

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Optimization of Protein Bioconjugation Reactions Using Combinatorial Peptide Libraries

Permalink

<https://escholarship.org/uc/item/07s5p9nk>

Author

Witus, Leah

Publication Date

2012

Peer reviewed|Thesis/dissertation

OPTIMIZATION OF PROTEIN BIOCONJUGATION REACTIONS USING COMBINATORIAL
PEPTIDE LIBRARIES

By

LEAH SUZANNE WITUS

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 Seung-Wuk Lee

Fall 2012

OPTIMIZATION OF PROTEIN BIOCONJUGATION REACTIONS USING COMBINATORIAL
PEPTIDE LIBRARIES

Copyright © 2012

By Leah Suzanne Witus

ABSTRACT

OPTIMIZATION OF PROTEIN BIOCONJUGATION REACTIONS USING COMBINATORIAL PEPTIDE LIBRARIES

BY

LEAH SUZANNE WITUS

DOCTOR OF PHILOSOPHY IN CHEMISTRY

UNIVERSITY OF CALIFORNIA, BERKELEY

PROFESSOR MATTHEW B. FRANCIS, CHAIR

The covalent attachment of chemical groups to proteins is an important tool for many areas of chemical biology, including the study of protein function and the creation of protein-based materials. In order to create well-defined protein bioconjugates, methods of site-specific protein modification are required. We have developed a high-throughput method of optimizing protein bioconjugation reactions using a combinatorial peptide library in which short peptides serve as a model for protein reactivity. The library screening was achieved using a one-bead-one-compound peptide library, a colorimetric detection scheme to identify high-yielding sequences, and a deconvolution method using a built-in peptide truncation ladder for rapid mass spectrometry sequencing of hit beads. We applied this screening platform in the context of a pyridoxal 5'-phosphate (PLP)-mediated transamination reaction specific for the N-terminus, and were able to identify an N-terminal sequence with high reactivity towards this reaction that can be incorporated into proteins of interest to result in high levels of modification. The screening platform was also used for the discovery and optimization of new N-terminal protein modification reactions.

Dedicated to my family.
“There is nothing worth sharing like the love that lets us share our name.”

TABLE OF CONTENTS

CHAPTER 1: SITE-SPECIFIC PROTEIN MODIFICATION...1

- 1.1 Protein Bioconjugates in Chemical Biology...2
- 1.2 Traditional Methods of Chemical Bioconjugation...3
- 1.3 Importance of Site-Specific Bioconjugation...4
- 1.4 Recent Advances in Site-Specific Bioconjugation...5
- 1.5 Secondary Bioconjugation Reactions...6
- 1.6 References...9

CHAPTER 2: SITE-SPECIFIC PROTEIN BIOCONJUGATION USING A PYRIDOXAL 5'-PHOSPHATE-MEDIATED N-TERMINAL TRANSAMINATION REACTION...13

- 2.1 N-Terminal Transamination...14
- 2.2 Strategic Planning for PLP-Mediated N-terminal Transamination...15
- 2.3 Protocol for Protein Labeling via PLP-Mediated N-Terminal Transamination...18
- 2.4 References...24

CHAPTER 3: DEVELOPMENT OF A COMBINATORIAL PEPTIDE LIBRARY SCREENING PLATFORM...26

- 3.1 Need for Combinatorial Approach...27
- 3.2 Library Synthesis...28
- 3.3 Colorimetric Detection...31
- 3.4 Re-calibration of Colorimetric Detection...33
- 3.5 Sequencing by Truncation Ladder...34
- 3.6 Materials and Methods...35
- 3.7 References...38

CHAPTER 4: IDENTIFICATION OF A HIGHLY REACTIVE N-TERMINAL SEQUENCE TOWARDS PLP-MEDIATED TRANSAMINATION...39

- 4.1 Library Screening...40
- 4.2 Resynthesis and Verification Using Peptides ...41
- 4.3 Positional Scanning...44
- 4.4 Application to Protein Substrates: AKT-Terminal Antifreeze Protein...45
- 4.5 Strategies to Minimize PLP Adduct Formation...47
- 4.6 Application to Protein Substrates: AKT-Terminal GFP...47
- 4.7 Library Screening with Short Reaction Time...50
- 4.8 Materials and Methods...51
- 4.9 References...54

CHAPTER 5: SITE- AND SEQUENCE-SPECIFIC PROTEIN TRANSAMINATION USING AN N-ALKYLPYRIDINIUM CARBOXYALDEHYDE...55

- 5.1 Introduction...56
- 5.2 Rapoport's Salt...57
- 5.3 Library Screening...58

5.4 Study of Rapoport's Salt-Mediated Transamination Using Peptides...	60
5.5 Rapoport's Salt-Mediated Transamination of Anti-HER2 Human IgG1...	61
5.6 Materials and Methods...	65
5.7 References...	72

CHAPTER 6: NEW AVENUES FOR N-TERMINAL REACTIVITY...74

6.1 Library Screening to Evaluate Potential Transamination Reagents...	75
6.2 Light-Mediated Enhancement of PLP Transamination...	77
6.3 Proline-Terminal Bioconjugation...	78
6.4 Turning the Adduct into the Desired Product...	79
6.5 Materials and Methods...	82
6.6 References...	84

ACKNOWLEDGEMENTS

There are so many people who have meant so much to me during graduate school, and I would like to try to express my gratitude here. I can't think of a way I would have rather spent the last five years than being a grad student in chemistry, at Berkeley, in the Francis lab, and that is entirely thanks to the people.

First I want to acknowledge the people whose influence helped initiate my interest in science and research and led me to attend graduate school. My high school chemistry teacher, Aimee Brazil, made my initial exposure to chemistry fun and exciting from her intense enthusiasm (and tons of demos). Trudy Crandell, who I tutored in chemistry while I was in high school, was also influential in my early experiences with chemistry. I appreciated her constant encouragement, and I admired her drive to go back to school to take classes such as chemistry simply for the sake of learning.

From college at Rice University, I want to thank my roommate Madeline Currie, and friend Alice Broussard, for their encouragement to stick with organic chemistry when I was ready to jump ship and change majors. Without them, this might be a history dissertation! Also involved in that turning-point period were Jeff Hartgerink and Seiichi Matsuda, the organic chemistry professors at the time, who I'd like to thank for their encouragement and advice. Once I started doing research in the Hartgerink lab, I was set on the path that led me to today. I was able to discover how interesting and fun research is thanks to Jeff, who is an outstanding research advisor. I learned more from working with Jeff than I have from anyone else, and I am lucky to have learned from him, because in addition to being immensely smart he is also a dedicated and caring teacher and person.

Now on to all the people who helped make graduate school such a good experience. The first person I want to thank is Matt Francis. As an advisor, Matt has created an extremely supportive, fun, and positive group environment. Matt has the ability to see and appreciate the good in everyone, and seems to value the unique craziness of each person in the group. That attitude has made the group feel more like a family than a collection of coworkers. As a role model, Matt shows it is possible to be a successful person in a competitive field yet remain nothing but an extremely nice, friendly, and decent person in every situation. I also appreciate the role Matt takes on in the department: he will listen to and support almost any student, graduate or undergraduate, in our group or others, with scientific or personal matters.

The Francis group theses have a tradition of acknowledging every group member with whom the writer has overlapped. I think this tradition goes to show what a close relationship the group has, so I'm going to do my best to carry it on here. When I first joined there were a number of established senior students. Dante Romanini set high standards for the new students, and although I was intimidated by him I appreciated his

true commitment to research and teaching. Pat Holder had an impressive amount of enthusiasm and he was always ready to drop what he was doing to help someone else out. Although it was brief, I was glad I got to overlap with Jesse McFarland for some helpful early scientific advice. I also felt lucky to get to know Andrew Presley, who was the first Francis group member I met. Andrew is one of the funniest story tellers I know, and really fun person to be around.

When I joined the lab, I remember being surprised at Rebecca and Rebekah's relationship, that labmates would also be such close friends. While I now understand this a lot better, I'm glad to have had them as an example, and grateful that they included me while I overlapped with them. Rebekah Miller was a great person to help show me the ways of the lab early on and a fun friend. I have to acknowledge my deep scientific gratitude to Rebecca Scheck. This entire thesis stems from the work she started, so I have truly been lucky enough to stand on the shoulders of giants. I looked to her as a role model, and admired her efficient approach to research. I also want to thank her for including me in her holiday celebrations, and for sharing her family's Passover Haggadah, which has been a huge hit at many Seders. The other classmate in this super-star year was Aaron Esser-Kahn. I was so glad to have Aaron as a labmate. Not only is he fun to be around, but he was also very inspiring in his scientific productivity and creativity. The chemistry holiday party skits were never as good after Aaron graduated, and it was awesome to have such a funny person to give my GRS introduction.

The next class was Nick Stephanopoulos. While I can't say following him to Northwestern was deliberate, I am glad I'll get scream if I see him on the streets of Chicago. Nick's end of grad school burst of productivity was inspiring. After Nick's class of one, was the class of six right above me- the lost boys. I was always impressed by Sonny Hsiao's ability to be a chemistry powerhouse and turn out an amazing amount of research. Although Zac Carrico seemed to be all business, he was a softie at heart, and I really appreciated when he would open up. I'm also amazed at his half marathon time after so many knee injuries, and I hope those are the last of his knee issues. Michel Dedeo, my parent's favorite freegan, earned his reputation by leading us to free food many times- thank you! He was also always willing to help younger students, and his calm demeanor helped balance the lab. Wesley Wu was my labmate in 743 and held down the male presence in that room for years. I want to thank him for putting up with us. I couldn't think of a sweeter person to have sat next to for five years than Wesley. Although he plays it modest, his hiking and sports skills put us all to shame. Kanna Palaniappan may have not-so-secretly wished he was 100% Bertozzi, but we were all glad he was part of the Francis group too. Kanna was a friend outside of lab as well, and made many parties a bit more memorable. I was impressed by Kanna's ability to keep up multiple projects and his accomplishments have been inspiring. Last but not least in this class is Gary Tong. Gary has been a great friend over the years, and was a huge contributor to making lab, and many meals and parties, a ton of fun. Gary has a true interest in what is going on in the lives of people around him. I really appreciated how some of Gary's favorite conversation topics kept lab from ever getting too serious.

At the same time, he also has a deep knowledge of chemistry and if he chooses to continue pursuing science I know he will be very successful.

In my year there were just two of us: me and Chris Behrens. I could not have asked for a better classmate. Where others would stress, Chris kept cool. His even temper provided a good balance in and out of lab, as long as he wasn't trying to walk home from parties.

The next class was another big one. I want to thank Kristen Seim for DJing 743 for so many years! I was glad to have our lab soundtrack put together by someone with such good taste in music. I've always enjoyed Kristen's pure enthusiasm for her favorite people, things and events. I also admire Kristen's lab skills, and only wish I had part of the experimental drive she has. Amy Twite provided some of the most fun memories from group trips, and I'm so glad I got an extra one in with her with our Yosemite trip. In addition to being hilarious she is an extremely capable scientist and I want to thank her for always being nicer to me than I deserve. If the time we spend wasting is not wasted time, then I want to hang out with Dan Finley all the time. He is always ready for joking around, laughing at almost nothing, lab dancing and creeping. I want to thank him for his awesome bacon baking skills, which brought us prizes in the Chem Bio Bake-Off two years in a row, one a first place win! Troy Moore was fun to express proteins with and work with early on. He was also a great ski buddy, and I really appreciated that he would never leave anyone behind. Mike Coyle was such an all around nice and good-natured person that he would be a good addition to any group. No wonder both we and the Groves group wanted him. Behind the Coyle Spill demeanor, he had some stealth awesome organizational skills, which were really helpful for softball, group jobs, and other lab events. And last in this class is Allie Obermeyer. I'm not sure how I would have gotten through graduate school without her. She has been the biggest support and best friend anyone could hope for. Allie is a great listener and from as soon as she arrived in the group, she was the best sounding board, for all things both scientific and personal. She has a deep and innate understanding of chemistry, well beyond what I could hope to one day achieve. If there is anyone to thank in these acknowledgements, it is Allie—thank you. You so good.

The class of five was next. It took a long time to get to know Stacy Capehart, but it was well worth it. She is such a joy to be around, and makes everyone who knows her happy. I am in awe of both her amazing running skills and dedicated work ethic. Katherine Mackenzie was another running buddy and friend. When she rotated with me I was amazed at her ease with lab techniques. Those abilities must translate from the lab to the kitchen, because she is an amazing cook and award-winning baker. Kareem El Muslemany is a dedicated scientist and sweet person. It has been wonderful to see his research progress so well lately. Jelly Netirojjanakul has an awesome dry sense of humor and amazing wardrobe. I'm not sure I've ever seen her wear the same thing twice. I admired her creativity to explore a new project area so early on. Jeff Glasgow was an impressive scientist and I'm glad he did not listen to the rest of us when we told him MS2 could not be reassembled. He also added hours of fun to group trips on how to

define the south.

While I liked being in a small class, Abby Knight took it to the next level since there was only one Francis group member in her year. Abby was such a positive and enthusiastic addition to the Francis group. Nobody gets more excited about colors than she does! I think her attitude will help keep science fun for the whole group.

Finally we come to the last class of Francis group members with whom I overlapped. Since he joined, Richard Kwant has been fun, friendly, and a little bit of a trouble maker. Excellent. In my interactions with Jake Jaffe he has been unfailingly nice and polite, and I'm sorry I won't be around to get to know him more. Ioanna Aanei I can only describe as fierce, in a good way. I can see her drive and know she will be very successful. Jim and Kanwal, I want to express my gratitude to you both. I am so excited that the N-terminal modification projects are in such good hands. Jim MacDonald has been a fun person to work and hang out with the last few months of my Ph.D. He is already so knowledgeable about and creative with chemistry, and, fitting in well with the Francis group, really into gossip. Kanwal Palla has also been a fun person to hang out with and get to know. She has an impressive drive and dedication. Thank you to you both for being so wonderful to work with at the end of my time in the Francis group.

We did not have many post-docs in the Francis group, but we had good ones. Praveena Garimella was such a nice person and so dedicated to the Francis group. Michelle Farkas has done incredibly impressive things since joining the lab, from getting an entire animal imaging endeavor up and running to bringing in the big bucks. Her attention to detail and meticulousness is unmatched. She has also been a great friend. I am endlessly grateful for her helping make my last months in California some of the most fun. The Yosemite trip she planned was amazing, as were Mt. Diablo, the Tourist club, and everything else—I think we got that bucket list taken care of!

I also want to thank other people with whom I have worked. Peter Robinson and Mike Blaisse were fun to work with during their rotations. They were both very talented, and very nice, and the groups they ended up joining were lucky to have them. My experience working with an undergraduate was nothing but positive thanks to the wonderful student I worked with, Ellen Muehl. Ellen was smart, talented, interested, and a very dedicated worker. I had fun getting to know her and watching her progress, and I am so grateful for the research she did that contributed to this thesis. I also want to thank Tony Iavarone, from the mass spec facility, for so much incredibly productive collaboration over the years. In addition to invaluable mass spectrometry data, I really appreciated Tony's enthusiasm and encouragement. There are also a few other people I would like to acknowledge even though I did not work with them directly. Emily Que, Ellen Sletten, Alison Narayan, and many others are friends that I have looked up to and admired. I've enjoyed their friendship and learned a lot from them. Alice Broussard's friendship has also been very important before, and during, graduate school. She is such a talented writer, caring friend, and interesting person. I really want to thank her for all of her incredible support through ups and downs. There are many other people I

would like to thank for their support and friendship. It doesn't seem possible to list them all, so I will mention just a few but I hope they all know grateful I am to have had them in my life. In particular, I want to thank friends Jess Wood and Heather Jensen, roommates Tara Yacovitch and Aaron Cantor, and running buddies John Weinstein, Hiedi Ferguson, and Callie Unruh.

Now that I have thanked my lab family, it's time to thank my actual family. My mom and dad are nothing but supportive and positive influences in my life. I am so grateful for all of their constant encouragement and glad that we have such a close relationship. I look up to and admire them for many reasons, including my dad's interest in so many different subjects and his enthusiastic love of learning new things, my mom's ability to get along with and befriend so many different types of people, and their love and dedication to each other. I also want to thank them for their solid values, which have been very helpful to me in graduate school and life in general. My brother I'd like to thank for having a really cool job that I can tell people about when describing what grad school is like seems too boring. But also I'd like to acknowledge how grateful I am to have a brother that I like and respect so much. Not only is he a really interesting, unique, and considerate person, but he is the least superficial person I know, which I admire. I'd also like to thank my entire extended family, particularly the California contingent, Ruth, Paul, Lauren and Daniel. It's been so nice to be close by. Finally I want to end my thanking and acknowledging my grandparents. All of my grandparents placed a high value on education and learning. I am especially inspired by my grandma Betty. Not only was she a trailblazer by being a rare woman in mathematics in the 1950s, but she also got her PhD while raising two girls as a single parent and working. Her inspiring example makes me realize how fortunate my circumstances are today. My experience in graduate school has been special, and I am so grateful to all of the people who made it that way.

CHAPTER 1:

SITE-SPECIFIC PROTEIN MODIFICATION

ABSTRACT

The chemical modification of proteins is an important tool for a wide range of endeavors, including the investigation of biological function, the construction of new biomaterials, and the development of novel therapeutics. The generation of a bioconjugate that is modified in a single, predicted site on the surface of a protein is a difficult task, but one that is often crucial for the success of a given application. Reviewed herein are a number of traditional protein bioconjugation reactions that modify native amino acid side chains, as well as some more recent advances in site-specific modification.

