibodies are not as easily fabricated due to their fragility. A key requirement is that proteins often need to remain hydrated so that they do not undergo surface-induced denaturation.³¹ So while proteins and antibodies cannot be incorporated into microarrays using the same technology as DNA microarrays, researchers have sought alternative methods. One strategy takes advantage of DNA-directed immobilization as a milder method for incorporating proteins and antibodies onto microarrays. By doing so, the use of DNA microarrays has been expanded to allow for applications such as proteomic studies and screening of pharmaceuticals.³²⁻³⁵

In one example, Becker *et al.* demonstrated the incorporation of DNA-protein conjugates into DNA microarrays for mass spectrometric analysis of protein-protein interactions.³⁶ The protein pair Ras/Ras-binding domain (RBD) was chosen because of its medicinal significance

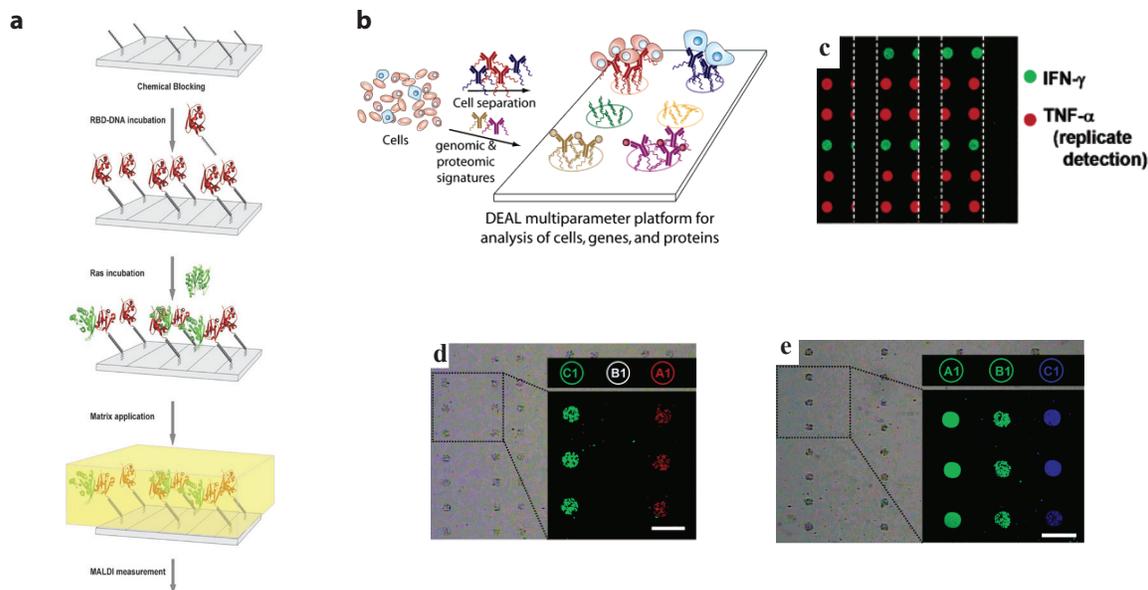


Figure 1-5. Protein and antibody microarrays via DNA-directed immobilization. **(a)** Protein microarrays were prepared for MALDI measurement to analyze protein-protein interactions. **(b)** Antibody microarrays are prepared for co-sorting/detection of gene sequences, proteins, and live cells via DNA-encoded antibody libraries (DEAL). **(c)** DEAL immunoassay shows the fluorescent co-detection of IFN- γ and TNF- α . **(d)** Brightfield and fluorescent microscopy images show separation and localization of T-cells expressing mRFP and B-cells expressing EGFP to discrete spots on a DNA microarray. **(e)** Microscopy images show the co-detection of FITC-labeled cDNA (A1), EGFP-expressing B-cells (B1), and fluorescent TNF- α immunosandwich (C1). [(a) from Becker *et al.*³⁶, (b-e) from Bailey *et al.*³⁷]

(mutations in Ras occur in 30% of all solid human cancers). First, the capture molecule, an RBD-DNA conjugate, was hybridized to a complement strand that was immobilized onto a silicon surface. Next, the surface was incubated with the analyte protein, Ras. A layer of sinapinic acid was then deposited on the microarrays and allowed to dry before analysis by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (Figure 1-5a). Analysis by MALDI showed a peak at 18862 m/z corresponding to the analyte, Ras. The study demonstrated this technique's utility in studying protein-protein interactions, and its utility in analyzing low concentrations of Ras in cell lysates.

In another example, researchers were able to transform a DNA microarray into a multiplexed platform for the co-detection of ssDNA, proteins, and whole cells. Bailey *et al.* were able to do this using DNA-encoded antibody libraries (DEAL). Using DEAL, antibodies were encoded by attachment of DNA oligomers of a specific sequence. The antibodies were then incubated in solution with the sample being analyzed, and then allowed to hybridize to a DNA microarray (Figure 1-5b). Upon hybridization, the sample was separated into its component biomarkers, allowing for multi-parameter analysis. In their report, Bailey *et al.* use both antibodies to proteins of interest for protein detection, as well as antibodies to cell-surface markers for cell sorting.³⁷

In order to demonstrate the detection of multiple proteins, the model proteins IFN- γ and TNF- α were chosen. First, non-fluorescently labeled DNA-encoded 1^o antibodies to the protein antigens were synthesized. Next, these 1^o antibodies were incubated with the antigens as well as a fluorescently-labeled (but not DNA-encoded) 2^o antibody against the antigens. After forming

these antibody-antigen-antibody sandwiches, the sample was allowed to self-assemble spatially to specific locations on a DNA microarray. Using this technique, they were able to visualize the IFN- γ and TNF- α sandwich complexes as separate fluorescent spots at precise locations on the microarray that were encoded by the DNA oligomers (Figure 1-5c).

For the cell sorting application of this platform, the researchers initially chose two immortalized cell types: T cells with a CD90.2 marker that expressed monomeric red fluorescent protein (mRFP), and B cells with a CD45R maker that expressed enhanced green fluorescent protein (EGFP). They incubated DNA-encoded antibodies against CD90.2 and CD45R with a 1:1 mixture of T and B cells, and then introduced the sample to the microarray. As shown in Figure 1-5d, mRFP-expressing T cells localized to spots A1 while EGFP-expressing B cells localized to spots C1, while no non-specific cell localization was observed on spots B1. This method was shown to be specific, and mild enough for sorting primary cells.

The authors demonstrated the sorting and co-detection of cDNA, protein, and cells using their DEAL-based platform. In this case, EGFP-expressing B cells were first localized onto the microarray in the manner described above. Then FITC-labeled DNA and the TNF- α immunosandwich were added to the same microarray and allowed to localize to their respective DNA spots. The microarray was imaged and the resulting brightfield and fluorescence microscopy images verify the detection of all three parameters on a single platform (Figure 1-5e).

1.3.2 Nanoscale self-assembly

Protein nanoarrays

With the discovery of self-assembled DNA nanostructures, a variety of shapes and sizes can be achieved solely from the programmable self-assembly of nucleic acids.^{38,39} These structures are highly programmable, and have proven their utility as scaffolds for nano-scale patterning and self-assembly. Again, the nature of this patterning and self-assembly arises due to Watson-Crick base pairing. Because of its simplicity, ordered and intricate protein complexes can be rationally designed on the nano-scale. One application of these DNA nanoarchitectures is in generating protein nanoarrays.⁴⁰⁻⁴² While protein microarrays like those described above have been useful for studying a variety of processes, there are advantages to miniaturizing them to the nano-scale. One advantage is that methods like atomic force microscopy (AFM) can be used to detect changes in height, shape, roughness, and hydrophobicity – all of which indicate a reaction or interaction on the nanoarray.⁴³

As a proof of concept, many researchers have used streptavidin as a model protein in creating protein nanoarrays. In one example, Kuzuya *et al.* utilized DNA origami – a specific type of DNA architecture that employs one long single strand of viral DNA and many short “staple” strands to generate two- and three-dimensional shapes developed by Paul Rothemund – as the scaffold for their protein nanoarray.^{39,44} They generated two-dimensional DNA origami strips with rectangular holes periodically spaced every 26 nm along the center of the strip. To incorporate proteins onto the DNA scaffold, staple strands modified with biotin adjacent to the rectangular holes were used in generating the scaffold (Figure 1-6). The biotin residues served as handles for streptavidin capture onto the DNA nanoarray due to the extremely strong biotin/streptavidin interaction. Using this strategy, they were able to generate nanoarrays of streptavidin in various patterns.

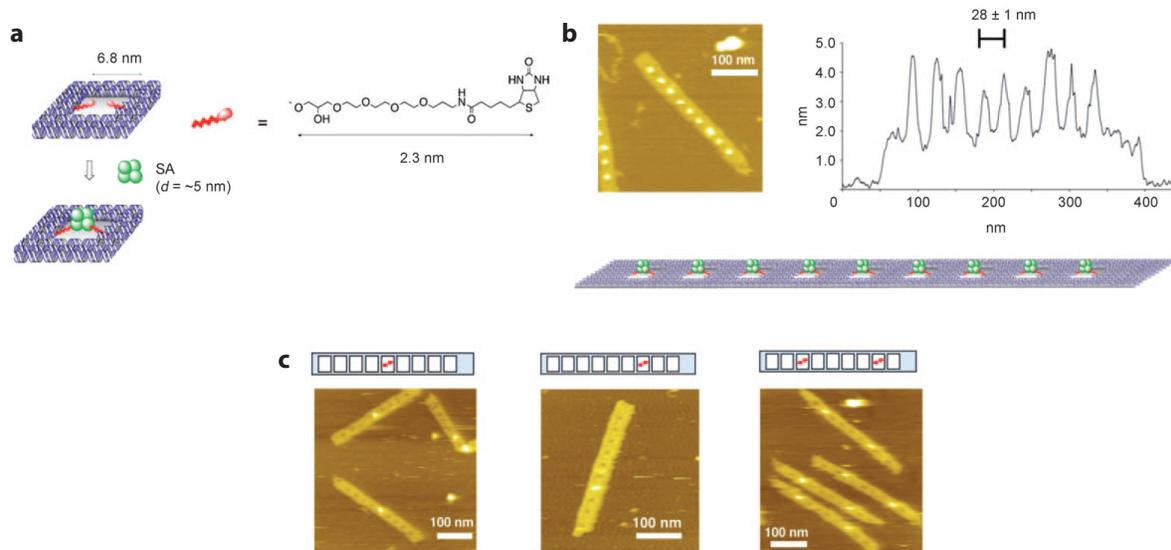


Figure 1-6. DNA origami as a scaffold for streptavidin nanoarrays. **(a)** Two biotin molecules are adjacent to each well in the DNA origami strip and anchor one streptavidin tetramer into each well. **(b)** An AFM image shows eight streptavidin tetramers immobilized on a DNA strip. **(c)** Various patterns of streptavidin can be generated on these nanoarrays by selective incorporation of biotin. [from Kuzuya *et al.*⁴⁴]

Enzyme complexes

Another example of self-assembled nanoarchitectures has been applied to the study of enzyme cascades and cofactor-mediated biocatalysis.⁴⁵ Wilner *et al.* designed self-assembled DNA strips composed of hexagons. In addition, un-hybridized tether DNA strands were incorporated into the DNA strips to allow biomolecules to associate with the strips in a sequence-specific manner. These DNA strips served as a scaffold onto which enzyme-DNA conjugates or cofactor-DNA conjugates were positioned, and could be assembled with differing widths (2- or 4-hexagon wide) to study the distance dependence of enzyme cascades and biocatalysis.

Using these scaffolds, the enzyme pair glucose oxidase (GOx) and horse radish peroxidase (HRP) was studied first. In this pair, GOx oxidizes glucose to gluconic acid, producing H_2O_2 as a byproduct, which then acts as a substrate for HRP in the oxidation of 2'2'-azino-bis[3-ethylbenzthiazoline-6-sulphonic acid], or ABTS²⁻, to the colored ABTS[•] radical. Both GOx and HRP were conjugated to DNA oligomers and hybridized to the two- and four-hexagon scaffolds, and the enzyme kinetics was measured upon addition of glucose (Figure 1-7a). In the absence of the DNA scaffold or in the presence of non-specific DNA, generation of the ABTS[•] radical was negligible. However in the presence of the DNA scaffold, the enzymatic oxidation of ABTS²⁻ was activated. The authors attributed this dramatic catalytic activation to higher local concentrations of H_2O_2 generated in the vicinity of HRP by GOx. Furthermore, the authors observed a 1.2-fold higher rate of oxidation with the two-hexagon scaffold (where GOx and HRP are closer in proximity) than with the four-hexagon scaffold that is very reproducible. They explain this small, but significant difference to the GOx/HRP pair being further apart in the four-hexagon scaffold, which allows more H_2O_2 to diffuse into the bulk solution and away from HRP.

In the same report, the authors applied the same system in order to study an enzyme/cofactor pair. In this example, they chose glucose dehydrogenase (GDH) and its cofactor NAD⁺ as the catalytic pair, and monitored the reaction by the colored compound methylene blue (MB⁺).

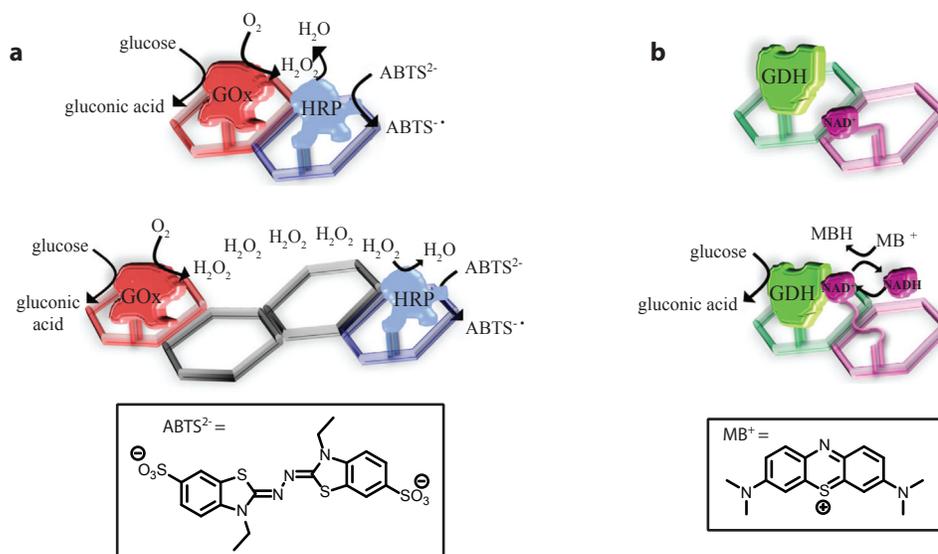


Figure 1-7. Assembly of enzyme complexes on DNA scaffolds. **(a)** GOx/HRP enzyme complexes self-assembled onto two-hexagon (top) and four-hexagon (bottom) scaffolds. **(b)** GDH/NAD⁺ system assembled onto two-hexagon scaffolds. When the NAD⁺ tether is too short (top), the cofactor cannot access GDH, therefore MB⁺ does not get reduced to MBH. When the tether is sufficiently long, the reaction proceeds in a tether length-dependent manner. [from Wilner *et al.*⁴⁵]

GDH oxidizes glucose to gluconic acid only in the presence of NAD⁺, reducing NAD⁺ to NADH. The resulting NADH then reduces MB⁺ to the colorless product MBH, which can be monitored by the disappearance of MB⁺ absorbance. In this case, only the two-hexagon scaffold was used, and the only thing varied was the length of the NAD⁺ tether (Figure 1-7b). The researchers hypothesized that at longer DNA tether lengths, the reaction would proceed more quickly given the accessibility of the NAD⁺ cofactor to GDH. And as they shortened this tether, the reaction would slow down and eventually become blocked due to NAD⁺ not being able to reach GDH. They were able to characterize this distance-dependant catalytic activity, and observed a decreasing reaction rate with decreasing tether length. When a 10-bp tether was used, the reaction rate matched that of the negative controls.

For both of the examples described above, precise spatial positioning was critical in order to observe the distance-dependence of these catalytic reactions. Few scaffolds provide the nano-scale features that DNA nanostructures provide, making them a great choice of scaffold for features in this size range. The DNA-enzyme conjugates that Wilner *et al.* employ are an elegant way of incorporating the catalytic functions of proteins with the nano-scale features of DNA scaffolds in a well-defined and self-assembled manner.

1.3.3 Artificial enzyme regulation

Some naturally occurring enzymes (as well as other classes of proteins) are intrasterically regulated.⁴⁶ In these enzymes, there is an intrasteric autoregulatory sequence near the N- or C-terminus that is dubbed a “pseudosubstrate” because it is similar to enzyme’s substrate and can occupy the active site, but it does not undergo the catalytic reaction. By residing in the active site, pseudosubstrates block the actual substrate from the enzyme active site, thus regulating the

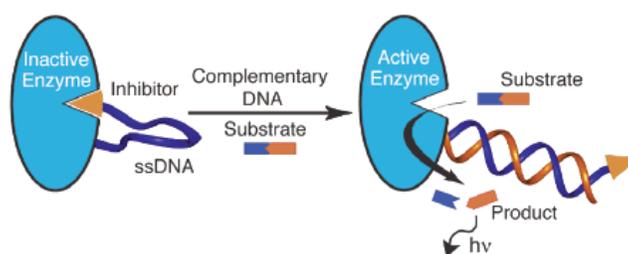


Figure 1-8. Representation of an inhibitor-DNA-enzyme (IDE) construct used for DNA detection. The enzyme is initially intrasterically inactivated by the tethered inhibitor. Upon hybridization of a target strand of complementary DNA to the ssDNA tether, the inhibitor is pulled away from the enzyme active site, and substrate turnover is observed via a fluorescent readout. [from Saghatelian *et al.*⁴⁷]

enzyme's activity. Usually, through allosteric activation, the pseudosubstrate will be pulled away from the active site to re-establish the enzyme's catalytic activity.

Researchers from the Ghadiri group have utilized DNA-protein conjugates to mimic intrasterically regulated enzymes by creating a system where the allosteric regulatory site is substituted by ssDNA and the pseudosubstrate is substituted by an enzyme inhibitor.⁴⁷ The three components, inhibitor-DNA-enzyme (IDE), were covalently linked as shown in Figure 1-8. The idea behind the design of their system is that in the absence of a complementary strand of DNA, the inhibitor will reside within the enzyme active site due to the flexibility and proximity of the ssDNA tether. However upon hybridization to a complementary sequence, the conformation of the DNA tether would be significantly altered such that the inhibitor would be pulled out of the active site and subsequently reconstitute enzyme activity. They applied this system for DNA detection, since enzyme-substrate turnover would inherently provide signal amplification and increase sensitivity for detection purposes.

To begin with, the *Cereus* neutral protease (CNP) and a small-molecule phosphoramidite inhibitor were chosen, and the two were linked together by a 24-mer oligonucleotide. The substrate for their IDE construct was DABCYL- β Ala-Ala-Gly-Leu-Ala- β Ala-EDANS, and was based off of CNP's natural substrate selectivity. The substrate was minimally fluorescent because the EDANS fluorescence was quenched by the proximity of DABCYL. Upon endolytic cleavage, the EDANS portion of the substrate would be released and generate a fluorescent signal. Therefore, CNP activity could be monitored by an increase in EDANS fluorescence.

To assess their IDE construct for sequence-specific DNA detection, the IDE construct was incubated with an oligonucleotide that was complementary to the DNA tether of their construct and measured the EDANS fluorescence over time. It was found that the rate of IDE activity correlated with the concentration of the complementary strand of DNA. Even in the case where a 100 pM concentration of complementary DNA (10 fmol) was used, a 5-fold rate increase over IDE was observed. Both in the case of non-complementary DNA and the use of an enzyme-inhibitor complex with no DNA tether, no signal was observed, which demonstrated sequence-specificity. To test the detection of even lower concentrations of DNA, reaction mixtures were allowed to incubate longer before imaging, allowing for more signal amplification. After extended incubation, they found that they could detect signal over IDE background for a 10 pM sample of DNA. Using the same concept, the Ghadiri group has also elaborated this system for developing molecular logic devices capable of AND, OR, and NOR logic operations.⁴⁸

1.4 Conclusion

This chapter provides some background on the ubiquitous biopolymers, DNA and protein, and highlights the benefits of merging the functionalities of these two biomolecules. Despite their natural origins, DNA and proteins are capable of many functions that are not limited to their biological roles. As a result, DNA-protein conjugates have been utilized in a variety of biomedical and nanobiotechnology applications. As these applications become increasingly complex, there is an increased need for well-defined and site-specific DNA-protein conjugates. The remainder of this dissertation will focus on the development of reactions for the site-specific incorporation of DNA onto proteins, as well as the application of these methods towards the synthesis of novel viral capsid DNA aptamer conjugates.

1.5 References

- (1) Crick, F. *Nature* **1970**, *227*, 561–563.
- (2) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737–738.
- (3) Auyeung, E.; Cutler, J. I.; Macfarlane, R. J.; Jones, M. R.; Wu, J.; Liu, G.; Zhang, K.; Osberg, K. D.; Mirkin, C. A. *Nat. Nanotech.* **2012**, *7*, 24–28.
- (4) Claridge, S. A.; Mastroianni, A. J.; Au, Y. B.; Liang, H. W.; Micheel, C. M.; Fréchet, J. M. J.; Alivisatos, A. P. *J. Am. Chem. Soc.* **2008**, *130*, 9598–9605.
- (5) Tyagi, S.; Kramer, F. R. *Nat. Biotech.* **1996**, *14*, 303–308.
- (6) Lee, J. H.; Wang, Z.; Liu, J.; Lu, Y. *J. Am. Chem. Soc.* **2008**, *130*, 14217–14226.
- (7) Chandra, R. A.; Douglas, E. S.; Mathies, R. A.; Bertozzi, C. R.; Francis, M. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 896–901.
- (8) Hsiao, S. C.; Shum, B. J.; Onoe, H.; Douglas, E. S.; Gartner, Z. J.; Mathies, R. A.; Bertozzi, C. R.; Francis, M. B. *Langmuir* **2009**, *25*, 6985–6991.
- (9) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818–822.
- (10) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505–510.
- (11) Shangguan, D.; Li, Y.; Tang, Z.; Cao, Z. C.; Chen, H. W.; Mallikaratchy, P.; Sefah, K.; Yang, C. J.; Tan, W. *Proc. Natl. Acad. Sci.* **2006**, *103*, 11838–11843.
- (12) Hedstrom, L. *Chem. Rev.* **2002**, *102*, 4501–4524.
- (13) Kester, W. R.; Matthews, B. W. *Biochemistry* **1977**, *16*, 2506–2516.

- (14) Flenniken, M. L.; Uchida, M.; Liepold, L. O.; Kang, S.; Young, M. J.; Douglas, T. *Curr. Top. Microbiol. and Immun.* **2009**, *327*, 71–93.
- (15) Douglas, T.; Young, M. *Science* **2006**, *312*, 873–875.
- (16) Evans, D. J. J. *Mater. Chem.* **2008**, *18*, 3746–3754.
- (17) Fischlechner, M.; Donath, E. *Angew. Chem. Int. Ed.* **2007**, *46*, 3184–3193.
- (18) Miller, R. A.; Presley, A. D.; Francis, M. B. *J. Am. Chem. Soc.* **2007**, *129*, 3104–3109.
- (19) Kovacs, E. W.; Hooker, J. M.; Romanini, D. W.; Holder, P. G.; Berry, K. E.; Francis, M. B. *Bioconjugate Chem.* **2007**, *18*, 1140–1147.
- (20) Hooker, J. M.; Datta, A.; Botta, M.; Raymond, K. N.; Francis, M. B. *Nano Lett.* **2007**, *7*, 2207–2210.
- (21) Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotech.* **2001**, *19*, 636–639.
- (22) Tomkins, J. M.; Nabbs, B. K.; Barnes, K.; Legido, M.; Blacker, A. J.; McKendry, R. A.; Abell, C. *ChemBioChem* **2001**, *2*, 375–378.
- (23) Niemeyer, C. M.; Bürger, W.; Peplies, J. *Angew. Chem. Int. Ed.* **1998**, *37*, 2265–2268.
- (24) Fong, R. B.; Ding, Z.; Long, C. J.; Hoffman, A. S.; Stayton, P. S. *Bioconjugate Chem.* **1999**, *10*, 720–725.
- (25) Niemeyer, C. M.; Koehler, J.; Wuerdemann, C. *ChemBioChem* **2002**, *3*, 242–245.
- (26) Woolley, A. T.; Guillemette, C.; Cheung, C. L.; Housman, D. E.; Lieber, C. M. *Nat. Biotech.* **2000**, *18*, 760–763.
- (27) Taton, T. A.; Mirkin, C. A. *Nat. Biotech.* **2000**, *18*, 713–713.
- (28) Lu, H.; Schöps, O.; Woggon, U.; Niemeyer, C. M. *J. Am. Chem. Soc.* **2008**, *130*, 4815–4827.
- (29) Niemeyer, C. M.; Ceyhan, B. *Angew. Chem. Int. Ed.* **2001**, *40*, 3685–3688.
- (30) Sano, T.; Smith, C. L.; Cantor, C. R. *Science* **1992**, *258*, 120–122.
- (31) Jonkheijm, P.; Weinrich, D.; Schröder, H.; Niemeyer, C. M.; Waldmann, H. *Angew. Chem. Int. Ed.* **2008**, *47*, 9618–9647.
- (32) Niemeyer, C. M. *Chem. Euro. J.* **2001**, *7*, 3188–3195.

- (33) Winssinger, N.; Harris, J. L.; Backes, B. J.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2001**, *40*, 3152–3155.
- (34) Wacker, R.; Niemeyer, C. M. *ChemBioChem* **2004**, *5*, 453–459.
- (35) Boozer, C.; Ladd, J.; Chen, S.; Jiang, S. *Anal. Chem.* **2006**, *78*, 1515–1519.
- (36) Becker, C. F. W.; Wacker, R.; Bouschen, W.; Seidel, R.; Kolaric, B.; Lang, P.; Schroeder, H.; Müller, O.; Niemeyer, C. M.; Spengler, B.; Goody, R. S.; Engelhard, M. *Angew. Chem. Int. Ed.* **2005**, *44*, 7635–7639.
- (37) Bailey, R. C.; Kwong, G. A.; Radu, C. G.; Witte, O. N.; Heath, J. R. *J. Am. Chem. Soc.* **2007**, *129*, 1959–1967.
- (38) Nadrian C, S. *Chemistry & Biology* **2003**, *10*, 1151–1159.
- (39) Rothmund, P. W. K. *Nature* **2006**, *440*, 297–302.
- (40) Bano, F.; Fruk, L.; Sanavio, B.; Glettenberg, M.; Casalis, L.; Niemeyer, C. M.; Scoles, G. *Nano Lett.* **2009**, *9*, 2614–2618.
- (41) Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H. *Science* **2003**, *301*, 1882–1884.
- (42) Chhabra, R.; Sharma, J.; Ke, Y.; Liu, Y.; Rinker, S.; Lindsay, S.; Yan, H. *J. Am. Chem. Soc.* **2007**, *129*, 10304–10305.
- (43) Lee, K.-B.; Park, S.-J.; Mirkin, C. A.; Smith, J. C.; Mrksich, M. *Science* **2002**, *295*, 1702–1705.
- (44) Kuzuya, A.; Kimura, M.; Numajiri, K.; Koshi, N.; Ohnishi, T.; Okada, F.; Komiyama, M. *ChemBioChem* **2009**, *10*, 1811–1815.
- (45) Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I. *Nat. Nanotech.* **2009**, *4*, 249–254.
- (46) Kobe, B.; Kemp, B. E. *Nature* **1999**, *402*, 373–376.
- (47) Saghatelian, A.; Guckian, K. M.; Thayer, D. A.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2002**, *125*, 344–345.
- (48) Gianneschi, N. C.; Ghadiri, M. R. *Angew. Chem. Int. Ed.* **2007**, *46*, 3955–3958.

Site-specific methods for DNA-protein conjugation

2.1 Abstract

While there have been a number of DNA-protein bioconjugation reactions reported in the literature, there is still a need for more reaction development – especially ones that are site-specific. First, this chapter provides a background on the current state of DNA-protein bioconjugation. These methods include noncovalent attachment, covalent attachment through native and non-native amino acids, enzyme-mediated attachment, and conjugation by expressed protein ligation. Then, the development of three new reactions for DNA-protein bioconjugation are described. The first reaction results from an alkoxyamine-bearing DNA, making DNA-protein conjugation possible with carbonyl-containing proteins. And the other two reactions result from the coupling of phenylenediamine- or *o*-aminophenol-modified DNA with aniline-containing proteins in the presence of NaIO₄. All three reactions are site-specific, and expand the current toolkit of conjugation chemistry for DNA and proteins.

Portions of this work have appeared in published form:
J. Am. Chem. Soc. **2009**, *131*, 11174-11178.

2.2 Introduction

2.2.1 Recent advances in methods for DNA-protein bioconjugation

For the types of applications described in Chapter 1, the strategy chosen for synthesizing hybrid DNA-protein conjugates is important. Some factors to take into consideration are covalent or noncovalent attachment of DNA, site-specificity of DNA attachment, whether the protein of interest is amenable to mutagenesis and protein expression, size of the overall conjugate, and others. Therefore, various methods have been developed for coupling DNA to proteins.

Noncovalent attachment

One approach that has been utilized for the noncovalent attachment of DNA to proteins takes advantage of the interaction between biotin and the tetrameric proteins avidin or streptavidin (STV). This interaction is highly specific, and has a dissociation constant that is on the order of 10^{-14} , approaching the strength of a covalent bond. In addition, STV is extremely chemically and thermally stable.¹ Therefore, this method has found use in a number of applications.^{2,3} Using this strategy, STV and biotin are conjugated to a protein of interest and DNA, respectively, and form a conjugate simply by incubation of the two substrates. However an intrinsic problem with this strategy is that STV is a homotetrameric protein, meaning that multiple DNA strands or proteins can be incorporated into the conjugate. This makes controlling the stoichiometry of the coupling difficult. To avoid this problem, genetically engineered monomeric versions of STV have been expressed. Another potential disadvantage is the size of the STV connector, and the resulting size of the conjugate. STV itself is a 60 kDa protein (but is found as a homotetramer), which means that applications where close spatial positioning is important are not amenable to STV-based conjugates.