1.1 PROTEIN BIOCONJUGATES IN CHEMICAL BIOLOGY

The chemical attachment of synthetic groups to proteins is a widely used method to add non-native functionalities to biomolecules. While proteins have evolved to perform a variety of sophisticated functions, there are many circumstances in which it is desirable to impart additional characteristics to them (Figure 1-1). The creation of such bioconjugates facilitates the study of protein function as well as the creation of novel protein-based materials.

The attachment of fluorophores to proteins as a visualization handle has been an enabling tool for the study of protein localization and function in both *in vitro* and *in vivo* systems. Protein-fluorophore conjugates have been used to investigate protein conformational changes through fluorescence resonance energy transfer (FRET) between two dyes, each attached to different domains of the same protein.¹ FRET between dyes on two different proteins of interest can also be used to study protein-protein interactions. For the imaging of cellular components, one of the traditional techniques for studying the localization of proteins and analytes of interest, immunolabeling, relies on protein bioconjugates. In this technique, a fluorophore is conjugated to a primary or secondary antibody in order to target the dye to the location of the analyte of interest.¹ Although modern biological imaging utilizes on a wide variety of techniques to label proteins of interest, including genetic fusions with fluorescent proteins, protein-dye bioconjugates continue to play an important role.^{2,3,4}

In addition to the study of biological function, the bioconjugation of synthetic groups to proteins has played a major role in medical applications. The development of novel protein therapeutics often involves the attachment of a polymer or drug molecule to a protein. The purpose of these conjugates is to improve the pharmacokinetic properties of the therapeutic component, which can be either the protein or the synthetic attachment. For therapeutic proteins, the conjugation to a biocompatible polymer such as poly(ethylene glycol) (PEG) can improve the serum half-life, decrease immunogenicity, and enhance stability of the protein.^{5,6,7} The pharmaceutical industry is also pursuing antibody-drug conjugates.^{8,9} In this class of biologics, the attachment is the therapeutic agent, often a small molecule drug, and the protein is used for its targeting capability. The specificities and affinities of antibodies for their targeted biological analytes in the diseased tissue enables highly potent cytotoxic drugs to be administered.

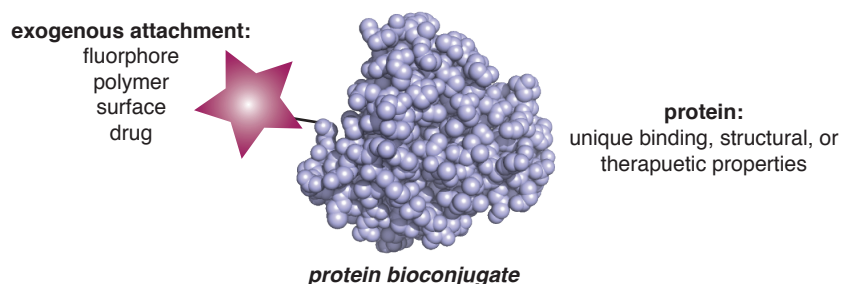


Figure 1-1. The attachment of synthetic groups to proteins to form protein bioconjugates is an important tool for many endeavours in chemical biology.

Protein bioconjugates have also been used in the generation of new protein-based materials. The attachment of synthetic components can be used in the creation of materials that utilize the novel structural or functional properties of proteins. Many binding proteins and enzymes, which have evolved over millions of years, possess functional abilities that are unmatched even by state-of-the-art synthetic compounds. One such example is the metallothionein class of metal binding proteins.¹⁰ These proteins have evolved to sequester toxic heavy metal ions specifically in the presence of much higher concentrations of innocuous metal ions. The conjugation of polymer chains to one such metallothionein to create a protein-polymer hybrid material allowed this protein to be used for detection of contaminated drinking water through a size change of the hydrogel.¹⁰ The unique structural properties of proteins can also be exploited for the creation of new materials by attaching functional components. The Tobacco Mosaic Virus (TMV) coat protein has multiple accessible self-assembly states, including disks and rods. This regular assembly has been used to align donor and acceptor chromophores precisely to build light harvesting devices.^{11,12} Many other types of protein-based materials have been generated as well by attaching proteins to surfaces, beads and nanoparticles.

1.2 TRADITIONAL METHODS OF CHEMICAL BIOCONJUGATION

To access the protein bioconjugates used in the aforementioned biological, medical, and material applications, chemical reactions capable of ligating a synthetic moiety of choice to a protein are required. Protein bioconjugation reactions present a considerable synthetic challenge. With hundreds of polar, unprotected functional groups, the protein substrate dwarfs other complex organic molecules. In order to avoid protein denaturation, the reaction conditions are generally limited to aqueous media, near neutral pHs, mild temperatures, and the reactions must be run at submillimolar substrate concentrations.

Several bioconjugation methods modify the native amino acid side chains possessing reactive functionalities.¹³ These residues include the thiol group of cysteine, the amine group of lysine and the carboxylates of aspartate and glutamate. A synthetic reagent of choice can be conjugated to the protein if it contains an appropriate functionality to react with one of these side chain functional groups. One example of the many modes of reactivity available for lysine, glutamate, aspartate, and cysteine residues is shown in Figure 1-2.¹⁴ Lysine side chain amines can be targeted with *N*-hydroxysuccinimide (NHS) ester-containing reagents (**1**) to form amide linkages. The carboxylic acid group of aspartic and glutamic acid residues can also be used to form an amide bond linkage with an amine-containing reagent (**2**) using water-soluble activating agents such as 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC). Cysteine modification can be achieved via alkylation with maleimide reagents (**3**) as well as disulfide formation with thiol-containing compounds.¹³

1.3 IMPORTANCE OF SITE-SPECIFIC BIOCONJUGATION

The use of native side-chain functionalities for protein bioconjugate formation often results in heterogeneous product distributions. This results when multiple copies of the targeted amino acid exist on the protein surface. For instance, if a protein contains multiple lysine residues on its surface (large proteins such as antibodies can have up to 100 surface accessible lysines⁹), it is difficult to control the number and location of the attachments. While there are some cases for which this does not matter, many applications require architecturally precise bioconjugates. Heterogeneous mixtures of therapeutic bioconjugates may have different pharmacokinetic properties, affecting both analysis and efficacy. Additionally, the indiscriminate attachment of multiple synthetic groups to a protein results in a higher likelihood that its native function will be disrupted due to blocking of the active site or other important residues. In order to form an attachment in a single, specific location on a protein surface, site-specific bioconjugation reactions are required.

Of the traditional bioconjugation techniques, cysteine modification can often result in site-specific attachment due to the low natural abundance of cysteine residues on the surfaces of proteins.¹⁵ Genetic mutation is a widely used technique to introduce a solvent-accessible cysteine in a desired location on the protein, giving a single conjugation site. However, there are cases in which it is inconvenient or impossible to introduce or modify a single copy of this residue without sacrificing protein function. For example, the metal binding ability of metallothioneins relies on multiple cysteine residues in their active sites,¹⁰ and antibodies are held together by structurally important disulfide bonds, which may not form properly with the introduction of an additional cysteine.⁹ Additionally, a growing number of applications, such as FRET-based sensors, require the modification of proteins in two distinct locations. While cysteine modification may be used to install the first group, compatible site-specific modification methods that can be used to install the second functional group are in short supply.

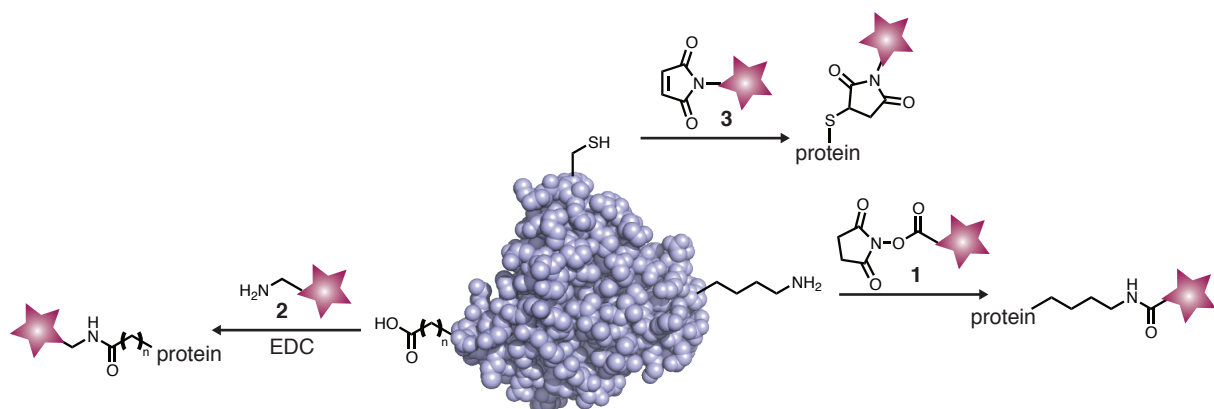


Figure 1-2. There are many reactions that target native amino acids such as lysine, glutamic and aspartic acid, and cysteine residues. However, it is not always possible to use these residues for bioconjugation in a site-specific manner.

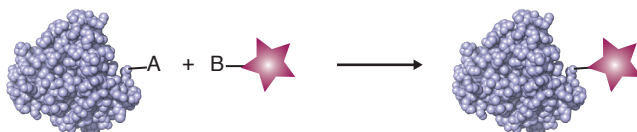
1.4 RECENT ADVANCES IN SITE-SPECIFIC BIOCONJUGATION

The challenges of protein bioconjugation reactions are exacerbated when site-selectivity is an additional requirement. Bioconjugation can be accomplished in a single step or by using two sequential reactions. In a one-step bioconjugation, the reaction entails the ligation of the reagent of interest to the protein in an A+B manner (Figure 1-3). In two-step bioconjugation strategies, the primary reaction introduces a unique functional group to the protein as a reactive handle, while the secondary reaction uses this functionality in the coupling reaction. To achieve site-specificity, the primary reaction must target a single location on the protein. Furthermore, the ligation reaction must be bioorthogonal, whether it takes place in the first or second step. This requires that no other native protein functionalities will react other than the desired coupling partner. Recent research efforts have resulted in the development of both types of methods for the production of site-specific protein bioconjugates.

One strategy for site-specific bioconjugation is the use of reactions that target native amino acids with a low natural abundance,¹⁵ so that mutagenesis may be able to introduce a unique solvent-accessible copy. Tryptophan and tyrosine fall into this category, and multiple bioorthogonal reactions have been developed for these residues. Tyrosine residues can be modified using aldehydes and anilines in a three-component Mannich type reaction,¹⁶ π -allylpalladium complexes,¹⁷ cyclic diazodicarboxamides,¹⁸ and anisidine derivatives using cerium(IV) ammonium nitrate,¹⁹ among others. Tryptophan residues can be targeted using rhodium carbenoids,²⁰ or phenylene diamine derivatives with cerium(IV) ammonium nitrate.¹⁹

Another widely-utilized method to achieve site-specificity is the targeting of native functionalities that occur only once in the protein sequence, namely the N- and C-termini. This strategy has been used to develop both one-step ligation reactions and to install functional handles for two-step bioconjugation schemes. One-step bioconjugation reactions utilizing these sites include the reaction of N-terminal tryptophan and histidine residues with aldehyde-containing reagents to form Pictet-Spengler products,²¹ and the use of N-terminal cysteine residues to form thioazolidines.²² Other bioconjugation

a) one-step bioconjugation strategies



b) two-step bioconjugation strategies

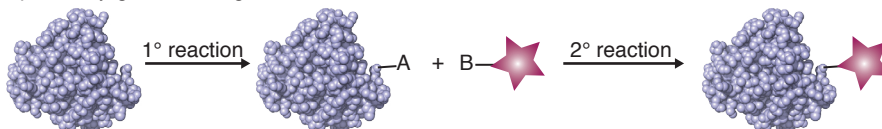


Figure 1-3. Site-specific bioconjugation can be performed in one or two steps. a) In a one-step bioconjugation reaction, the ligation takes place between a single native protein functional group and a exogenous reagent bearing the appropriate coupling partner. b) In two-step bioconjugation strategies, a reactive handle is first incorporated into the protein in a single location. This group is used for a bioorthogonal ligation in the second step.

strategies that target the termini introduce bioorthogonal handles in these positions. A thioester can be installed at the C-terminus through expressed protein ligation (EPL), and can then react chemoselectively with cysteine derivatives in a native chemical ligation reaction.^{23,24} If the N-terminal residue is a serine or threonine, the 2-amino alcohol group can be oxidized with sodium periodate to result in an aldehyde on the N-terminus.²⁵ This aldehyde functionality can then be used for a secondary bioorthogonal ligation, although the use of sodium periodate can oxidize sugars and sulfur-containing residues on the protein.²⁶ An alternative method of introducing a keto-group at the N-terminus that is widely applicable to many N-terminal residues and utilizes mild reaction conditions that do not affect other protein functionalities is pyridoxal 5'-phosphate-mediated N-terminal transamination,^{27,28,29,30,31} which is discussed in depth in Chapter 2.

An alternate way to achieve site-specificity among native, highly abundant amino acids is the use of a unique recognition sequence for enzymatic labeling. Although enzymatic reactions are not as easily scalable as chemical reactions, enzymatic labeling reactions targeting specific recognition sequences can be highly useful, especially for modifying a single protein in a complex mixture. This can be achieved using biotin ligase and formylglycine-generating enzymes, which have been used to install ketones and aldehydes, respectively, on their recognition sequences.^{32,33} Appropriately functionalized substrates can also be directly conjugated to the recognition sequence using enzymatic techniques, as seen with bacterial sortases that capitalize on the tolerance of the enzymes for non-natural oligoglycine substrates.³⁴ Guanosine derivatives can also be used to modify specific domains of fusion proteins using the “SNAP” tagging strategy.³⁵ These reactions provide just a few examples of the rapidly growing set of enzymatic labeling techniques that are now available.^{2,36}

In addition to chemical and enzymatic reactions, another method of installing reactive handles is through the incorporation of an unnatural amino acid into the protein sequence during expression. This can be performed during *in vivo* protein expression by using the amber codon suppression developed by the Schultz lab, or by using *in vitro* expression.^{37,38} The unnatural amino acid labeling technique is powerful because the reactive handle can be introduced in any desired location within the protein sequence using genetic methods, and a wide variety of functional groups can be incorporated. However, this is balanced with the complexity of this method. The choice of reactive handles for incorporation into the protein sequence is informed by the available arsenal of secondary bioorthogonal ligation reactions.

1.5 SECONDARY BIOCONJUGATION REACTIONS

To achieve site-specific protein bioconjugation, the exogenous group of choice can be directly coupled to native protein functionalities, but more often a reactive handle is installed in a single location first, followed by a secondary bioorthogonal ligation. These ligations must fit the criteria of all bioconjugation reactions: proceed in aqueous media with low substrate concentrations, and use mild pH and temperature conditions. The

ligations are generally A+B type couplings, and must be highly chemoselective so that the hundreds of other polar spectator groups present on the protein are not modified. Additionally, the coupling partner that is incorporated onto the protein in the first step must be inert and small. Although there are many recent developments in this field, a few well-established and emerging ligation methods are described below.

One widely known ligation chemistry is the cycloaddition between an alkyne (**4**) and an azide (**5**) to form a triazol linkage (**6**), shown in Figure 1-4a. Although this reaction is often referred to as ‘click chemistry’, it is actually one of several such reactions described in a 2001 review by Sharpless and coworkers. The authors defined click reactions as reactions that are fast, high yielding, chemoselective, and compatible with mild reaction conditions.³⁹ The alkyne-azide cycloaddition is thus one of many click reactions that are suitable for use on biomolecules. The reaction is orthogonal to biomolecule functionalities, and the coupling partners are inert in the absence of copper catalysts or activating groups. Both alkynes and azides can be site-specifically incorporated into proteins using artificial amino acids, and there are a number of commercially available azide- and alkyne-functionalized dyes and polymers. Although the required copper catalyst poses an issue for certain biological applications, there have been recent advances in the development

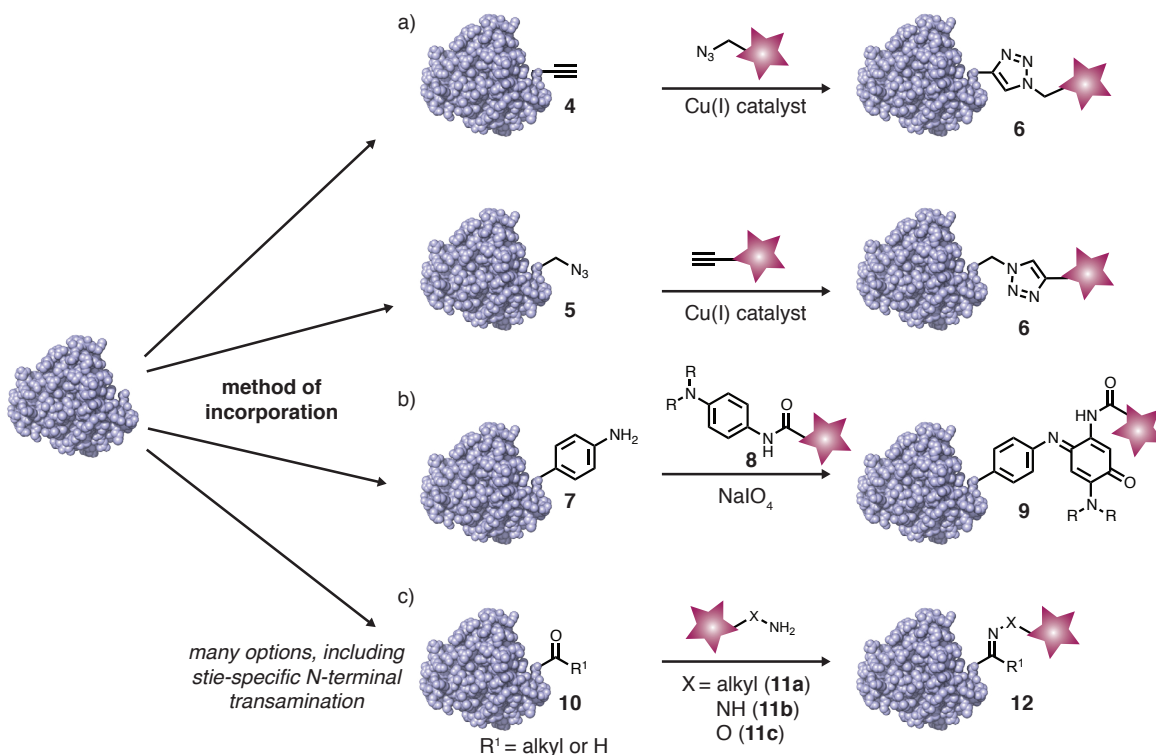


Figure 1-4. Bioorthogonal ligation reactions. After an appropriate coupling partner has been site-specifically introduced into the protein surface, a secondary ligation reaction can attach the reagent of interest. a) Either an alkyne or an azide can be incorporated into the biomolecule for the formation of triazol linkages. b) A recently developed oxidative coupling reaction targets proteins bearing aniline functionalities. c) The reaction between a carbonyl and a hydrazide or alkoxyamine to form a hydrazone or oxime linkage is a widely used ligation strategy. There are multiple ways to incorporate a carbonyl into the protein including site-specific N-terminal transamination, which is described in depth in Chapters 2, 4, and 5.

of water-soluble copper ligands⁴⁰ and copper-free ligations using cyclooctynes.⁴¹

To add to the toolbox of available reactions, a new chemoselective “click”-type bioconjugation reaction has been recently developed in the Francis group.⁴² This reaction is a sodium periodate-mediated oxidative coupling between an aniline (**7**) and phenylene diamine (**8**) or animophenol derivatives. It was developed based on the finding that anilines reacted rapidly and cleanly with *N*-acyl phenylene diamines in the presence of sodium periodate. The resulting “Bandrowski bases”⁴³ were stable with respect to hydrolysis across a wide pH range, and could withstand strong reductants, oxidants, nucleophiles, and high temperatures. Although the original reaction yielded a trimeric species, it was found that the product symmetry could be broken by adding two alkyl groups to the phenylene diamine group.⁴² This prevented it from reacting with itself, but still allowed an additional aniline in solution to couple in an A+B fashion with very high efficiency. The aniline coupling partner can be incorporated into proteins using unnatural amino acid techniques. This oxidative coupling strategy can be used in the presence of the natural amino acids and nucleic acids without interference and can take place in aqueous solution at low micromolar concentrations, thus offering great potential for the attachment of exogenous groups to proteins.⁴⁴

Another class of bioorthogonal ligation widely used in chemical biology is the reaction of carbonyls with hydrazides and alkoxyamines to form hydrazones and oximes, respectively. Carbonyl groups are not naturally found in proteins and after introduction as a chemical handle, aldehyde and ketone carbonyls are largely inert to native biological functionalities. As an additional advantage, carbonyl condensation reactions do not require potentially toxic metal additives to proceed and the linkages are small (as little as 3-4 atoms), so as not to disturb the biological function of interest.

Imine ligations between carbonyls and amines are common synthetic transformations; however, the case of primary amines (**11a**), the equilibrium position in water lies toward the side of the carbonyl (**10**) and thus leads only to transient product formation (Figure 1-4d). In contrast, for amines with adjacent electronegative atoms such as nitrogen (hydrazides, **11b**) and oxygen (alkoxyamines, **11c**), the equilibrium lies toward the side of the imine (**12b-c**, $K_{eq} = 10^4$ to 10^8 M⁻¹).⁴⁵ The increased product stability of the α -effect imines can be attributed to the inductive effect of the electronegative heteroatom, which lowers the pK_a of the imine nitrogen and increases its resistance to hydrolysis.⁴⁶ The formation of these α -effect imines are highly chemoselective for aldehydes and ketones, proceed under mild aqueous reaction conditions, and lead to single adducts with no byproducts aside from water.

One potential limitation of hydrazone and oxime ligations in biological applications is their rate profile. Imine ligations proceed with optimal kinetics at slightly acidic pH,⁴⁷ but they are more sluggish at the neutral pH conditions that many biomolecules require. In order to compensate for this, high concentrations of the exogenous α -effect imine reagent can be added to the carbonyl-containing protein, since most proteins are not soluble or even available at concentrations above 100 μ M. Additionally, to achieve higher coupling rates,

Dawson and coworkers have reported the use of aniline as a nucleophilic catalyst for imine ligations.⁴⁸ The authors demonstrated significant rate enhancements of oxime and hydrazone ligations at pH 7 with <100 μ M concentration of reactants. The mechanism of catalysis was proposed to involve the formation of a highly reactive protonated aniline imine as an intermediate. Aniline catalysis can thus significantly improve the utility of α -effect imine ligations for a number of biological applications.

One versatile aspect of oxime and hydrazone ligations is the range of product stabilities that can result. Raines and co-workers studied the hydrolytic stability of α -effect imines in detail.⁴⁶ The order of product stability at pH 7.0 was found to be oxime > semicarbazone > acylhydrazone > primary hydrazone > secondary hydrazone. The range of hydrolytic half-lives was significant, from over 25 days for oximes to approximately 30 min for secondary hydrazones. The differences observed in hydrolytic stability were attributed to the affect of the various α -heteroatoms on the protonation state of the imine nitrogen. Thus, oximes are the optimal choice for the creation of stable bioconjugates. However, the lability of hydrazone linkages can be an asset in certain applications, such as the coupling of prodrugs to carriers, dynamic covalent chemistry,⁴⁹ and for proteomic approaches in which cleavable linkers are used to liberate biomolecules for identification after capture using affinity resins.⁵⁰ This comparative study underscores the versatility of α -effect imine linkages—once a carbonyl coupling partner has been introduced into the biomolecule of interest, a highly stable or highly dynamic linkage may be installed through the selection of the α -effect amine.

One of the most attractive aspects of using hydrazone and oxime chemistry in biological applications is the abundance of methods available for the site-specific introduction of ketones and aldehydes into proteins. This can be accomplished using chemical, enzymatic, and unnatural amino acid methods. One chemical method which achieves this goal is a transamination reaction that installs a carbonyl group at the N-terminus of proteins. Given the importance of the production of protein bioconjugates to applications in chemical biology, drug production, and materials science, there remains a need to expand the available bioconjugation tool box. The subsequent chapters present our progress in the development and optimization of N-terminal transamination reactions.

1.6 REFERENCES

1. Giepmans, B.N.G., Adams, S.R., Ellisman, M.H., Tsien, R.Y. The Fluorescent Toolbox for Assessing Protein Location and Function. *Science* **2006**, 312, 217–224.
2. O'Hare, H.M., Johnsson, K., Gautier, A. Chemical probes Shed Light on Protein Function. *Curr. Opin. Struct. Biol.* **2007**, 17, 488–494.
3. Johnsson, N., Johnsson, K. Chemical Tools for Biomolecular Imaging. *ACS Chem. Biol.* **2007**, 2, 31–38.
4. Marks, K.M., Nolan, G.P. Chemical Labeling Strategies for Cell Biology. *Nat. Meth.* **2006**, 3, 591–596.
5. Gao, W., Liu, W., Mackay, J.A., Zalutsky, M.R., Toone, E.J., Chilkoti, A. In Situ Growth of a Stoichiometric PEG-Like Conjugate at a Protein's N-Terminus with Significantly Improved Pharmacokinetics. *Proc. Natl. Acad. Sci.* **2009**, 106, 15231–15236.

6. Zalipsky, S. Functionalized Poly(ethylene glycols) for Preparation of Biologically Relevant Conjugates. *Bioconj. Chem.* **1995**, *6*, 150–165.
7. Baker, D.P., Lin, E.Y., Lin, K., Pellegrini, M., Petter, R.C., Chen, L.L., Arduini, R.M., Brickelmaier, M., Wen, D., Hess, D.M., Chen, L., Grant, D., Whitty, A., Gill, A., Lindner, D.J., Pepinsky, R.B. N-Terminally PE-Glylated Human Interferon- β -1a with Improved Pharmacokinetic Properties and in Vivo Efficacy in a Melanoma Angiogenesis Model. *Bioconj. Chem.* **2005**, *17*, 179–188.
8. Wu, A.M., Senter, P.D. Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotech.* **2005**, *23*, 1137–1146.
9. Alley, S.C., Okeley, N.M., Senter, P.D. Antibody-Drug Conjugates: Targeted Drug Delivery for Cancer. *Curr. Opin. Chem. Biol.* **2010**, *14*, 529–537.
10. Esser-Kahn, A.P., Iavarone, A.T., Francis, M.B. Metallothionein-Cross-Linked Hydrogels for the Selective Removal of Heavy Metals from Water. *J. Am. Chem. Soc.* **2008**, *130*, 15820–15822.
11. Miller, R.A., Presley, A.D., Francis, M.B. Self-Assembling Light-Harvesting Systems from Synthetically Modified Tobacco Mosaic Virus Coat Proteins. *J. Am. Chem. Soc.* **2007**, *129*, 3104–3109.
12. Dedeo, M.T., Duderstadt, K.E., Berger, J.M., Francis, M.B. Nanoscale Protein Assemblies from a Circular Permutant of the Tobacco Mosaic Virus. *Nano Lett.* **2010**, *10*, 181–186.
13. Hermanson G.T. Bioconjugate Techniques. Academic Press. **1996**, 1st Ed.
14. Tilley, S.D., Joshi, N.S., Francis, M.B., Begley, T.P., Proteins: Chemistry and Chemical Reactivity, in: Wiley Encyclopedia of Chemical Biology. John Wiley & Sons, Inc. **2007**.
15. Doolittle, R.F. R. Redundancies in Protein Sequence, in Prediction of Protein Structures and the Principles of Protein Conformation. Plenum Press. **1989**.
16. Joshi, N.S., Whitaker, L.R., Francis, M.B. A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J. Am. Chem. Soc.* **2004**, *126*, 15942–15943.
17. Tilley, S.D., Francis, M.B. Tyrosine-Selective Protein Alkylation Using π -Allylpalladium Complexes. *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081.
18. Ban, H., Gavriluk, J., Barbas. Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **2010**, *132*, 1523–1525.
19. Seim, K.L., Obermeyer, A.C., Francis, M.B. Oxidative Modification of Native Protein Residues Using Cerium(IV) Ammonium Nitrate. *J. Am. Chem. Soc.* **2011**, *133*, 16970–16976.
20. Antos, J.M., McFarland, J.M., Iavarone, A.T., Francis, M.B. Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *J. Am. Chem. Soc.* **2009**, *131*, 6301–6308.
21. Li, X., Zhang, L., Hall, S.E., Tam, J.P. A New Ligation Method for N-Terminal Tryptophan-Containing Peptides Using the Pictet-Spengler Reaction. *Tetrahedron Lett.* **2000**, *41*, 4069–4073.
22. Tam, J.P., Miao, Z. Stereospecific Pseudoproline Ligation of N-Terminal Serine, Threonine, or Cysteine-Containing Unprotected Peptides. *J. Am. Chem. Soc.* **1999**, *121*, 9013–9022.
23. Dawson, P., Muir, T., Clark-Lewis, I., Kent, S. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776–779.
24. Muir, T.W. Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
25. Geoghegan, K.F., Stroh, J.G. Site-Directed Conjugation of Nonpeptide Groups to Peptides and Proteins via Periodate Oxidation of a 2-Amino Alcohol: Application to Modification at N-Terminal Serine. *Bioconj. Chem.* **1992**, *3*, 138–146.
26. Zeng, Y., Ramya, T.N.C., Dirksen, A., Dawson, P.E., Paulson, J.C. High-Efficiency Labeling of Sialylated Glycoproteins on Living Cells. *Nat. Meth.* **2009**, *6*, 207–209.

27. Witus L.S., Francis M. Site-Specific Protein Bioconjugation via a Pyridoxal 5'-Phosphate-Mediated N-Terminal Transamination Reaction. *Current Protocols in Chemical Biology*. John Wiley & Sons, Inc. **2010**. 125-134.
28. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S., Francis M.B. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **2006**, *45*, 5307-5311.
29. Scheck R.A., Francis M.B. Regioselective Labeling of Antibodies Through N-Terminal Transamination. *ACS Chem. Biol.* **2007**, *2*, 247-251.
30. Scheck R.A., Dedeo M.T., Iavarone A.T., Francis M.B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, *130*, 11762-11770.
31. Witus L.S. Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., Francis, M.B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, *132*, 16812-16817.
32. Chen, I., Howarth, M., Lin, W., Ting, A.Y. Site-Specific Labeling of Cell Surface Proteins with Biophysical Probes Using Biotin Ligase. *Nat. Meth.* **2005**, *2*, 99-104.
33. Carrico, I.S., Carlson, B.L., Bertozzi, C.R. Introducing Genetically Encoded Aldehydes into Proteins. *Nat. Chem. Biol.* **2007**, *3*, 321-322.
34. Mao, H., Hart, S.A., Schink, A., Pollok, B.A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004**, *126*, 2670-2671.
35. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., Johnsson, K. A General Method for the Covalent Labeling of Fusion Proteins with Small Molecules In Vivo. *Nat. Biotech.* **2003**, *21*, 86-89.
36. Sletten, E.M., Bertozzi, C.R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem. Int. Ed.* **2009**, *48*, 6974-6998.
37. Wang, L., Brock, A., Herberich, B., Schultz, P.G. Expanding the Genetic Code of Escherichia Coli. *Science* **2001**, *292*, 498-500.
38. Link, A.J., Mock, M.L., Tirrell, D.A. Non-Canonical Amino Acids in Protein Engineering. *Curr. Opin. Biotech.* **2003**, *14*, 603-609.
39. Kolb, H.C., Finn, M.G., Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021.
40. Hong, V., Presolski, S.I., Ma, C., Finn, M.G. Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation. *Angew. Chem. Int. Ed.* **2009**, *48*, 9879-9883.
41. Jewett, J.C., Bertozzi, C.R. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chem. Soc. Rev.* **2010**, *39*, 1272-1279.
42. Hooker, J.M., Esser-Kahn, A.P., Francis, M.B. Modification of Aniline Containing Proteins Using an Oxidative Coupling Strategy. *J. Am. Chem. Soc.* **2006**, *128*, 15558-15559.
43. Bandrowski, E. Über die Oxidation des Paraphenylendiamins. *Chem. Ber.* **1894**, *27*, 480.
44. Behrens, C.R., Hooker, J.M., Obermeyer, A.C., Romanini, D.W., Katz, E.M., Francis, M.B. Rapid Chemoselective Bioconjugation through Oxidative Coupling of Anilines and Aminophenols. *J. Am. Chem. Soc.* **2011**, *133*, 16398-16401.
45. Dirksen, A., Dawson, P.E. Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling. *Bioconj. Chem.* **2008**, *19*, 2543-2548.
46. Kalia, J., Raines, R.T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed.* **2008**, *47*, 7523-7526.
47. Jencks, W.P. Studies on the Mechanism of Oxime and Semicarbazone Formation. *J. Am. Chem. Soc.* **1959**, *81*, 475-481.

48. Dirksen, A., Hackeng, T.M., Dawson, P.E. Nucleophilic Catalysis of Oxime Ligation. *Angew. Chem. Int. Ed.* **2006**, 118, 7743–7746.
49. Dirksen, A., Dirksen, S., Hackeng, T.M., Dawson, P.E. Nucleophilic Catalysis of Hydrazone Formation and Transimination: Implications for Dynamic Covalent Chemistry. *J. Am. Chem. Soc.* **2006**, 128, 15602–15603.
50. Park, K.D., Liu, R., Kohn, H. Useful Tools for Biomolecule Isolation, Detection, and Identification: Acylhydrazone-Based Cleavable Linkers. *Chem. Biol.* **2009**, 16, 763–772.

CHAPTER 2:

SITE-SPECIFIC PROTEIN BIOCONJUGATION USING A PYRIDOXAL 5'-PHOSPHATE-MEDIATED N-TERMINAL TRANSAMINATION REACTION

ABSTRACT

The covalent attachment of chemical groups to proteins is a critically important tool for the study of protein function and the creation of protein-based materials. Methods of site-specific protein modification are necessary for the generation of well defined bioconjugates possessing a new functional group in a single position in the amino acid sequence. This chapter describes the protocol for modifying a protein using a pyridoxal 5'-phosphate (PLP)-mediated transamination reaction that is specific for the N-terminus. The reaction oxidizes the N-terminal amine to a ketone or an aldehyde, which can form a stable oxime linkage with an alkoxyamine reagent of choice. Screening studies have identified the most reactive N-terminal residues, facilitating the use of site-directed mutagenesis to achieve high levels of conversion.