Another noncovalent method utilizes the affinity of hexahistidine (His_6) tags for Ni(II) ions that are complexed by nitrilotriacetic acid (NTA) ligands.^{4,5} For this approach, the oligonucleotide is modified with the NTA ligand, incubated with Ni(II), and then introduced to a protein of interest containing a His_6 tag. This method is convenient because it avoids the size issue of the biotin/STV method. In addition, proteins are often already expressed with His_6 tags for ease of purification, so this strategy would allow for the conjugation of DNA without additional engineering of the protein of interest. However the conjugates resulting from the Ni-NTA/ His_6 interaction are prone to dissociation and re-association, making them more labile than the biotin/STV conjugates. To address this, Goodman *et al.* synthesized DNA oligonucleotides with one, two, or three NTA groups and showed there was an increase in the strength of the interaction with increasing NTA ligands.⁶

Covalent attachment to native amino acids

As with small molecule bioconjugation reactions, lysine and cysteine are commonly targeted amino acids for DNA attachment to proteins. Lysine modification involves targeting the nucleophilic side chain with *N*-hydroxysuccinimidyl (NHS) esters, and for the purpose of attaching DNA to proteins, various NHS ester-bearing linkers have been synthesized for DNA-modification and subsequent protein bioconjugation.^{7,8} Figure 2-1d shows a few examples of synthetic linkers used for DNA attachment. For cysteine modification, the most common functional groups used are maleimides and pyridyl disulfides, which undergo chemoselective Michael additions and disulfide exchange reactions, respectively, with cysteines.⁹⁻¹⁴ Various linkers have also been synthesized to

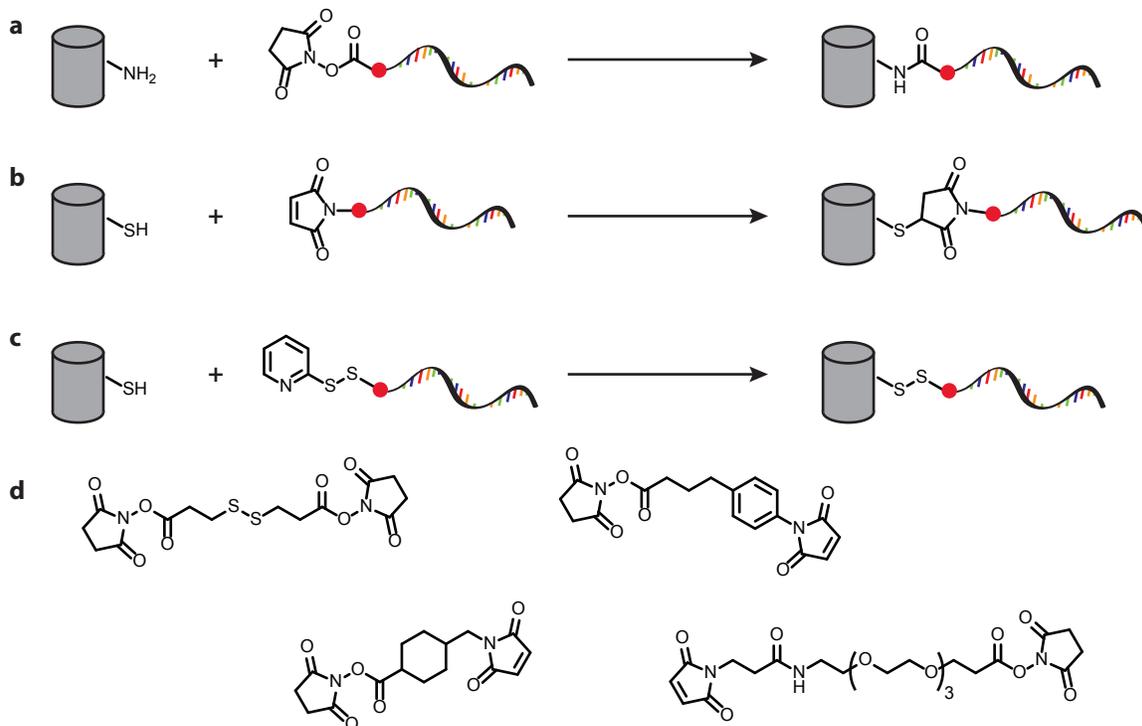


Figure 2-1. DNA conjugation to lysine and cysteine residues. **(a)** Lysine residues are most commonly targeted with NHS esters. Oligonucleotides can be modified with a variety of small-molecule linkers to incorporate a terminal NHS ester. Cysteine residues are most commonly alkylated with maleimides **(b)** or undergo a disulfide exchange with pyridyl disulfides **(c)**. **(d)** Both homobifunctional and heterobifunctional small-molecule linkers, most commonly containing amine- and thiol-reactive functional groups, have been used to crosslink proteins and DNA.

incorporate these functional groups onto DNA for subsequent protein bioconjugation. A number of these linkers are hetero-bifunctional linkers containing both amine- and thiol-reactive moieties, and in these cases, the linker can be used either with amine-DNA to modify cysteines, or thiol-DNA to modify lysines.

Covalent attachment to non-native amino acids

In some cases, it is desirable to modify proteins by targeting non-native functional groups. This strategy can be beneficial if site-specific modification is difficult using natural amino acids (i.e. no unique solvent-exposed cysteines) or if multiple well-defined modifications are desired. In order to use this strategy, a non-native functional group must first be incorporated into the protein of interest.

For the purposes of DNA attachment to proteins, the most common functional group incorporated has been the azide. Azides can be incorporated into proteins after protein translation, during protein translation as non-natural amino acids, or through chemical modification. For example, after protein translation, proteins ending with the consensus sequence CVIA can be incubated with protein farnesyl transferase (PFTase) to append an azido-farnesyl group off of the cysteine residue of the consensus sequence. This strategy has been used to site-specifically incorporate an azide at the C-terminus of GFP.¹⁵ Another way to incorporate azides is through post-translational glycosylation. The Bertozzi group has demonstrated that feeding live cells azide-

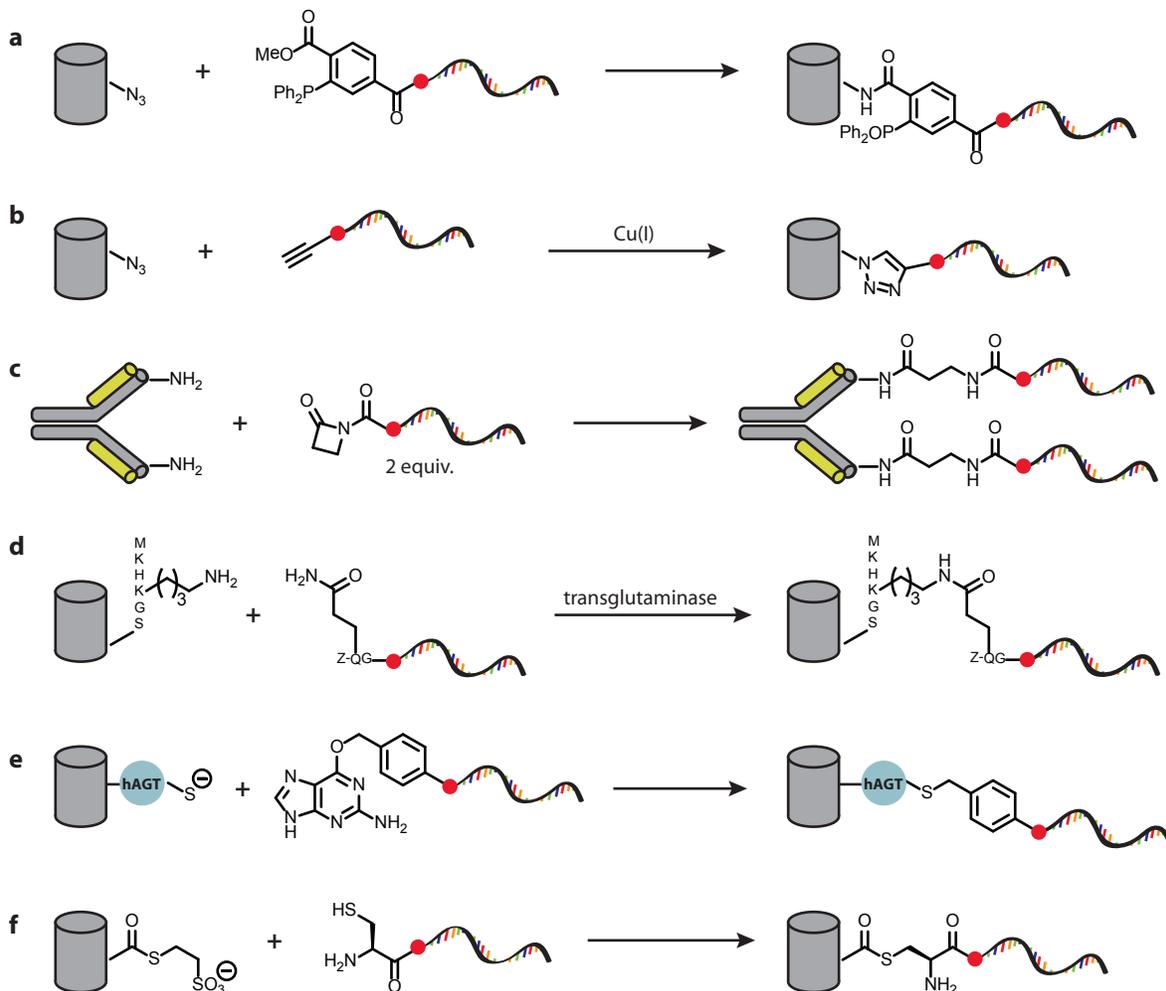


Figure 2-2. Additional strategies for DNA conjugation to proteins. Proteins which have been incorporated with azide moieties can be conjugated to DNA either using the Staudinger ligation (**a**) or the Cu(I)-catalyzed azide-alkyne cycloaddition reaction (**b**). (**c**) The catalytic antibody, 38C2, can be site-specifically labeled with two oligonucleotide strands. (**d**) Transglutaminase catalyzes the crosslinking of a lysine residue on the protein of interest to a glutamine residue incorporated onto the DNA strand. DNA conjugation can also be achieved with fusion proteins. hAGT fusion proteins can be labeled with benzylguanine-modified DNA strands (**e**), and intein fusion proteins yield C-terminal thioesters which can be modified with cysteine-modified DNA strands (**f**).

containing sugars results in the incorporation of azides into glycoproteins.¹⁶ For incorporation during protein translation, the amber stop codon suppression method developed by the Schultz lab and the residue-specific incorporation method developed by the Tirrell lab have been shown to incorporate azide-containing non-natural amino acids, among many others, successfully into many proteins.^{17,18} For chemical modification of proteins with azides, generally NHS esters are most commonly used.

Azide-bearing proteins prepared by these methods have been modified with DNA using both the Staudinger ligation and the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction (Figure 2-2a,b).^{16,19} The Staudinger ligation has proven to be very useful for cell-surface azido-protein labeling; however its reaction rate is not high enough for high levels of protein

modification.²⁰ CuAAC has also been demonstrated for DNA-protein conjugation in a number of examples in the literature.^{15,21,22} The reaction proceeds rapidly and is convenient because it generally does not require a large excess of the modifying substrate. And though it is an oxygen-sensitive reaction, much research has been devoted to optimizing reaction conditions for protein bioconjugation.

Enzyme-based conjugation

Enzyme-based strategies for conjugating DNA to proteins can often yield site-specific bioconjugates. One example is the monoclonal catalytic antibody, 38C2, which has aldolase activity.²³ This aldol activity is due to a particularly nucleophilic lysine residue located in the variable region of the antibody. In aldolase reactions, this lysine residue reversibly forms enamines. However, when incubated with a β -lactam-containing molecule, 38C2 can be irreversibly and covalently attached to the molecule by ring-opening of the β -lactam and amide-bond formation. This concept has been demonstrated with RNA aptamers by the Barbas lab for the synthesis of chemically programmed antibodies.²⁴

Another enzyme-based strategy utilizes transglutaminase enzymes, which catalyzes the acyl transfer reaction between the γ -carboxamide (acyl donor) group of glutamine residues and primary amines (acyl acceptor) such as those from lysine. Using the microbial transglutaminase from *Streptomyces mobaraensis*, Tominaga *et al.* demonstrated the utility of transglutaminase for synthesizing site-specific DNA-protein conjugates.²⁵ They were able to do this by first modifying DNA with *N*-carbonyloxyglutaminyglycine (Z-QG) to introduce the acyl donor onto the DNA. Then a short lysine-containing peptide (MKHKGS), the acyl acceptor, was genetically fused either to the N- or C-terminus of model proteins. In the presence of transglutaminase, the reaction proceeded with good yield (50-60%) and gave rise to a singly-modified protein product.

A more recent method involves the fusion of human *O*⁶-alkylguanine-DNA-alkyltransferase (hAGT) to proteins of interest. hAGT is a self-labeling protein that has been shown to be useful as a general method for labeling proteins of interest with a variety of substrates.²⁶ The modification arises because hAGT irreversibly transfers alkyl groups from its substrate, *O*⁶-benzylguanine, to one of its cysteine residues. This strategy has been demonstrated for DNA-protein conjugation by Jongsma *et al.* using an *O*⁶-benzylguanine-containing oligonucleotide.²⁷

Expressed protein ligation

Proteins can be site-specifically modified at the C-terminus by “expressed protein ligation.” This ligation strategy is based on the ligation of recombinant proteins that contain a C-terminal thioester with compounds containing an N-terminal cysteine.²⁸⁻³⁰ Proteins of interest are expressed as an intein-fusion protein, and cleaved with a small thiol such as 2-mercaptoethanesulfonate (MESNa) to generate a C-terminal thioester. Incubation with a cysteine-containing oligonucleotide results in a transthioesterification, followed by an *S*-to-*N* acyl shift, resulting in the formation of an amide linkage between the protein and DNA.

2.2.2 Development of new methods for DNA-protein bioconjugation

While a number of methods have been developed for DNA-protein bioconjugation, there still remains a need for more methods that are site-specific. This is because the current methods for site-specific DNA-protein conjugation often involve extensive genetic manipulation of the

protein, require high substrate concentrations, or are known to cause protein precipitation. The development of new methods is important because as applications for DNA-protein conjugates become more sophisticated, the substrates required often need to be increasingly well-defined. As an example, for diagnostic protein arrays, the specific orientation of proteins is important for retaining bioactivity, and therefore multiple modifications at multiple sites on a protein can lead to reduced bioactivity.³¹

When developing new methods for DNA-protein conjugation, there are a few desirable criteria. Ideally, the reaction would occur rapidly at relatively low concentrations. This is both because biomolecules are not always stable at high concentrations, and because there is an added expense associated with the use of large excesses of nucleic acids. The reaction would also need to be compatible with the structure and function of both proteins and DNA, meaning undergoing reactions in aqueous buffers and at neutral pH. For site-specific methods, the reaction would result in a single modification at a single site on the protein and yield a stable bioconjugate.

Herein we report three new reactions for attaching DNA to proteins. The first reaction adapts a commonly used bioconjugation strategy, oxime formation, to the attachment of DNA to proteins. For this strategy, commercially-available amine-modified DNA was modified with an allyloxycarbonyl-protected alkoxyamine, which could then be deprotected with a water-soluble palladium catalyst in the presence of a carbonyl-containing protein of interest to form an oxime. The second reaction involves a previously reported NaIO_4 -mediated oxidative coupling reaction that couples phenylenediamines to aniline-containing proteins. This reaction was adapted for DNA-protein conjugation and demonstrated on bacteriophage MS2 as a model protein. The last reaction is an updated version of the oxidative coupling reaction, using *o*-aminophenol instead of phenylenediamine as the reactive partner. This updated version proceeds more rapidly and gives rise to higher yields.

2.3 Results and discussion

2.3.1 Development of alkoxyamine-DNA for oxime formation

Oxime formation between carbonyl-containing proteins and small molecule alkoxyamines is a popular method for modifying biomolecules and has found widespread use in chemical biology.³²⁻³⁴ And due to its widespread applicability, various methods have been developed for

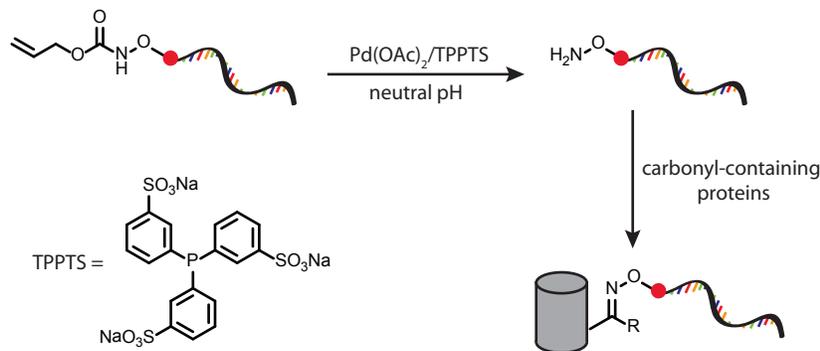


Figure 2-3. Generation of alkoxyamine-DNA with a water-soluble palladium complex. Alloc-protected alkoxyamine-DNA can be deprotected under biomolecule-compatible reaction conditions to generate free alkoxyamine-DNA, which can subsequently undergo oxime formation to form a DNA-protein conjugate.

incorporating non-native carbonyl groups into proteins of interest. Therefore, we sought to develop a method for synthesizing stable, well-defined DNA-protein conjugates via this route.

In order to adapt this strategy for use with commercially available amine-terminated DNA, a synthetic linker containing an NHS ester on one end for DNA functionalization and a protected alkoxyamine on the other end for oxime formation was required. An allyloxycarbonyl (Alloc) protecting group was chosen for this purpose because of a previously reported tyrosine-selective bioconjugation reaction involving π -allylpalladium complexes.³⁵ In this report, Tilley *et al.* described a water-soluble palladium catalyst, pre-formed by reacting Pd(OAc)₂ and triphenylphosphine trisulfonate (TPPTS), that could displace an allylic leaving group to form a π -allylpalladium complex, which could subsequently alkylate tyrosine residues (Figure 2-3). The reactions conditions for this reaction were well-suited to conditions favorable for proteins and DNA (i.e. neutral pH), and thus we proceeded with the Alloc group for masking the alkoxyamine.

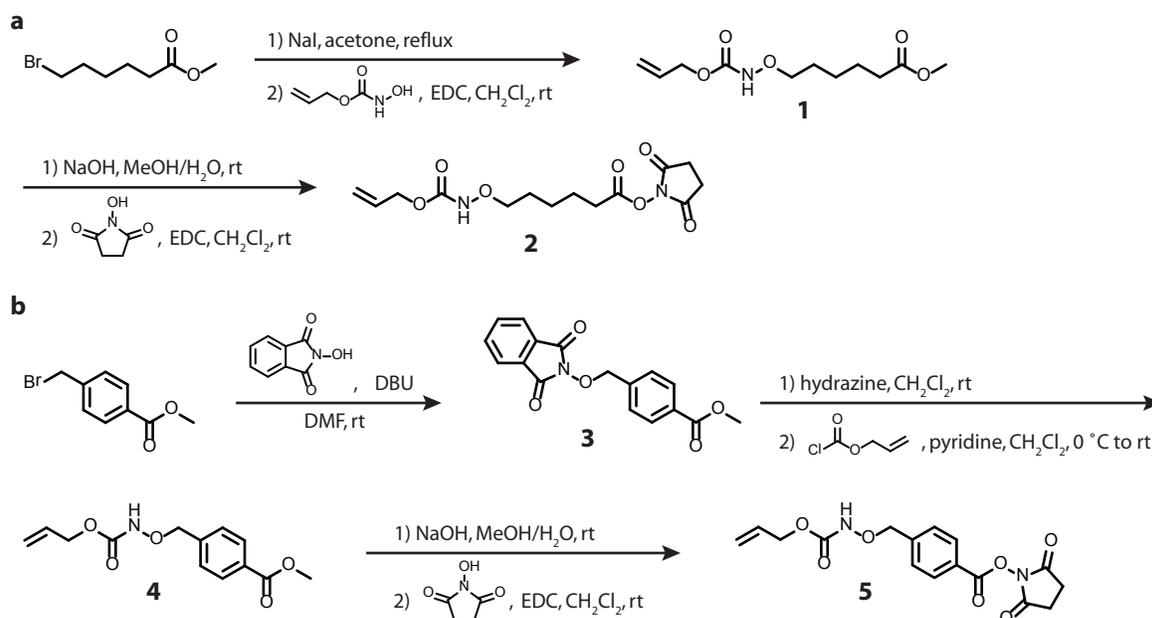


Figure 2-4. Synthesis of heterobifunctional linkers for the incorporation of an Alloc-protected alkoxyamine onto DNA. Two different linkers were synthesized: **(a)** one containing an alkyl spacer, **2**, and **(b)** one containing a benzyl spacer, **5**.

Initially, two linkers were synthesized (Figure 2-4) – one with an alkyl spacer (**2**) and one with a benzyl spacer (**5**). The synthesis of **5** was particularly convenient because the final compound as well as the intermediates were crystalline. **2.A** and **5.A** were generated by acylation of amine-terminated 20-mer DNA oligonucleotide strand **A**, and were then tested for their ability to be deprotected under aqueous reaction conditions. For the deprotection reactions, the modified-DNA was incubated with 2.7 equivalents of the palladium catalyst at room temperature for 2 hours, and the reaction was analyzed by MALDI-TOF mass spectrometry. However, we hypothesized that the deprotected alkoxyamine could react directly with the resulting π -allyl complex, which would render the alkoxyamine un-reactive towards carbonyls. Therefore, we added hydroxylamine into the reaction as a scavenger to reduce this unwanted side-reaction. As shown in Figure 2-5,

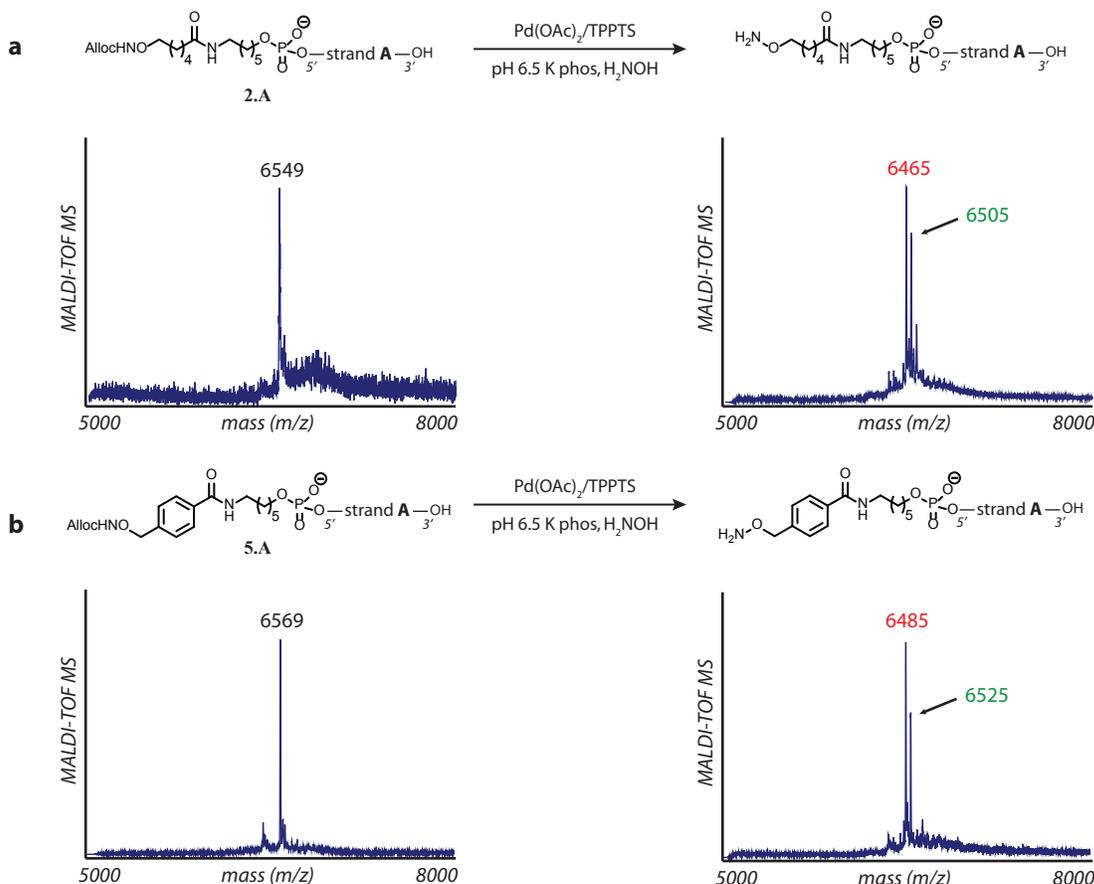


Figure 2-5. Generation of alkoxyamine-bearing DNA. MALDI-TOF MS analysis confirmed the generation of alkoxyamine-DNA from **2.A** (**a**) and **5.A** (**b**) in the presence of Pd(OAc)₂ and TPPTS to yield free alkoxyamine-DNA, as indicated by the ion corresponding to a loss of 84 amu from the protected DNA (red). Alkylation of the free alkoxyamine by the π -allyl fragment was also observed and is indicated by the ion corresponding to a loss of 44 amu (green).

alkoxyamine-DNA was generated upon addition of the palladium catalyst, but alkylation by the allyl fragment was observed as well. But the generation of an observable free alkoxyamine-DNA highlighted the potential for oxime formation with a carbonyl-containing protein.

To test for oxime formation on a protein, ketone-bearing GFP (k-GFP) was chosen as a model protein. The protein was prepared by expressed protein ligation to incorporate a single ketone group at the C-terminus, and was previously reported by our group.³³ Both **2.A** and **5.A** were tested for oxime formation, along with a commercially available “iLinker”-modified DNA, which is believed to contain a hydrazide functional group. For this initial experiment, **2.A** and **5.A** were deprotected *in situ* with the palladium complex and the reaction was allowed to proceed overnight and at 37 °C. Upon analysis by SDS-PAGE followed by Coomassie staining, a single band corresponding to a GFP-DNA conjugate originating from **2.A** and **5.A** was observed, but not for the iLinker-functionalized DNA (Figure 2-6). Furthermore, reaction conversion could be determined using densitometry after Coomassie staining. Since **5.A** appeared to result in a higher yield of the GFP-DNA conjugate over **2.A** and its synthesis was more convenient, we chose to proceed with **5.A** for further experiments.

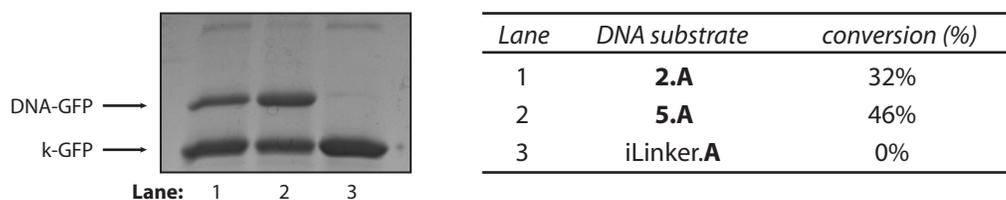


Figure 2-6. Synthesis of a DNA-GFP conjugate *via* oxime formation. SDS-PAGE analysis showed both **2.A** and **5.A** underwent oxime formation after Alloc-deprotection to form a DNA-GFP conjugate. Commercially available iLinker-modified DNA (purported to contain a hydrazide group for condensation with carbonyls) did not form a stable DNA-GFP conjugate. The percent conversion to the DNA-GFP product was determined by optical densitometry.

Next, the concentrations of the **5.A** and the palladium complex were screened to determine their effects on modification yield. The number of equivalents of DNA relative to protein was varied from 5-40 equivalents, and the number of equivalents of the palladium complex relative to DNA was varied from 0.25-4. Figure 2-7 shows the results of the screen. In this experiment, dimedone was added as the π -allyl scavenger (Lanes 1-8). However, Lane 9, which was run under the same reaction conditions as Lane 7 but without any scavenger, showed the highest amount of the GFP-DNA conjugate. Initially, the scavenger was included because of a concern we had that the π -allyl fragment could react with other nucleophilic residues on the protein surface. However, the palladium-catalyzed allylation reaction has only been reported on tyrosine residues at pH 9 or above, whereas these deprotection reactions were performed at pH 6.5. Therefore, the scavenger was omitted in all future experiments as this led to higher conversion.

Following the concentration screen, the use of aniline as a catalyst was investigated as it has been demonstrated in the literature to catalyze oxime formation.³⁶ To test the effects of aniline, 100 mM aniline was added to the reaction buffer as anilinium acetate at pH 6.5. The same protocol

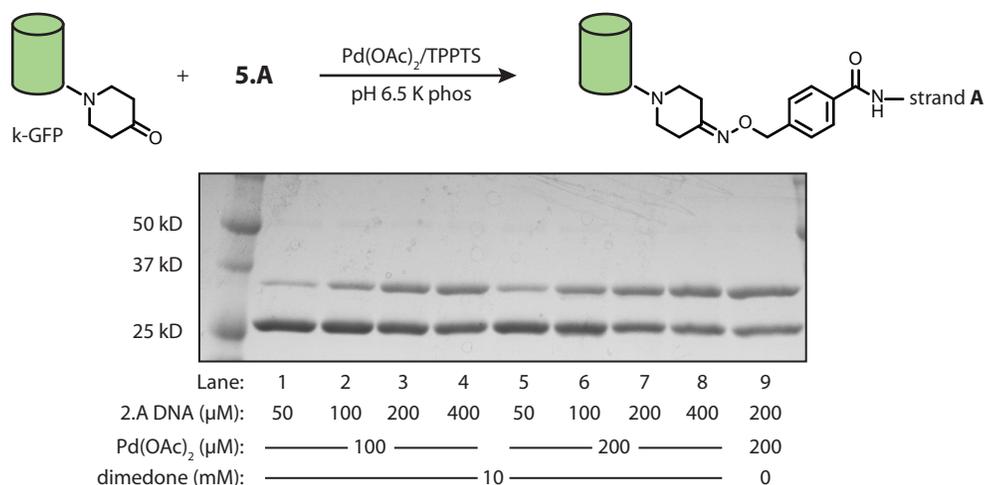


Figure 2-7. Optimization of reaction conditions for **5.A** and k-GFP. The general reaction scheme for the modification of k-GFP with **6.A** in the presence of the palladium catalyst is shown on top. The concentrations of DNA and palladium were varied and reactions were analyzed by SDS-PAGE. Optimal conversion was obtained with 200 μ M **5.A** and 200 μ M Pd(OAc)₂. Additionally, omitting dimedone as a π -allyl scavenger led to higher conversion (lane 9 compared to lane 7).

as previous experiments was used. Unfortunately, SDS-PAGE analysis showed no conversion to the DNA-GFP conjugate. From previous experience in our lab, it was known that k-GFP maintains reactivity in aniline-containing buffers. Therefore we hypothesized that the aniline was poisoning the palladium catalyst, and therefore inhibiting the Alloc deprotection reaction. To test this hypothesis, we deprotected the DNA with the palladium complex separately, and then added the DNA/palladium mixture to the reaction mixture containing k-GFP and aniline. When deprotected *ex situ*, alkoxyamine-DNA was able to form the desired product, demonstrating that this reaction is amenable to aniline catalysis.

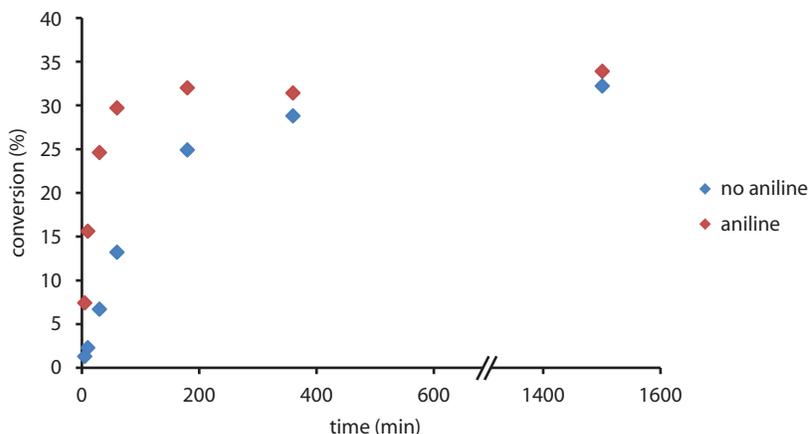


Figure 2-8. Effect of aniline catalysis on k-GFP conversion to the DNA-GFP conjugate. A large excess of the water-soluble aryl aldehyde, 2-formylbenzenesulfonate, was used to quench reactivity at each time point, and samples were analyzed by SDS-PAGE. Conversion to DNA-GFP conjugate was quantified using optical densitometry. The presence of aniline significantly increased the rate of conversion to the DNA-GFP conjugate. However at 1500 min, the percent conversion of reactions run with and without aniline were comparable. All reactions were run at 9 μM k-GFP, 200 μM 5.A, and 200 μM Pd(OAc)₂.

To test the effects of aniline catalysis on the reaction, a time trial was performed. Two reactions were run in parallel – one with and one without aniline in the reaction buffer. At specified time points, the reactions were quenched by the addition of a large excess of 2-formylbenzenesulfonate in order to cap any free alkoxyamine-DNA, and the samples were analyzed by SDS-PAGE. Figure 2-8 shows a significant rate enhancement was indeed observed in the case where aniline was added to the reaction.

These experiments demonstrate the successful conjugation of DNA to proteins via oxime formation. Similar to oxime formation with small molecules, this DNA-protein oxime formation reaction does require longer reaction times. However, this strategy is amenable to aniline catalysis and offers a new way to functionalize carbonyl-containing proteins site-specifically with DNA.

2.3.2 An oxidative approach for coupling DNA to aniline-containing proteins

In a previous publication from our group, a NaIO₄-mediated oxidative reaction that coupled together phenylenediamines and anilines was reported.³⁷ In this reaction, an *N,N*-diethyl-*N'*-acylphenylenediamine coupled rapidly and chemoselectively to anilines on the surface of bacteriophage MS2 (Figure 2-9). We sought to adapt this bioconjugation reaction for coupling

DNA to proteins given that the chemical structure of a native DNA oligonucleotide is inert to NaIO_4 . However, a potential concern was that the DNA bases A, G, and C contain aryl amines and that they might react with the phenylenediamine coupling partner. Therefore, it was important not only to test for conversion to the DNA-protein conjugate, but also that the DNA bases would not be affected by this reaction.

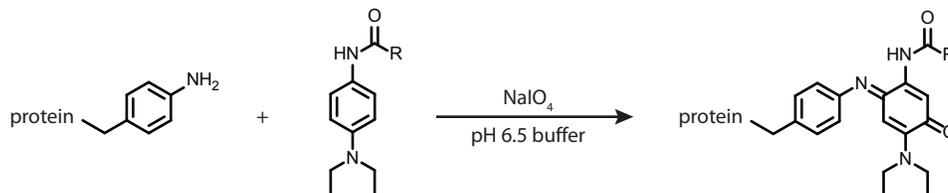


Figure 2-9. A sodium periodate-mediated oxidative coupling strategy for protein bioconjugation. In the presence of NaIO_4 , anilines on proteins can be chemoselectively modified with *N,N*-dialkyl-*N'*-acyl-*p*-phenylenediamines.

For this strategy, the aniline and phenylenediamine coupling partners needed to be incorporated onto protein and DNA, respectively. Again, bacteriophage MS2 was chosen as the protein substrate for investigating this oxidative coupling approach. Our lab previously reported on the incorporation of the aniline coupling partners into the MS2 coat protein in the form of an unnatural amino acid, *p*-aminophenylalanine (*paF*).³⁸ The mutation was introduced at position 19 on the exterior of the MS2 capsid (MS2-T19*paF*) using the amber stop codon suppression method. This afforded MS2 capsids with 180 anilines on the exterior surface. To incorporate the phenylenediamine functionality onto DNA, a heterobifunctional linker containing the phenylenediamine moiety and an NHS ester for amine acylation (**6**) was reacted with amine-terminal strand **A** to generate **6.A**.

We were unable to observe the oxidative coupling of toluidine to **6.A** by MALDI, and therefore we proceeded directly to modifying proteins. In the presence of NaIO_4 , MS2-T19*paF* and **6.A** gave rise to a single DNA-protein conjugate which could be identified from SDS-PAGE by a single gel-shifted band. As a negative control, the same reaction was run using wild-type MS2 (without *paF*) and no DNA attachment was observed, confirming the chemoselectivity and site-specificity of this reaction. It is important to note that any amount of the phenylenediamine linker remaining in solution inhibited the formation of the DNA-protein conjugate, and instead gave rise to only a small amount of MS2 dimer, presumably through aniline-aniline cross coupling.

To optimize the reaction conditions, DNA concentration, periodate concentration, reaction time, and ionic strength were screened (Figure 2-10). The reaction gave the highest conversion at 10-20 equivalents of the phenylenediamine DNA relative to the MS2 concentration, 250 equivalents of periodate, 1 h reaction time, and 150 mM NaCl. The increase in coupling efficiency we observed at higher ionic strengths is likely due to shielding the buildup of negative charge density on the capsid surface as more DNA is attached. With strand **A**, optical densitometry indicated that 32% of the capsid monomers were modified with a single strand of DNA, corresponding to 55 strands of DNA on the surface of each MS2 capsid.

To ensure that the conjugation reaction did not adversely affect either the protein or the DNA, a number of experiments were run (Figure 2-11). To test for capsid integrity, **A**-modified MS2 was imaged by transmission electron microscopy (TEM) and analyzed by dynamic light

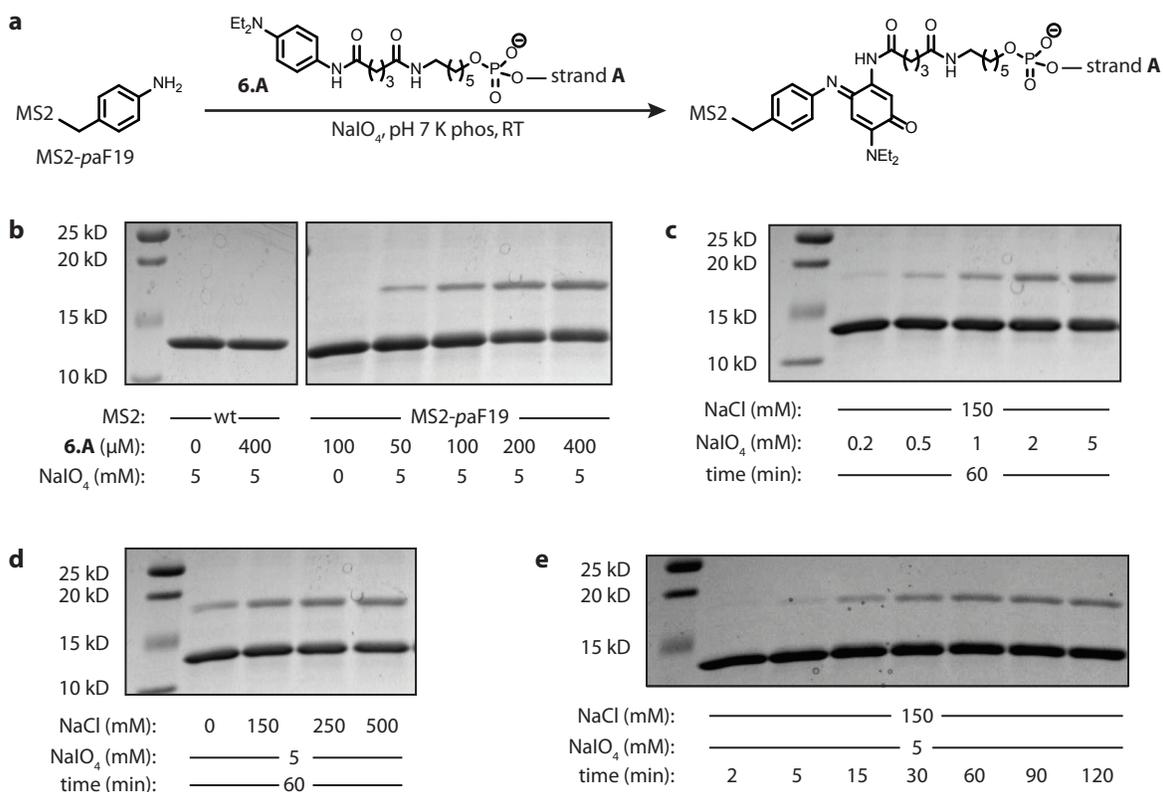


Figure 2-10. Optimizing reaction conditions for the oxidative coupling of **6.A** to MS2-paF19. The general scheme for the reaction between **6.A** and MS2-paF19 is shown in (a). No DNA-MS2 conjugate was observed when wild-type MS2 was used as the protein substrate, or when NaIO₄ was not present (b). However with both MS2-paF19 and periodate, a singly modified DNA-MS2 conjugate was observed. To screen reaction conditions, the concentration of substrate DNA (b), NaIO₄ (c), NaCl (d), and reaction time (e) were varied. Optimal conversion was obtained when 20 μM MS2-paF19 was treated with 400 μM **6.A**, 150 mM NaCl and 5 mM NaIO₄ for 1 h.

scattering (DLS). TEM images show that the capsids remained intact after DNA conjugation, and DLS confirmed this as well. In addition, DLS results showed an increase of 10.5 ± 0.7 nm in the hydrodynamic diameter of the modified capsids. To test for DNA integrity, two hybridization assays were used. First, the ability of the DNA to hybridize was analyzed by using a gel-shift assay. A 40-mer strand of DNA that was complementary to 20-mer strand A and contained an additional 20 adenine bases for an increased electrophoretic shift was incubated with the DNA. Following incubation, SDS-PAGE analysis confirmed DNA hybridization while attached to the MS2 conjugate by a gel-shift. DNA hybridization was also confirmed by DLS, where A-modified capsids that were hybridized to a complementary 20-mer exhibited a hydrodynamic diameter of 42.5 ± 0.9 nm. The difference in diameters between the hybridized DNA-MS2 conjugate and unmodified MS2 was 14.4 nm, which corresponds precisely to the 3.6 nm per 10 base pairs reported in the literature for solution structures of DNA double helices.³⁹ These data indicated that the MS2 capsids and DNA remained intact and functional under the reaction conditions.

In addition to MS2, this reaction has also been demonstrated on aniline-containing eGFP. Again, the protein substrate was prepared by expressed protein ligation as previously reported.³⁷

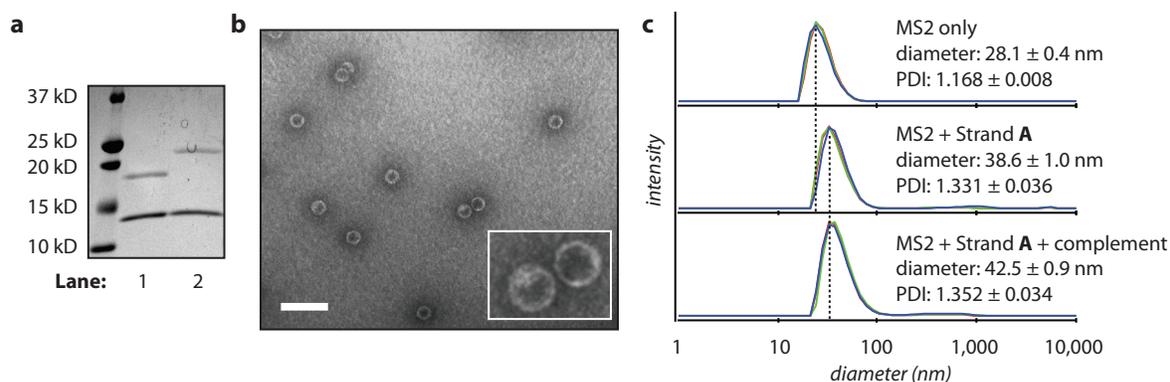


Figure 2-11. Analysis of protein structure and DNA base-pairing after exterior attachment of DNA to MS2. **(a)** DNA base-pairing after conjugation to MS2 was confirmed by a gel-shift assay. Lane 2 was incubated with a complementary sequence to strand A with an additional 20 adenine bases for increasing the electrophoretic shift, while lane 1 was not incubated with additional DNA. Transmission electron micrograph images **(b)** and dynamic light scattering analysis **(c)** showed MS2 capsids remained intact after DNA attachment. In addition, DLS showed an increase in the diameter after conjugation of strand A, as well as an additional increase upon incubation with a 20-base complementary sequence. The scale bar in the TEM image represents 100 nm.

The conversion to the DNA-GFP conjugate was significantly higher using optimized reaction conditions (55%), which is likely due to the fact that eGFP is a monomeric protein, which eliminates the electrostatic repulsion introduced when attaching multiple DNA strands onto the MS2 capsid. And the reaction conditions are mild enough such that the DNA-GFP conjugate maintained its fluorescence. Overall, this oxidative coupling reaction has been shown to modify proteins to good yield within 1 h, and is compatible with the structures and functions of proteins and DNA.

2.3.3 Oxidative coupling of DNA to proteins *via o*-aminophenols

After the oxidative coupling reaction between anilines and phenylenediamines was published, a follow-up reaction describing the coupling of anilines and *o*-aminophenols was reported.⁴⁰ In this updated version, *o*-aminophenols were able to form a stable, covalent adduct with anilines within 2 min (Figure 2-12). The reaction conditions were similar to those used in the oxidative coupling of phenylenediamines and anilines. Therefore, we sought to test this oxidative coupling reaction for the conjugation of DNA to MS2.

MS2-T19*paF* was used as the protein substrate again, however the *o*-aminophenol moiety

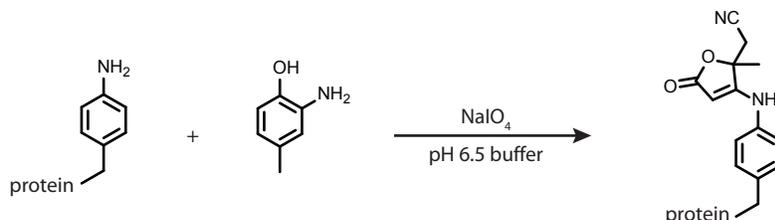


Figure 2-12. Rapid coupling of anilines to *o*-aminophenols in the presence of sodium periodate. This reaction has been demonstrated for protein bioconjugation, and can reach high conversion within 5 min.

needed to be incorporated onto synthetic DNA. To do this, strand **A** was functionalized with a linker containing an *o*-nitrophenol moiety on one end and an NHS ester for amine acylation on the other (**7**). Following acylation, the *o*-nitrophenol on the DNA was reduced with sodium dithionite to yield the corresponding aminophenol, **7.A**. The DNA-MS2 conjugate was then generated by reacting **7.A** with MS2-T19paF in the presence of NaIO₄. Under the same reaction conditions previously optimized for the oxidative coupling of phenylenediamines, the aminophenol-DNA reacted with MS2-T19paF more quickly and to higher conversion to yield a single DNA-MS2 conjugate by SDS-PAGE analysis (Figure 2-13). This resulted in the attachment of ~100 strands of DNA to the MS2 capsid in under five minutes.

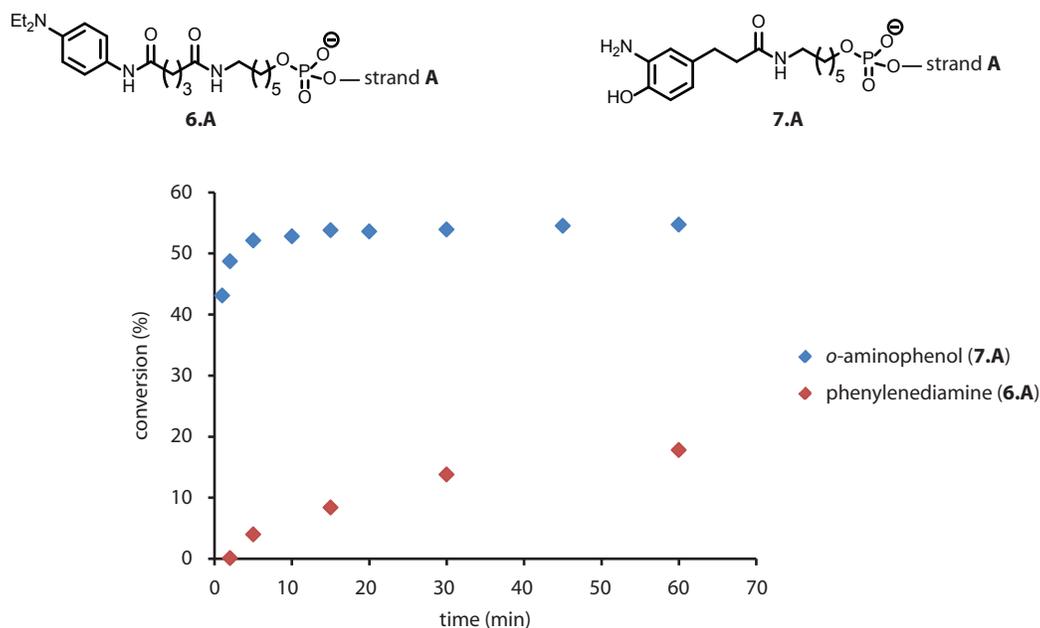


Figure 2-13. Time course comparison of *o*-aminophenol- and phenylenediamine-functionalized DNA. The structures of the DNA substrates are shown above. The time course experiment is shown below, where reactions were run in parallel, and were quenched at each time point by the addition of an excess of TCEP. Analysis by SDS-PAGE showed conversion to the DNA-MS2 conjugate was significantly higher and faster with **7.A** than with **6.A**.

This oxidative coupling strategy using *o*-aminophenols has a number of advantages. Not only does it proceed rapidly and to high yield, it is more compatible with a number of functional molecules that may be sensitive to NaIO₄. For example, our lab has observed that some fluorophores are increasingly quenched due to NaIO₄ exposure. The significantly reduced reaction times required for the *o*-aminophenol coupling reaction would mitigate this NaIO₄-dependent quenching. In addition, the higher coupling yields obtained from this reaction meant that fewer equivalents of synthetic DNA were required to achieve conversion comparable to that of the coupling of phenylenediamines.

2.4 Conclusion

This chapter seeks to expand the toolkit of methods for DNA-protein bioconjugation. To this end, three reactions are described. The first reaction involves the generation of an alkoxyamine-bearing DNA, which is amenable to oxime formation with carbonyl-containing proteins. While this reaction requires longer reaction times, it can be catalyzed by the addition of aniline, and takes advantage of a number of methods for carbonyl incorporation into proteins. Next, an oxidative coupling strategy between phenylenediamine-DNA and aniline-containing proteins is described. This reaction goes to good yield within 1 h, and is mild enough to maintain the structure and function of both protein substrates tested as well as the DNA. Lastly, an updated oxidative coupling reaction involving anilines and *o*-aminophenols was adapted to DNA-protein bioconjugation, which gave rise to shorter reaction times and higher conversion to the desired conjugate.

2.5 Materials and methods

2.5.1 General procedures and materials

Unless otherwise noted, all chemicals and solvents were of analytical grade and used as received from commercial sources. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F254 plates with visualization by ultraviolet (UV) irradiation at 254 nm or potassium permanganate stain. All organic solvents were removed under reduced pressure using a rotary evaporator. Dichloromethane (CH₂Cl₂) and tetrahydrofuran (THF) were distilled under a nitrogen atmosphere from calcium hydride. Water (dd-H₂O) used in biological procedures or as the reaction solvent was deionized using a NANOpure purification system (Barnstead, USA). 4-(4-diethylamino-phenylcarbamoyl)-butyric acid succinimidyl ester (**6**) was prepared using the previously reported method.³⁸ 3-(4-hydroxy-3-nitrophenyl)propanoic acid succinimidyl ester (**7**) was prepared using the previously reported method.⁴⁰ Allyl hydroxycarbamate was prepared using the previously reported method.⁴¹ k-GFP and MS2-T19paF were prepared using previously reported methods.^{33,38}

All oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). Samples were purified by reverse-phase HPLC or NAP-5 gel filtration columns (GE Healthcare). Samples were lyophilized using a LAB CONCO Freezone 4.5 (Lab Conco). Lyophilized oligonucleotides were re-suspended in the appropriate buffer and the concentration was determined by measuring the absorbance at 260 nm. The sequence of strand A reads 5'–TCATACGACTCACTCTAGGGA–3'

2.5.2 Instrumentation and sample analysis

NMR. ¹H and ¹³C spectra were measured with a Bruker AVQ-400 (400 MHz) spectrometer. Chemical shifts are reported as δ in units of parts per million (ppm) relative to chloroform-d (δ 7.26, s). Multiplicities are reported as follows: s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublets), p (pentet), m (multiplet), br (broadened), or app (apparent). Coupling constants are reported as a J value in Hertz (Hz). The number of protons (n) for a given resonance is indicated nH, and is based on spectral integration values.

Mass Spectrometry. Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DETM system (PerSeptive Biosystems, USA). Prior to MALDI-TOF MS analysis, samples were desalted using C18 ZipTip®

pipettips (Millipore, USA). Oligonucleotide samples were co-crystallized using a 3-hydroxypicolinic acid:ammonium citrate solution (45 mg/mL:5 mg/mL in 4.5:5.5 MeCN:ddH₂O). All MS data for oligonucleotides and protein samples were found to be within 0.1% of the expected values.

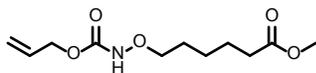
High Performance Liquid Chromatography (HPLC). HPLC was performed on an Agilent 1100 Series HPLC System (Agilent Technologies, USA). Sample analysis for all HPLC experiments was achieved with an inline diode array detector (DAD). Both analytical and preparative reverse-phase HPLC of oligonucleotides was accomplished using a C18 stationary phase and a MeCN/100 mM triethylammonium acetate (TEAA, pH = 7.0) gradient.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus from Bio-Rad (Hercules, CA), following the protocol of Laemmli. All protein electrophoresis samples were heated for 10 minutes at 100 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of any disulfide bonds. Gels were run for 5 minutes at 30 V and 70-90 minutes at 120 V to ensure separation of bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad).

Dynamic Light Scattering. DLS measurements were obtained using a Malvern Instruments Zetasizer Nano ZS, usage courtesy of Professor Jean Fréchet. Data plots and standard deviations are calculated from an average of three measurements, each of which consists of 10 runs of 45 seconds each. Measurement data are presented as an intensity plot, which weights larger dimensions by a factor of 106 more than smaller dimensions. Samples were taken in 10 mM pH 7.0 phosphate buffer.

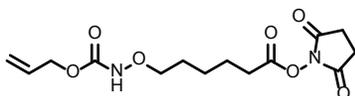
Transmission Electron Microscopy (TEM). TEM images were obtained at the UC-Berkeley Electron Microscope Lab (www.em-lab.berkeley.edu) using a FEI Tecnai 12 transmission electron microscope with 100 kV accelerating voltage. TEM grids were prepared by charging carbon-coated, formvar-supported copper mesh grids with argon plasma (40 mA at 0.1 mbar for 30 s) in a Cressington 108 Auto Sputter Coater. Protein samples were prepared for TEM analysis by pipetting 5 µL samples onto these grids and allowing them to equilibrate for 3 minutes. The samples were then wicked with filter paper and rinsed with ddH₂O. The grids were then exposed to 5 µL of a 1% (w/v) aqueous solution of uranyl acetate for 90 s as a negative stain. After excess stain was removed, the grid was allowed to dry in air.

2.5.3 Experimental



Methyl 6-(allyloxycarbonyl-aminoxy)hexanoate (1). Methyl 6-bromohexanoate (4.00 g, 19.1 mmol) and NaI (5.73 g, 38.3 mmol) and 50 mL of acetone were added to a 100 mL round bottom flask and stirred and refluxed. After refluxing for 4 h, the solvent was removed under reduced pressure and the residue was re-dissolved in EtOAc. The organic layer was washed twice with water, once with brine, and then dried over sodium sulfate. The product was purified by silica gel chromatography and the resulting methyl 6-iodohexanoate was carried onto the next step. Next NaH (0.40 g, 10.0 mmol) was added to a flame-dried 50 mL round bottom flask under nitrogen atmosphere. 15 mL of THF was added and the flask was placed in an ice bath. Allyl hydroxycarbamate (1.17 g, 10.0 mmol) was separately dissolved in 5 mL of THF and was added to the NaH suspension, resulting in the evolution of hydrogen gas. The reaction mixture was allowed

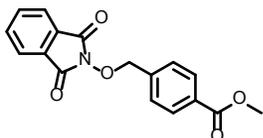
to warm to room temperature for 5 min, before methyl 6-iodohexanoate (2.14 g, 8.4 mmol) was added as a solution in 5 mL of THF. The reaction was allowed to stir at room temperature overnight. The reaction mixture was then concentrated under reduced pressure and EtOAc and water were added. Some 1 M HCl was added to separate the layers. The organic layer was then washed once with brine, dried over sodium sulfate, and purified by silica gel chromatography. The product was isolated as a pale yellow oil (0.65 g, 16% over two steps). $^1\text{H NMR}$ (400 MHz, CDCl_3) δ (ppm) 7.36 (s, 1H), 5.92 (m, 1H), 5.33 (dd, 1H), 5.24 (dd, 1H), 4.64 (d, 2H), 3.88 (t, 2H), 3.66 (s, 3H), 2.32 (t, 2H), 1.64 (m, 4H), 1.42 (m, 2H).



6-(allyloxycarbonyl-aminoxy)hexanoic acid succinimidyl ester

(2). 1 (0.12 g, 0.50 mmol) was dissolved in 2.5 mL of MeOH, and then 2.5 mL of an aqueous 1 M NaOH solution was added to the reaction mixture. The reaction was stirred at room temperature for

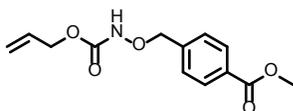
3 h. The pH of the reaction mixture was then carefully neutralized to pH 5.5-6 with concentrated HCl. The MeOH was removed under reduced volume and the remaining aqueous solution was flash-frozen and lyophilized before being carried onto the next step. The resulting solid was dissolved in 2 mL of dichloromethane and NHS (86.3 mg, 0.75 mmol) was added. A suspension of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (115.0 mg, 0.60 mmol) in 4 mL of dichloromethane was added to the reaction mixture, and the reaction was stirred at room temperature for 3 h. The organic layer was washed once with water, once with saturated sodium bicarbonate, and then dried over sodium sulfate. The product was purified by silica gel chromatography to yield a clear and colorless oil (56.5 mg, 34% over two steps). $^1\text{H NMR}$ (400 MHz, CDCl_3) δ (ppm) 7.45 (s, 1H), 5.92 (m, 1H), 5.33 (dd, 1H), 5.25 (dd, 1H), 4.64 (d, 2H), 3.89 (t, 2H), 2.83 (br, 4H), 2.62 (t, 2H), 1.80 (m, 2H), 1.67 (m, 2H), 1.52 (m, 2H).



O-phthalimidyl-4-(hydroxymethyl)benzoic acid methyl ester

(3). *N*-hydroxyphthalimide (0.98 g, 5.50 mmol) was added to a flame-dried scintillation vial under nitrogen atmosphere, and dissolved in 6 mL anhydrous DMF. DBU (0.82 mL, 5.50 mmol) was added to the reaction mixture, resulting in a dark orange solution. Methyl 4-(bromomethyl)

benzoate (1.15 g, 5.00 mmol) was dissolved in 4 mL of anhydrous DMF and the solution was added to the reaction mixture. A white precipitate formed within seconds. After 10 min, TLC showed the complete disappearance of starting material. The DMF was removed under reduced pressure, and the resulting residue was dissolved in dichloromethane. The organic layer was washed twice with water, once with saturated sodium bicarbonate, and dried over sodium sulfate. TLC showed only one spot, therefore the solvent was removed under reduced pressure and a white solid was collected (1.41 g, 90%). $^1\text{H NMR}$ (300 MHz, CDCl_3) δ (ppm) 8.05 (d, 2H), 7.81 (m, 2H), 7.74 (m, 2H), 7.62 (d, 2H), 5.26 (s, 2H), 3.92 (s, 3H).