Portions of the work described in this chapter have been reported in a separate publication.¹

2.1 N-TERMINAL TRANSAMINATION

The generation of a bioconjugate that is modified in a single predicted site on the surface of a protein is a difficult task, but one that is often crucial for the success of a given application. It is frequently difficult to target a unique instance of a particular amino acid side chain, resulting in indiscriminate and heterogeneous modification. As one solution to this challenge, a protein bioconjugation reaction has been developed that achieves site-specificity by targeting the N-terminus, a unique position in the sequence. When incubated with pyridoxal 5'-phosphate (PLP), the N-terminal amine of a protein undergoes a transamination reaction that installs a ketone or an aldehyde in that position without modifying lysine side chain amines (Figure 2-1a).^{1,2} This newly installed functionality can be further derivatized through the formation of a stable oxime bond using a diverse array of alkoxyamine probes (Figure 2-1b). This technique provides a convenient and readily scalable way to install a single functional group in a single location.

The N-terminal transamination of proteins was originally reported by Dixon and coworkers.³ In this work, transamination was effected by incubation of the protein with glyoxylic acid, a divalent heavy metal cation such as Cu^{2+} , a high concentration of base such as acetate or pyridine, and in some cases high temperatures (up to 100 °C). The N-terminal transamination was used for the stepwise removal of the N-terminal amino acid in order to allow the study of the importance of that residue on protein function, which was achieved by following the transamination with a subsequent reaction with appropriate dinucleophiles. However, damage to the protein from the reaction conditions limited the number of protein substrates on which this reaction could be used.⁴

The N-terminal transamination of proteins under much milder conditions, and the application of this reactivity towards the creation of site-specific bioconjugates, was reported by the Francis group. As reported in Gilmore and Scheck *et al.*, N-terminal transamination and decarboxylation was observed upon incubation of the peptide angiotensin with glyoxylic acid.² A number of aldehydes were subsequently screened for similar reactivity and PLP was found to result in the highest conversion. Notably, PLP-mediated transamination was

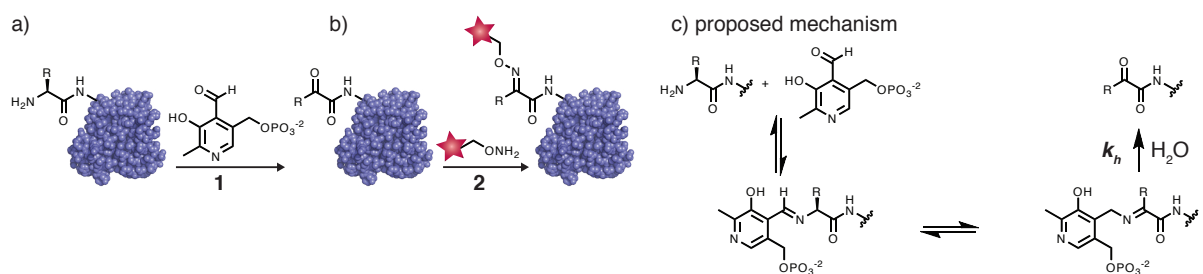


Figure 2-1. The general scheme of PLP-mediated bioconjugation. a) In the first step, a protein is incubated with PLP (1) under mild, aqueous conditions. This oxidizes the N-terminus of the protein to a ketone or an aldehyde, providing a unique functional group for further modification. b) In the second step the ketone is conjugated to an alkoxyamine-bearing reagent (2) through oxime formation. c) The proposed mechanism begins with Schiff base formation between the N-terminal amine and the PLP aldehyde. Tautomerization, followed by hydrolysis, affords the keto-protein product. Based on this mechanism, the reaction rate is expected to depend on both the concentration of protein and PLP.

found to proceed under very mild reaction conditions without the use of base or metal ions. PLP, also known as Vitamin B₆, is a common biological cofactor.^{5,6} It is used in enzymes to catalyze a number of transformations, including transamination, racemization, decarboxylation, beta elimination, and others.⁵ PLP was found to have nonenzymatic transamination activity in the 1940s when Snell heated PLP with amino acids and observed the formation of pyridoxamine (PMP).⁷ The report from the Francis group demonstrated that PLP could also transaminate full-length intact proteins in the absence of an enzyme.²

The mechanisms of enzymatic PLP-mediated transformations have been studied extensively.⁵ The non-enzymatic N-terminal PLP-mediated transamination is proposed to follow a similar reaction pathway (Figure 2-1c). The first step of the proposed mechanism is reversible Schiff base formation between the N-terminal amine and the PLP aldehyde. The next step is likely the deprotonation at the imine α position, which is stabilized by resonance delocalization. The pyridine nitrogen serves as an electron sink to drive this step. Subsequent hydrolysis of PMP leaves a carbonyl group on the protein. This mechanism offers an explanation for the specificity of the transamination for the N-terminal amine over lysine side chain amines. Lysine amines may form Schiff bases with PLP but are unlikely to undergo the subsequent deprotonation since the pK_a of the α proton of an aliphatic imine is much higher (~ 40) than that of an amino acid (29 for glycine).⁵

Since the initial observation of its transamination reactivity on angiotensin, PLP has been used to transaminate a number of intact proteins for the generation of site-specific bioconjugates with a variety of exogenous groups including fluorophores,⁸ polymer chains,⁹ polymer initiators,^{10,11} and surfaces.^{12,13} Due to its compatibility with other bioconjugation reactions, PLP-mediated transamination has been used in conjunction with other bioconjugation techniques, such as cysteine modification and expressed protein ligation, for dual protein modification.^{14,15} PLP-mediated bioconjugation proceeds at mild pH and temperature, and is tolerant of a number of buffers and conditions. Peptide screening studies have shown that this reaction proceeds with high yield for a number of N-terminal residues (Ala, Gly, Asn, Glu, and Asp, in particular), and to a lesser degree for others.¹⁶ Additionally, a highly reactive N-terminal sequence has been identified, which consists of an alanine as the N-terminal residue and a positive charge in the second or third position.¹⁷ As such, the reaction has often proven successful for native protein sequences and, in cases where it is not, standard molecular biology techniques can typically be used to change the N-terminal sequence to obtain increased levels of conversion.^{16,17}

2.2 STRATEGIC PLANNING FOR PLP-MEDIATED N-TERMINAL TRANSAMINATION

A wide variety of proteins with both known and unknown sequences have been successfully modified using PLP-mediated bioconjugation. Therefore, a new target should initially be evaluated empirically. If the protein is recombinantly expressed, its suitability for N-terminal modification can often be improved using the methodology described herein. First, the identity and accessibility of the N-terminal residues should be taken into consideration. In general, buried and/or highly hydrophobic N-terminal sequences react poorly, most likely

because they do not participate in imine formation with PLP. In these cases, installation of one to three “spacer” residues can lead to enhanced reactivity (for a specific example, see Scheck *et al.*, 2008). The reaction is also not appropriate for proteins (such as actin) that have been acylated at the N-terminal amino group through post-translational modifications.

Although its side chain is not thought to participate directly in the reaction mechanism, the influence of the identity of the N-terminal residue on the reaction outcome was investigated in peptide studies.¹⁶ As shown in Figure 2-2, a library of tetrapeptides including all 20 natural amino acids as the N-terminal residue was exposed to transamination and oxime formation under identical reaction conditions (transamination conditions were 10 mM PLP, 18 h, pH 6.5, room temperature). The resulting products included the desired oxime product, transaminated ‘keto-peptides’ that formed but did not participate in subsequent oxime formation, and covalent adducts of PLP to the N-terminal group. The N-terminal residues can be grouped by yield into the categories of *high conversion*, *intermediate conversion*, or *byproduct formation*. The *high conversion* category included alanine, glycine, aspartate, glutamate, and asparagine, which were found to result in the highest yields of oxime product. Cysteine, arginine, threonine, tyrosine, leucine, serine,

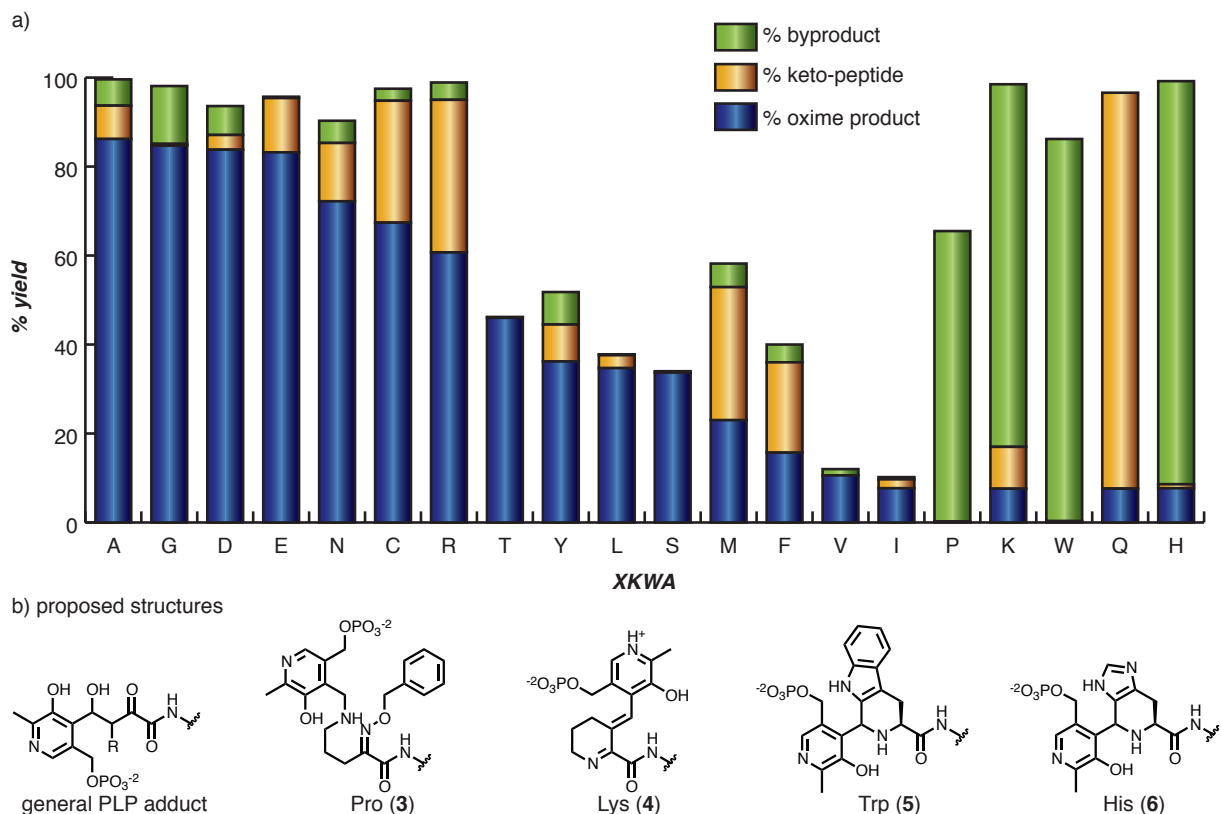


Figure 2-2. Although the side chain does not directly participate in the reaction mechanism, the identity of the N-terminal residue can significantly influence the transamination yield. a) Tetrapeptides of the form XKWA, were synthesized where X was varied to include the 20 natural amino acids. Each peptide was treated with 10 mM PLP for 18 h at pH 6.5, followed by oxime formation with benzyl alkoxyamine. The products were analyzed by liquid chromatography-mass spectrometry and the resulting yields are shown. b) The byproducts had various structures, described in the text. (Figure adapted from reference 16).

methionine, phenylalanine, and valine all resulted in intermediate levels of conversion.

The *byproduct formation* category included various species of byproducts arising from N-terminal glutamine, histidine, tryptophan, lysine, and proline residues. Glutamine was found to transaminate to produce a ketone, but this structure was resistant to further oxime formation. Histidine and tryptophan (**5** and **6**) underwent a previously reported Pictet-Spengler reaction.¹⁸ Lysine and proline each formed unique types of PLP adducts, and in these cases the formation of the covalent PLP adducts did not preclude oxime formation (although they have been found to be slightly less reactive than the keto groups formed on the N-termini). The lysine-terminal PLP adduct (**4**) presumably forms from transamination followed by the loss of water to form a cyclic imine.¹⁶ The nucleophilic enamine tautomer of this structure undergoes aldol addition to additional molecules of PLP to form a covalent adduct. The proline-terminal adduct (**3**) observed involves the covalent attachment of PLP through an as-yet unknown pathway, which by mass may correspond to a ring opening.

Surprisingly, in addition to the N-terminal residue, it was found that the identity and order of internal residues could affect transamination as well.¹⁶ As shown in Figure 2-3, the reactivity of peptides with the same N-terminal residue was significantly altered depending on the amino acids in the second and third positions. The terminal and internal residues were found to have a complex synergistic effect on overall reactivity. To identify short N-terminal sequences that were highly reactive towards PLP-mediated transamination, a combinatorial peptide library screen was performed (discussed in depth in Chapters 3 and 4). A reactive motif consisting of alanine in the N-terminal position and a positive charge in the second or third position was identified. A three-amino acid tag, Ala-Lys-Thr, was genetically incorporated into the N-terminus of a number of protein substrates and resulted in high levels of conversion for all the proteins studied.¹⁷

The intended application and empirical analysis of new target proteins will dictate the need

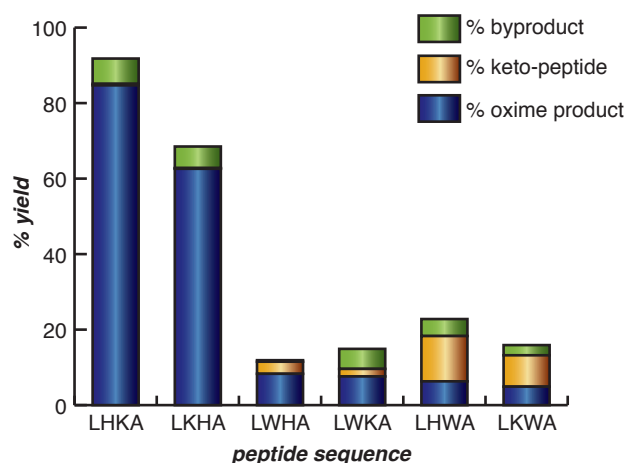


Figure 2-3. In addition to the N-terminal residue, internal residues were also found to affect the reactivity of a sequence towards PLP-mediated transamination significantly. The order and the identity of internal residues were found to influence yield, indicating that the N-terminal sequence can have a complex synergistic effect on the reactivity of a substrate. (Figure adapted from reference 16).

for mutagenesis. In many cases, sufficient reactivity can be achieved using the native terminus, and in some cases lower levels of conversion are tolerable for the application (e.g., surface attachment or radiolabeling). In other cases, standard molecular biology protocols can be used to change the N-terminus of proteins to the highly reactive Ala-Lys motif. As noted above, extending the N-terminus by one to three residues at the same time as converting the N-terminal residues may improve the outcome by increasing the accessibility of the N-terminus. Similarly, if a protein that already terminates in a *high conversion* residue is found to have low yield, mutagenesis to extend the N-terminus and increase accessibility may help to improve the reactivity. If the protein cannot tolerate a positive charge near the N-terminus, rendering an Ala-Lys terminal protein inaccessible, a glutamate-terminal sequence can be installed instead for Rapoport's salt-mediated transamination (discussed in Chapter 5). If the location of the modification on the protein is crucial, circular permutation has been used in one case to change the location of the N-terminus.¹⁵ In this technique, the gene is reorganized such that new N- and C-termini are created in a desired location.¹⁹

2.3 PROTOCOL FOR PROTEIN LABELING VIA PLP-MEDIATED N-TERMINAL TRANSAMINATION

This protocol describes a two-step process for conjugating a chemical group of interest to the N-terminus of a protein. The first step is the PLP-mediated transamination, which installs a reactive carbonyl (keto) group on the N-terminus that can be used for further derivatization. The second step describes typical conditions for the reaction of this site with a small molecule alkoxyamine reagent to form an oxime linkage, which is stable under physiological conditions. A similar strategy may be used to attach hydrazide-functionalized compounds to form a hydrazone linkage, although these groups are more labile.²⁰ The analysis technique used to verify the modification will vary according to the chemical moiety being conjugated to the protein: the attachment of small molecules can be confirmed using mass spectrometry, while the attachment of fluorescent or high molecular weight compounds can be detected by UV-Vis spectrometry, HPLC, or SDS-PAGE. If the bioconjugation probe is in short supply or is difficult to detect using these analytical techniques, a proxy alkoxyamine reagent (such as commercially-available benzyloxyamine or aminooxyacetic acid) can be used to confirm reactivity and optimize the transamination.

Materials

- Target protein stock solution (see recipe)
- PLP stock solution (see recipe)
- Alkoxyamine solution (see recipe)

Additional reagents and equipment for protein separations and analysis of protein modification, including ESI-MS or MALDI-TOF MS methods²¹

PLP-mediated N-terminal transamination

1. Combine the protein solution and the PLP stock solution in a 1.5-ml microcentrifuge tube to give a final concentration of 100 mM PLP and 10 to 500 μ M protein at pH 6.5.

Carefully check the pH of this solution. If the pH of the PLP stock was not adjusted (see Reagents and Solutions) this solution may be overly acidic, leading to protein precipitation or suboptimal levels of conversion.