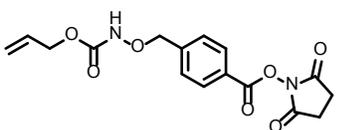


4-(N-allyloxycarbonyl-aminoxy-methyl)benzoic acid methyl ester

(4). 3 (0.53 g, 1.70 mmol) was dissolved in 15 mL dichloromethane, and hydrazine monohydrate (0.41 mL, 8.50 mmol) was added to the reaction mixture. The reaction was stirred at room temperature and a

white precipitate formed in less than a minute. The reaction was allowed to stir for 1 h and then filtered through glass wool. The filtrate was dried over sodium sulfate, and passed through a plug

of silica gel, eluting with EtOAc. The solvent was removed under reduced pressure yielding the free alkoxyamine as a white solid that was carried on to the next step. The alkoxyamine (0.30 g, 1.63 mmol) was dissolved in 8 mL of dichloromethane, and then pyridine (0.14 mL, 1.71 mmol) was added. The reaction mixture was placed in an ice bath. Allyl chloroformate (0.18 mL, 1.71 mmol) was added to the reaction mixture and a white precipitate formed immediately. The reaction mixture was taken out of the ice bath and allowed to warm to room temperature, at which point the precipitate dissolved. Reaction was allowed to stir at room temperature for 16 h. The reaction mixture was washed once with 1M aqueous sodium bisulfate, once with water, once with brine, and then dried over sodium sulfate. The product was isolated as a white solid (0.36 g, 80% over two steps). ¹H NMR (400 MHz, CDCl₃) δ (ppm) 8.04 (d, 2H), 7.46 (d, 2H), 7.31 (s, 1H), 5.91 (m, 1H), 5.33 (dd, 1H), 5.25 (dd, 1H), 4.94 (s, 2H), 4.65 (d, 2H), 3.93 (s, 3H).



4-(N-allyloxycarbonyl-aminoxy-methyl)benzoic acid succinimidyl ester (5). **4** (0.36 g, 1.37 mmol) was dissolved in 7 mL of MeOH. The resulting solution was dissolved but slightly murky. 7 mL of aqueous 1 M NaOH was added to the reaction and the reaction mixture was stirred at room temperature. After 30 min, the MeOH was removed

under reduced pressure and 1M sodium bisulfate was added to the remaining aqueous layer. The addition of sodium bisulfate (aqueous layer was ~ pH 2) caused a white precipitate to form. The suspension was extracted with dichloromethane five times, and the combined organic layers were dried over sodium sulfate. The solvent was removed under reduced pressure yielding the free acid as a white solid that was carried onto the next step. The acid and NHS (0.24 g, 2.05 mmol) were suspended in 5 mL of dichloromethane and stirred under a nitrogen atmosphere. EDC (0.32 g, 1.65 mmol) was added to the reaction mixture as a suspension in 10 mL of dichloromethane. Upon addition of EDC to the reaction mixture, everything dissolved and the reaction mixture became clear. The reaction was stirred at room temperature under a nitrogen atmosphere for 2 h. The organic layer was washed once with water, once with saturated sodium bicarbonate, once with brine, and dried over sodium sulfate. The solvent was removed under reduced pressure and purified by silica gel chromatography to yield a clear viscous oil (0.24 g, 49% over two steps). ¹H NMR (400 MHz, CDCl₃) δ (ppm) 8.14 (d, 2H), 7.54 (d, 2H), 7.35 (s, 1H), 5.91 (m, 1H), 5.33 (dd, 1H), 5.26 (dd, 1H), 4.97 (s, 2H), 4.65 (d, 2H), 2.91 (s, 4H).

General Procedure for the modification DNA oligonucleotides with NHS ester linkers.

DNA oligonucleotides were purchased containing a primary amine on the 5'-end. A typical reaction was as follows: DNA at a concentration 300 μM was reacted with an NHS ester-containing heterobifunctional linker (60-120 eq) in a 1:1 solution of DMF and 50 mM pH 8.0 phosphate buffer. The reaction mixture was briefly vortexed and then allowed to react at rt for 2 h. Either RP-HPLC or commercially available gel filtration columns was used to purify the small molecule from DNA, following the commercially provided protocol. Following purification, the DNA-containing solution was lyophilized and then re-suspended in the desired buffer. Concentration was determined by measuring the absorbance at 260 nm.

Preparation of catalyst/ligand solution for deprotection reactions. Commercially available TPPTS contained 10% phosphine oxide as determined by ³¹P NMR, and a stoichiometric amount of DMSO as determined by ¹H NMR. Taking these two impurities into account, the mass of TPPTS

was measured such that 5 equivalents of unoxidized TPPTS was mixed with Pd(OAc)₂. Generally, 3 mg of Pd(OAc)₂ and 48 mg of Strem TPPTS were added to a vial and dissolved in 1.67 mL of distilled water. The mixture was briefly sparged with nitrogen and then sonicated until the mixture became homogeneous. The homogeneous solution was then sparged again for about 30 min. The final solution contains 8 mM Pd(OAc)₂ and 40 mM TPPTS. The catalyst was either stored under argon as small aliquots in sealed ampules for future use, or prepared before an experiment and used immediately.

General procedure for deprotection reaction. **2.A** or **5.A** (300 μM) were incubated with hydroxylamine (3-33 eq) as a scavenger and the catalyst solution (800 μM) in a total volume of 10 μL. The reaction was briefly vortexed and allowed to proceed at room temperature for 2 h. Reactions were analyzed by MALDI-TOF MS.

General procedure for protein modification with 5.A. A 0.6-mL Eppendorf tube was charged with k-GFP (10 μM), **5.A** (50-200 μM), dimesone (10 mM or none), and lastly the palladium catalyst solution (100-200 μM). If necessary, 25 mM pH 6.5 potassium phosphate buffer was added to reach a final volume of 10 μL (this adjustment was calculated and done before catalyst solution was added). The reaction was briefly vortexed, and allowed to stand at room temperature for 18-20 h. Samples were analyzed by SDS-PAGE.

Procedure for aniline catalysis time trial with 5.A. k-GFP (17.1 μL from a 32 μM stock) was diluted in 100 mM pH 6.5 sodium phosphate buffer (30 μL) either containing 200 mM aniline or no aniline at all. Separately, **5.A** (11.4 μL from a 1050 μM stock) and the catalyst solution (1.5 μL from an 8 mM stock) were premixed, briefly vortexed, and allowed to stand at room temperature for 15 min. After 15 minutes, the DNA/catalyst solution was added to the k-GFP solutions and allowed to react at room temperature. At specified time points, 8 μL of the reaction mixture was quenched by the addition of 2-formylbenzenesulfonate sodium salt (1 μL from a 500 mM stock). After all time points were quenched, the samples were analyzed by SDS-PAGE.

General procedure for conjugation of 6.A to MS2. An Eppendorf tube was charged with MS2-T19paF (20 μM), phenylene diamine-containing oligonucleotide (200-400 μM), and NaIO₄ (5 mM). The reaction was carried out in 50 mM pH 7.0 phosphate buffer containing 150 mM NaCl. The reaction was briefly vortexed and allowed to react at rt for 1 h. After an hour, for a 50 μL reaction, the reaction was quenched by the addition of 5 μL of 500 mM tris(2-carboxyethyl) phosphine hydrochloride (TCEP). For purification, the sample was first buffer-exchanged by gel filtration (NAP-5) into the desired buffer. The excess DNA was then removed by successive centrifugal filtration using 100k molecular weight cutoff filters (Millipore).

General procedure for generation of 7.A. DNA oligonucleotides were purchased containing a primary amine on the 5'-end. A typical reaction was as follows: 250 μL of a solution of strand A (600 μM stock) in 50 mM pH 8.0 potassium phosphate buffer was added to 250 μL of a DMF solution containing 3-(4-hydroxy-3-nitrophenyl)propanoic acid succinimidyl ester (**7**, 24 μL from a 0.4 M stock in DMSO). The reaction was briefly vortexed and allowed to stand at room temperature for 2 h. Occasionally, a precipitate slowly formed over the course of the reaction. If so, the reaction was filtered to remove precipitate. The reaction mixture was then passed through

a NAP-5 gel filtration column equilibrated with water, resulting in a 1 mL solution containing the acylated DNA. Sodium dithionite (45 μ L from a 20 mg/mL stock dissolved in 200 mM pH 6.5 potassium phosphate) was added to the 1 mL DNA solution, briefly vortexed, and allowed to stand at room temperature for 15 min. The reaction mixture was passed through a NAP-10 gel filtration column equilibrated with water, and the eluate was flash-frozen and lyophilized to yield a white solid. The solid was re-suspended in the desired buffer at a low volume (high concentration) and stored as a stock solution at -20 °C.

Procedure for oxidative coupling time trial with 6.A and 7.A. A 50 μ L reaction was prepared containing MS2-T19paF (20 μ M), **6.A** or **7.A** (200 μ M), NaCl (150 mM) and sodium periodate (5 mM). The reaction was briefly vortexed, and allowed to stand at room temperature. At specified time points, 5 μ L was taken out of the reaction mixture and quenched with an excess of TCEP (1 μ L from a 330 mM stock that was pH adjusted to \sim pH 7). After all time points were quenched, the samples were analyzed by SDS-PAGE.

2.6 References

- (1) Green, N. M. *Adv. Protein Chem.* **1975**, *29*, 85–133.
- (2) Sano, T.; Smith, C. L.; Cantor, C. R. *Science* **1992**, *258*, 120–122.
- (3) Niemeyer, C. M.; Adler, M.; Pignataro, B.; Lenhert, S.; Gao, S.; Lifeng, C.; Fuchs, H.; Dietmar, B. *Nucl. Acids Res.* **1999**, *27*, 4553–4561.
- (4) Meredith, G. D.; Wu, H. Y.; Allbritton, N. L. *Bioconjugate Chem.* **2004**, *15*, 969–982.
- (5) Shen, W.; Zhong, H.; Neff, D.; Norton, M. L. *J. Am. Chem. Soc.* **2009**, *131*, 6660–6661.
- (6) Goodman, R. P.; Erben, C. M.; Malo, J.; Ho, W. M.; McKee, M. L.; Kapanidis, A. N.; Turberfield, A. J. *ChemBioChem* **2009**, *10*, 1551–1557.
- (7) Hsiao, S. C.; Shum, B. J.; Onoe, H.; Douglas, E. S.; Gartner, Z. J.; Mathies, R. A.; Bertozzi, C. R.; Francis, M. B. *Langmuir* **2009**, *25*, 6985–6991.
- (8) Strable, E.; Johnson, J. E.; Finn, M. G. *Nano Lett.* **2004**, *4*, 1385–1389.
- (9) Gianneschi, N. C.; Ghadiri, M. R. *Angew. Chem. Int. Ed.* **2007**, *46*, 3955–3958.
- (10) Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotech.* **2001**, *19*, 636–639.
- (11) Niemeyer, C. M.; Sano, T.; Smith, C. L.; Cantor, C. R. *Nucl. Acids Res.* **1994**, *22*, 5530–5539.
- (12) Fong, R. B.; Ding, Z.; Long, C. J.; Hoffman, A. S.; Stayton, P. S. *Bioconjugate Chem.* **1999**, *10*, 720–725.

- (13) Fruk, L.; Müller, J.; Weber, G.; Narváez, A.; Domínguez, E.; Niemeyer, C. M. *Chem. Euro. J.* **2007**, *13*, 5223–5231.
- (14) Bano, F.; Fruk, L.; Sanavio, B.; Glettenberg, M.; Casalis, L.; Niemeyer, C. M.; Scoles, G. *Nano Lett.* **2009**, *9*, 2614–2618.
- (15) Duckworth, B. P.; Chen, Y.; Wollack, J. W.; Sham, Y.; Mueller, J. D.; Taton, T. A.; Distefano, M. D. *Angew. Chem. Int. Ed.* **2007**, *46*, 8819–8822.
- (16) Saxon, E.; Bertozzi, C. R. *Science* **2000**, *287*, 2007–2010.
- (17) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498–500.
- (18) Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235–237.
- (19) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596–2599.
- (20) Chandra, R. A.; Douglas, E. S.; Mathies, R. A.; Bertozzi, C. R.; Francis, M. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 896–901.
- (21) Gupta, S. S.; Kuzelka, J.; Singh, P.; Lewis, W. G.; Manchester, M.; Finn, M. G. *Bioconjugate Chem.* **2005**, *16*, 1572–1579.
- (22) Humenik, M.; Huang, Y.; Wang, Y.; Sprinzl, M. *ChemBioChem* **2007**, *8*, 1103–1106.
- (23) Wagner, J.; Lerner, R. A.; Barbas, C. F. *Science* **1995**, *270*, 1797–1800.
- (24) Wuellner, U.; Gavriilyuk, J. I.; Barbas III, C. F. *Angew. Chem. Int. Ed.* **2010**, *49*, 5934–5937.
- (25) Tominaga, J.; Kemori, Y.; Tanaka, Y.; Maruyama, T.; Kamiya, N.; Goto, M. *Chem. Commun.* 401–403.
- (26) Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. *Nat. Biotechnology* **2002**, *21*, 86–89.
- (27) Jongsma, M. A.; Litjens, R. H. G. M. *Proteomics* **2006**, *6*, 2650–2655.
- (28) Becker, C. F. W.; Wacker, R.; Bouschen, W.; Seidel, R.; Kolaric, B.; Lang, P.; Schroeder, H.; Müller, O.; Niemeyer, C. M.; Spengler, B.; Goody, R. S.; Engelhard, M. *Angew. Chem. Int. Ed.* **2005**, *44*, 7635–7639.
- (29) Lovrinovic, M.; Seidel, R.; Wacker, R.; Schroeder, H.; Seitz, O.; Engelhard, M.; Goody, R. S.; Niemeyer, C. M. *Chem. Commun.* **2003**, 822–823.

- (30) Burbulis, I.; Yamaguchi, K.; Gordon, A.; Carlson, R.; Brent, R. *Nature Methods* **2004**, *2*, 31–37.
- (31) Christman, K. L.; Broyer, R. M.; Tolstyka, Z. P.; Maynard, H. D. *J. Mater. Chem.* **17**, 2021–2027.
- (32) Scheck, R. A.; Francis, M. B. *ACS Chem. Biol.* **2007**, *2*, 247–251.
- (33) Esser-Kahn, A. P.; Francis, M. B. *Angew. Chem. Int. Ed.* **2008**, *47*, 3751–3754.
- (34) Carrico, I. S.; Carlson, B. L.; Bertozzi, C. R. *Nat. Chem. Bio.* **2007**, *3*, 321–322.
- (35) Tilley, S. D.; Francis, M. B. *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081.
- (36) Dirksen, A.; Hackeng, T. M.; Dawson, P. E. *Angew. Chem.* **2006**, *118*, 7743–7746.
- (37) Hooker, J. M.; Esser-Kahn, A. P.; Francis, M. B. *J. Am. Chem. Soc.* **2006**, *128*, 15558–15559.
- (38) Carrico, Z. M.; Romanini, D. W.; Mehl, R. A.; Francis, M. B. *Chem. Commun.* **2008**, 1205–1207.
- (39) Nelson, D. L.; Cox, M. M. *Principles of Biochemistry; 4th ed.*; W.H. Freeman and Company: New York, **2005**.
- (40) Behrens, C. R.; Hooker, J. M.; Obermeyer, A. C.; Romanini, D. W.; Katz, E. M.; Francis, M. B. *J. Am. Chem. Soc.* **2011**, *133*, 16398–16401.
- (41) Greck, C.; Bischoff, L.; Ferreira, F.; Genet, J. P. *J. Org. Chem.* **1995**, *60*, 7010–7012.

Viral capsid DNA aptamer conjugates for cell-targeted imaging and drug delivery

3.1 Abstract

Using orthogonal bioconjugation reactions, we demonstrated the dual surface functionalization of bacteriophage MS2 in order to generate multivalent cell-targeting vehicles. Using our dual surface modification strategy, we have successfully incorporated imaging and therapeutic molecules on the interior of MS2, and DNA aptamers for targeting purposes on the exterior. Aptamer-modified capsids possessing interior fluorophores were used to verify cell-targeting and internalization. In addition, targeted positron emission tomography (PET) agents were produced by radiolabeling capsids modified on the interior with chelating agents for Cu-64. Lastly, we describe a multivalent aptamer-MS2 photodynamic agent that specifically directs singlet oxygen generation to targeted cells. This agent is shown to be selectively kill these cells in a mixture containing both receptor-positive and receptive-negative cells.

Portions of this work have appeared in published form:

J. Am. Chem. Soc. **2009**, *131*, 11174-11178.

ACS Nano **2010**, *4*, 6014-6020.

3.2 Introduction

3.2.1 Multivalent targeted agents for diagnostic imaging and drug delivery

With the advent of molecular medicine, there is a wealth of knowledge regarding specific molecular signatures, or biomarkers, related to individual diseases. As a result, there is great interest in the ability to create biomedical agents that target these biomarkers for the purposes of both detecting diseases at an earlier stage and reducing the side-effects of cytotoxic drugs. For diagnostic imaging, these agents would target a specific biomarker that is over-expressed on the surface of disease-related cells and result in the accumulation of signal only at the site where the biomarker is present. In contrast, one of the imaging agents currently used for the detection of cancer is F-18-fluorodeoxyglucose (FDG). FDG is an analog of glucose, and its use in imaging is based on its accumulation in highly metabolic cells. While high metabolism is correlated with cancer, it is not exclusive to cancer. In addition, active targeting would provide insight into the types of cancer cells present by virtue of their biomarkers. For drug delivery purposes, a goal would be to actively target drugs to diseased tissues and avoid the exposure of cytotoxic drugs to healthy tissues.

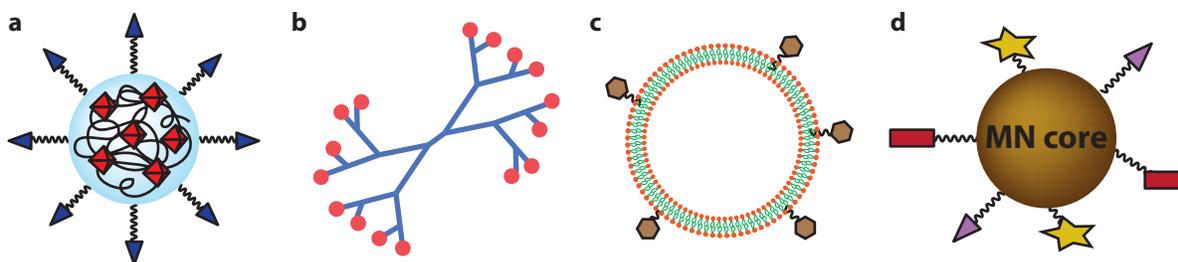


Figure 3-1. Core scaffolds that have been used for the construction of multivalent nanoscale delivery vehicles. **(a)** Spherical polymer nanoparticles have been used to encapsulate platinum (IV) prodrug complexes (red octahedra) by nanoprecipitation, and they have been endowed with targeting capabilities by the conjugation of aptamers on the exterior (blue triangles).³ **(b)** Dendrimers are branched, synthetic polymers with a variety of possible nanoarchitectures. They have been used in a variety of drug delivery applications due to their multivalency for drug conjugation (red circles). **(c)** Liposomes are vesicles composed of a lipid bilayer and are capable of encapsulating both hydrophilic and hydrophobic components. They have been endowed with targeting capabilities by the conjugation of molecules such as folic acid (brown hexagon).⁶ **(d)** Inorganic nanoparticles, such as magnetic nanoparticles (MN), have been used as scaffolds for functionalization with near-infrared fluorophores (yellow stars), siRNA (purple triangles), and membrane translocation peptides (red rectangles).⁸

Biomedical agents that are targeted and multivalent have additional advantages. For targeting purposes, a multivalent display of biomarker-binding groups has the potential to impart a lower apparent dissociation constant, which would result in improved binding properties. For imaging and therapeutic agents, multivalency has the potential to increase signal and drug payload, respectively. To take advantage of these properties, a number of core scaffolds to which targeting groups and payload molecules have been attached have been used (Figure 3-1), including polymers,¹⁻³ dendrimers,^{4,5} liposomes,⁶ and inorganic nanoparticles.⁷⁻⁹ The most common targeting groups that have been explored include folic acid,^{10,11} cobalamin,^{12,13} carbohydrates,¹⁴ peptides and

antibodies,^{15,16} and nucleic acid aptamers.¹⁷ These particular systems have the added advantage of possessing nanoscale dimensions. Compared to small molecules, agents in the size range of tens to hundreds of nanometers have been shown to exhibit longer circulation times, and often have greater access to tumor vasculature due to the enhanced permeation and retention (EPR) effect.¹⁸

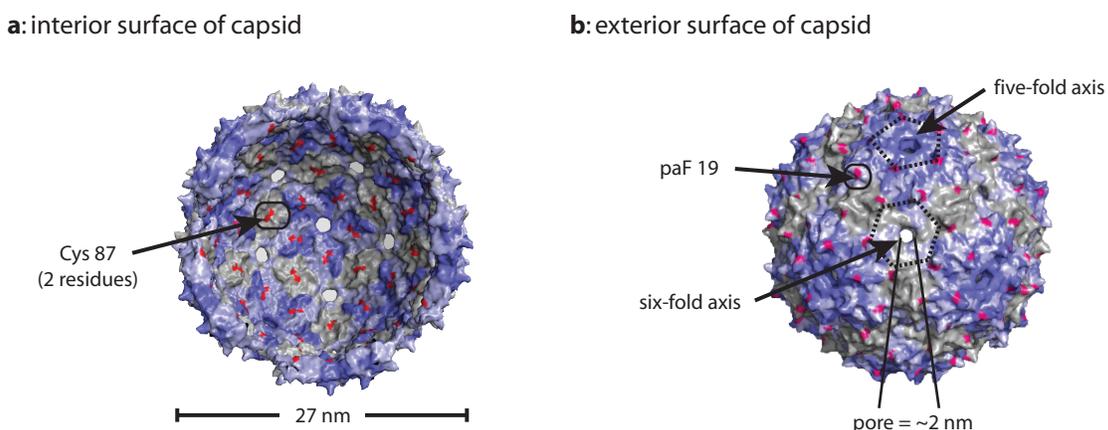


Figure 3-2. Bacteriophage MS2 as a self-assembled nanoscale scaffold. The viral capsid of bacteriophage MS2 self-assembles from 180 sequence-identical monomers into a hollow, spherical protein shell with a 27 nm diameter. Access to the interior is afforded by 32 pores that are ~2 nm across. By mutating two of the amino acid residues present on the capsid, it is possible to functionalize the two surfaces of the capsid orthogonally. In this chapter, the interior of the capsid will be modified at (a) Cys 87, and the exterior of the capsid will be modified by targeting (b) the unnatural amino acid *paF19*.³⁹

3.2.2 Development of a viral capsid-based multivalent cell-targeting vehicle

We chose bacteriophage MS2 as the scaffold for our multivalent cell-targeting construct. The coat protein of MS2 consists of 180 sequence-identical monomers that self-assemble into a hollow spherical structure.^{19,20} The resulting spherical capsid is nontoxic, robust, biodegradable, and easily expressed in *E. coli*. In addition, there are 32 pores that allow access to the interior of the capsid, which makes functionalization of the interior and exterior surfaces of MS2 possible (Figure 3-2). In previous reports, our lab has demonstrated the use of MS2 as a scaffold for the incorporation of F-18 PET tracers,²¹ Gd-based magnetic resonance imaging contrast agents,^{22,23} and the mitotic inhibitor paclitaxel.²⁴

Following the work described in Chapter 2, we chose DNA aptamers to be the receptor-binding groups for our MS2-based construct. DNA aptamers are oligonucleotides that bind to specific molecular targets with affinities comparable to (or in some cases better than) that of antibodies. However, DNA aptamers are much smaller in size and can be readily synthesized using automated solid-phase synthesis techniques (IgG antibodies ~150 kDa, 40-nt DNA aptamer ~13 kDa). Their amenability to solid-phase synthesis allows for the incorporation of modified bases and backbones in order to improve serum stability or impart novel functionalities (Figure 3-3).²⁵⁻²⁷ In addition, with the development of SELEX (systematic evolution of ligands by exponential enrichment) and diversification techniques such as error-prone PCR, hypermutagenic PCR,²⁸ and

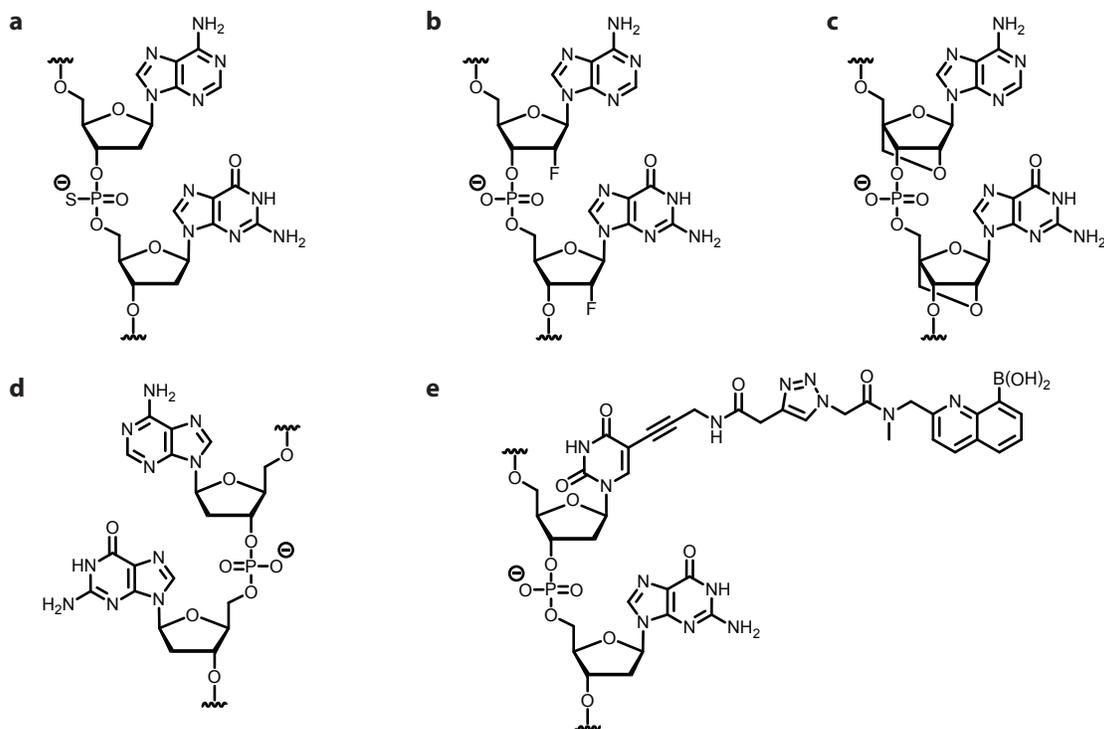


Figure 3-3. Synthetic modifications to nucleic acids impart novel properties. **(a)** Phosphorothioates are nuclease-resistant oligonucleotides where one oxygen of the phosphodiester backbone has been substituted with a sulfur atom. **(b)** RNA aptamers are often conferred with nuclease-stability by replacing the 2'-OH group with an electronically similar fluorine atom. **(c)** Locked nucleic acids (LNAs) are the result of a methylene bridge connecting the 2'-oxygen and 4'-carbon of RNA. The resulting nucleic acids are also nuclease stable. Furthermore, due to their locked ribose conformation, they exhibit enhanced hybridization properties. **(d)** Spiegelmers are enantiomers of aptamers, built from L-ribonucleotides rather than the naturally occurring D-ribonucleotides. Because they consist of only L-ribonucleotides, they exhibit high enzymatic stability. **(e)** Boronic acid moieties have been incorporated into DNA bases in order to target aptamer selection towards glycosylation sites.

recombination,²⁹ DNA aptamers can be generated to bind virtually any target.^{30,31} For the generation of cell-specific aptamers, the Tan group has developed a process called cell-SELEX, which evolves aptamers based on binding to live cells (Figure 3-4).³² For our multivalent cell-targeting vehicle, we chose a 41-nt aptamer selected from this process that was previously reported as *sgc8c*.³³ This aptamer was demonstrated to bind to Jurkat T leukemia cells, as well as a number of other leukemia cell lines. Its molecular binding partner was determined to be protein tyrosine kinase 7 (PTK7), which is a transmembrane protein that has been proposed as a potential biomarker for T cell acute lymphoblastic leukemia.³⁴

In order to incorporate both payload cargo and DNA aptamers onto the interior and exterior surfaces of MS2, respectively, two orthogonal bioconjugation reactions were required. This dual-surface modification strategy is outlined in Figure 3-5. Using this strategy, viral capsid DNA aptamer conjugates were synthesized with either fluorophores or PET agents on the interior, and DNA on the exterior. Furthermore, we demonstrated the selective targeting of Jurkat T leukemia cells by our aptamer-MS2 conjugates, as well as their cellular internalization.

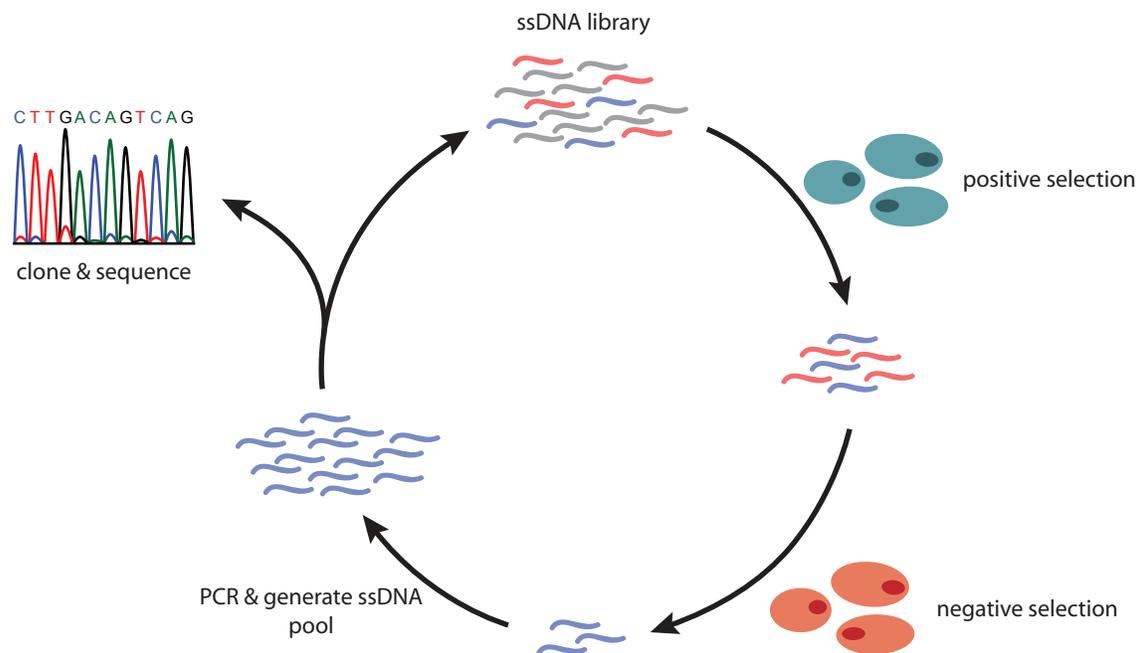


Figure 3-4. Cell-SELEX adapts the traditional evolution of nucleic acid-based aptamers to live-cell selection. A ssDNA library is screened against a targeted cell line, where ssDNA sequences that bind to the cell line are retained while non-binding sequences are removed from the selection process. The reduced pool then undergoes a round of negative selection to remove any sequences that bind to the negative cell line. The resulting pool from the negative selection is then PCR amplified, and the process is repeated with increasingly stringent selection conditions. To identify the sequences of high affinity library members, DNA aptamers are cloned into a plasmid and sequenced using standard plasmid sequencing technology.

3.2.3 Development of a targeted MS2-based agent for photodynamic therapy

To apply our concept towards drug delivery, we also designed an MS2-based agent capable of delivering therapeutic agents to specific cell types. As a model, we chose to construct a targeted agent for photodynamic therapy (PDT). Photodynamic therapy is a treatment that has been applied to a wide range of medical conditions and is comprised of three main components: a photosensitizer, a light source, and oxygen. In PDT, the light source is used to excite the photosensitizer, which in turn, transfers that energy to nearby ground state triplet oxygen ($^3\text{O}_2$) in order to generate singlet oxygen ($^1\text{O}_2$). Whereas $^3\text{O}_2$ is unreactive with most organic molecules, due to the fact that any reaction between $^3\text{O}_2$ and organic molecules (most commonly found in the singlet state) would be spin-forbidden, $^1\text{O}_2$ is highly reactive and can cause oxidative damage with many components in a biological system. In fact, $^1\text{O}_2$ generated for photodynamic therapy has been shown to damage cell membranes, DNA, lysosomes, and mitochondria.³⁵ In addition, due to its high reactivity, $^1\text{O}_2$ only has an effective damage radius of ~ 100 nm,³⁶ meaning targeted PDT should not affect the surrounding non-targeted tissues.

To create our targeted photodynamic agent, porphyrins were selected as a model photosensitizer for incorporation into MS2 capsids, and the exterior was again modified with the PTK7-binding DNA aptamer. The resulting construct was shown to generate $^1\text{O}_2$ upon illumination

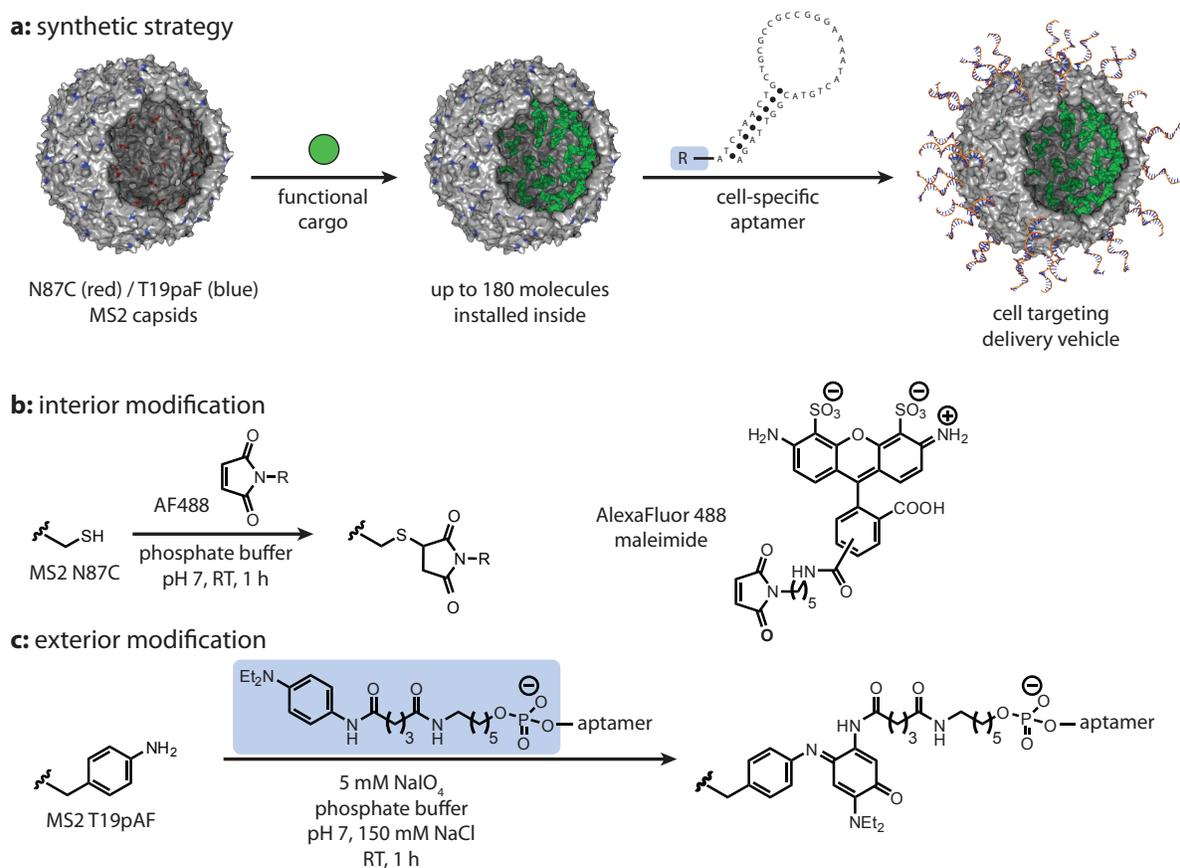


Figure 3-5. Dual-surface modification of MS2 capsids for targeted delivery. The overall synthetic strategy is shown in (a). (b) For interior surface modification, an N87C mutation on the MS2 coat protein allows for site specific alkylation. Up to 180 cargo molecules can be installed in these locations. (c) For exterior surface modification, the aptamer is first modified with a phenylene diamine group. A T19paF mutation on the capsid allows for the attachment of the modified DNA to the exterior surface of MS2 by the NaIO₄ mediated oxidative coupling reaction described in Chapter 2.

with 415 nm light, and the capsids were not damaged in the process. Furthermore, we demonstrated the selective killing of Jurkat cells with our multivalent photodynamic agent. This example highlights the utility of our MS2-based system, combining the multivalent display of both cell-specific aptamers and catalytic ¹O₂-generating porphyrins, for targeted drug delivery.

3.3 Results and discussion

3.3.1 Synthesis of a dual-surface modified viral capsid DNA aptamer conjugate

We chose a standard cysteine alkylation reaction in order to functionalize the interior of MS2. While MS2 does contain two native cysteine residues, these cysteines have proven to be unreactive under normal maleimide bioconjugation reaction conditions. Therefore, we were able to introduce a uniquely reactive cysteine residue at position 87 using site-directed mutagenesis. This N87C mutation, along with the previously described T19paF mutation, yielded the double mutant

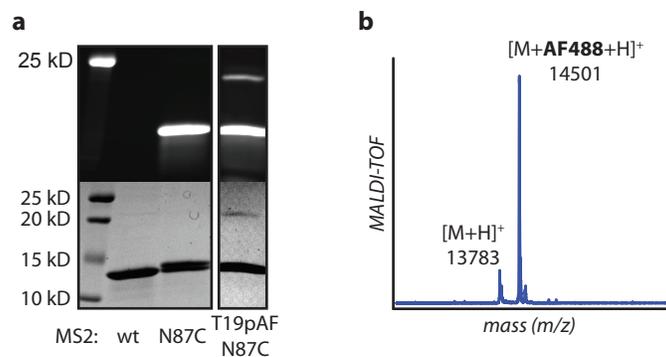


Figure 3-6. Interior surface modification of MS2 is achieved through the introduction of a cysteine mutant at position 87. **(a)** SDS-PAGE followed by visualization by fluorescence imaging (top) and Coomassie staining (bottom) showed the absence of any modification of wild type MS2 with AF488 maleimide, despite having two native cysteines. A fluorescent band was only observed upon introduction of the N87C mutation. Capsids modified with AF488 on the interior can also be subsequently modified on the exterior with strand **B**. **(b)** Interior modification with AF488 maleimide resulted only in singly-modified product, as shown by MALDI-TOF-MS. Expected $[M+H]^+=13,783$ amu; $[M+AF488+H]^+=14502$ amu.

MS2-T19paF-N87C. MS2-T19paF-N87C was reacted with Alexa Fluor 488 maleimide (AF488) to generate AF488-MS2, and the reactivity of the N87C mutation is shown in Figure 3-6. Despite having two native cysteines, MS2 lacking the N87C mutation showed no dye modification. In addition, near-quantitative conversion to the singly modified product was confirmed by MALDI-TOF MS.

For the exterior functionalization of our delivery vehicle, we chose the oxidative coupling

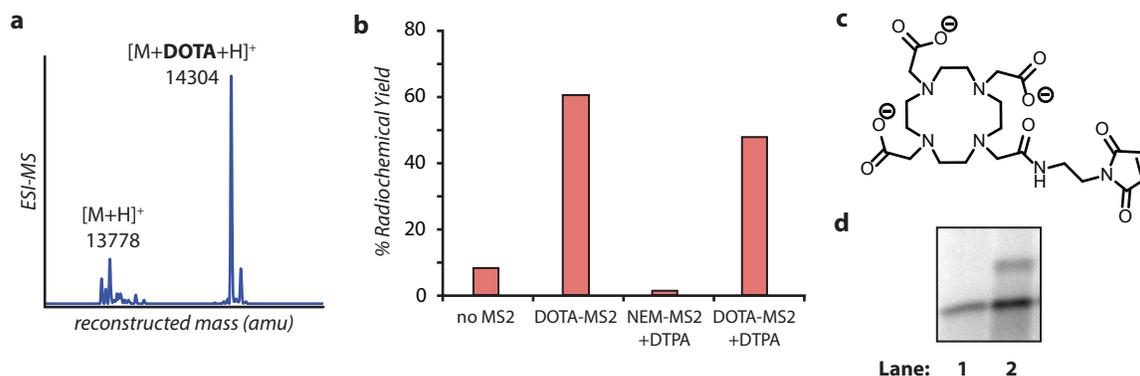


Figure 3-7. Interior surface modification of MS2 with DOTA-maleimide for Cu-64 chelation. **(a)** ESI-MS analysis showed that modification with DOTA-maleimide resulted in a singly-modified product. Expected $[M+H]^+=13,783$ amu; $[M+DOTA+H]^+=14309$ amu. **(b)** DOTA-MS2 capsids were radiolabeled with Cu-64, and the radiochemical yield was determined using SEC on gel filtration columns. The highest radiochemical yield is obtained when Cu-64 is incubated with DOTA-MS2. Without any MS2, only a small amount of Cu-64 is able to elute from the gel filtration column. MS2 modified with *N*-ethylmaleimide (NEM-MS2) showed negligible Cu-64 chelation, confirming that chelation was DOTA-dependent. Furthermore, the addition of DTPA was used to compete away any Cu-64 that was non-specifically bound to the MS2 capsid. The chemical structure of the DOTA-maleimide used to functionalize MS2-paF19-N87C is shown in **(c)**. **(d)** Analysis by SDS-PAGE and visualization through autoradiography showed that DOTA-MS2-A capsids (functionalized with strand **A** on the exterior) are also capable of being radiolabeled. Lane 1 corresponds to DOTA-MS2 and lane 2 corresponds to DOTA-MS2-A.

reaction between anilines and phenylenediamines described in Chapter 2. The PTK7-binding aptamer, or strand **B**, was first labeled with the diethyl phenylenediamine linker. It was then attached to the exterior surface of MS2 by reacting it with MS2-T19paF-N87C in the presence of NaIO₄ to yield MS2-**B**. SDS-PAGE and densitometry analysis showed that 20-40 copies of strand **B** were attached to the exterior surface of MS2. The slightly lower conversion with strand **B** compared to the 20-nt strand **A** used in Chapter 2 is likely due to increased steric and electrostatic effects.

To generate the entire cell-targeting construct, AF488-MS2 was first synthesized, followed by the oxidative coupling of the aptamer onto the exterior surface to yield AF488-MS2-**B**. Figure 3-6a confirms the dual-surface modification; the upper fluorescent band in the polyacrylamide gel corresponds to MS2 monomers modified with AF488 that are shifted by the attachment of strand **B**. The resulting construct would be amenable to optical imaging techniques. Due to the modular design of the aptamer-MS2 construct, the interior can also be modified with a variety of other cysteine-reactive molecules.

As a complement to optical imaging techniques, the chelation of Cu-64 would allow us to conduct positron emission tomography (PET) imaging experiments with our modified capsids. Therefore, we have also demonstrated the conjugation of maleimide-functionalized 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA) on the interior of MS2 for Cu-64 chelation (DOTA-MS2, Figure 3-7c). Again, modification of MS2-T19paF-N87C with DOTA-maleimide gave rise to only a singly-modified product as shown in Figure 3-7a. The resulting DOTA-MS2 capsids were radiolabeled with Cu-64 and their radiochemical yields were determined by size-exclusion chromatography (SEC). SEC was chosen because Cu-64 associated with MS2 capsids should elute more quickly than free Cu-64. In addition, any Cu-64 that may be non-specifically associated with the capsids would be competed off by the addition of excess diethylene triamine pentaacetic acid (DTPA). For the DOTA-MS2 capsids, we were able to

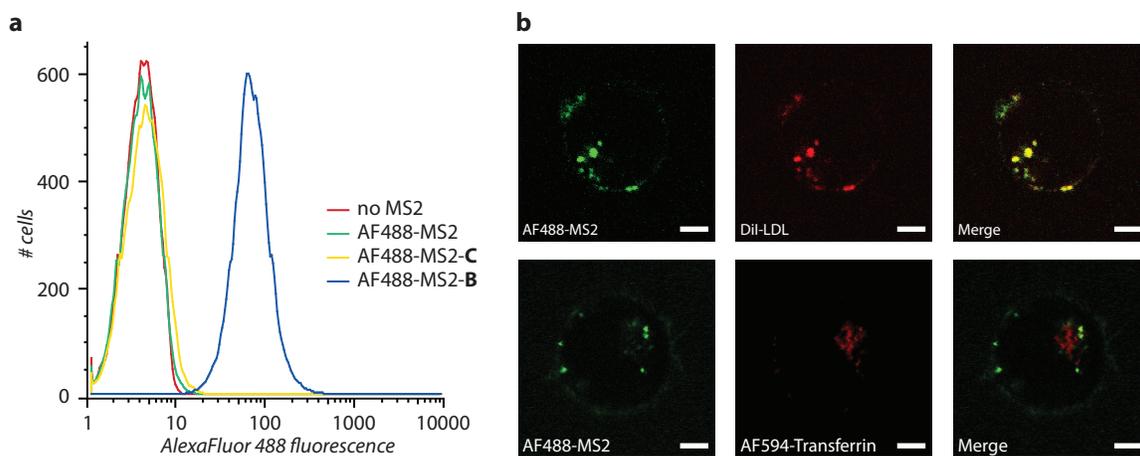


Figure 3-8. Cellular targeting and uptake with aptamer-labeled capsids. **(a)** Cell targeting was confirmed using flow cytometry. Only MS2 capsids modified with aptamer **B** bound to Jurkat cells (blue trace). Capsids that were unmodified on the exterior (green trace) and capsids modified with strand **C** (yellow trace) showed only background autofluorescence (red trace) similar to untreated cells. **(b)** Live-cell confocal images showed colocalization of **B**-labeled capsids with LDL-labeled endosomes, but not with transferrin-labeled endosomes. Scale bars represent 3 μ m.

achieve a radiochemical yield of 47% following labeling with Cu-64 and treatment with DTPA. As a negative control, capsids were also modified with *N*-ethylmaleimide (NEM-MS2) and subjected to the same radiolabeling conditions. For NEM-MS2, Cu-64 labeling resulted in a radiochemical yield of ~2%, indicating that Cu-64 labeling was DOTA-dependent. DOTA-MS2 was modified on the exterior with strand **A** (used as a proof of concept) to yield DOTA-MS2-**A**, which was also incubated with Cu-64. Figure 3-7d shows the analysis of DOTA-MS2-**A** by SDS-PAGE and visualization by autoradiography, confirming the dual modification with DOTA and DNA, as well as radiolabeling with Cu-64. The DOTA-MS2-**A** example highlights the potential range of imaging applications to which the dual-modified MS2 platform can be extended.

3.3.2 Cell-binding and internalization of a viral capsid DNA aptamer conjugate

In order to test the aptamer-MS2 cell-targeting construct in cell-binding assays, AF488-modified capsids were used. The targeting capabilities of AF488-MS2-**B** capsids were tested with Jurkat cells based on strand **B**'s reported specificity, and was evaluated by flow cytometry. Upon

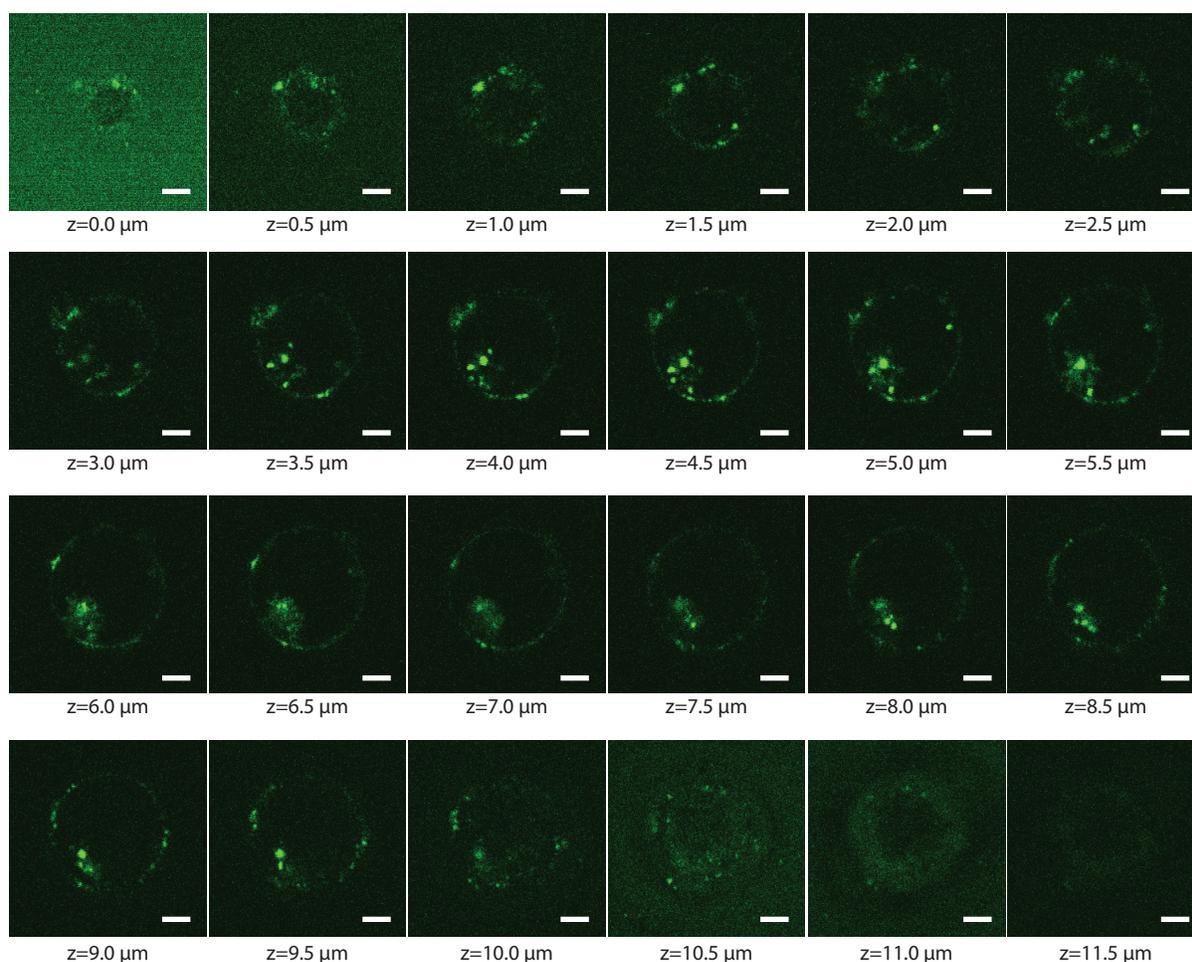


Figure 3-9. Z-stack of live cell confocal images showing the internalization of AF488-MS2-**B** capsids in Jurkat cells. Scale bars represent 3 μm.

incubation of AF488-MS2-**B** with Jurkat cells at 37 °C for 30-60 min, a significant increase in the mean fluorescence intensity of the Jurkat cells was observed (Figure 3-8a). AF488-MS2 (no exterior functionalization) was used as a control for the non-specific binding of the MS2 capsids to Jurkat cells, and did not give rise to an increase in the mean fluorescence intensity compared to background cellular autofluorescence. AF488-MS2-**C** (synthesized using a 41-nt strand of a random sequence, strand **C**) was tested as an additional negative control, confirming the specificity of the aptamer sequence for cell binding.

Following validation of the specific cell-targeting ability of AF488-MS2-**B** by flow cytometry, we investigated the agent's cellular internalization by confocal microscopy. AF488-MS2-**B** was incubated with Jurkat cells at 37 °C for 30-60 min prior to imaging. A z-stack of live cell confocal images was taken, where AF488 fluorescence was detected by bright, punctate signals throughout the inside of the cell (Figure 3-9). In addition, AF488 fluorescence was detected on the cell membrane, likely due to the presence of PTK7 on the cell surface. To determine in which cellular compartments the **B**-labeled capsids were accumulating, Jurkat cells were co-stained with the capsids and one of two endocytic markers. The two markers chosen were fluorescently labeled low-density lipoprotein (LDL) particles and transferrin. While both LDL and transferrin are known markers for endocytosis, they have been shown to traffic differently once inside the cell. LDL particles accumulate in endosomes that eventually traffic to lysosomes, whereas transferrin-associated vesicles are directed back to the cell-surface through a recycling pathway.³⁷ AF488-MS2-**B** capsids were found to co-localize with fluorescently labeled LDL particles, suggesting a lysosomal fate (Figure 3-8b). Due to the acidity of lysosomes, utilizing capsids modified with **B** would be a good strategy for the targeted drug delivery of acid-labile prodrugs.

3.3.3 Development of a multivalent, cancer-targeted, photodynamic agent

The results discussed in the previous section highlight the feasibility of creating an MS2-based agent capable of delivering therapeutic molecules to specific cell types. As a proof of

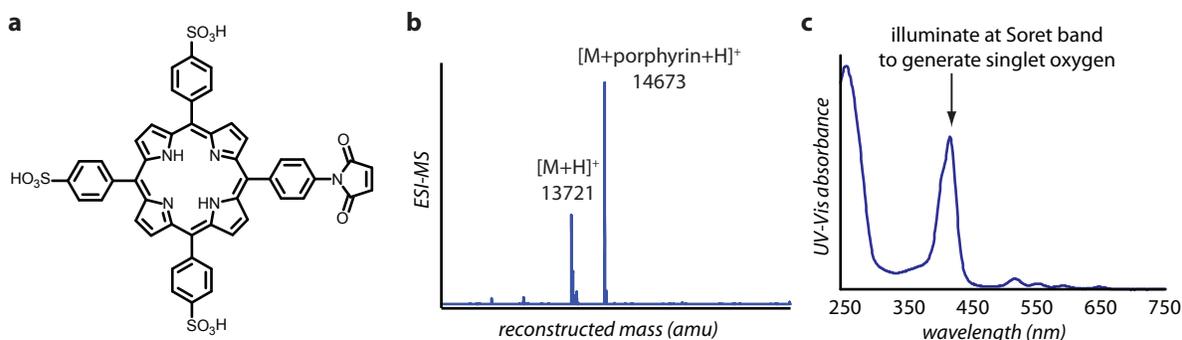


Figure 3-10. Conjugation of porphyrins to the interior surface of MS2 for photodynamic therapeutic applications. The chemical structure of the asymmetric porphyrin-maleimide used for cysteine alkylation is shown in (a). (b) Modification of MS2-N87C with the porphyrin-maleimide gave rise to a singly-modified protein bioconjugate, as shown by ESI-MS. Expected $[M+H]^+ = 13,718$ amu; $[M+porphyrin+H]^+ = 14668$ amu. (c) UV-vis spectroscopy of porphyrin-MS2 clearly showed the Soret band of the porphyrin. The calculated conversion to the porphyrin-MS2 product as determined by UV-vis spectroscopy was nearly quantitative. The calculated conversion by UV-vis deviates from the conversion determined by ESI-MS (75%), likely due to poor ionization of the porphyrin-MS2 conjugate.

concept, we chose to generate an agent for targeted photodynamic therapy. Benefiting from the modular design of our aptamer-MS2 construct, we were able to create a multivalent photodynamic agent by modifying the interior of MS2 with a tri-sulfonated water-soluble porphyrin containing a maleimide functional group. The porphyrin modification was confirmed by ESI-MS (Figure 3-10b) and gave rise to a singly modified product. Percent conversion to the porphyrin-modified product was estimated at 75% by ESI-MS. However, our group has previously observed that proteins modified with anionic dyes have exhibited suppressed ionization under electrospray ionization conditions, and therefore we estimated that the modification was likely higher. To this end, we characterized the conjugate by UV-vis spectroscopy (Figure 3-10c). By UV-vis, the conjugate exhibited a strong characteristic Soret band corresponding to the porphyrin. The absorbance of the Soret band relative to protein absorbance was used to calculate percent conversion, and indicated near quantitative yields. Therefore, we were confident that capsids incorporated 135 to 180 copies of the porphyrin on the interior.

The exterior surface of the porphyrin-labeled capsids were modified in the same manner with aptamer **B** using the oxidative coupling strategy as described in the previous section, and the

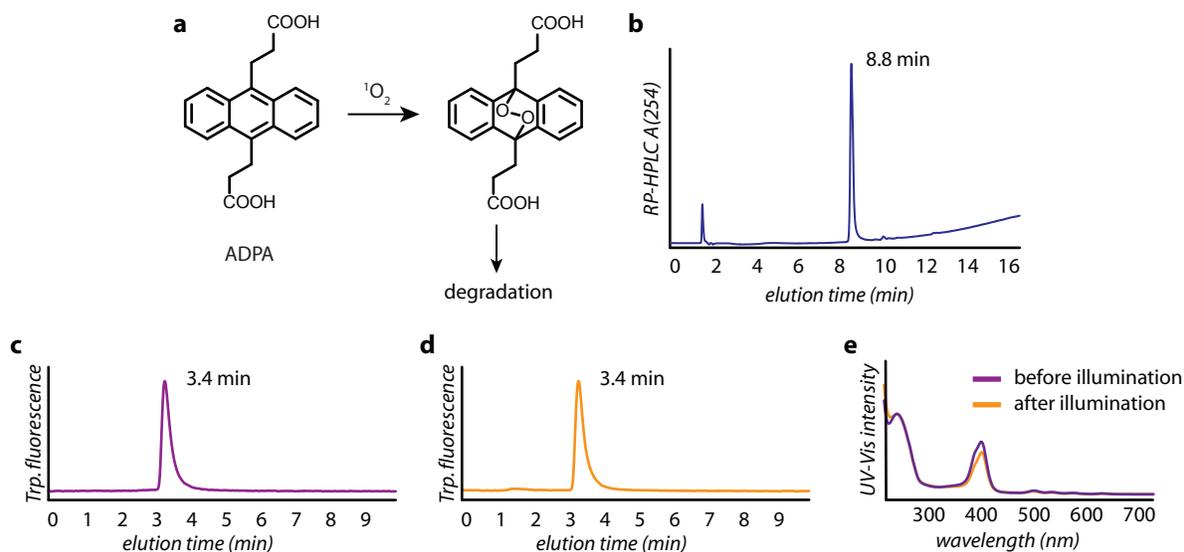


Figure 3-11. Generation of singlet oxygen with porphyrin-modified MS2 capsids. **(a)** ADPA is used as an indicator for the generation of singlet oxygen because it undergoes a Diels-Alder type cycloaddition reaction with singlet oxygen. The resulting peroxo compound degrades into various byproducts, resulting in the disappearance of the characteristic absorbance of ADPA at 254 nm, which can be monitored by RP-HPLC **(b)**. The integrity of the porphyrin-MS2-**B** capsids was characterized by SEC before **(c)** and after **(d)** singlet oxygen generation, and the chromatograms indicated that capsids remained intact following dual modification and 20 minutes of illumination at 415 nm, respectively. **(e)** Porphyrin integrity was also assessed using UV-vis spectroscopy. Only minimal porphyrin bleaching by singlet oxygen was observed, as indicated by a slight reduction in the intensity of the Soret band after illumination.

resulting porphyrin-MS2-**B** sample was tested for its ability to generate singlet oxygen. In order to detect singlet oxygen, anthracene-9,10-dipropionic acid (ADPA) was used as an indicator. ADPA functions as an indicator for singlet oxygen by undergoing a Diels-Alder type cycloaddition with the internal diene of ADPA to form a peroxo species that then degrades into various fragments.

Therefore, a reverse-phase HPLC assay was used, where the disappearance of ADPA correlated with the generation of singlet oxygen (Figure 3-11b). We found that porphyrin-MS2-**B** capsids could indeed generate singlet oxygen, and did so only upon illumination with a commercially available 415 nm LED lamp.

Having determined porphyrin-MS2-**B** capsids could generate singlet oxygen, we tested the conjugates for capsid stability prior to and following illumination. The integrity of the capsids was determined by SEC, and it was observed that after 20 minutes of illumination, capsids remained intact, due to their detection at the same elution time and comparable peak areas before and after illumination (Figure 3-11c and d). At longer illumination times (90 minutes), the area of the peak corresponding to the intact capsids did decrease to some extent, indicating loss of intact protein. However there was minimal degradation of the conjugates at 20 minutes. In addition, the conjugate was analyzed by UV-vis spectroscopy before and after illumination to assess porphyrin stability. A small reduction in the absorbance of the Soret band indicated porphyrin bleaching (Figure 3-11e). However, at 20 minutes of illumination, the bleaching was minimal.