2. Incubate at 37 °C for 1 to 2 h.

The reaction conditions listed here represent a starting point for optimization with each new protein target. Factors that may be varied include the concentration of PLP, the incubation time, the reaction temperature, and the pH. For Ala-Lys terminal proteins, 100 mM PLP at pH 6.5 for 1-2 h are conditions that have been found to lead to high yields of transamination with low amounts of byproduct formation. Another standard reaction condition that has been used successfully is 10 mM PLP at pH 6.5 for 18 h. The stability of the protein may dictate the range of acceptable temperature and pH conditions. Consult the Troubleshooting section for details on how these parameters may affect reaction outcome.

3. Remove PLP using one of a variety of size-exclusion methods (such as dialysis, buffer exchange, or gel filtration²¹), capitalizing on the difference in the mass of the protein relative to PLP.

If the target protein or peptide has a low molecular weight, reversed-phase separation²¹ can be used. If excess PLP is not removed, the aldehyde of the PLP molecule will quench the ketoreactive group in the next step, significantly lowering its effective concentration.

Derivatization via oxime-formation

4. Using a freshly prepared solution of keto-protein from step 3, add the alkoxyamine solution such that the alkoxyamine is in 10- to 1000-fold molar excess of the protein.

Since oxime formation reactions occur at an optimal pH of 4.5,^{22,23} prepare the alkoxyamine solution such that the final pH will be as acidic as possible within the stability requirements of the protein. For larger proteins, it is common to carry out this reaction step between pH 5.5 and 6.5. If the protein is not stable at acidic pH, or if it is not possible to use a large excess of alkoxyamine, then aniline or anisidine catalysis of oxime formation may be used to increase the reaction rate.^{24,25}

5. Incubate at room temperature for 18 to 48 h.

6. Analyze conversion and remove excess alkoxyamine, if necessary.

If analyzing by mass spectrometry, small amounts of PLP adduct species may be observed. For identification of common product masses, including those with PLP, see Troubleshooting.

Reagents and Solutions

Use deionized, distilled water in all recipes and protocol steps.

Alkoxyamine solution

The alkoxyamine should be prepared as a concentrated stock solution. The storage conditions will vary with each individual reagent, but exposure to adventitious carbonyl groups, such as acetone, should be strictly avoided. Commercially available alkoxyamines include small molecules and dyes, and the nomenclature may include the names “hydroxylamine,” “aminooxy,” or “alkoxyamine.” Small molecule alkoxyamines that have been used to determine conversion of a protein by mass spectrometry include benzylalkoxyamine (*O*-benzylhydroxylamine hydrochloride) and (aminooxy) acetic acid (*O*-(carboxymethyl)hydroxylamine hemihydrochloride). Poly(ethylene glycol) alkoxyamines can be synthesized²⁶ for measurement of protein conversion by SDS-PAGE shifts.

PLP stock solution

The PLP stock solution should be prepared immediately before use, as this compound has been reported to degrade in aqueous solution.²⁷ The PLP solution should be made in the same buffer in which the transamination reaction will be run, typically 10 to 50 mM phosphate buffer at pH 6.5. After the addition of PLP to the buffer, it is important to check and adjust the pH, as the phosphate group of PLP may significantly alter the pH of the buffer solution. It is often convenient to make the PLP stock solution at two times the desired concentration for the reaction, assuming that it will be added to the protein solution in a 1:1 volume ratio. For a transamination reaction run with 100 mM PLP, the following guidelines may be followed to make 1 ml of a 200 mM (2x) PLP stock solution. (1) Add 53 mg of pyridoxal 5'-phosphate monohydrate (Sigma) to 1 ml of 25 mM phosphate buffer, pH 6.5, followed by the addition of 60 μ l of 5 M NaOH and brief sonication and vortexing. (2) Check the pH with pH paper or a microelectrode and adjust to 6.5 if necessary.

Protein Stock Solution

The protein stock solution should be prepared in the buffer in which the transamination reaction will be run. The standard buffer is 10 to 50 mM potassium phosphate at pH 6.5; however, other buffers, such as Tris and HEPES, will also work.² The presence of glycerol and other polyalcohols should be avoided, as they may form acetals with the PLP aldehyde group and thus lower its effective concentration. Note that many spin concentrators are packed in glycerol solutions, and can serve as unexpected sources of this impurity.

Critical Parameters

It is important that the PLP solution be prepared according to the guidelines described in the Reagents and Solutions section. Using a fresh solution of PLP, checking and adjusting the pH, and storing the solid PLP at 4 °C until use will yield optimal results.

As noted above, glycerol and other polyhydroxylated compounds should be removed before the transamination step. It should be noted that many commercially available spin concentrators are packed in glycerol solutions. Adventitious ketones (such as acetone) should be avoided when handling the alkoxyamine compounds.

Troubleshooting

Regardless of N-terminal residue identity, it is possible to vary the reaction conditions for each protein target to optimize conversion. A proxy alkoxyamine that is commercially available and/or readily analyzed can be used while the optimal conditions are being screened for a new target protein. See Reagents and Solutions for examples. To confirm the compatibility of a protein to the overall reaction conditions, it is often beneficial to run a control reaction that lacks PLP, but in which all of the other parameters (including the addition of alkoxyamine) are held constant.

During the PLP-mediated transamination reaction, PLP aldol addition products can form from the reversible keto-enol tautomerization of the keto-protein product, as shown in Figure 2-4. Both species have been observed to form oximes upon exposure to the alkoxyamine reagent in the next step; however, the presence of both intermediates creates a complex product mixture. If the goal is simply ligation of the alkoxyamine reagent to the protein, the presence of the PLP adduct species may be tolerable, but if the PLP adduct is detrimental, the following changes in the reaction conditions can often be used to minimize the amount that is formed.

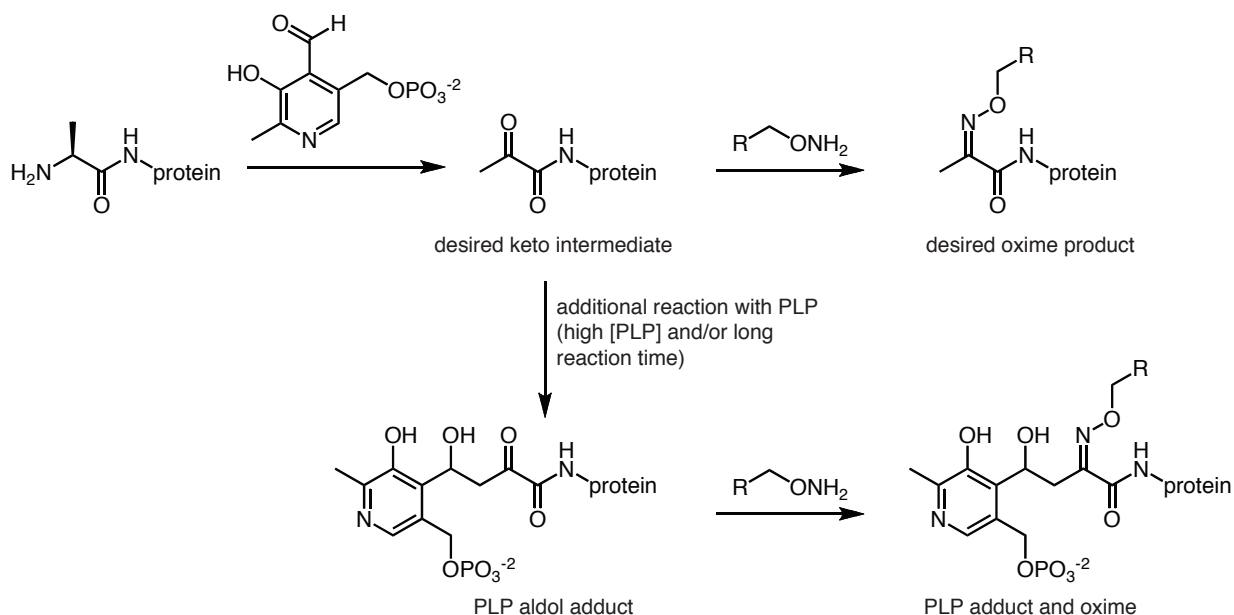


Figure 2-4. Proposed pathway for the formation of PLP adducts. In addition to the desired pathway, the enol form of the keto-protein intermediate can react with the aldehyde of PLP to form an aldol adduct. The PLP adduct structure still contains a ketone and is able to undergo oxime formation in some cases. Because the aldol reaction is bimolecular, reaction conditions found to result in higher purity oxime product and minimal PLP adduct formation were (1) the use of lower concentrations of PLP for longer reaction times, or (2) the use of higher concentrations of PLP for shorter times.

As would be expected for a bimolecular aldol reaction, the yield of the PLP addition product increases with higher PLP concentration and longer reaction times—two parameters that also influence the transamination yield. Thus, a balance between these reaction pathways must be struck. Higher concentrations of PLP (30 to 100 mM) with shorter reaction times (1 to 6 hr) or lower concentrations of PLP (5 to 10 mM) with longer reaction times (8 to 20 hr) are conditions that will help maximize the yield of the desired oxime product while keeping PLP addition products to a minimum. The optimal strategy will depend on the specific protein, so exploring multiple sets of conditions is recommended during screening.

Table 1 lists suggested strategies for overcoming potential issues encountered during PLP-mediated bioconjugation. When analyzing the mass spectrum of a PLP-modified

Table 1. Troubleshooting PLP-Mediated Bioconjugation

Problem	Possible Cause	Solution
Poor conversion to oxime product	Low yield of transamination reaction	Increase the concentration of PLP, screening from 10 mM to 100 mM. Increase the incubation time or the temperature of the transamination reaction.
	Low yield of oxime formation reaction	Increase the equivalents of alkoxyamine. Lower the pH during oxime formation step. Use aniline catalysis, as reported by Dirksen et al. Check quality of alkoxyamine reagent by NMR.
High conversion to undesired PLP adduct	Byproduct formation	Use lower concentrations of PLP. Use shorter reaction times. To identify common byproducts, see the table of common mass adducts.
Nominal reactivity	Inaccessible or unreactive N-terminal residues	Consider mutagenesis.
Protein precipitation	PLP solution is too acidic	Check and adjust pH of PLP solution before mixing with protein.
	Protein insolubility	The transamination reaction can be run under buffer, pH, and temperature conditions that vary from those listed in the Basic Protocol. Alter these parameters as dictated by the target protein's stability.

Table 2. Identifying Observed Species in Mass Spectrometry of PLP-Modified Proteins^a

Observed Mass	N-terminus	Possible Product ^b
M - 1	Any	Keto-protein
M + Alk - 19	Any	Oxime
M - 44	Asp	Decarboxylation
M - 16	Ser	Beta elimination
M - 32	Cys	Beta elimination
M + 247	Any	PLP aldol addition
M + 229	Any	PLP aldol addition with dehydration
M + 229	His, Trp	Pictet-Spengler addition of PLP
M - 19	Lys	Cyclic enamine
M + 293	Pro	Ring opening with PLP addition

^a M = unmodified protein mass. Alk = alkoxyamine mass.

^b Note that many of these species can also form oximes.

protein, the information presented in Table 2 can be used to identify the observed species. In some cases, a species may produce more than one of these mass changes (e.g., a PLP aldol addition product may also form an oxime).

Anticipated Results

For many protein targets, optimization of the basic protocol can result in 50% to 90% conversion to the desired oxime product. Site-directed mutagenesis may improve the conversion of initially unreactive target proteins. The mild reaction conditions are not expected to denature most proteins, and in most cases they retain their desired activity after modification.

Time Considerations

N-terminal protein modification using PLP-mediated transamination and subsequent oxime formation can be completed in 2-3 days, with minimal hands-on time. On the first day, the PLP transamination reaction is run for 1-2 h (although the reaction may run overnight). Following transamination, the excess small molecule is removed and the oxime-forming reaction is started. After overnight incubation, the conversion can be analyzed and the protein conjugate is ready for use. The incubation times of both the PLP transamination and the oxime formation may be reduced in some cases. Heating may reduce the conversion time for the transamination reaction, and an excess of alkoxyamine will increase the rate of the oxime-formation step. Considering the small amount of active time, it is quite feasible to modify multiple proteins in parallel or to screen multiple reaction conditions at once. A gradient program on a PCR thermocycler is convenient for screening conditions.

2.4 REFERENCES

1. Witus L.S., Francis M. Site-Specific Protein Bioconjugation via a Pyridoxal 5'-Phosphate-Mediated N-Terminal Transamination Reaction, in *Current Protocols in Chemical Biology*. John Wiley & Sons, Inc. **2010**. 125-134.
2. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S., Francis M.B. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **2006**, 45, 5307-5311.
3. Dixon, H.B.F., Fields, R. Specific Modification of NH₂-Terminal Residues by Transamination, in: *Part B: Enzyme Structure*. Academic Press. **1972**, 409-419.
4. Dixon, H.B.F. N-Terminal Modification of Proteins—a Review. *J. Prot. Chem.* **1984**, 3, 99-108.
5. Toney, M.D. Reaction Specificity in Pyridoxal Phosphate Enzymes. *Arch. Biochem. Biophys.* **2005**, 433, 279-287.
6. Eliot, A.C., Kirsch, J.F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **2004**, 73, 383-415.
7. Snell, E.E. The Vitamin B6 Group- The Reversible Interconversion of Pyridoxal and Pyridoxamine by Transamination Reactions. *J. Am. Chem. Soc.* **1945**, 67, 194-197.
8. Scheck R.A., Francis M.B. Regioselective Labeling of Antibodies Through N-Terminal Transamination. *ACS Chem. Biol.* **2007**, 2, 247-251.
9. Esser-Kahn, A.P., Francis, M.B. Protein-Cross-Linked Polymeric Materials through Site-Selective Bioconjugation. *Angew. Chem. Int. Ed.* **2008**, 120, 3811-3814.
10. Heredia, K.L., Maynard, H.D. Synthesis of Protein-Polymer Conjugates. *Org. Biomol. Chem.* **2007**, 5, 45-53.
11. Gao, W., Liu, W., Mackay, J.A., Zalutsky, M.R., Toone, E.J., Chilkoti, A. In Situ Growth of a Stoichiometric Peg-Like Conjugate at a Protein's N-Terminus with Significantly Improved Pharmacokinetics. *Proc. Natl. Acad. Sci.* **2009**, 106, 15231-15236.
12. Christman, K.L., Broyer, R.M., Tolstyka, Z.P., Maynard, H.D. Site-specific Protein Immobilization Through N-Terminal Oxime Linkages. *J. Mater. Chem.* **2007**, 17, 2021-2027.
13. Lempens, E.H.M., Helms, B.A., Merckx, M., Meijer, E.W. Efficient and Chemoselective Surface Immobilization of Proteins by Using Aniline-Catalyzed Oxime Chemistry. *ChemBioChem* **2009**, 10, 658-662.
14. Esser-Kahn, A.P., Iavarone, A.T., Francis, M.B. Metallothionein-Cross-Linked Hydrogels for the Selective Removal of Heavy Metals from Water. *J. Am. Chem. Soc.* **2008**, 130, 15820-15822.
15. Dedeo, M.T., Duderstadt, K.E., Berger, J.M., Francis, M.B. Nanoscale Protein Assemblies from a Circular Permutant of the Tobacco Mosaic Virus. *Nano Lett.* **2010**, 10, 181-186.
16. Scheck R.A., Dedeo M.T., Iavarone A.T., Francis M.B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, 130, 11762-11770.
17. Witus L.S. Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., Francis, M.B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, 132, 16812-16817.
18. Li, X., Zhang, L., Hall, S.E., Tam, J.P. A New Ligation Method for N-Terminal Tryptophan-Containing Peptides Using the Pictet-Spengler Reaction. *Tetrahedron Lett.* **2000**, 41, 4069-4073.
19. Regan, L. Protein Redesign. *Curr. Opin. Struct. Biol.* **1999**, 9, 494-499.
20. Kalia, J., Raines, R.T. Hydrolytic Stability of Hydrazones and Oximes. *Angew. Chem. Int. Ed.* **2008**, 47, 7523-7526.
21. Coligan, J.E., Dunn, B.M., Speicher, D.W., Wingfield, P.T. *Current Protocols in Protein Science*. John

Wiley & Sons, **2010**.

22. Jencks, W.P. Studies on the Mechanism of Oxime and Semicarbazone Formation. *J. Am. Chem. Soc.* **1959**, *81*, 475–481.
23. Carey, F.A., Sundberg, R.J. Advanced Organic Chemistry. Plenum. **1977**.
24. Dirksen, A., Dawson, P.E. Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling. *Bioconj. Chem.* **2008**, *19*, 2543–2548.
25. Dirksen, A., Dirksen, S., Hackeng, T.M., Dawson, P.E. Nucleophilic Catalysis of Hydrazone Formation and Transimination: Implications for Dynamic Covalent Chemistry. *J. Am. Chem. Soc.* **2006**, *128*, 15602–15603.
26. Schlick, T.L., Ding, Z., Kovacs, E.W., Francis, M.B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.
27. Ball, G.F.M. Vitamins in foods. CRC Press. **2006**.

CHAPTER 3:

DEVELOPMENT OF A COMBINATORIAL PEPTIDE LIBRARY SCREENING PLATFORM

ABSTRACT

Site-specific protein bioconjugation using pyridoxal 5'-phosphate (PLP)-mediated transamination has been used to generate bioconjugates for a number of protein targets. However, peptide studies had indicated that the N-terminal sequence could have a significant influence on the overall yield. Since internal as well as the N-terminal residue was found to have an effect, we sought to develop a method of screening all combinations of N-terminal residues at once in order to identify highly reactive sequences. This chapter describes the design of a combinatorial peptide library screening platform suitable for optimizing protein bioconjugation reactions, including the development of a colorimetric screening scheme and a high-throughput method of sequencing beads based on a built-in truncation ladder.