Next, the photodynamic cytotoxicity of the porphyrin-MS2-**B** capsids was assessed *in vitro*. Jurkat cells were first incubated with porphyrin-MS2-**B** at a concentration of 7.2 nM in capsids on ice and in the dark. After incubation, the cells were washed in order to remove unbound capsids, and were illuminated for 20 minutes with 415 nm light. Cells were stained with both Annexin

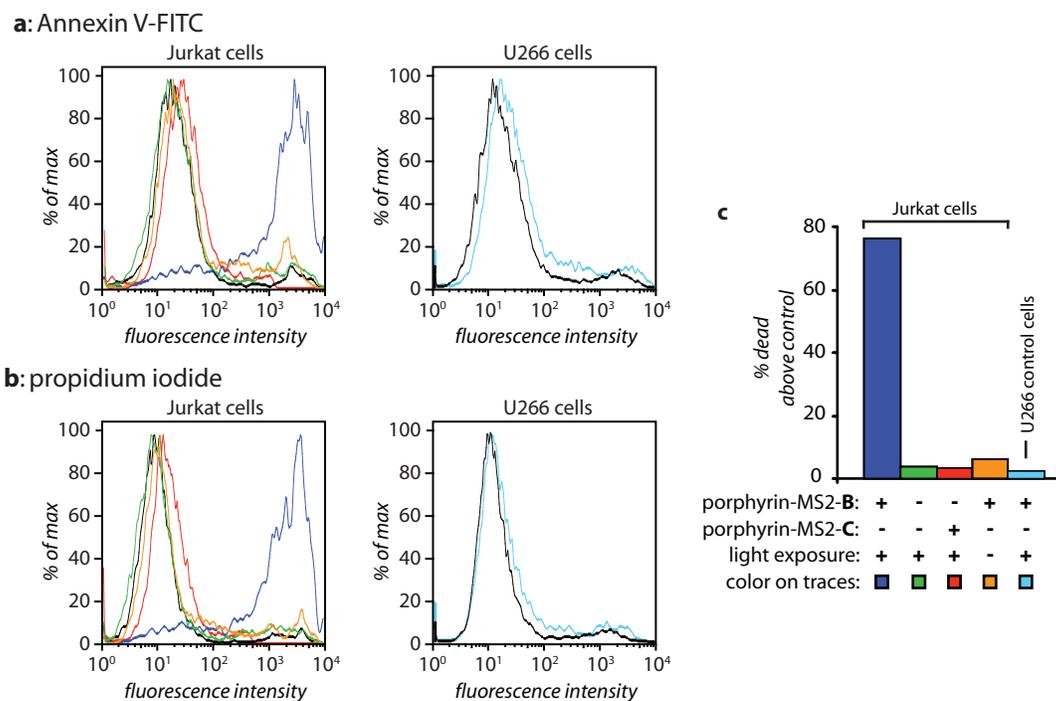


Figure 3-12. Flow cytometry analysis of the efficiency of targeted cell killing by porphyrin-MS2-**B** capsids. Cells were treated with photodynamic agents (7.2 nM in capsids) on ice for 30 minutes before illumination with 415 nm light for 20 minutes. After co-staining with the cell viability reagents Annexin V-FITC and propidium iodide, cells were analyzed by flow cytometry (**a,b**). The flow cytometry data are summarized in (**c**). Only Jurkat cells that were incubated with the porphyrin- and PTK7 aptamer-modified capsids (dark blue trace) showed significant cell death compared to the untreated cells (black trace). Capsids bearing a random DNA sequence, samples that were not exposed to 415 nm light, and using receptor-negative U266 cells resulted in no more than 3-6% cell death.

V-FITC and propidium iodide (PI) for flow cytometry analysis in order to assess cell viability. Binding of Annexin V to cells is an indicator of apoptosis because it binds phosphatidylserine lipids that translocate to outer cell membranes when cells are apoptotic. Intercalation of propidium iodide into genomic DNA is indicative of cell membranes that have been compromised, either due to late-stage apoptosis or necrosis. Together, these two reagents are a commonly used assay for studying cell viability.

Flow cytometry analysis using these two reagents showed that exposure of Jurkat cells to porphyrin-MS2-**B** capsids and 415 nm light resulted in at least 76% cell death relative to the untreated control (Figure 3-12). Additional controls run for this experiment included Jurkat cells treated with **B**-labeled capsids without illumination, untreated Jurkat cells that were illuminated, Jurkat cells treated with porphyrin-MS2-**C** capsids and illuminated, and U266 cells (a PTK7-negative cell line reported in the initial publication on aptamer **B**) treated with **B**-labeled capsids and illumination. In all controls, no more than 6% cell death was observed. It is important to note that the concentration of the capsids used for these experiments had a significant effect on the efficiency and selectivity of cytotoxicity. Treatment of cells with targeted capsids at a concentration that was too low resulted in increased cell viability. However, treatment with much higher concentrations began to reduce cell viability in control reactions. Therefore, by adjusting the conditions appropriately, we were able to demonstrate the efficiency and targeting of the multivalent photodynamic agents.

Lastly, porphyrin-MS2-**B** capsids were tested for cell-specific cytotoxicity in a heterogeneous cell mixture containing Jurkat cells and red blood cells (RBCs). Because both cell types are non-adherent, they were immobilized onto a glass surface for ease of analysis by optical microscopy using a previously reported method, where cells were modified with DNA on the cell surface and then hybridized to a complementary DNA-coated glass slide.³⁸ Red blood cells were chosen because they have a distinctly different morphology from Jurkats, making them easily distinguishable by eye. In order to visualize cell death, cells were stained with trypan blue after treatment with capsids and illumination. Trypan blue is only able to pass through the membranes of dead cells, resulting in a blue appearance under a microscope. The cell mixture was treated by

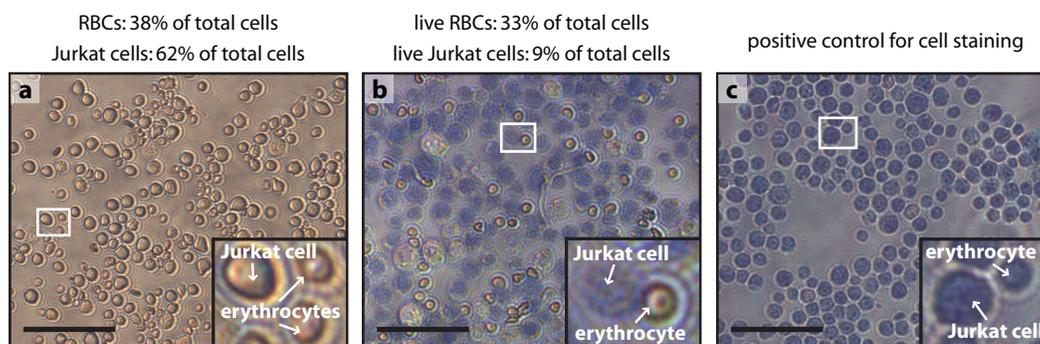


Figure 3-13. Targeted cell killing in a heterogeneous cell mixture. Both Jurkat cells and red blood cells (RBCs) were immobilized onto glass slides using a DNA-directed immobilization technique described by Hsiao *et al.*³⁸ The initial cell population consisted of 38% red blood cells and 62% Jurkat cells (**a**). After incubation with porphyrin-MS2-**B** capsids and illumination with 415 nm light, cells were stained with the vital dye trypan blue in order to determine live from dead cells. A large percentage of cell death was observed for the targeted Jurkat cells, however not for the non-targeted RBCs (**b**). Both cell types stain equally well with trypan blue, as determined by a positive control where the cell mixture was treated with 30% ethanol and then stained with trypan blue (**c**).

the targeted photodynamic capsids and illuminated to generate singlet oxygen. Figure 3-13 shows the microscope images following treatment and staining. These images show that Jurkat cells are killed selectively in the presence of red blood cells, despite the entire sample being treated with the same photodynamic agent. The observed selectivity in spite of such close proximity between the cell types is most likely due to the short diffusion distance of singlet oxygen (~100 nm). Upon quantification, it was determined that live red blood cells comprised 38% and 33% of total cells in the mixture before and after treatment, respectively. In contrast, live Jurkat cells represented 62% and 9% of total cells before and after treatment, respectively. This experiment also confirms the cell-selectivity of our multivalent photodynamic therapeutic agent. In addition, this particular agent highlights the potential of this class of agents to be used in blood-based dialysis systems.

3.4 Conclusion

Bacteriophage MS2 has been demonstrated as a scaffold for the construction of multivalent heterobimolecular structures, carrying functional molecules for imaging or therapeutics, and DNA aptamers for targeting. In this chapter, we demonstrate the use of a particular aptamer (targeted to PTK7) for endowing targeting capabilities to MS2 capsids. Capsids generated with a fluorescent moiety on the interior and aptamers on the exterior were shown to specifically bind to Jurkat cells, and internalized into endosomes that eventually traffic to lysosomes.

The great advantage of this system is that its design is modular. The interior can be modified with a variety of maleimide-containing functional molecules (so long as they are able to pass through the 2 nm pores of the MS2 capsid), and the exterior can be modified with virtually any DNA (or RNA) aptamer in order to target different receptors or cell types. In one example, we demonstrated the incorporation of DOTA into MS2 capsids. These capsids were then modified with DNA on the exterior and were successfully radiolabeled with Cu-64. An MS2-based construct would be particularly useful as a PET imaging agent, since multivalency is anticipated to increase the specific activity on a per capsid basis. In another example, we demonstrated the synthesis of a multivalent, targeted photodynamic agent where MS2 capsids were labeled with singlet oxygen-generating porphyrins on the interior, and PTK7-binding DNA aptamer on the exterior. The resulting construct was shown to generate singlet oxygen with minimal damage to the agent itself, and was able to target Jurkat cells and kill them in a cell-specific manner.

Together, these efforts represent the first report of targeted imaging and therapeutic treatment of cancer cells with a viral capsid-based scaffold in the Francis group. Moving forward, a number of extensions to this concept are possible. For diagnostic imaging, near-infrared dyes with greater tissue penetration and a variety of other chelating agents for different radionuclides can be tested. While not necessary for imaging applications, cellular internalization of MS2-based agents would be advantageous for drug delivery purposes, and especially in the use of acid-labile prodrugs. New therapeutic molecules can be incorporated into these multivalent cell-targeted vehicles, with the goal of reducing exposure of healthy tissues to cytotoxic agents, or even to redirect drugs traditionally used for one type of cancer to another. These viral capsid DNA aptamer conjugates represent a promising platform for novel applications in bionanotechnology requiring active targeting capabilities.

3.5 Materials and methods

3.5.1 General procedures and materials

Unless otherwise noted, all chemicals and solvents were of analytical grade and used as received from commercial sources. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F254 plates with visualization by ultraviolet (UV) irradiation at 254 nm or potassium permanganate stain. All organic solvents were removed under reduced pressure using a rotary evaporator. Dichloromethane (CH_2Cl_2) was distilled under a nitrogen atmosphere from calcium hydride. Water ($\text{dd-H}_2\text{O}$) used in biological procedures or as the reaction solvent was deionized using a NANOpure purification system (Barnstead, USA). Cu-64 was purchased as a solution of copper (II) chloride in 0.1 N HCl. 4-(4-diethylamino-phenylcarbamoyl)-butyric acid succinimidyl ester was prepared using the previously reported method.³⁹ The porphyrin-maleimide was prepared using a previously reported method and stored as a 50 mM stock solution in DMSO at $-20\text{ }^\circ\text{C}$.^{40,41}

All oligonucleotides were obtained from Integrated DNA Technologies (Coralville, IA). Samples were purified by reverse-phase HPLC or NAP-5 gel filtration columns (GE Healthcare). Samples were lyophilized using a LAB CONCO Freezezone 4.5 (Lab Conco). Lyophilized oligonucleotides were re-suspended in the appropriate buffer and the concentration was determined by measuring the absorbance at 260 nm.

All cell culture reagents were obtained from Gibco/Invitrogen Corp (Carlsbad, CA) unless otherwise noted. Cell culture was conducted using ATCC recommended guidelines. Jurkat cells and U266 cells were grown in T-25 culture flasks (Corning, USA) in RPMI Medium 1640 supplemented with 10% (v/v) fetal bovine serum (FBS, HyClone) and 1% penicillin/streptomycin (P/S, Sigma). Cell viability was assayed using an Annexin V-FITC/propidium iodide kit (BD Biosciences, USA) or by trypan blue.

3.5.2 Instrumentation and sample analysis

Mass Spectrometry. Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DETM system (PerSeptive Biosystems, USA). Prior to MALDI-TOF MS analysis, samples were desalted using C18 ZipTip® pipettips (Millipore, USA). Oligonucleotide samples were co-crystallized using a 3-hydroxypicolinic acid:ammonium citrate solution (45 mg/mL:5 mg/mL in 4.5:5.5 MeCN:ddH₂O). For electrospray ionization mass spectrometry (ESI-MS) oligonucleotide conjugates were analyzed using an LTQ Orbitrap XL mass spectrometer equipped with an Ion Max electrospray ionization (ESI) source (Thermo Fisher Scientific, Waltham, MA). Sample solutions were infused into the ESI probe at a flow rate of 5 $\mu\text{L}/\text{min}$ using a syringe pump. The voltages applied to the ion optics were adjusted automatically for optimum desolvation and transmission of the ions of interest using Tune Plus software (version 2.4, Thermo). Mass spectra were recorded in the negative ion mode over the range $m/z = 500$ to 2000 for a period of two minutes. Mass spectra were processed using Xcalibur software (version 4.1, Thermo) and the measured charge state distributions were deconvoluted using ProMass software (version 2.5 SR-1, Novatia, Monmouth Junction, NJ). Prior to ESI-MS analysis, oligonucleotides were prepared as previously described.⁴² All MS data for oligonucleotides and protein samples were found to be within 0.1% of the expected values.

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

was achieved with an inline diode array detector (DAD). Both analytical and preparative reverse-phase HPLC of oligonucleotides was accomplished using a C18 stationary phase and a MeCN/100 mM triethylammonium acetate (TEAA, pH = 7.0) gradient.

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus from Bio-Rad (Hercules, CA). All protein electrophoresis samples were heated for 10 minutes at 100 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of any disulfide bonds. Gels were run for 5 minutes at 30 V and 70-90 minutes at 120 V to allow good separation of bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad). For fluorescent protein conjugates, visualization was accomplished on a UV backlight. Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA).

UV-Vis. UV-Vis spectroscopic measurements were conducted on a Cary 50 Scan bench-top spectrophotometer (Varian, Inc., USA).

Illumination at 415 nm. In order to illuminate the Soret band of the porphyrin we used a commercially available blue LED lamp from www.acnelamp.com (Dima-Tech Inc., USA). The lamp contained 70 blue LED's with a peak emission at 415 nm and an intensity of 26,273 $\mu\text{W}/\text{cm}^2$.

3.5.3 Experimental

General Procedure for the addition of phenylene diamine to oligonucleotides. DNA oligonucleotides were purchased containing a primary amine on the 5'-end. A typical reaction was carried out as follows: DNA at a concentration of 300 μM was reacted with 4-(4-diethylamino-phenylcarbamoyl)-butyric acid succinimidyl ester (60-120 eq) in a 1:1 solution of DMF and 50 mM pH 8.0 phosphate buffer. The reaction mixture was briefly vortexed and then allowed to react at rt for 2 h. Either RP-HPLC or commercially available gel filtration columns (GE Healthcare) were used to purify the small molecule from the DNA, following the commercially provided protocol. Following purification, the DNA-containing solution was lyophilized and then re-suspended in the desired buffer. Concentration was determined by measuring the absorbance at 260 nm.

The sequence identities of **A**, **B**, and **C** are as follows:

A: 5'-TCATACGACTCACTCTAGGGA-3'

B: 5'-ATCTAACTGCTGCGCCGCGGGAAAATACTGTACGGTTAGA-3'

C: 5'-CCCTAGAGTGAGTCGTATGACCCTAGAGTGAGTCGTATGAA-3'

General Procedure for DNA conjugation to MS2. An Eppendorf tube was charged with either MS2-T19paF or MS2-T19paF-N87C (20 μM), phenylene diamine-containing oligonucleotide (200-400 μM), and NaIO_4 (5 mM). The reaction was carried out in 50 mM pH 7.0 phosphate buffer containing 150 mM NaCl. The reaction was briefly vortexed and allowed to react at rt for 1 h. After an hour, for a 50 μL reaction, the reaction was quenched by the addition of 5 μL of 500 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP). For purification, the sample was first buffer-exchanged by gel filtration (NAP-5) into the desired buffer. The excess DNA was then removed by successive centrifugal filtration using 100k molecular weight cutoff filters (Millipore).

Cloning and Expression of MS2 mutants. The pBAD-MS2-T19paF plasmid production

and growth has been previously reported.³⁹ We would like to thank the Peter Schultz lab (Scripps Research Institute, LaJolla, CA) for the tRNA- and tRNA-synthetase-encoding plasmids necessary for *p*-aminophenylalanine (*paF*) incorporation. Position 87 was mutated into a cysteine using the following forward and reverse primers:

Forward: 5'-AGCCGCATGGCGTTCGTA CTTATGTATGGA ACTA ACCATTC-3'

Reverse: 5'-GAATGGTTAGTTCCATACATAAGTACGAACGCCATGCGGCT-3'

Growth and purification of MS2-T19*paF*-N87C was identical to that of MS2-T19*paF*, although a lower yield was obtained for MS2-T19*paF*-N87C (~1-10 mg/L as compared to ~20 mg/L for MS2-T19*paF*).

Dual-Surface modification of MS2-T19paF-N87C with AF488 maleimide. MS2-T19*paF*-N87C is first modified on the interior cysteine. For the cysteine alkylation reaction, Alexa Fluor 488 maleimide (Invitrogen; 15 μ L of a 19 mM solution in DMSO) is added to MS2-T19*paF*-N87C (285 μ L of a 100 μ M solution in 10 mM pH 7.2 phosphate buffer). The reaction is briefly vortexed and allowed to react at rt for 1 h. Excess small molecules were removed by gel filtration (NAP-5) and the remaining protein was concentrated using centrifugal filtration. It was important to note that centrifugal filters were pre-rinsed before use as we found this to prevent problems with the oxidative coupling step. The exterior modification was performed as described above.

Flow Cytometry of Fluorescent Capsids. Flow cytometry analysis was acquired on a FACSCalibur flow cytometer (BD Biosciences, USA) using a standard 488 Ar laser. Data were collected for at least 10,000 live cells for all experiments. Jurkat cells (1×10^6 cells in 250 μ L) were treated with fluorescent capsids (2 μ M) in culture media and incubated either on ice or at 37 $^{\circ}$ C for 30-60 min. After incubation, the cells were washed twice with 1 mL of fresh media, and then analyzed on the flow cytometer.

Confocal Microscopy. Confocal fluorescence imaging was carried out on a Zeiss LSM510 META/NLO Axioimager using a 63x Achroplan IR oil-immersion objective lens. Double labeling of Jurkat cells was performed by incubating cells with modified MS2 (20 μ M) and DiI-LDL (15 μ g/mL, Invitrogen) or Alexa Fluor 594-Transferrin (25 μ g/mL, Invitrogen) for 30-60 min at 37 $^{\circ}$ C. Modified MS2 (fluorescently-labeled with Alexa Fluor 488) was excited at 488 nm and emission was collected between 495-530 nm. Both DiI-LDL and AF594-Transferrin were excited with a 543 nm line and emission was collected between 590-625 nm and 590-655 nm, respectively.

Dual-Surface modification of MS2-T19paF-N87C with DOTA maleimide. MS2-T19*paF*-N87C (38 μ L of a 100 μ M stock solution in 10 mM pH 7.2 potassium phosphate buffer), EDTA (1 μ L of an 80 mM aqueous stock adjusted to pH 7.2), and DOTA maleimide (1 μ L of a 38 mM stock solution in DMSO) were combined in an Eppendorf tube. The mixture was briefly vortexed and allowed to stand at room temperature for 1.5 h. The reaction mixture was passed through a NAP-5 gel filtration column, eluted into 100 mM pH 5.5 ammonium citrate buffer, and the resulting eluate was spin-concentrated 3 times using a 100 kDa MWCO spin-concentrator. The exterior modification was performed as described above.

Radiolabeling DOTA-MS2 with Cu-64. DOTA-MS2 (37.5 μL of a 170 μM stock in 100 mM pH 5.5 ammonium citrate buffer), 100 mM pH 6.0 ammonium citrate buffer (7.5 μL), and Cu-64 (5 μL , 105 μCi) were combined in an Eppendorf tube for a total volume of 50 μL . Because the Cu-64 solution was dissolved in 0.1 N HCl, it was important to test the overall pH of the reaction before actual radiolabeling to ensure that the resulting pH would not cause protein precipitation. The reaction mixture was incubated for 80 min in a heat block at 37 $^{\circ}\text{C}$. An aqueous solution of DTPA (5.5 μL of a 10 mM pH 6.0 stock solution in water) was added to the reaction mixture before purification on microspin G-25 gel filtration columns. The activity of flow-through (DOTA-MS2 containing fraction) and the column (retained Cu-64) were determined in order to calculate radiochemical yield.

Dual-Surface modification of MS2-T19paF-N87C with porphyrin maleimide. MS2-T19paF-N87C is first modified on the interior cysteine. To a solution of MS2-T19paF-N87C (80 μM in 10 mM phosphate buffer, pH 7) was added 20 equivalents of porphyrin maleimide as a 50 mM solution in DMSO. The reaction mixture was vortexed briefly, then incubated at room temperature for 2 h in the dark. The mixture was then passed through a NAP-5 column equilibrated with 10 mM phosphate buffer, pH 7, to remove excess porphyrin. The capsids were further concentrated using a 100 kDa molecular weight cutoff filter. The conversion of porphyrin was determined by comparing the absorbance of the porphyrin Soret band ($\epsilon = 266,000 \text{ M}^{-1}\text{cm}^{-1}$) to the A_{260} of the protein ($\epsilon = 172,000 \text{ M}^{-1}\text{cm}^{-1}$) and assuming negligible porphyrin absorbance at 260 nm. The exterior modification was performed as described above.

Flow cytometry for photodynamic therapy experiments. To quantify cell death, we used flow cytometry to analyze the Annexin V-FITC and PI staining intensities of the treated samples. For all experiments, we used 5×10^5 cells (either Jurkat or U266) suspended in 1 mL of PBS containing 1% FBS. To these cells was added 200 μL of porphyrin-MS2-**B** (or porphyrin-MS2-**C**) at a concentration of 1.3 μM in protein monomer, or 7.2 nM in capsid (roughly 1.75×10^6 capsids/cell). The cells were incubated with the capsids for 30 min on ice in the dark, then washed three times by centrifugation (at 240 rcf) followed by resuspension in 1 mL of PBS containing 1% FBS. They were then placed in glass-bottomed wells and illuminated for 20 min using a 415 nm LED lamp.

Following illumination, the cells were washed once by centrifugation, resuspended in the apoptosis kit binding buffer, and stained with Annexin-V FITC and PI. Following staining, the cells were analyzed by flow cytometry to determine the amount of FITC and PI fluorescence. To determine the amount of cell death, the flow cytometry results were gated using a control sample of live, healthy cells that had not been treated with anything. For each experiment, a population of cells not exposed to any treatment was also used to control for variability in cell viability from day to day. To determine the proper fluorescence ranges for dead cells, we induced apoptosis with camptothecin and necrosis with Tween 20, then stained with Annexin V-FITC or PI (respectively) and analyzed the cells by flow cytometry. When possible, 10,000 cells were counted.

Photodynamic treatment of a surface-immobilized cell mixture. A mixed population of Jurkat cells and red blood cells was immobilized on a glass slide as previously described.³⁸ Briefly, fresh samples of red blood cells were obtained from a blood sample of a healthy human and stored in a 1% citric acid solution at room temperature. Cells were used within 1 h. Both Jurkat cells

and red blood cells were exposed to the same sample of single-stranded DNA modified with an NHS ester in order to attach DNA on the cell surface proteins. This mixed population was then immobilized on a glass slide bearing the complementary DNA sequence. Because both cell types were modified with the same DNA strand, the immobilized population was heterogeneous.

Next, the surface-immobilized cells were exposed to 200 μ L of a solution of porphyrin-MS2-**B** (7.2 nM in capsid) for 30 min, on ice in the dark. The slides were rinsed with three portions of 1 mL of PBS containing 1% FBS and then illuminated with a 415 nm lamp for 20 min. Following illumination, cell viability was assayed using trypan blue to detect cell death. A negative control slide, consisting of mixed cells not exposed to capsids, was also stained by trypan blue to show cell viability. Finally, a positive control slide, consisting of mixed cells exposed to 30% ethanol, was used to demonstrate that both cell types stained blue upon death.

3.6 References

- (1) Liu, S.; Maheshwari, R.; Kiick, K. L. *Macromolecules* **2009**, *42*, 3–13.
- (2) Farokhzad, O. C.; Cheng, J.; Teply, B. A.; Sherifi, I.; Jon, S.; Kantoff, P. W.; Richie, J. P.; Langer, R. *Proc. Natl. Acad. Sci.* **2006**, *103*, 6315–6320.
- (3) Dhar, S.; Gu, F. X.; Langer, R.; Farokhzad, O. C.; Lippard, S. J. *Proc. Natl. Acad. Sci.* **2008**, *105*, 17356–17361.
- (4) Lee, C. C.; MacKay, J. A.; Frechet, J. M. J.; Szoka, F. C. *Nat. Biotech.* **2005**, *23*, 1517–1526.
- (5) Almutairi, A.; Rossin, R.; Shokeen, M.; Hagooley, A.; Ananth, A.; Capoccia, B.; Guillaudeu, S.; Abendschein, D.; Anderson, C. J.; Welch, M. J.; Fréchet, J. M. J. *Proc. Natl. Acad. Sci.* **2009**, *106*, 685–690.
- (6) Lee, R. J.; Low, P. S. *J. Biol. Chem.* **1994**, *269*, 3198–3204.
- (7) Liong, M.; Lu, J.; Kovichich, M.; Xia, T.; Ruehm, S. G.; Nel, A. E.; Tamanoi, F.; Zink, J. I. *ACS Nano* **2008**, *2*, 889–896.
- (8) Medarova, Z.; Pham, W.; Farrar, C.; Petkova, V.; Moore, A. *Nat. Med.* **2007**, *13*, 372–377.
- (9) Javier, D. J.; Nitin, N.; Levy, M.; Ellington, A.; Richards-Kortum, R. *Bioconjug. Chem.* **2008**, *19*, 1309–1312.
- (10) Hilgenbrink, A. R.; Low, P. S. *J. Pharm. Sci.* **2005**, *94*, 2135–2146.
- (11) Sudimack, J.; Lee, R. J. *Adv. Drug Deliv. Rev.* **2000**, *41*, 147–162.
- (12) Petrus, A. K.; Fairchild, T. J.; Doyle, R. P. *Angew. Chem. Int. Ed.* **2009**, *48*, 1022–1028.

- (13) Gupta, Y.; Kohli, D.; Jain, S. *Crit. Rev. Therap. Drug Carrier Syst.* **2008**, *48*, 347–379.
- (14) Darbre, T.; Reymond, J. *Curr. Top. Med. Chem.* **2008**, *8*, 1286–1293.
- (15) Brown, K. C. *Curr. Op. Chem. Biol.* **2000**, *4*, 16–21.
- (16) Schrama, D.; Reisfeld, R. A.; Becker, J. C. *Nat. Rev. Drug Discovery* **2006**, *5*, 147–159.
- (17) Lee, J. F.; Stovall, G. M.; Ellington, A. D. *Curr. Op. Chem. Biol.* **2006**, *10*, 282–289.
- (18) Allen, T. M.; Cullis, P. R. *Science* **2004**, *303*, 1818–1822.
- (19) Valegård, K.; Liljas, L.; Fridborg, K.; Unge, T. *Nature* **1990**, *345*, 36–41.
- (20) Mastico, R. A.; Talbot, S. J.; Stockley, P. G. *J Gen Virol* **1993**, *74*, 541–548.
- (21) Hooker, J. M.; O’Neil, J. P.; Romanini, D. W.; Taylor, S. E.; Francis, M. B. *Mol. Imaging Biol.* **2008**, *10*, 182–191.
- (22) Hooker, J. M.; Datta, A.; Botta, M.; Raymond, K. N.; Francis, M. B. *Nano Lett.* **2007**, *7*, 2207–2210.
- (23) Garimella, P. D.; Datta, A.; Romanini, D. W.; Raymond, K. N.; Francis, M. B. *J. Am. Chem. Soc.* **2011**, *133*, 14704–14709.
- (24) Wu, W.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B. *Angew. Chem. Int. Ed.* **2009**, *48*, 9493–9497.
- (25) Keefe, A. D.; Cload, S. T. *Curr. Op. Chem. Biol.* **2008**, *12*, 448–456.
- (26) Eulberg, D.; Klusmann, S. *ChemBioChem* **2003**, *4*, 979–983.
- (27) Li, M.; Lin, N.; Huang, Z.; Du, L.; Altier, C.; Fang, H.; Wang, B. *J. Am. Chem. Soc.* **2008**, *130*, 12636–12638.
- (28) Vartanian, J.-P.; Henry, M.; Wain-Hobson, S. *Nucl. Acids Res.* **1996**, *24*, 2627–2631.
- (29) Bittker, J. A.; Le, B. V.; Liu, D. R. *Nat. Biotech.* **2002**, *20*, 1024–1029.
- (30) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818–822.
- (31) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505–510.
- (32) Shangguan, D.; Li, Y.; Tang, Z.; Cao, Z. C.; Chen, H. W.; Mallikaratchy, P.; Sefah, K.; Yang, C. J.; Tan, W. *Proc. Natl. Acad. Sci.* **2006**, *103*, 11838–11843.

- (33) Shangguan, D.; Tang, Z.; Mallikaratchy, P.; Xiao, Z.; Tan, W. *Chembiochem* **2007**, *8*, 603–606.
- (34) Shangguan, D.; Cao, Z.; Meng, L.; Mallikaratchy, P.; Sefah, K.; Wang, H.; Li, Y.; Tan, W. *J. Proteome Res.* **2008**, *7*, 2133–2139.
- (35) Macdonald, I. J.; Dougherty, T. J. *J. Porphyrins and Phthalocyanines* **2001**, *5*, 105–129.
- (36) Hatz, S.; Poulsen, L.; Ogilby, P. R. *Photochemistry and Photobiology* **2008**, *84*, 1284–1290.
- (37) Maxfield, F. R.; McGraw, T. E. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 121–132.
- (38) Hsiao, S. C.; Shum, B. J.; Onoe, H.; Douglas, E. S.; Gartner, Z. J.; Mathies, R. A.; Bertozzi, C. R.; Francis, M. B. *Langmuir* **2009**, *25*, 6985–6991.
- (39) Carrico, Z. M.; Romanini, D. W.; Mehl, R. A.; Francis, M. B. *Chem. Commun.* **2008**, 1205–1207.
- (40) Endo, M.; Fujitsuka, M.; Majima, T. *Chem. Euro. J.* **2007**, *13*, 8660–8666.
- (41) Chen, Y.; Parr, T.; Holmes, A. E.; Nakanishi, K. *Bioconjugate Chem.* **2007**, *19*, 5–9.
- (42) Shah, S.; Friedman, S. H. *Nat. Protocols* **2008**, *3*, 351–356.

Development of a palladium-catalyzed reaction for the conjugation of pyrene to proteins surfaces

4.1 Abstract

An amphiphilic pyrene taurine carbamate substrate was synthesized and characterized for the development of a palladium-catalyzed tyrosine-selective bioconjugation reaction, and was demonstrated to be reactive with a number of protein substrates. In particular, the reactivity was developed on the protein H-NOX. Mass spectrometry revealed that H-NOX was highly reactive and was modified up to three times with pyrene. In addition, an H-NOX mutant, K186W H-NOX, was expressed and tested under the same reaction conditions. K186W H-NOX was hypothesized to be more reactive due to the introduction of a tryptophan residue adjacent to a surface-exposed tyrosine. However, the mutant displayed lower reactivity than wild-type H-NOX. To determine the exact sites of modification on wild-type H-NOX, in-gel digests of pyrene-modified H-NOX samples were analyzed by tandem mass spectrometry. Lysine 120 was identified as one of the sites of modification, suggesting that unique lysines can be reactive under the conditions for the palladium-catalyzed alkylation reaction.

4.2 Introduction

As demonstrated in Chapter 3 and in many other reports, proteins have emerged as versatile scaffolds for the development of materials that require well-defined, nanoscale structures.¹⁻³ These scaffolds are provided by the orientation of the various protein amino acid side chains, which can serve as reactive handles for the bioconjugation of small molecules. These protein conjugates have been developed into a variety of light-harvesting, imaging, drug-delivery and electronic materials.⁴⁻⁶

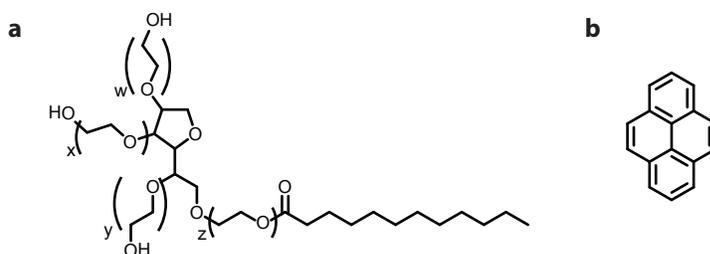


Figure 4-1. Common molecules for conjugation to carbon nanotubes: (a) Tween 20 ($w+x+y+z=20$) and (b) pyrene.

An emerging field of materials research involves the coupling of biomolecules to carbon nanotubes to create highly sensitive field-effect transistors.⁷⁻¹¹ The basis for making protein-nanotube sensors stems from the extreme sensitivity of nanotube conductivity to the surrounding environment. Thus, protein binding events and conformational changes are expected to give rise to significant changes in electrical readouts when coupled to nanotube-based devices. Early efforts have indeed shown that proteins bound to nanotubes through non-specific interactions can function as sensors; however in these reports it was difficult to control the orientation of proteins on the surface of the nanotubes.^{12,13} In addition, not all proteins are capable of non-specifically binding to nanotube surfaces.¹⁴ As an alternative more specific approach, *N*-hydroxysuccinimidyl

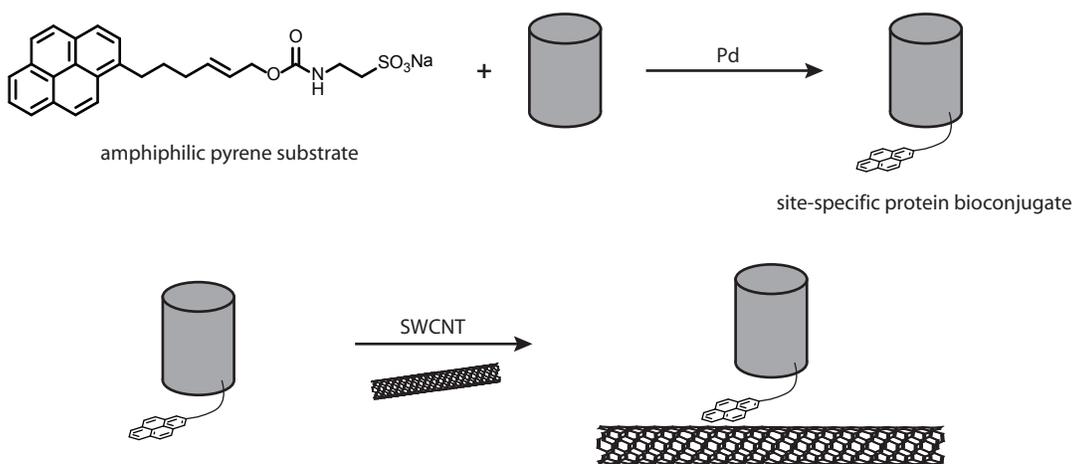


Figure 4-2. Pyrene modification of proteins for conjugation to carbon nanotubes. A water-soluble pyrene substrate can be used to attach pyrene to proteins, and the resulting bioconjugate should adsorb to the surface of carbon nanotubes through π - π stacking interactions.

(NHS) esters have been synthesized either directly on the surface of nanotubes¹⁵ or coupled to Tween 20 and pyrene (Figure 5-1).^{9,14} These latter two molecules are known to adsorb strongly to the surface of carbon nanotubes through hydrophobic effects and π - π stacking interactions, and thus serve as stable anchors. Subsequent coupling of proteins to the NHS esters through their lysine residues has yielded functional devices. However, lysine chemistry is generally not site-selective, and it can still be difficult to control protein orientation. Therefore, we sought to develop a more site-selective approach to install pyrene onto the surface of proteins for the development of homogeneous protein-nanotube hybrids (Figure 5-2).

In order to attach pyrene groups to proteins in a more well-defined manner, site-selective bioconjugation methods are required. Typically, bioconjugation reactions target nucleophilic amino acid side chains such as lysine, aspartic acid, glutamic acid and cysteine. Unfortunately, there are often many lysine, aspartic acid and glutamic acid residues at various locations on the surface of most proteins, making it difficult to control the location and number of modifications. In contrast, cysteine modification can be extremely site-selective due to its low occurrence as a free thiol on protein surfaces. However site-directed mutagenesis is often required to introduce cysteines onto the surface of proteins, and in some cases, these mutations at the genetic level can lead to complications in protein expression, purification, and storage. In an effort to expand the available set of site-selective bioconjugation strategies, the Francis group has developed several additional reactions that can selectively target the N-termini of proteins^{16,17} and aromatic amino acids that occur less frequently on the surface of proteins.¹⁸⁻²⁰ Commonly, there are fewer aromatic amino acids than lysines on the surfaces of proteins, and in the case of certain proteins, these

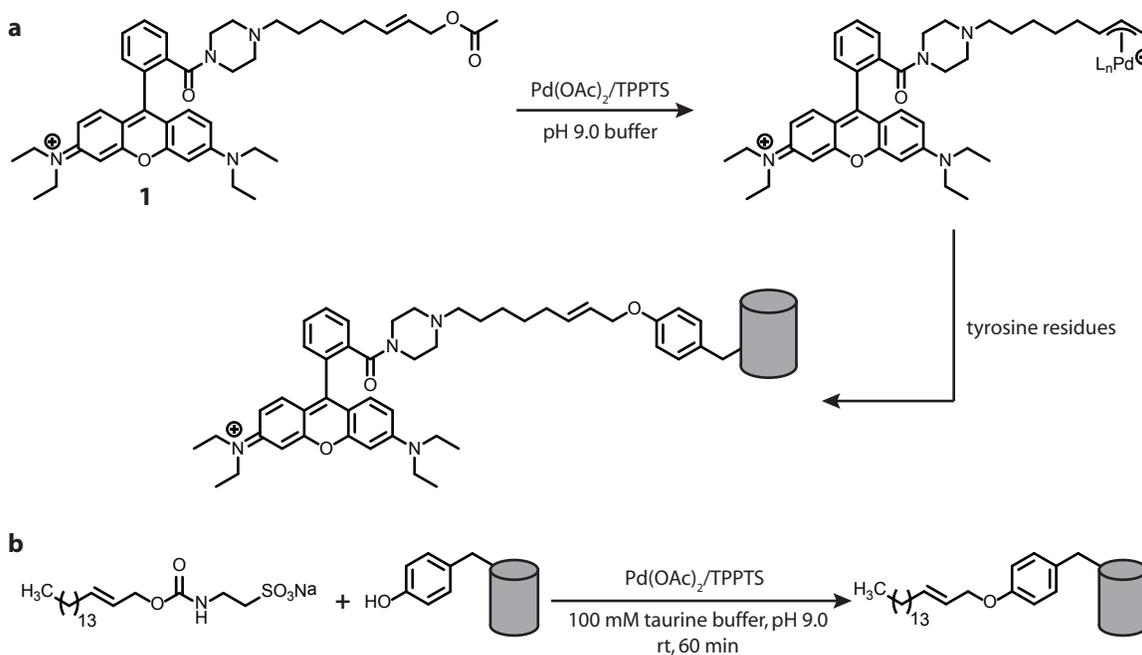


Figure 4-3. A general scheme showing the palladium-catalyzed alkylation of tyrosine residues. **(a)** The reaction proceeds *via* a π -allylpalladium intermediate, and has been demonstrated for the modification of proteins with rhodamine allyl acetate (**1**). **(b)** This strategy has also been used for the generation of semi-synthetic lipoproteins. Amphiphilic molecules containing a water-solubilizing leaving group can act as substrates for the palladium-catalyzed reaction to install entirely hydrophobic groups onto the surface of proteins.

modifications can be site-selective.²⁰

Among these reactions that target aromatic amino acids is the tyrosine-selective alkylation reaction briefly mentioned in Chapter 2. This reaction involves a π -allylpalladium intermediate that is generated from an allylic substrate. Previous reports had demonstrated that a π -allylpalladium complex could mediate the *O*-allylation of phenolic groups in organic solvent.²¹ The reaction could tolerate a variety of substrates, as long as they contained a leaving group in the allylic position that could be displaced upon addition of the active Pd⁰ catalyst. For proteins, the analogous *O*-allylation would take place on tyrosine residues. Using a water-soluble palladium complex, the reaction was demonstrated by our group to proceed in aqueous solution, and was able to modify proteins with a rhodamine allyl acetate dye (**1**, Figure 5-3a).²² Most notably, the advantage of this reaction is its ability to switch the solubility of the substrate. Recently, our group has shown that the allylic modification can consist of a completely hydrophobic chain as long as the π -allyl precursor contains a water-solubilizing leaving group (Figure 5-3b). This strategy for installing hydrophobic groups onto proteins is extremely useful, since protein reactions are run in aqueous solutions and are limited by substrate solubility. As such, this reaction is well-suited for the addition of hydrophobic pyrene groups onto proteins. To explore this possibility, we herein report the synthesis of an amphiphilic pyrene substrate and its use in a palladium-catalyzed alkylation of a number of protein targets, including a protein that has potential for the development of nanotube-based biosensors – the protein H-NOX.

4.3 Results & Discussion

In designing a substrate, it was important to choose an allylic leaving group that would be sufficiently polar to solubilize pyrene in aqueous solution. Taurine was chosen as the water-solubilizing component because it was previously shown that the negatively charged sulfonate group imparted a substantial amount of water-solubility to extremely hydrophobic molecules.²² Starting from pyrenebutyric acid the allyl taurine carbamate-functionalized pyrene substrate (**7**)

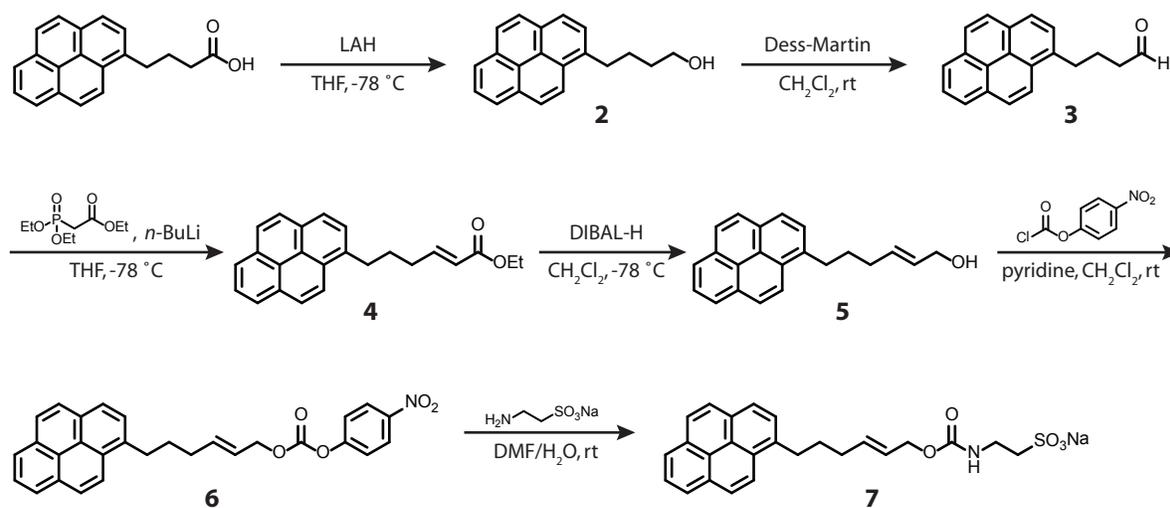


Figure 4-4. Synthesis of a pyrene-containing substrate (**7**) for the palladium-catalyzed alkylation reaction. The molecule was made water-soluble by the incorporation of an allylic taurine carbamate moiety.

was synthesized in six steps (Figure 5-4). The final product was successively recrystallized from ethanol. However, not all of the taurine could be removed, with 12 mol% remaining in the final product. However, since the palladium-catalyzed reaction has been demonstrated to proceed in taurine buffer, we decided to continue with the crude isolated product.

Following synthesis, the palladium allylation methodology was first tested on the protein substrate chymotrypsinogen A (ChyA). This model protein was chosen because it was previously shown to possess tyrosine residues that participate in this reaction. Generally, the palladium-catalyzed reaction was run in a 100 mM taurine buffer at pH 9.0 for 1 hour at room temperature, and the active catalyst was prepared from Pd(OAc)₂ and five equivalents of water-soluble triphenylphosphine trisulfonate (TPPTS). For initial experiments, **7** was used as the substrate for the palladium-catalyzed alkylation using the same general procedure. A stock solution of **7** was prepared in DMSO such that upon dilution into the reaction mixture, there was generally no more than 5% co-solvent. Since pyrene is a fluorescent molecule, conversion to the modified product was qualitatively determined by visualizing the fluorescence of the protein band after SDS-PAGE. We were pleased to find that **7** indeed reacted with ChyA, yielding similar results to previously run lipidation experiments.²²

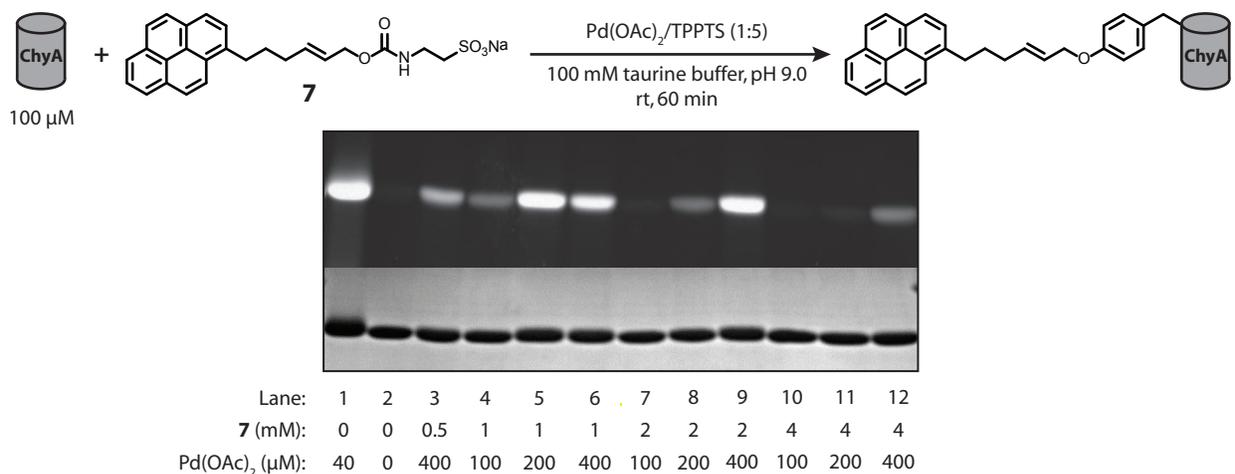


Figure 4-5. Optimization of reaction conditions for **7** and chymotrypsinogen A (ChyA). The general reaction scheme for the modification of ChyA with **7** in the presence of the palladium catalyst is shown on top. The concentration of **7** and palladium were varied and reactions were analyzed by SDS-PAGE. The top panel shows the fluorescence image of the gel, and the bottom panel shows the Coomassie-stained gel. Optimal conversion is obtained with 100 μM chymotrypsinogen A, 1 mM **7** and 200 μM Pd(OAc)₂. Reactions analyzed in lanes 1 and 2 used 1 mM rhodamine allyl acetate (**1**) and represent positive and negative controls, respectively.

Next, the reaction conditions were screened to optimize the conversion to the pyrene-modified product. While keeping the protein concentration and 1:5 ratio of Pd(OAc)₂ to TPPTS constant, the Pd(OAc)₂ concentration and the concentration of **7** were varied to determine which conditions would lead to the highest amount of conversion. Fluorescence visualization after SDS-PAGE showed that the highest conversion was obtained when reactions were run with 1 mM **7** and 200 μM Pd(OAc)₂, or 2 mM **7** and 400 μM Pd(OAc)₂ (Figure 5-5, lanes 5 and 9). Based on this information, 1 mM **7** and 200 μM Pd(OAc)₂ were chosen as the general reaction conditions for the

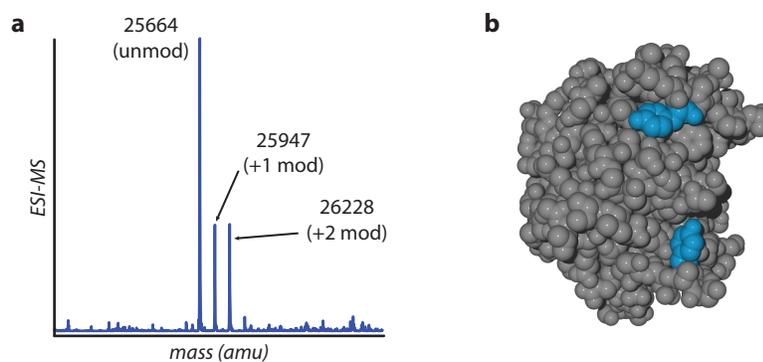


Figure 4-6. Analysis of pyrene-modified chymotrypsinogen A. **(a)** The ESI-MS analysis of ChyA modified with **7** shows peaks corresponding to unmodified, singly, and doubly modified products. Expected masses for the unmodified, singly and doubly modified conjugates are 25664, 25946, and 26228 amu, respectively. The crystal structure of ChyA is shown in **(b)**. The two most solvent-accessible tyrosines are highlighted in blue, Y146 and Y171, and are presumed to be the sites of pyrene-modification.

modification of ChyA because the lower concentrations of the reagents minimized the amount of co-solvent used in the reaction.

To characterize the protein conjugate further, the product was analyzed by mass spectrometry. ESI-LCMS analysis revealed peaks corresponding to the unmodified protein, as well as peaks corresponding to singly and doubly modified species (Figure 5-6a). These results were analogous to results previously obtained from the lipidation of ChyA with other hydrophobic substrates, in which proteolytic digest and subsequent MALDI-TOF MS confirmed the site of modification.²¹ Therefore the sites of modification with **7** were also presumed to be the two solvent-accessible tyrosine residues, Y171 and Y146 (Figure 5-6b).

Following the successful modification of ChyA, we screened a variety of proteins to

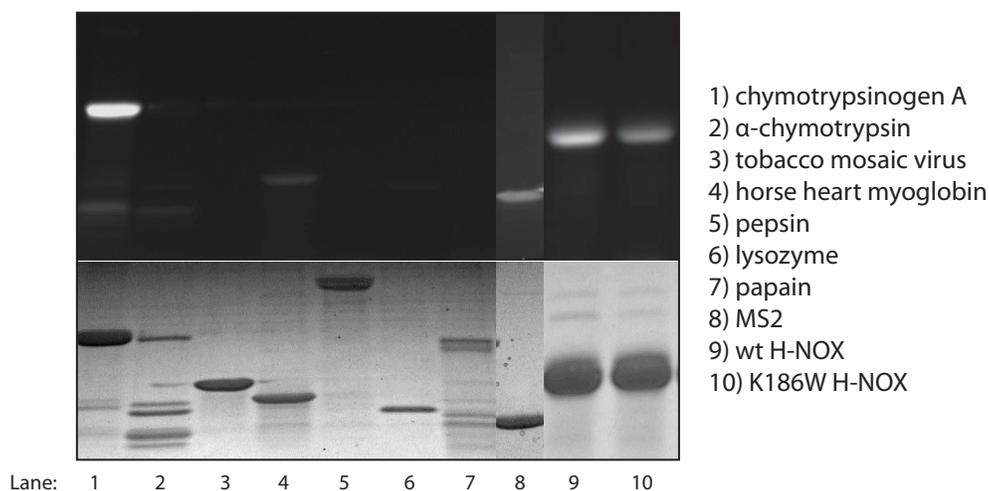


Figure 4-7. Reactivity of **7** with various protein substrates. The reactivity was analyzed by SDS-PAGE. The top panel shows the fluorescence image while the bottom panel shows the Coomassie-stained gel. All reactions were run at room temperature with 200 μ M protein, 1 mM **7**, and 400 μ M Pd(OAc)₂, with the exceptions of lanes 8, 9 and 10, which were run with 100 μ M protein. Precipitation was observed in the reaction corresponding to lane 6.

determine the scope of the palladium-catalyzed alkylation with **7**. Pyrene fluorescence allowed for the convenient use of a gel-based assay to easily screen multiple reaction conditions and protein substrates. As shown in Figure 5-7, horse heart myoglobin, bacteriophage MS2, H-NOX and an H-NOX point mutant (K186W) all underwent modification with **7**, however the reactivity of horse heart myoglobin was very low and bacteriophage MS2 showed only moderate modification.

Among the proteins screened was H-NOX (**H**eme-**N**itric oxide/**O**xygen-binding domain), a heme-containing protein that is known to discriminate between NO and O₂ ligands.²³ Isolated from *Thermoanaerobacter tengcongensis*, H-NOX is a trans-membrane protein of 602 amino acids that contains an N-terminal heme binding domain that can be expressed in *E. coli* as a 188 amino acid water-soluble protein.²⁴ The protein can be easily purified using a heat denaturation protocol due to its thermophilic nature. Denaturation followed by size exclusion and anion-exchange chromatography yielded a pure solution of H-NOX whose concentration was determined by the absorbance of the heme's Soret band. Due to its ability to differentiate between different gases,

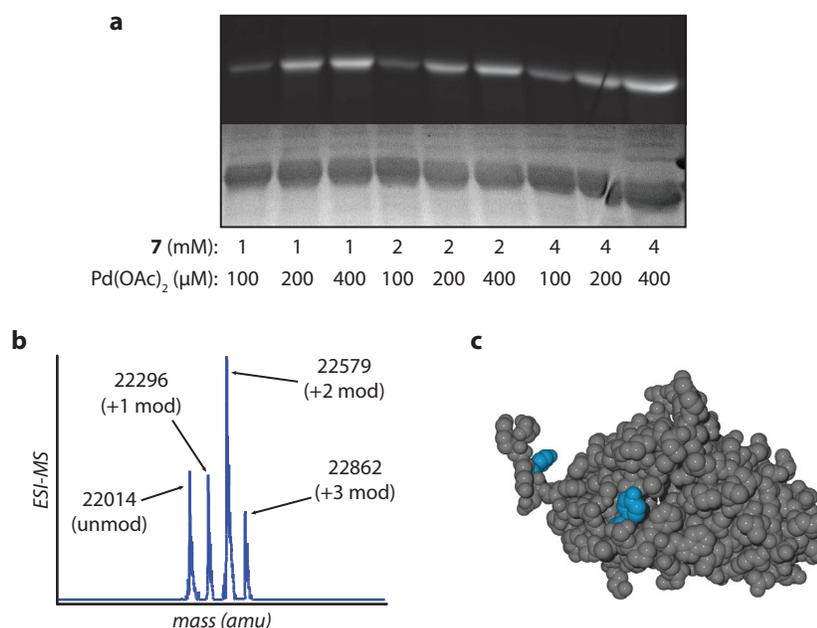


Figure 4-8. SDS-PAGE and ESI-MS analysis of **7** with H-NOX. **(a)** The highest level of conversion, determined by SDS-PAGE analysis followed by fluorescence imaging of protein bands, was achieved at 100 μM H-NOX, 1 mM **7**, and 400 μM Pd(OAc)₂. **(b)** ESI-MS analysis showed up to three additions of pyrene onto H-NOX. The expected masses for the unmodified, singly, doubly, and triply modified bioconjugates are 22014, 22296, 22578, and 22860 amu, respectively. **(c)** The crystal structure of H-NOX, however, shows only two solvent-accessible tyrosine residues, Y85 and Y185.

H-NOX was chosen as a potential candidate for attachment to nanotubes for the development of electrical biosensors.

Also screened alongside with H-NOX was an H-NOX K186W point mutant. ChyA and GFP had previously been shown to be particularly reactive substrates for the palladium-catalyzed allylation reaction, and it was hypothesized that their high reactivity could be due to the micro-environment around the targeted tyrosine residues. Upon inspection of their crystal structures, it was noticed that tyrosines on both Chy A and GFP have adjacent tryptophan residues, an amino acid that is rarely found on the surface of proteins due to its hydrophobic nature. It was

hypothesized that the presence of these surface-exposed tryptophan residue provided a binding site for hydrophobic small molecules, effectively leading to higher local reagent concentration. Thus, we reasoned that the introduction of tryptophan residues adjacent to tyrosines residues could be a general scheme for enhancing tyrosine reactivity with hydrophobic substrates using this reaction. H-NOX provided an interesting system in which to test this theory because its crystal structure showed a highly solvent-accessible tyrosine, Y185.²⁵ The adjacent lysine residue was mutated to a tryptophan residue by site-directed mutagenesis and the K186W mutant was expressed. However when the K186W mutant and wt H-NOX were subjected to the same reaction conditions, it was observed that the mutant displayed decreased reactivity (Figure 5-7, lanes 9 and 10). These data suggest that the presence of an adjacent tryptophan residue may not be solely responsible for increased reactivity.

We chose to focus on the modification of H-NOX due to its potential utility in developing nanotube-based sensors. Again, we screened the reaction conditions to optimize the conversion (Figure 5-8a). The reaction conditions of 1 mM **7** with 400 μ M Pd(OAc)₂ showed the highest levels

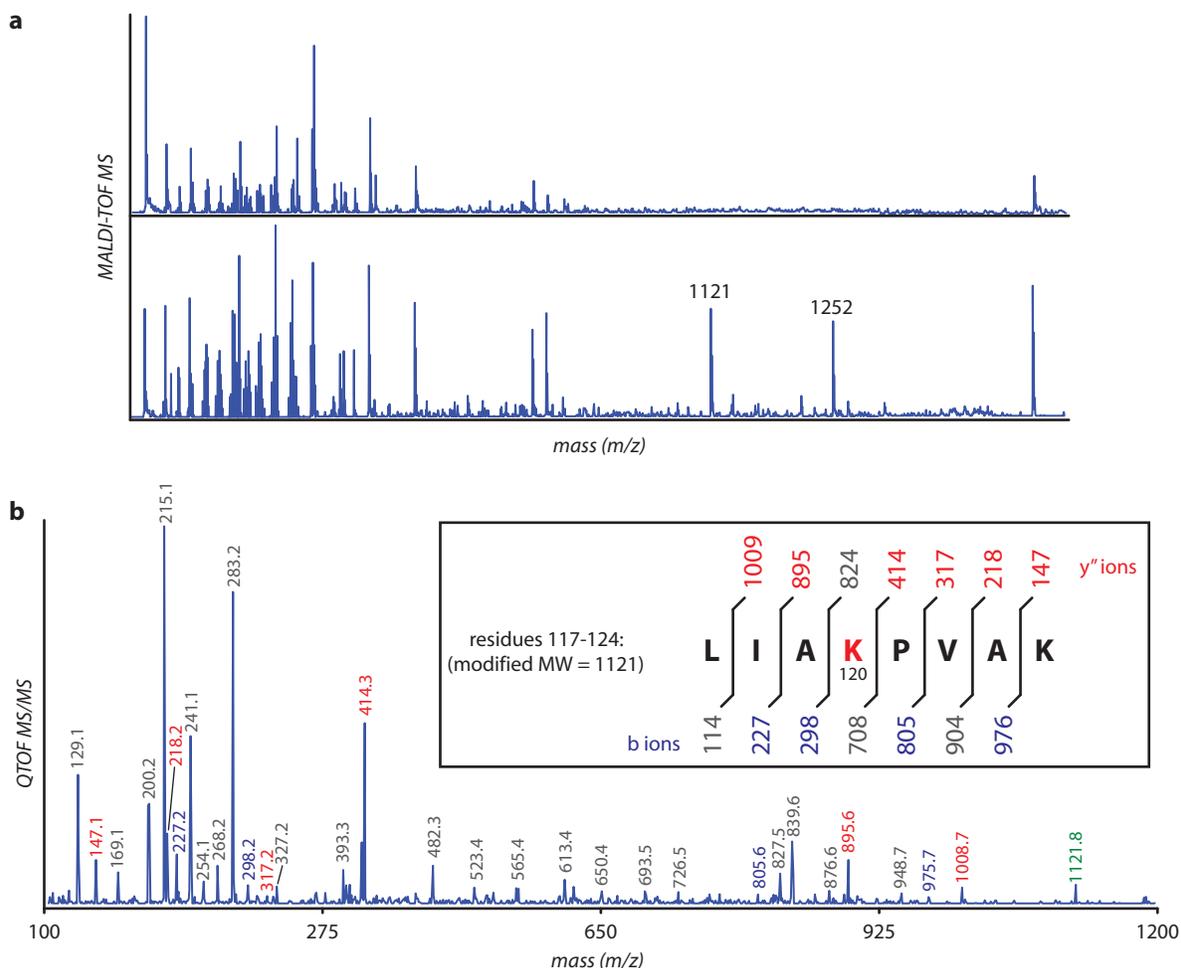


Figure 4-9. Determination of the site of H-NOX modification with **7**. **(a)** Both unmodified (top) and modified (bottom) samples of H-NOX were digested with trypsin. Analysis by MALDI-TOF MS identified the 1121 m/z and 1252 m/z fragments as potentially modified peptides. **(b)** MS/MS data for the 1121 m/z fragment identified lysine 120 as the modification site. The identified b series ions are marked in blue and the identified y series ions are marked in red. The precursor ion is marked in green.

of conversion with minimal co-solvent. We observed that the optimal modification of H-NOX required twice as much catalyst as ChyA. Nonetheless, a sample of H-NOX modified under these conditions was analyzed by ESI-MS, revealing a mixture of unmodified, singly-, doubly- and triply-modified products (Figure 5-8b). These results are surprising due to the fact that the crystal structure of H-NOX shows only two solvent-accessible tyrosine residues. Therefore, we sought to determine the sites of pyrene modification on H-NOX by proteolytic digestion of modified samples.