Portions of the work described in this chapter have been reported in a separate publication.¹

3.1 NEED FOR COMBINATORIAL APPROACH

The ability to generate site-specific protein bioconjugates is crucial for the success of many chemical biology applications. Since its initial report, pyridoxal 5'-mediated (PLP)-mediated transamination had been used to modify a wide-range of protein substrates both in the Francis group and other laboratories. These proteins included GFP, viral coat proteins, monoclonal antibodies, RNase, metallothionines, and myoglobin, among others.^{2,3} The successful production of these bioconjugates confirmed that the PLP-mediated transamination technique was a general bioconjugation strategy, applicable to a wide range of substrates. However, the transamination yield was found to be variable and unpredictable. Some proteins did not result in modification over 30 or 40%, and it was difficult to predict the extent of modification expected for new protein targets.⁴ Although low levels of attachment are suitable for some applications (such as surface or radio-labeling), we sought to improve the reliability and yield of this reaction.

Peptide studies carried out in the Francis group have showed that the identity of the N-terminal residue can significantly effect the transamination yield.⁴ Thus, the N-terminal residue could be used to predict, and improve through mutagenesis, the yield of new protein targets. However, the peptide studies also indicated that the second and third residues could have a great effect on the reactivity of a peptide sequence as well. Therefore, we sought to identify a highly reactive combination of N-terminal residues that could be incorporated at the N-terminus of new protein targets to ensure reliable high yields. To study the influence of terminal and internal residues, a large substrate space would have to be explored. A short reactive sequence was desired to allow facile genetic incorporation into recombinant proteins, but even the variation of just the three N-terminal residues to include each of the 20 natural amino acids encompasses 8,000 different sequences. As such, the synthesis of individual peptide sequences and the mass spectrometry analysis of their modification was no longer practical. Therefore, in order to identify highly reactive combinations of N-terminal sequences we turned to a combinatorial peptide library screening platform.¹

Combinatorial libraries allow the exploration of a large chemical space. Combinatorial approaches have been applied many types of molecules, including peptides, DNA, peptoids and small molecules.^{5,6} These libraries have been used for drug discovery, for the identification of inhibitors or ligands for target proteins, and for materials applications. To study the effects of N-terminal sequence on reactivity towards PLP-mediated transamination, a combinatorial peptide library was used. Although PLP-mediated transamination is used primarily as a protein bioconjugation strategy, we used a peptide library to serve as a model for protein reactivity. The N-terminus of a protein is influenced by its complex three-dimensional environment in addition to its sequence; however, previous studies had demonstrated that the sequence effects found on peptides correlated with protein reactivity.⁴

There are many types of combinatorial peptide libraries in common usage, including phage display libraries, spatially addressable libraries, and one-bead-one-compound libraries.

The choice of library type is informed by the application at hand; for example biological display techniques are useful for large libraries with long randomization sequences, while spatially addressable libraries are more suitable for small library sizes.⁵ For this application, a one-bead-one-compound (OBOC) library was used. Because the peptides are still attached to the solid-phase resin in a OBOC library, it is possible to screen the library components concurrently but independently, and the detection and removal of active sequences is facilitated by the ease of working with macroscale beads. Screening a combinatorial library involves the following steps: library synthesis, an assay for the desired characteristics, and deconvolution of the active components. To create a library screening platform that would be suitable to study protein bioconjugation reactivity, we used established techniques as well as adapting and developing new methods.

3.2 LIBRARY SYNTHESIS

As our objective was to find a short genetically encodable sequence that could be conveniently introduced through mutagenesis, we limited our randomization to the terminal three residues, which were followed by an invariant region. The invariant region consisted of the pentapeptide WSNAG, resulting in a library of the form $X_3X_2X_1$ WSNAG. This base sequence was used in order to provide the library members with enough mass to separate their signals from matrix noise peaks during MALDI-TOF MS analysis; however, in future studies this region could include purification handles or other sequences of interest. The resulting library included all 20 natural amino acids in each of the three N-terminal positions, corresponding to 8,000 members.

The library was synthesized on resin beads using a split-and-pool technique (Figure 3-1).⁵ Additional steps were performed after the coupling of each variable position to cap a portion of the growing chain to facilitate sequencing by mass spectrometry, as described in Section 3.5. Tentagel, a PEGylated polystyrene resin, was used in order to provide compatibility with both the organic solvents needed for peptide synthesis and the aqueous conditions of transamination screening (Tentagel S OH, Advanced ChemTech). To facilitate the manual handling of individual beads during the screening phase, 130 μ m beads were used. The scale of library synthesis was 500 mg of resin per batch in order to allow for multiple screening experiments. This scale was based on the statistical calculation that in each screening experiment, for 99% confidence that 95% of the library

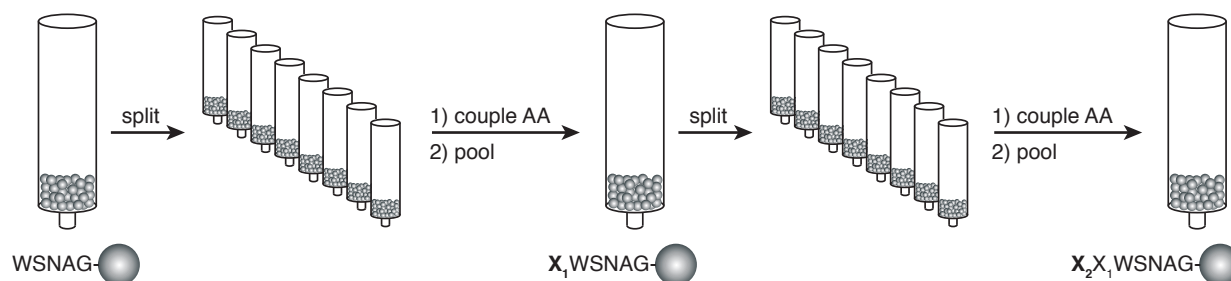


Figure 3-1. General scheme for split-and-pool one-bead-one-compound combinatorial library synthesis.

members are represented, the number of beads used must be three times the number of library members.⁷ For an 8,000 member library this corresponds to 24,000 beads. Since 130 μm Tentagel beads have a count of 890,000/g, approximately 27 mg of library would be screened in each experiment.

The C-terminal WSNAG invariant sequence was synthesized on the resin according to standard Fmoc solid-phase peptide chemistry procedures.¹ To attach the first amino acid through an ester bond, 10 equivalents of Fmoc-Gly-OH, the C-terminal amino acid, was dissolved in *N,N*-dimethylformamide (DMF) and dichloromethane (DCM) and preactivated with 5 equivalents of diisopropylcarbodiimide (DIC) on ice for 20 min. The equivalents were calculated relative to the resin loading (0.15 mmol, using 500 mg resin with a theoretical loading of 0.3 mmol/g). The dry resin was swelled by adding 10 mL DMF and subjecting to gentle rotation for 20-30 min in a 20 mL polypropylene spin column equipped with a frit at the bottom (Econo-pac Chromatography Column, BioRad). After resin swelling, the preactivated amino acid was added to the resin with 0.1 equivalents of *N,N*-dimethylaminopyridine (DMAP) as an additive, and allowed to rotate on a LabQuake shaker at room temperature for 1.5 h. After coupling, the general procedure for rinsing excess reagents away from the resin was performed, which was to drain the solution from the resin, add approximately 10 mL DMF, shake briefly, drain, and repeat for a total of 5 rinses.

To verify the attachment of the first residue to the resin and quantify loading, an Fmoc quantification was performed. The resin was dried by rinsing with DCM then applying vacuum for 10-20 min. A 2-6 mg portion of resin was removed and weighed, and 3 mL of the Fmoc deprotection solution (20% by volume piperidine in DMF) was added. After 5 min of rotation, the absorbance at 290 nm was measured using the piperidine/DMF solution as a blank. Division of the absorbance at 290 nm by 1.65 multiplied by the mass of the resin in milligrams gave the loading in mmol/g.⁸

After the attachment of the first amino acid to the resin, iterative Fmoc deprotection and amino acid coupling steps were performed to synthesize the desired sequence. The general rinsing protocol was performed after each of the deprotection and coupling steps. The amino acids used were listed in Table 1, which shows an example of a worksheet to guide the synthesis of a given peptide. The Fmoc deprotection was accomplished by incubation (with rotation on a LabQuake shaker) with 10 mL of Fmoc deprotection solution (described above) for 20 minutes (or incubation for 5 minutes followed by replacement of the deprotection solution and a second 2.5 min incubation). Coupling reactions were carried out using 10 equivalents of amino acid with 10 equivalents of 2-(6-chloro-1-*H*-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU)⁹ and 20 equivalents of *N,N*-diisopropylethylamine (DIPEA) in DMF (total volume 10 mL) for 10 minutes of rotation on a LabQuake shaker. If a particular amino acid was used more than once in a peptide sequence, a stock solution was made in DMF and aliquots were added for each coupling.

The variable positions were added to the base sequence using a split-and-pool technique

with an additional capping step before and after each variable position. To the WSNAG sequence, capping was performed by coupling 0.15 equivalents of bromobenzoic acid using 0.1 equivalents of HCTU and 1 equivalent of DIPEA. The resin was then split into 20 reaction vials by aliquoting an isopicnic suspension of the resin made with DCM and DMF. Each vial was exposed to coupling conditions using one of the twenty amino acids (using equivalents based on a scale of 25 mg resin per reaction and a total reaction volume of 1 mL). After coupling and rinsing, the resins were then recombined, with the exception of the vials corresponding to Gln or Leu, which were kept separate. After Fmoc deprotection, the next capping step was performed. Bromobenzoic acid was attached to the mixed resin sample and methylbromobenzoic acid was used to cap the Gln and Leu samples to distinguish them from isobaric residues during sequencing. The Gln and Leu samples were then mixed with the rest of the resin before splitting into separate vials for the next coupling reaction. This sequence was repeated until the desired number of variable positions were attached, ending with a capping step.

After completion of the coupling steps, side-chain deprotection was achieved using a 1-2 h incubation with a 95:2.5:2.5 ratio of trifluoroacetic acid (TFA) to H₂O to triisopropylsilane (TIPS-H), followed by equilibration in 50 mM phosphate buffer (pH 6.5) using three 5 min exposures. For peptides without variable positions, the general procedure for cleaving

Table 1: Solid-phase peptide synthesis worksheet^a

Amino Acid	Equivalents	mmol per reaction ^b	molecular weight	mg per reaction	number of occurrences	total mg needed	stock solution total volume	volume per occurrence (mL)
Fmoc-Ala-OH	10	1.5	311.1	466.7	1	466.7	2	2
Fmoc-Cys(Trt)-OH	10	1.5	585.7	878.6	0	0	0	0
Fmoc-Asp(OtBu)-OH	10	1.5	411.5	617.3	0	0	0	0
Fmoc-Glu(OtBu)-OH	10	1.5	425.5	638.3	0	0	0	0
Fmoc-Phe-OH	10	1.5	387.4	581.1	0	0	0	0
Fmoc-Gly-OH	10	1.5	297.3	446.0	0	0	0	0
Fmoc-His(Trt)-OH	10	1.5	619.7	929.6	0	0	0	0
Fmoc-Ile-OH	10	1.5	353.4	530.1	0	0	0	0
Fmoc-Lys(Boc)-OH	10	1.5	468.5	702.8	0	0	0	0
Fmoc-Leu-OH	10	1.5	353.4	530.1	0	0	0	0
Fmoc-Met-OH	10	1.5	371.5	557.3	0	0	0	0
Fmoc-Asn(Trt)-OH	10	1.5	596.7	895.1	1	895.1	2	2
Fmoc-Pro-OH	10	1.5	337.4	506.1	0	0	0	0
Fmoc-Gln(Trt)-OH	10	1.5	610.7	916.1	0	0	0	0
Fmoc-Arg(Pbf)-OH	10	1.5	648.8	973.2	0	0	0	0
Fmoc-Ser(tBu)-OH	10	1.5	383.4	575.1	1	575.1	2	2
Fmoc-Thr(tBu)-OH	10	1.5	397.5	596.3	0	0	0	0
Fmoc-Val-OH	10	1.5	339.4	509.1	0	0	0	0
Fmoc-Trp(Boc)-OH	10	1.5	526.6	789.9	1	789.9	2	2
Fmoc-Tyr(tBu)-OH	10	1.5	459.6	689.4	0	0	0	0

^a filled in for the synthesis of a WSNAG peptide, after C-terminal residue already attached to resin

^b based on a 500 mg (0.15 mmol) resin scale

the peptide from the resin (for approximately 10 mg resin) was a 30 min incubation with 300 μL of 100 mM NaOH. The filtrate was collected and 700 μL of 50 mM phosphate buffer pH 6.5 was added to neutralize the solution. If mass spectrometry analysis by liquid chromatography-mass spectrometry (LC-MS) was to be performed this solution was then diluted 20-100 fold in water. Library members were cleaved individually, as described in Section 3.5.

3.3 COLORIMETRIC DETECTION

To assay the reactivity of the N-terminal sequences in the library towards PLP-mediated transamination, a colorimetric detection method was developed. Previously, the transamination efficiency for various peptide and protein substrates had been measured by the oxime product formed after incubation with an alkoxyamine reagent.⁴ Using the oxime product to report the transamination conversion is necessary for substrates such as proteins, as the keto-protein species (a mass change of 1 Da) is difficult to measure directly. Even for peptide substrates for which it is possible to measure the transaminated species by mass spectrometry, it is desirable to measure the extent of oxime formation possible after transamination since the oxime linkage is the ultimate goal. Therefore, in previous studies the conversion of individual peptides had been determined by LC-MS analysis following transamination and benzyl oxime formation. This method has been quite powerful, as it allowed quantification of the products formed. However, LC-MS analysis of each individual peptide would be far too time-intensive to be practical for screening an entire combinatorial library. Therefore, a new, high-throughput method was needed to survey the transamination activity of various sequences.

As shown in Figure 3-2, the screening scheme consisted of subjecting the entire library, on the solid-phase resin, to the transamination reagent. Stringent transamination conditions (such as low concentrations of PLP) could be used so that only the most highly reactive sequences would undergo transamination. To distinguish the beads that contained

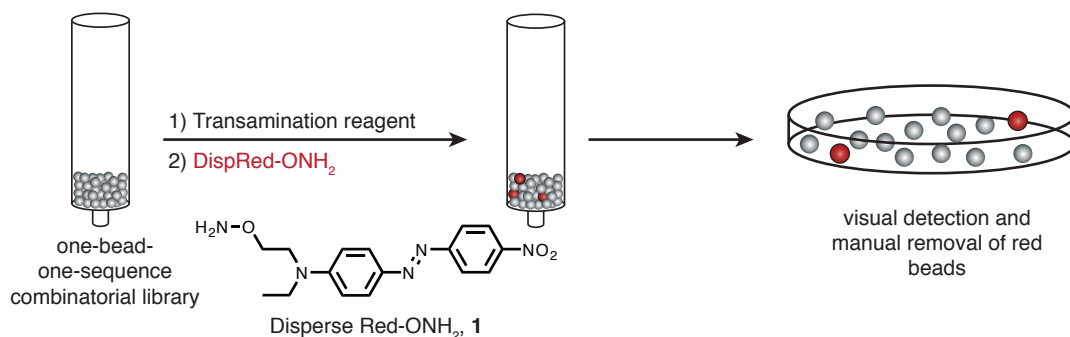


Figure 3-2. Library screening scheme. The 8,000-member combinatorial peptide library of the form XXXWSNAG was subjected to PLP-mediated transamination and subsequent oxime formation with a Disperse Red alkoxyamine (**1**). Stringent reaction conditions were used during the transamination to target only sequences with high reactivity. These active sequences that resulted in a keto-peptide product were colorimetrically distinguished from the others during the next step, where the Disperse Red dye formed a covalent oxime linkage with the transaminated sequences. The library was then examined under a microscope, and the red beads were manually removed for sequencing.

the ketone and aldehyde products such that they could be identified and sequenced, oxime formation was used as a reporter. By synthesizing a colorful alkoxyamine reagent, Disperse Red alkoxyamine (DispRed-ONH₂, 1), the beads containing sequences that formed an oxime product would have a red color. To determine whether the red color of the beads after incubation with DispRed-ONH₂ was due to covalent attachment of the dye to the peptide via oxime formation, a peptide of known sequence was tested before the dye was used on the library. A sample of unmodified and transaminated (10 mM PLP, 18 h, pH 6.5) EGKWSNAG beads were incubated with 10 mM DispRed-ONH₂ for 3 h in DMF, rinsed with DCM and DMF three times each, and then mixed to see if the two samples could be differentiated. A clear contrast between beads with and without covalent modification was observed (Figure 3-3). This demonstrated that a visually detectable red color could be used to identify the formation of oxime product, thus allowing a simple and high-throughput method of screening the reactivity of the entire combinatorial library simultaneously.

After the transamination and oxime formation reactions, spatial separation of the library components allowed visual inspection for red beads. The library beads were rinsed with ethanol (EtOH) for transfer to a Petri dish for examination under a simple light microscope (Leica S6D). The red beads were manually removed for determination of the corresponding sequences using a pipette set to 3 μ L. The excess ethanol was removed with a pipette and then the bead was dried under a flow of N₂. The peptide on the bead was then cleaved from the resin for sequencing by MALDI-TOF analysis by incubation with 10 μ L of 100 mM NaOH for 1 h. After cleavage, The solution was desalted using Ziptips with a 0.2 μ L C18 resin. The peptide was eluted from the Ziptip with 2 μ L of matrix solution directly onto a MALDI sample plate.