Samples of unmodified H-NOX and pyrene-modified H-NOX were digested with trypsin, and the resulting peptides were analyzed by MALDI-TOF MS (Figure 5-9a). In comparing the mass spectra for both samples, peptides corresponding to the masses 1121 m/z and 1252 m/z were identified as potentially modified fragments corresponding to the 839 Da and 970 Da predicted tryptic fragments for H-NOX, respectively. Using quadrupole time of flight mass spectrometry (QTOF-MS), we were able to isolate the 1121 m/z precursor ion and sequence it by tandem mass spectrometry (MS/MS). The MS/MS spectrum is shown in Figure 5-9b and revealed that modification for the 1121 m/z fragment occurred on lysine 120. While the 1252 m/z fragment could not be isolated for MS/MS, the corresponding unmodified peptide fragment also contains a lysine but not tyrosine. Therefore, it is likely pyrene modification on the 1252 m/z fragment is occurring at a lysine residue as well. This reaction had previously been demonstrated to be tyrosine-specific, so we were surprised to discover lysine modification. In addition, all reactions were run in 100 mM pH 9.0 taurine buffer (a large excess of primary amines), yet this did not inhibit product formation. Therefore, these data suggest that while lysine modification on H-NOX occurs with this reaction, these lysines appear to be uniquely reactive.

4.4 Conclusion

A pyrene-containing substrate for a previously developed palladium-catalyzed reaction was synthesized and shown to allow the successful attachment of pyrene to the surface of a number of protein substrates. The incorporation of pyrene into proteins was motivated by the construction of protein-nanotube hybrids for developing highly sensitive biosensors. To this end, the protein H-NOX was expressed and optimized for reactivity with the pyrene substrate. H-NOX was chosen because it is innately able to discriminate between NO and O₂ ligands, and therefore could be a good model protein for conferring sensing properties to protein-nanotube-based biosensors. However, while we were able to optimize reaction conditions for modifying H-NOX with pyrene, we also discovered that pyrene modification of H-NOX was not tyrosine-specific. Using MS/MS, we confirmed that at least one modification was occurring at lysine 120, which we believe to be a uniquely reactive lysine residue.

While the palladium-catalyzed alkylation reaction was not as site-selective for our purposes as was previously described, it did serve as the foundation for the work regarding palladium deprotection described in Chapter 2. In addition, there are other applications for this reaction, especially with regards to “solubility-switching.” For example, the semi-synthesis of lipidated proteins still remains challenging. One could envision incorporating amphiphilic molecules containing a water-solubilizing leaving group onto proteins in a site-specific manner, and then using the palladium catalyst to displace the leaving group. This would result in the incorporation of an entirely hydrophobic group on the protein. Therefore, the fundamental properties of this reaction may still prove useful for a variety of other applications.

4.5 Materials and methods

4.5.1 General procedures and materials

Unless otherwise noted, all chemicals were obtained from commercial sources and used without further purification. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F254 plates with visualization by ultraviolet (UV) irradiation at 254 nm and/or staining with potassium permanganate or dinitrophenylhydrazine. Purifications by flash chromatography were performed using EM silica gel 60 (230-400 mesh). The eluting system for each purification was determined by TLC analysis. Chromatography solvents were used without distillation. All reactions were carried out under a nitrogen atmosphere in oven dried glassware unless otherwise noted. All organic solvents were removed under reduced pressure using a rotary evaporator. Dichloromethane (CH_2Cl_2), tetrahydrofuran (THF), and acetonitrile (MeCN) were distilled under a nitrogen atmosphere from calcium hydride. Water (ddH_2O) used in biological procedures or as a reaction solvent was de-ionized using a NANOpure™ purification system (Barnstead, USA). Pyrenebutyric acid was purchased from Aldrich.

4.5.2 Instrumentation and sample analysis

NMR. ^1H and ^{13}C spectra were measured with a Bruker AV-300 (300 MHz) spectrometer or a Bruker AVB-400 (400 MHz) spectrometer as noted. ^1H NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to CDCl_3 (δ 7.26, singlet) and dimethyl sulfoxide- d_6 (δ 2.50, pentet). Multiplicities are reported as follows: s (singlet), d (doublet), t (triplet), q (quartet), dd (doublet of doublets), dt (doublet of triplets), or m (multiplet). Coupling constants are reported as a J value in Hertz (Hz). The number of protons (n) for a given resonance is indicated as nH, and is based on spectral integration values. ^{13}C NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to CDCl_3 (δ 77.2, triplet) and dimethyl sulfoxide- d_6 (δ 39.5, septet).

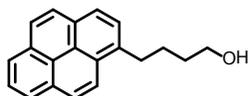
Mass Spectrometry. Fast Atom Bombardment (FAB), Electron Ionization (EI) and QTOF mass spectra were obtained at the UC Berkeley Mass Spectrometry Facility. Electrospray LC/MS analysis was performed using an API 150EX system (Applied Biosystems, USA) equipped with a Turbospray source and an Agilent 1100 series LC pump. Protein chromatography was performed using a Agilent ZORBAX® 300SB-C8 reversed phase column (2.1 mm x 50 mm) with a MeCN:ddH₂O gradient mobile phase containing 0.1% formic acid (250 $\mu\text{L}/\text{min}$) or with a Phenomenex Jupiter™ 300 5 μ C5 300Å reversed-phase column (2.0 mm x 150 mm) with a MeCN:ddH₂O gradient mobile phase containing 0.1% formic acid (250 $\mu\text{L}/\text{min}$). Protein mass reconstruction was performed on the charge ladder with Analyst software (version 1.3.1, Applied Biosystems). Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DETM system (PerSeptive Biosystems, USA). Prior to MS analysis, biological samples were desalted and/or separated from small molecule contaminants using $\mu\text{C}18$ ZipTip® pipet tips (Millipore, USA), NAP-5™ gel filtration columns (Amersham Biosciences, USA), or Strata C-18ETM reversed-phase columns (Phenomenex, USA) as indicated below. All MS data for protein and peptide samples were found to be within 0.1% of the expected values.

High Performance Liquid Chromatography. HPLC was performed on an Agilent 1100

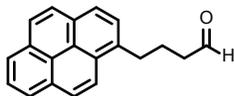
Series HPLC System (Agilent Technologies, USA). Sample analysis for all HPLC experiments was achieved with an in-line diode array detector (DAD) and an in-line fluorescence detector (FLD).

Gel Analyses. For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus from Bio-Rad (Hercules, CA). All protein electrophoresis samples were heated for 10 minutes at 100 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of any disulfide bonds. Gels were run for 5 minutes at 30 V and 70-90 minutes at 120 V to ensure separation of bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250 (Bio-Rad).

4.5.3 Experimental

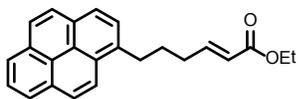


4-Pyren-1-yl-butan-1-ol (2). To a stirred solution of lithium aluminum hydride (21 mL, 1.0 M in THF, 21 mmol, 3.0 equiv) at -78 °C was added 20 mL of a solution of pyrenebutyric acid (2.02 g, 7.00 mmol, 1.0 equiv) in THF via syringe. The reaction was stirred at -78 °C for 1.5 h and then warmed to rt and stirred for an additional 2 h. The reaction was quenched with 10 mL of water and 10 mL of 1 M sulfuric acid, resulting in a cloudy, yellow mixture and some gas evolution. The mixture was diluted with dichloromethane and the resulting layers were separated. The aqueous layer was extracted twice with dichloromethane and then the combined organic layers were washed once with 5% NaHCO₃, once with water, and once with brine. The solution was dried over sodium sulfate, filtered, and concentrated under reduced pressure. The resulting material was purified by silica gel chromatography (30-50% EtOAc in hexanes), yielding 1.71 g of a beige-colored solid (89%). ¹H NMR (400 MHz, CDCl₃) δ (ppm) 1.55 (s, 1H), 1.71 (m, 2H), 1.91 (m, 2H), 3.34 (t, 2H, J = 7.7 Hz), 3.66 (t, 2H, J = 6.6 Hz), 7.84 (d, 1H, J = 7.8 Hz), 8.01 (m, 3H), 8.09 (m, 2H), 8.16 (m, 2H), 8.26 (d, 1H, J = 9.2 Hz). ¹³C NMR (100 MHz, CDCl₃) δ (ppm) 28.1, 32.8, 33.3, 62.9, 123.5, 124.8, 124.9, 125.0, 125.17, 125.22, 125.9, 126.7, 127.35, 127.37, 127.66, 128.74, 129.93, 131.04, 131.57, 136.80. IR (cm⁻¹) 3350, 3041, 1603, 1459, 1434, 1060. mp 79-80 °C. HRMS (EI⁺) calculated for C₂₀H₁₈O ([M]⁺) 274.1358, found 274.1353.

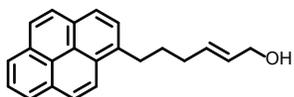


4-Pyren-1-yl-butyraldehyde (3). To a solution of **2** (1.52 g, 5.55 mmol, 1.0 equiv) in 50 mL of undistilled dichloromethane was added 1,1,1-tris(acetyloxy)-1,1-dihydro-1,2-benziodoxol-3-(1H)-one (Dess-Martin periodinane, 2.83 g, 6.66 mmol, 1.2 equiv), and the reaction was stirred for 40 min at rt. The reaction was quenched with 20 mL of a 60 mM aqueous sodium thiosulfate solution (0.3 g, 2 mmol, 0.4 equiv). The reaction mixture was basified with 1 M NaOH and the layers were separated. The aqueous layer was washed twice with dichloromethane, and then the combined organic layers were washed once with water and once with brine. The organic solution was dried over sodium sulfate, filtered, and concentrated to yield a brown solid. Silica gel chromatography (15-30% EtOAc in hexanes) yielded the product as a white solid (1.28 g, 85%). ¹H NMR (400 MHz, CDCl₃) δ (ppm) 2.16 (m, 2H), 2.51 (dt, 2H, J = 1.3, 7.2 Hz), 3.33 (t, 2H, J = 7.7 Hz), 7.81 (d, 1H, J = 7.8 Hz), 8.01 (m, 3H), 8.11 (m, 2H), 8.18 (m, 2H), 8.26 (d, 1H, J = 9.2 Hz), 9.76 (t, 1H, J = 1.3 Hz). ¹³C NMR (100 MHz, CDCl₃) δ (ppm) 24.0, 32.6, 43.5, 123.3, 124.9, 125.1, 125.2, 126.0, 126.9, 127.4, 127.57, 127.59,

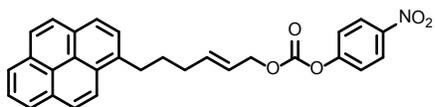
128.8, 130.1, 131.0, 131.5, 135.6, 202.3. IR (cm⁻¹) 3041, 2822, 2722, 1721, 1603, 1459. mp 71-72 °C. HRMS (EI⁺) calculated for C₂₀H₁₆O ([M]⁺) 272.1201, found 272.1204.



(E)-6-Pyren-1-yl-hex-2-enoic acid, ethyl ester (4). To a solution of triethylphosphonoacetate (1.1 mL, 5.4 mmol, 1.2 equiv) in 2 mL of dry THF at -78 °C was added *n*-butyllithium (2.3 mL, 2.2 M in hexanes, 5.1 mmol, 1.1 equiv) dropwise, and then the reaction mixture was warmed to rt. The reaction mixture remained colorless, and so *n*-BuLi was added dropwise until the solution became bright orange. The reaction mixture was cooled to -78 °C and to it was added dropwise a solution of **3** (1.27 g, 4.66 mmol, 1.0 equiv) in 15 mL of THF, resulting in a pale yellow reaction mixture. The reaction was stirred for 1.5 h at -78 °C and then diluted with diethyl ether. The resulting solution was transferred to a separatory funnel and washed twice with distilled water. The combined aqueous layers were extracted with dichloromethane, and then the combined organic layers were washed once with brine and dried over sodium sulfate. Purification on silica gel (1-10% EtOAc in hexanes) and concentration *in vacuo* overnight yielded product **4** as a pale yellow solid (0.989 g, 62%). ¹H NMR (300 MHz, CDCl₃) δ (ppm) 1.32 (t, 3H, J = 7.1 Hz), 2.03 (m, 2H), 2.36 (m, 2H), 3.35 (t, 2H, J = 7.7 Hz), 4.23 (q, 2H, J = 7.1 Hz), 5.91 (dt, 1H, J = 1.5, 15.7 Hz), 7.07 (dt, 1H, J = 6.9, 15.6 Hz), 7.84 (d, 1H, J = 7.8 Hz), 7.97-8.25 (m, 8H). ¹³C NMR (75 MHz, CDCl₃) δ (ppm) 14.5, 30.1, 32.2, 33.0, 60.4, 122.0, 123.4, 124.9, 125.08, 125.14, 125.2, 126.0, 126.8, 127.4, 127.5, 127.6, 128.8, 130.1, 131.0, 131.6, 136.1, 148.8, 166.8. IR (cm⁻¹) 3041, 1715, 1654, 1603, 1462, 1268, 1183, 1043. mp 39-40 °C. HRMS (FAB⁺) calculated for C₂₄H₂₂O₂ ([M]⁺) 342.1620, found 342.1612.

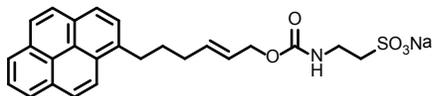


(E)-6-Pyren-1-yl-hex-2-en-1-ol (5). To a solution of **4** (0.984 g, 2.87 mmol, 1.0 equiv) in 10 mL of dichloromethane at -78 °C was added dropwise a solution of DIBAL-H (8.6 mL, 1.0 M in hexanes, 8.6 mmol, 3.0 equiv), and the reaction mixture was stirred for 1 h at -78 °C. The reaction was quenched with saturated NH₄Cl, resulting in a gelatinous solid. After warming to room temperature, the biphasic mixture was filtered through Celite and the layers were separated. The organic layer was washed once with brine, dried over sodium sulfate, then concentrated *in vacuo*. Purification on silica gel (35-50% EtOAc in hexanes) yielded product **5** as an off-white solid (0.660 g, 77%). ¹H NMR (400 MHz, CDCl₃) δ (ppm) 1.56 (s, 1H), 1.95 (m, 2H), 2.22 (m, 2H), 3.34 (t, 2H, J = 7.8 Hz), 4.11 (m, 2H), 5.73 (m, 2H), 7.85 (d, 1H, J = 7.8 Hz), 8.01 (m, 3H), 8.10 (m, 2H), 8.17 (m, 2H), 8.26 (d, 1H, J = 9.3 Hz). ¹³C NMR (100 MHz, CDCl₃) δ (ppm) 31.3, 32.3, 33.1, 63.9, 123.6, 124.9, 124.95, 125.03, 125.2, 125.3, 126.0, 126.8, 127.36, 127.43, 127.7, 128.8, 129.8, 130.0, 131.1, 131.6, 132.8, 136.8. IR (cm⁻¹) 3351, 3040, 1669, 1603, 1458, 1435, 1183, 1090. mp 72-73 °C. HRMS (FAB⁺) calculated for C₂₂H₂₀O ([M]⁺) 300.1514, found 300.1516.



Carbonic acid 4-nitro-phenyl ester (E)-6-Pyren-1-yl-hex-2-enyl ester (6). To a solution of 4-nitrophenyl chloroformate (0.455 g, 2.23 mmol, 1.1 equiv) in 10 mL of dichloromethane under nitrogen was added dropwise a solution of **5** (0.610 g, 2.03 mmol, 1.0 equiv) and distilled pyridine (0.4 mL, 5 mmol, 2.5 equiv) in 10 mL of dichloromethane. After stirring for 5 min at rt, the reaction mixture was washed three times with 0.1 M HCl, twice with saturated NaHCO₃, once with water, once with brine, and then

dried over sodium sulfate. Purification on silica gel (5-20% EtOAc in hexanes) yielded the product **6** as a yellow solid (0.615 g, 65%). ¹H NMR (300 MHz, CDCl₃) δ (ppm) 2.00 (m, 2H), 2.28 (m, 2H), 3.36 (t, 2H, J = 7.7 Hz), 4.74 (d, 2H, J = 6.7 Hz), 5.72 (m, 1H), 6.00 (m, 1H), 7.35 (d, 2H, J = 9.2 Hz), 7.86 (d, 1H, J = 7.8 Hz), 7.96-8.28 (m, 10H). ¹³C NMR (75 MHz, CDCl₃) δ (ppm) 30.9, 32.3, 33.1, 70.1, 121.8, 121.9, 123.1, 123.5, 124.9, 125.0, 125.1, 125.2, 125.3, 125.4, 125.7, 126.0, 126.8, 127.43, 127.45, 127.7, 128.8, 130.0, 131.0, 131.6, 136.5, 138.5, 145.5, 152.5, 155.7. IR (cm⁻¹) 3052, 1767, 1527, 1348, 1272, 1205. mp 92-94 °C. HRMS (FAB⁺) calculated for C₂₉H₂₃NO₅ ([M]⁺) 465.1576, found 465.1565.



(E)-6-Pyren-1-yl-hex-2-enyl taurine carbamate (7).

Taurine (0.287 g, 2.29 mmol, 2.2 equiv) was dissolved in a concentrated aqueous solution of NaOH such that a stoichiometric amount of NaOH relative to taurine was present. The solution was subsequently diluted to 2.5 mL with distilled water. The taurinate solution was added dropwise to a rapidly stirred solution of **6** (0.495 g, 1.06 mmol, 1.0 equiv) in 10 mL of dry DMF. Upon completion of the addition, the reaction mixture became cloudy and yellow and the reaction was observed to be complete by TLC. The solvents were removed under reduced pressure and the yellow material was sonicated in MeCN to yield a suspension. The suspension was loaded onto a Celite column (with a frit at the top) and washed with several portions of MeCN until the yellow color was removed. The product was eluted with a solution of acetic acid (25%) in MeCN until the eluant was no longer fluorescent. The solvents were removed under reduced pressure and the material was recrystallized three times from hot ethanol. Product **7** was recovered via filtration as an off-white solid (0.133 g, 26%). ¹H NMR (400 MHz, DMSO-*d*₆) δ (ppm) 1.85 (m, 2H), 2.18 (m, 2H), 2.56 (t, 2H, J = 7.3 Hz), 3.25 (m, 2H), 4.42 (d, 2H, J = 6.0 Hz), 5.59 (m, 1H), 5.83 (m, 1H), 6.95 (t, 1H, J = 5.4 Hz), 7.95 (d, 1H, J = 7.8 Hz), 8.05 (t, 1H, J = 7.6 Hz), 8.12 (m, 2H), 8.20-8.36 (m, 5H). ¹³C NMR (125 MHz, DMSO-*d*₆) δ (ppm) 30.9, 31.5, 32.1, 37.3, 50.7, 64.2, 123.5, 124.17, 124.23, 124.8, 124.9, 125.0, 125.7, 126.1, 126.5, 127.3, 127.48, 127.51, 128.1, 129.3, 130.4, 130.9, 134.2, 136.7, 155.7. IR (cm⁻¹) 3041, 1699, 1534, 1274, 1206, 1064. mp 190-197 °C. HRMS (FAB⁺) calculated for C₂₅H₂₅NNaO₅S ([M+H]⁺) 474.1351, found 474.1346.

Preparation of catalyst/ligand solution for protein modification reactions. Commercially available TPPTS contained 10% phosphine oxide as determined by ³¹P NMR, and a stoichiometric amount of DMSO as determined by ¹H NMR. Taking these two impurities into account, the mass of TPPTS was measured such that 5 equivalents of unoxidized TPPTS was mixed with Pd(OAc)₂. Generally, 3 mg of Pd(OAc)₂ and 48 mg of Strem TPPTS were added to a vial and dissolved in 1.67 mL of distilled water. The mixture was briefly sparged with nitrogen and then sonicated until the mixture became homogeneous. The homogeneous solution was then sparged again for about 30 min. The final solution contains 8 mM Pd(OAc)₂ and 40 mM TPPTS. The catalyst solution was stored under argon as small aliquots in sealed ampules for future use.

Typical protein reaction with pyrene taurine carbamate 7. Generally, the palladium-catalyzed protein reactions were run on a 10 μL scale. In addition to pre-prepared catalyst solution, a 20 mM solution of **7** was prepared in DMSO. For the protein reaction, 2.7 μL of a 370 μM solution of H-NOX (final concentration = 100 μM, 1 equiv) in 100 mM taurine buffer, pH 9.0, were first diluted with 6.3 μL of additional taurine buffer. 0.5 μL (final concentration = 1 mM, 10

equiv, 5% DMSO) of the solution of **7** and 0.5 μL of the catalyst solution (final concentration = 400 μM $\text{Pd}(\text{OAc})_2$, 4 equiv; 2 mM TPPTS, 20 equiv) were added to the protein reaction mixture. The reaction mixture was briefly vortexed and then incubated at rt for 1 h. 2 μL of the reaction mixture was diluted into 18 μL of loading buffer containing 10 mM DTT for SDS-PAGE analysis.

Preparation of pyrene labeled H-NOX for MS analysis. A sample of modified H-NOX was prepared as described above. The reaction mixture was purified from small molecules using a NAP-5 column. To do so, the reaction mixture was first diluted to 200 μL with taurine buffer and loaded onto a pre-equilibrated NAP-5 column. An additional 300 μL of taurine buffer were added. Finally, the protein was eluted with 1 mL of taurine buffer and the eluent was concentrated by centrifugation using 30K MWCO spin concentrators. The final protein solution was purified by RP-HPLC, and the purified protein was analyzed by QTOF-MS.

Expression and purification of wild type H-NOX. H-NOX was expressed according to a modified literature procedure;²⁴ the specific modification was the buffer used for FPLC purification. The same buffer (referred to hereafter as H-NOX buffer) was used for both size-exclusion chromatography (SEC) and anion-exchange chromatography: 50 mM triethanolamine (TEA), pH 7.5, 20 mM NaCl, 5% glycerol, and 5 mM dithiothreitol (DTT). Tuner DE3 plysS competent cells were transformed via heat shock with a pET-20b expression plasmid containing the H-NOX gene. The cells were plated on and selected from chloramphenicol/ampicillin plates for culture in Luria Bertani broth at 37 °C. When the cultures reached an $\text{OD}_{600} \sim 0.5-0.8$, the cultures were allowed to cool to 27 °C before adding amino levulinic acid (1 mM) and IPTG (10 μM) for induction. The cultures were incubated at 25 °C for 12-18 h, after which the cells were collected by centrifugation (15 min; 7,000 rpm; 4 °C) and the pellets were stored at -80 °C.

For purification, the pellets were thawed and re-suspended in H-NOX buffer (with 1 mM Pefablock). DNase and lysozyme were added and the cells were lysed by sonication. The lysate was centrifuged in a Ti-45 rotor (45 min; 42,000 rpm; 4 °C), and the supernatant was collected and heat denatured at 75 °C for 40 min. Again, the sample was centrifuged (1 hr; 42,000 rpm; 4 °C), followed by spin-concentration of the supernatant. The volume was reduced, and the concentrated protein was sterile filtered through a 0.22-0.45 μm Millipore filter. The filtrate was then purified by SEC and anion-exchange chromatography using H-NOX buffer as the eluent. The resulting protein was spin-concentrated yielding a dark-red solution. For experiments, the protein concentration was determined by the absorbance of the Soret band.

QuikChange mutagenesis of H-NOX. For generating the K186W plasmid, QuikChange mutagenesis was performed. The following primers were used:

forward: 5'-CCCCGTTTTTGAGTATTGGAAAAATTGAGCGGCCG-3'
reverse: 5'-CGGCCGCTCAATTTTTCCAATACTCAAAAACGGGG-3'

The mutant was expressed and purified in the same manner as described above for wt H-NOX, and the point mutation was confirmed by ESI-MS. For experiments, the protein concentration was determined by the absorbance of the Soret band.

Protocol for in-gel digest of H-NOX. The UCSF in-gel digestion protocol was followed.

Briefly, a modified sample of H-NOX was separated by SDS-PAGE. The fluorescent pyrene-modified H-NOX band was excised from the gel, and diced into small 1 mm² pieces and placed into a 0.6 mL eppendorf tube. A 25 mM NH₄HCO₃/50% MeCN solution (~100 μL) was added and the mixture was vortexed for 10 min. The supernatant was discarded using a gel loading pipet tip, and the rinsing was repeated two more times. The resulting gel pieces were placed in a Speed Vac and dried to completeness. Trypsin (12.5 ng/μL in 25 mM NH₄HCO₃) was added to just barely cover the gel pieces, and the gel was rehydrated on ice for 10 min. The tube was spun, and additional 25 mM NH₄HCO₃ buffer was added as needed to cover the gel pieces. The mixture was briefly vortexed, and then incubated at 37 °C overnight. After incubation, the digest solution was transferred to a clean 0.6 mL Eppendorf tube. A 50% MeCN/5% formic acid solution (30 μL) was added to the gel pieces, the mixture was vortexed for 30 min, and then sonicated for 5 min. This sequence was repeated one more time, and the solution was extracted and combined with the digest solution. The resulting solution was placed on a Speed Vac to reduce the volume. The resulting solution was then analyzed by QTOF MS/MS.

4.6 References

- (1) Douglas, T.; Young, M. *Nature* **1998**, *393*, 152.
- (2) Douglas, T.; Young, M. *Adv. Mater.* **1999**, *11*, 679.
- (3) Schlick, T.L.; Ding, Z.; Kovacs, E.W.; Francis, M.B. *J. Am. Chem. Soc.* **2005**, *127*, 3719.
- (4) Miller, R.A.; Presley, A.D.; Francis, M.B. *J. Am. Chem. Soc.* **2007**, *129*, 3104.
- (5) Lee, L.A.; Wang, Q. *Nanomedicine* **2006**, *2*, 137.
- (6) Douglas, T.; Young, M. *Science* **2006**, *312*, 873.
- (7) Star, A.; Gabriel, J.P.; Bradley, K.; Grüner, G. *Nano Lett.* **2003**, *3*, 459.
- (8) Bradley, K.; Briman, M.; Star, A.; Grüner, G. *Nano Lett.* **2004**, *4*, 253.
- (9) Chen, R.J.; Bangsaruntip, S.; Drouvalakis, K.A.; Wong, N.; Shim, M.; Li, Y.; Kim, W.; Utz, P.J.; Dai, H. *Proc. Natl. Acad. Sci.* **2003**, *100*, 4984.
- (10) Besteman, K.; Lee, J.; Wiertz, F.G.M.; Heering, H.A.; Dekker, C. *Nano Lett.* **2003**, *3*, 727.
- (11) Keren, K.; Berman, R.S.; Buchstab, E.; Sivan, U.; Braun, E. *Science* **2003**, *302*, 1380.
- (12) Guo, Z.; Sadler, P.J.; Tsang, S.C. *Adv. Mater.* **1998**, *10*, 701.
- (13) Chen, R.J.; Choi, H.C.; Bangsaruntip, S.; Yenilmez, E.; Tang, X.; Wang, Q.; Chang, Y.; Dai, H. *J. Am. Chem. Soc.* **2004**, *126*, 1563.

- (14) Chen, R.J.; Zhang, Y.; Wang, D.; Dai, H. *J. Am. Chem. Soc.* **2001**, *123*, 3838.
- (15) Jiang, K.; Schadler, L.S.; Siegel, R.W.; Zhang, X.; Zhang, H.; Terrones, M. *J. Mater. Chem.* **2004**, *14*, 37.
- (16) Gilmore, J.M.; Scheck, R.A.; Esser-Kahn, A.P.; Joshi, N.S.; Francis, M.B. *Angew. Chem. Int. Ed.* **2006**, *45*, 5307.
- (17) Scheck, R.A.; Francis, M.B. *ACS Chem. Biol.* **2007**, *4*, 247.
- (18) Antos, J.M.; Francis, M.B. *J. Am. Chem. Soc.* **2004**, *126*, 10256.
- (19) Joshi, N.S.; Whitaker, L.R.; Francis, M.B. *J. Am. Chem. Soc.* **2004**, *126*, 15942.
- (20) Hooker, J.M.; Kovacs, E.W.; Francis, M.B. *J. Am. Chem. Soc.* **2004**, *126*, 3718.
- (21) Trost, B.M.; Toste, F.D. *J. Am. Chem. Soc.* **1998**, *120*, 815.
- (22) Tilley, S.D.; Francis, M.B. *J. Am. Chem. Soc.* **2006**, *128*, 1080.
- (23) Karow, D.S.; Pan, D.; Tran, R.; Pellicena, P.; Presley, A.; Mathies, R.A.; Marletta, M.A. *Biochemistry* **2004**, *43*, 10203.
- (24) Boon, E.M.; Davis, J.H.; Tran, R.; Karow, D.S.; Huang, S.H.; Pan, D.; Miazgowicz, M.M.; Mathies, R.A.; Marletta, M.A. *J. Biol. Chem.* **2006**, *281*, 21892.
- (25) Pellicena, P.; Karow, D.S.; Boon, E.M.; Marletta, M.A.; Kuriyan, J. *Proc. Natl. Acad. Sci.* **2004**, *101*, 12854.