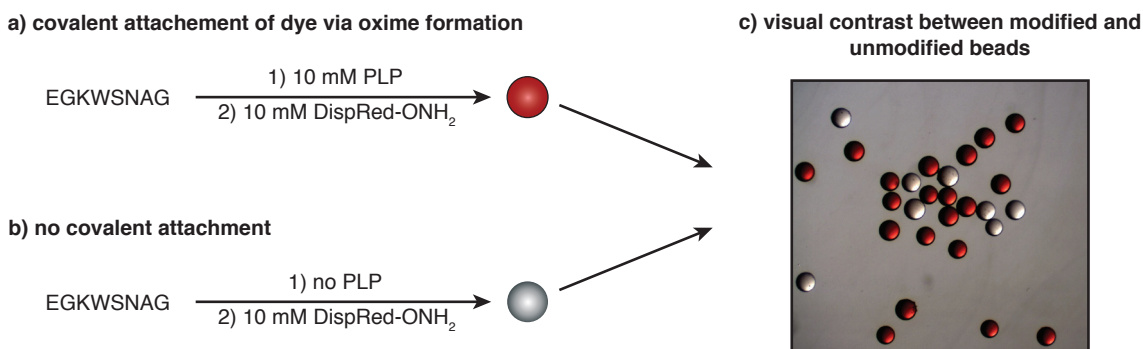


Figure 3-3. Testing contrast for modified and unmodified samples. (a) A peptide known to give high yields under standard transamination conditions, EGKWSNAG, was subjected to standard PLP transamination conditions to install a keto group. The beads were then incubated with Disperse Red alkoxyamine to form an oxime with the dye, resulting in a visually red bead. (b) The same peptide was subjected to Disperse Red alkoxyamine without prior transamination to check for non-specific binding of the dye to the bead. The resulting colorless beads confirmed that there was no association of the dye in the absence of a covalent linkage. (c) The beads with and without covalent attachment of Disperse Red were mixed and viewed under a light microscope. The contrast was sufficient to distinguish beads bearing modified and unmodified peptides easily.

3.4 RE-CALIBRATION OF COLORIMETRIC DETECTION

When the colorimetric detection system was used to screen the library, the sequences identified from red beads were resynthesized to verify the results (discussed in depth in Chapter 4). The red color of the beads indicated the formation of oxime product but it was found that the Disperse Red oxime was not amenable to mass spectrometry. Therefore, the resynthesized sequences were subjected to transamination, followed by oxime formation with a benzyl alkoxyamine for quantification of the benzyl oxime yield by LC-MS. These experiments revealed that the red color on the beads saturated at ~30% oxime yield.¹ This was sufficient to identify highly reactive sequences, since the sequences that resulted in red beads under the stringent transamination conditions used to screen the library led to higher yields under more generous transamination conditions. However, identification of the high yielding sequences was difficult.

To use the colorimetric screening method to distinguish between high and low levels of oxime formation, we recalibrated the detection scheme such that a visually red bead would correspond to a high oxime yield (Figure 3-4). To do this, reaction conditions were established that would result in beads with a low and high level of modification using a known peptide sequence. The peptide used was AOTWSNAG (where O stands for ornithine, which was found to have similar reactivity to the AKT sequence described in Chapter 4). With this sequence, incubation with 100 μ M PLP for 18 h resulted in 20% oxime yield (representing a low modification case), as verified by LC-MS quantification of the corresponding benzyloxime. Incubation with 10 mM PLP for 18 h resulted in 75% yield, representing a high conversion case. Then, DispRed-ONH₂ was combined with different ratios of a colorless alkoxyamine (BnONH₂) to find a ratio that resulted in the greatest visual contrast between low and high oxime yield. With a high proportion (67%) of DispRed-ONH₂, both cases resulted in a red color, and with a low ratio (1%) neither beads with high nor low levels of modification had a red color. An intermediate ratio was found (20% DispRed-ONH₂, using a solution that was 22.5 mM BnONH₂ and 5 mM DispRed-ONH₂) where only the high oxime yield beads had a red color. Thus, this ratio of alkoxyamines was used for subsequent library screening to identify transamination

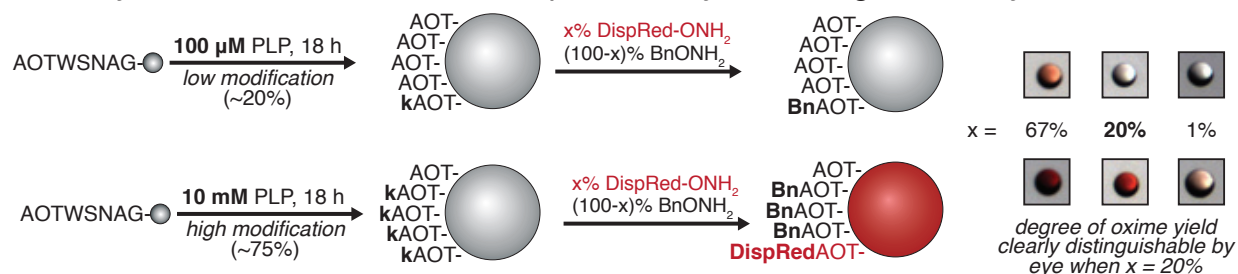


Figure 3-4. Calibration of colorimetric detection for a combinatorial peptide library screening platform. To adjust the sensitivity of the colorimetric detection method, a colorless alkoxyamine (BnONH₂) was combined with DispRed-ONH₂ in varying ratios. PLP-mediated transamination of a peptide sequence (AOT-terminal, with similar reactivity to AKT-termini) was used to create beads with high (75%) and low (20%) levels of modification. The greatest visual distinction between beads bearing high and low modification levels was observed when using 20% DispRed-ONH₂.

conditions that resulted in maximal yields.

3.5 SEQUENCING BY TRUNCATION LADDER

A key step of the library screening scheme was the development of a simple and rapid method of determining the sequence on active beads. To this end, we created a sequencing technique based on a built-in truncation ladder.¹⁰ As compared to *de novo* sequencing based on tandem MS/MS fragmentation data, this approach allowed for facile identification of the peptide sequence using only intact ions. To create a truncation ladder, a small portion (<10%) of the growing peptide on each bead was capped with bromobenzoic acid during synthesis of the variable positions. Figure 3-5 shows an LC-MS total ion count chromatogram of a representative library member, in which the relative amounts of peptide and capped species on an individual bead can be seen. Since each individual bead contains only a few hundred picomoles of material, the amount of capping had to be sufficient for MALDI-TOF signal of each truncated species, but low enough that each bead retained some full length peptide for reaction and oxime formation. Fortunately the truncated species, lacking an N-terminus, were innocuous spectators during the transamination of the full length peptide.

Once a red bead was selected from the library screening and cleaved, all of the peptide species on the bead were examined by MALDI-TOF mass spectrometry. The unique isotope patterns of the bromine-containing caps on the truncated species enabled their unambiguous identification from other noise. Thus, the sequence on each bead could be determined by simple calculation of the mass differences between the molecular ions of the capped species (Figure 3-6). Additionally, isobaric amino acids (isoleucine and leucine) and nearly isobaric amino acids (glutamine and lysine) were differentiated by

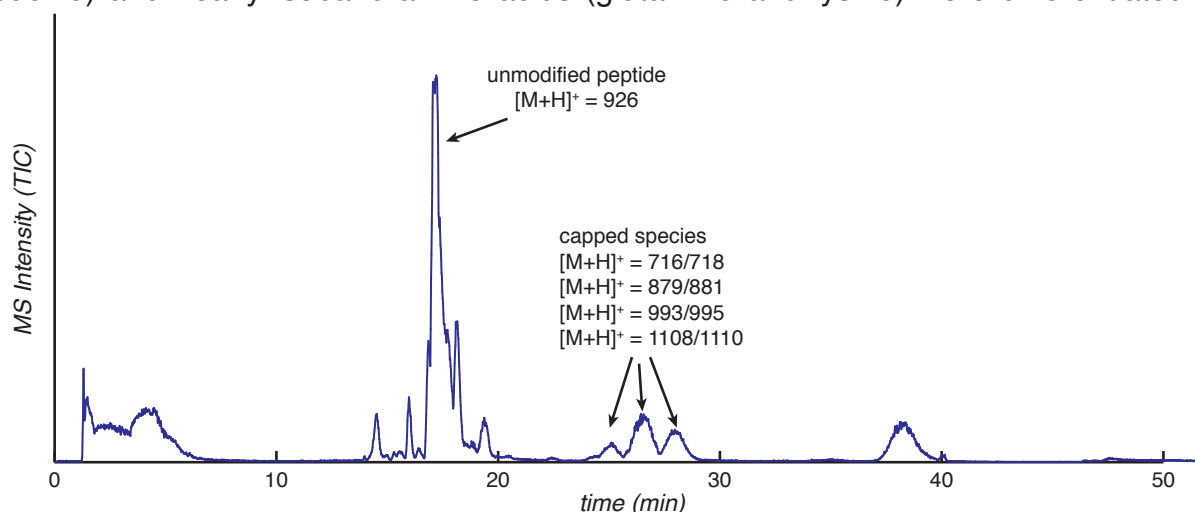


Figure 3-5. ESI-QTOF MS analysis of the peptide mixture released from a single bead from the library. The main species identified was the unmodified peptide. The capped species (identified by their bromine isotope patterns) and other impurities comprised a small portion of the total material on the bead. The mass differences observed between the capped species corresponded to the amino acids DNY, and the sequence assignment was confirmed when the calculated mass of the peptide (DNYWSNAG $[M+H]^+ = 926.31$ m/z) matched the observed $[M+H]^+ = 926.25$ m/z .

capping one of each pair with methylbromobenzoic acid (**3**) during the synthesis.

In summary, we have developed a high-throughput method of optimizing protein bioconjugation reactions using a combinatorial peptide library in which short peptides serve as a model for protein reactivity. The screening platform developed involved a one-bead-one-compound peptide library, a colorimetric detection scheme to identify high-yielding sequences, and a deconvolution method using a built-in peptide truncation ladder for the rapid sequencing of hit beads using mass spectrometry. This screening platform can be used for the discovery and optimization of N-terminal protein modification reactions.

3.6 MATERIALS AND METHODS

General Procedures. Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H₂O) used as reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Pyridoxal 5'-phosphate monohydrate was obtained from Aldrich. All Fmoc-protected amino acids were obtained from Novabiochem (EMD, Germany). Tentagel S OH resin was obtained from Advanced ChemTech (Louisville, KY). Centrifugations were conducted with an Eppendorf Mini Spin Plus (Eppendorf, Hauppauge, NY).

General Procedure for Solid-Phase Peptide Synthesis. See Section 3.2.

Split-and-Pool Combinatorial Library Synthesis. See Section 3.2.

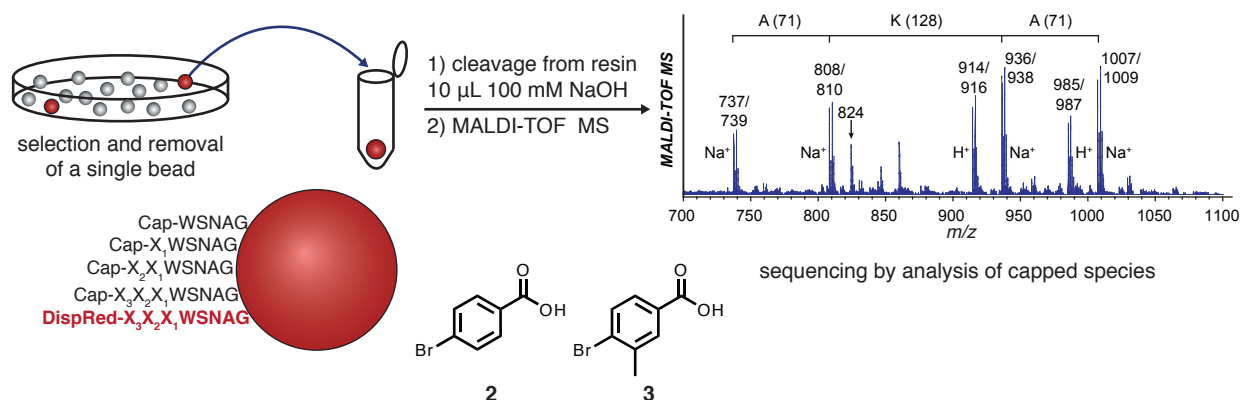


Figure 3-6. Each bead contained a truncation ladder in addition to the full-length peptide that had participated in the reaction. Sequencing was performed by cleaving the peptide species from selected beads and identifying the ladder peptides using MALDI-TOF MS. The mass differences between adjacent capped species corresponded to the amino acid in that position. Capped species were easily identified by the unique isotope pattern of the bromine atoms in the cap (**2** and **3**). The mass of the uncapped peptide can be identified in the mass spectrum at 824 m/z , confirming the assignment of a representative bead. Isobaric residues lysine and glutamine, and isoleucine and leucine, were distinguished by capping one of each pair with methylbromobenzoic acid (**3**) instead of bromobenzoic acid (**2**).

General Procedure for Library Screening. Portions of resin-bound peptides (appx. 25 mg each) were treated with 1 mL of PLP solutions of varying concentration in 50 mM potassium phosphate buffer, pH 6.5 with 10% DMF to facilitate bead swelling and 0.02% NaN_3 for preservation of the buffer. Prior to use, the pH of each PLP solution was adjusted to 6.5 with NaOH following the addition of the PLP. After 18 h of reaction time, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual PLP. The peptides were then incubated with 1 mL of a 10 mM solution of Disperse Red alkoxyamine (**1**) in DMF for 3 h at rt, or the specified ratio of DispRed- ONH_2 to BnONH_2 in a solution of DMF and water. The excess disperse red alkoxyamine was removed by rinsing with three portions of dichloromethane (DCM), followed by three portions of DMF. The beads were then rinsed with ethanol and transferred to a Petri dish for visual inspection. The beads were examined using a Leica S6D Microscope and L2 Light Source (Leica, Germany) equipped with a Moticam 2300 3.0 MP camera using Motic Image Plus 2.0 ML Software for capturing images. Individual red beads were manually removed using a Pipet-Lite LTS L-20 pipet (Rainin, Oakland, CA) and transferred to PCR tubes for sequencing. The residual ethanol in the tubes was removed by pipetting.

General Procedure for Library Sequencing. See Section 3.3 for the cleavage conditions of the peptide from a single bead. MALDI-TOF analysis was performed on a Voyager-DE instrument (Applied Biosystems), and all spectra were analyzed using Data Explorer software. The matrix solution was a saturated α -cyano-4-hydroxycinnamic acid solution in 50% acetonitrile, 50% water, 0.1 % TFA.

General Procedure for the Modification of Resin-Bound Peptide Substrates. Portions of resin-bound peptides (appx. 10 mg each) were treated with 1 mL of PLP solutions of varying concentration in 50 mM potassium phosphate buffer, pH 6.5 with 10% DMF to facilitate bead swelling and 0.02% NaN_3 for preservation of the buffer. Prior to use, the pH of each PLP solution was adjusted to 6.5 with NaOH following the addition of the PLP. After 18 h of reaction time, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual PLP. The peptides were then incubated with 1 mL of a 250 mM *O*-benzylhydroxylamine hydrochloride (BnONH_2) solution in water for 3 h at rt. Excess alkoxyamine was removed by rinsing with three portions of deionized water, followed by three portions of DMF. The peptides were then cleaved from the resin through incubation with 300 μL of a 100 mM sodium hydroxide solution for 30 min. The resulting peptide solution was added to 700 μL of 50 mM phosphate buffer (pH 6.5). The resulting solution was diluted 20-fold into dd- H_2O for mass spectrometry analysis.

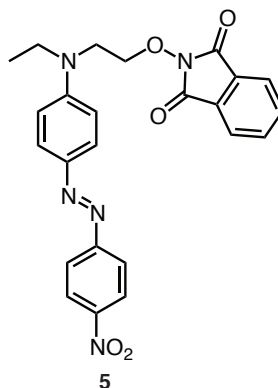
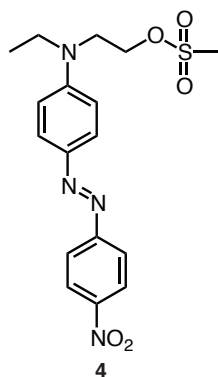
Synthesis of Disperse Red Alkoxyamine.

NMR. ^1H and ^{13}C spectra were measured with a Bruker AVQ-400 (400 MHz) or AV-300 (300 MHz) spectrometer. ^1H NMR 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), 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 nH , and is based on spectral integration values. ^{13}C NMR spectra are reported as δ in units of part per million (ppm) relative to chloroform- d (77.23, t).

Disperse Red mesylate (4). To an oven-dried 500 mL round bottom flask was added Disperse Red 1 (2.0 g, 6.4 mmol, 1.0 equiv) and distilled methylene chloride (70 mL). The mixture was stirred in an ice bath under an atmosphere of nitrogen until the Disperse Red had dissolved (<5 min). *N,N*-Diisopropylethylamine (DIPEA) (2.5 mL, 14.6 mmol, 2.3 equiv) was added dropwise, followed by the dropwise addition of mesyl chloride (590 μ L, 7.6 mmol, 1.2 equiv). The reaction was allowed to warm to room temperature and was stirred for an additional 2 h. The mixture was washed with saturated ammonium chloride and brine, and the organic layer was dried over anhydrous sodium sulfate, filtered through glass wool and concentrated under reduced pressure. The resulting red solid was used in subsequent experiments without further purification (2.5 g, quant. yield). High resolution FAB-MS M^+ : calculated 392.1154, found 392.1149. 1H -NMR (300 MHz, $CDCl_3$) δ 8.31 (d, 2H, $J=9.2$) 7.90 (m, 4H) 6.77 (d, 2H, $J=9.3$) 4.39 (t, 2H, $J=6.0$) 3.79 (t, 2H, $J=6.0$) 3.54 (q, 2H, $J=7.2$) 3.06 (s, 3H) 1.25 (t, 3H, $J=7.2$) ^{13}C -NMR (400 MHz, $CDCl_3$) δ 156.6, 150.7, 147.5, 144.1, 126.2, 124.7, 122.7, 111.5, 66.0, 53.4, 49.3, 46.0, 37.6, 12.3.

Disperse Red phthalimide (5). To a 200 mL round bottom flask was added **4** (2.5 g, 6.4 mmol, 1.0 equiv), *N,N*-dimethylformamide (DMF) (20 mL) and *N*-hydroxyphthalimide (1.25 g, 7.6 mmol, 1.2 equiv). To this solution was added 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) (1.14 mL, 7.6 mmol, 1.2 equiv), and the mixture was stirred under nitrogen and heated (50 $^{\circ}C$) for 3 h. Following the reaction, the DMF was evaporated under reduced pressure. The mixture was dissolved in a 2:1 mixture of dichloromethane and isopropyl alcohol and eluted through a silica gel plug. The resulting solution was transferred to a separatory funnel and washed with sodium bicarbonate solution, water, and brine. The organic layer was then dried over sodium sulfate, filtered through glass wool, and concentrated under reduced pressure (1.9 g, 66% yield). High resolution FAB-MS M^+ : calculated 459.1543, found 459.1540. 1H -NMR (400 MHz, $CDCl_3$) δ 8.31 (d, 2H, $J=9.2$) 7.90 (m, 4H) 7.83 (q, 2H, $J=2.8$) 7.75 (q, 2H, $J=2.8$) 6.78 (d, 2H, $J=9.2$) 4.41 (t, 2H, $J=6.0$) 3.87 (t, 2H, $J=6.0$) 3.64 (q, 2H, $J=7.2$) 1.29 (t, 3H, $J=7.2$) ^{13}C -NMR (400 MHz, $CDCl_3$) δ 163.4, 156.7, 151.0, 147.4, 143.9, 134.7, 128.7, 126.3, 124.7, 123.7, 122.6, 111.5, 75.9, 48.7, 46.1, 12.3.



Disperse Red alkoxyamine (1). To an oven-dried 500 mL round bottom flask was added **5** (1.9 g, 4.2 mmol, 1.0 equiv), methylene chloride (50 mL), and hydrazine hydrate (1.02 mL, 21.1 mmol, 5.0 equiv). The reaction was stirred for 1 h at rt, during which time an off-white precipitate formed. The white precipitate was removed by filtration and the red filtrate was concentrated under reduced pressure and used without further purification (0.89 g, 68% yield). High resolution FAB-MS M^+ : calculated 329.1488, found 329.1494. $^1\text{H-NMR}$ (400 MHz, CDCl_3) δ 8.31 (d, 2H, $J=9.2$) 7.89 (t, 4H, $J=9.6$) 6.78 (d, 2H, $J=9.2$) 3.86 (t, 2H, $J=6.0$) 3.65 (t, 2H, $J=6.0$) 3.52 (q, 2H, $J=7.2$) 1.23 (t, 3H, $J=7.3$). $^{13}\text{C-NMR}$ (400 MHz, CDCl_3) δ 157.1, 151.7, 147.5, 143.8, 126.5, 124.9 122.8, 111.5, 72.8, 49.1, 46.0, 12.4.

3.7 REFERENCES

1. Witus L.S. Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., Francis, M.B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, 132, 16812-16817.
2. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S., Francis M.B. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **2006**, 45, 5307-5311.
3. Scheck R.A., Francis M.B. Regioselective Labeling of Antibodies Through N-Terminal Transamination. *ACS Chem. Biol.* **2007**, 2, 247-251.
4. Scheck R.A., Dedeo M.T., Iavarone A.T., Francis M.B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, 130, 11762-11770.
5. Lam, K.S., Lebl, M., Krchňák, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, 97, 411-448.
6. Lam, K.S., Salmon, S.E., Hersh, E.M., Hruby, V.J., Kazmierski, W.M., Knapp, R.J. A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **1991**, 354, 82-84.
7. Burgess, K., Liaw, A.I., Wang, N. Combinatorial Technologies Involving Reiterative Division/Coupling/Recombination: Statistical Considerations. *J. Med. Chem.* **1994**, 37, 2985-2987.
8. NovaBiochem Catalog and Peptide Synthesis Handbook. NovaBioChem. **2007**.
9. Hood, C.A., Fuentes, G., Patel, H., Page, K., Menakuru, M., Park, J.H. Fast Conventional Fmoc Solid-Phase Peptide Synthesis with HCTU. *J. Pept. Sci.* **2008**, 14, 97-101.
10. Youngquist, R.S., Fuentes, G.R., Lacey, M.P., Keough, T. Generation and Screening of Combinatorial Peptide Libraries Designed for Rapid Sequencing by Mass Spectrometry. *J. Am. Chem. Soc.* **1995**, 117, 3900-3906.

CHAPTER 4:

IDENTIFICATION OF A HIGHLY REACTIVE N-TERMINAL SEQUENCE TOWARDS PLP-MEDIATED TRANSAMINATION

ABSTRACT

Chemical reactions that facilitate the attachment of synthetic groups to proteins are useful tools for the field of chemical biology. We have previously reported a pyridoxal 5'-phosphate (PLP)-mediated reaction that site-specifically oxidizes the N-terminal amine of a protein to afford a ketone or an aldehyde. This unique functional group can then be used to attach a reagent of choice through oxime formation. Since its initial report, we have found that the N-terminal sequence of the protein can significantly influence the overall success of this strategy. To obtain short sequences that lead to optimal conversion levels, a combinatorial peptide library screening platform was used. A highly reactive alanine-lysine motif emerged, which was confirmed to promote the modification of peptide substrates with PLP. This sequence was also tested on two protein substrates, leading to substantial increases in reactivity relative to their wild-type termini. This readily encodable tripeptide thus appears to provide a significant improvement in the reliability with which the PLP-mediated bioconjugation reaction can be used. This study also provides an important first example of how synthetic peptide libraries can accelerate the discovery and optimization of protein bioconjugation strategies.

Portions of the work described in this chapter have been reported in a separate publication.¹

4.1 LIBRARY SCREENING

Pyridoxal 5'-phosphate (PLP)-mediated bioconjugation has been used to modify a number of different proteins in applications ranging from surface and polymer attachment^{2,3} to the dual modification of viral coat proteins with different chromophores for light harvesting systems.⁴ Since the initial report of this method for protein functionalization,⁵ our group has sought to explore the scope of this reaction and to improve its reliability for new protein targets. Our previous work has shown that the transamination reactivity of a protein is dependent on the identity of the N-terminal residue, as well as those in the second and third positions.⁶ The terminal and internal residues were found to have a complex synergistic effect on the overall reactivity, quickly outpacing our ability to screen sequence candidates individually. We sought to identify the most highly reactive sequences of amino acids from the chemical space spanning all combinations of three N-terminal residues. To identify such sequences, we developed a combinatorial peptide library screening platform based on a colorimetric detection of desired reactivity, as described in Chapter 3.

In order to probe the reactivity of the sequences in the combinatorial library towards PLP-mediated transamination, the library was initially subjected to standard peptide transamination conditions (10 mM PLP, 18 h, room temp, pH 6.5). The colorimetric read-out of reactivity in the assay corresponded well with our previous experience. The majority of the library members were modified by the alkoxyamine dye to some extent but varying degrees of conversion for different sequences could be observed in the various shades of red ranging from nearly colorless to light orange to deep red (Figure 4-1b). In order to access sequences with greater than average reactivity, we next applied more stringent reaction conditions. The stringent reaction were conditions accessed by lowering the concentration of PLP (and reducing the time of the reaction, as described in Section 4.7).

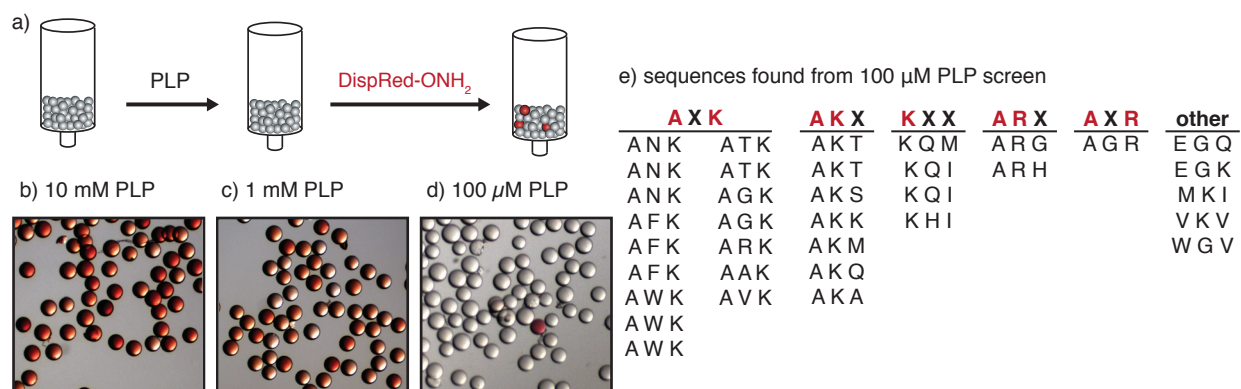


Figure 4-1. Visual evaluation of the library at different screening conditions, including sequences identified from a 100 μM PLP reaction. a) The library screening procedure consisted of subjecting the one-bead-one-sequence combinatorial library to transamination conditions followed by a Disperse Red alkoxyamine such that beads containing sequences that formed the desired oxime product would result in a red color. Stringent transamination conditions (accessed by lowering the concentration of PLP) were used to target highly reactive sequences. b) When subjected to standard transamination conditions, 10 mM PLP, the majority of the library members turned red, confirming previous results that many sequences exhibit at least some degree of conversion. c) Using 1 mM PLP, fewer library members turned red, and most showed a lighter color. d) Using only 100 μM PLP, most of the sequences showed no conversion (as evidenced by the colorless beads) but a small number of red beads were still identified. e) The sequences corresponding to the red beads at 100 μM PLP revealed AXK and AKX as the predominant N-terminal motifs.

Sequences with high reactivity, which resulted in oxime product when other sequences did not, were observed upon treatment of the library with low concentrations of PLP for the standard reaction time (18 h). Subjecting the library to 1 mM PLP resulted in fewer and lighter red beads than the standard conditions (Figure 4-1c). When the library was treated with 100 μ M PLP, a stark contrast in relative reactivity was seen: almost all of the beads were colorless except for a very small number (approximately 1 in 300) that were bright red (Figure 4-1d). The library was screened multiple times with these conditions.

The sequences that were identified after selection and sequencing of the red beads are listed in Figure 4-1e. These sequences showed a high degree of consensus and were categorized by a few convergent patterns. The preferred N-terminal amino acid showed a striking preference for alanine, which was typically followed by a positively charged amino acid in the second or third position, with lysine occurring more often than arginine. There were also several lysine terminal sequences, as well as other sequences that did not match this motif. The prevalence of N-terminal alanine fit our previous observations of this reaction,⁶ as alanine had been identified as a successful N-terminal residue in peptide and protein substrates. However, the benefits of the adjacent positive charge had not been formally observed. Interestingly, the previous individual peptide reactivity studies also contained a lysine residue in the second position. This was included for the purpose of increased solubility and to check for lysine modification in addition to N-terminal transamination. Although it was not modified directly, it now seems that this residue was more than an innocent bystander in these reactions.

4.2 RESYNTHESIS AND VERIFICATION USING PEPTIDES

Once a consensus sequence was identified from library screening, the reactivity of the motif containing an alanine followed by a positive charge was verified by quantifying the reaction yield on a resynthesized peptide. Although the red color of the beads identified during library screening was indicative of the oxime product, the MALDI-TOF mass spectra of the peptide mixtures cleaved from individual beads were not sufficient for quantification of the reaction yield. Instead, quantification of the reaction conversion was achieved by treating the resynthesized peptides with PLP followed by benzylalkoxyamine and analysis by liquid chromatography-mass spectrometry (LC-MS). Benzylalkoxyamine was used because we found that the Disperse Red oxime suppressed ionization. After analysis by LC-QTOF MS, integration of the extracted ion chromatograms of the reaction products was used to obtain the relative yields of each species.⁶ As confirmation of these values, the tryptophan fluorescence ratios of the products were also compared in some cases and similar results were obtained.

The general procedure for the modification of peptide substrates was performed as follows. Portions of resin-bound peptides (appx. 10 mg each) were treated with 1 mL of PLP solutions of varying concentration in 50 mM potassium phosphate buffer, pH 6.5 with 10% DMF to facilitate bead swelling and 0.02% NaN_3 for preservation of the buffer. The

PLP solution was freshly prepared for each experiment and the pH of the solution was adjusted to 6.5 with NaOH following the addition of the PLP. After 18 h of gentle rotation at room temperature, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual PLP. The peptides were then incubated with 1 mL of a 250 mM *O*-benzylhydroxylamine hydrochloride (BnONH₂) solution in water for 3 h at rt. Excess alkoxyamine was removed by rinsing with three portions of deionized water, followed by three portions of DMF. The peptides were then cleaved from the resin through incubation with 300 μ L of a 100 mM sodium hydroxide solution for 30 min. The resulting peptide solution was added to 700 μ L of 50 mM phosphate buffer (pH 6.5). The resulting solution was diluted 20-fold into dd-H₂O for mass spectrometry analysis.

To obtain a more detailed understanding of the reactivity of the alanine plus positive charge motif, we resynthesized an AKTWSNAG peptide. AKT was selected as a sequence for testing because lysine came up more frequently in library screening than arginine, and this particular sequence was identified multiple times. Upon subjecting beads bearing the AKT-terminal peptide to conditions identical to those that were used to screen the library (100 μ M PLP, 18 h, room temperature), they were visually found to react with the Disperse Red alkoxyamine much as they had during the library screen. However, the benzyl oxime yield was found to be approximately 30%, Figure 4-2. This indicated that the red color of a bead saturated at a relatively low yield, prompting a recalibration of the colorimetric detection (described in Chapter 3). However, in this case the higher reactivity of the Ala-Lys motif was also observed after transamination with other concentrations of PLP and led to higher yields at more generous transamination conditions (Figure 4-2).

To verify the synergistic effect of the Ala-Lys combination, peptides containing only one of these components (i.e., AET- and LKT-terminal peptides) were synthesized. Under the transamination conditions used to screen the library, only beads bearing the consensus motif resulted in red color. Using LC-MS quantification, these peptides were found to be unreactive after treatment with the low concentration of PLP. At higher PLP

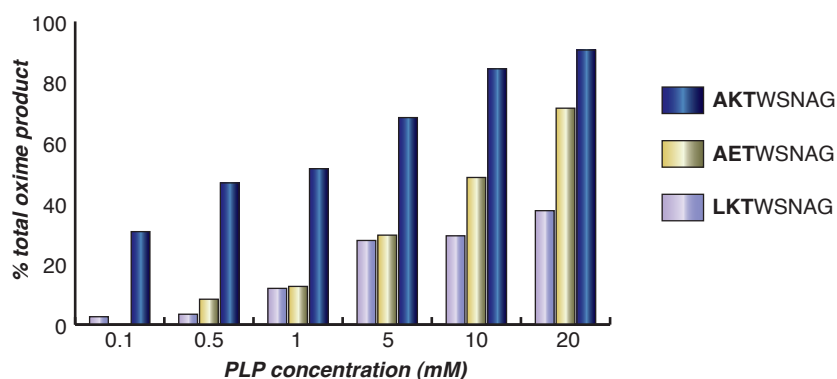


Figure 4-2. Analogs of the reactive Ala-Lys motif identified during library screening were synthesized to verify the reactivity of this sequence. AKT was compared to LKT to confirm the importance of alanine as the first residue and AET was used to probe the role of a positive charge in the second position. At 0.1 mM PLP, the concentration of library screening, AKT showed a significantly higher yield than the other sequences. AKT also resulted in notably improved yields at higher concentrations of PLP.

concentrations, the yield of oxime product increased for all three peptides, with the AKT sequence consistently producing higher reactivity. This indicated that the Ala-Lys motif had enhanced reactivity and could produce high yields of the desired product.

Although lysine was identified more frequently than arginine in the library screening, it was unclear whether the lysine functionality was uniquely important to the reactive motif or whether the effect could be achieved with any positive charge. To study this, peptides with a variety of positively charged functional groups in the second position were synthesized. These peptides were of the form AXTWSNAG where X included lysine, ornithine, arginine and dimethyl lysine as the positive charges, and alanine and glutamate for comparison. Lysine and ornithine both contain a primary amine functional group, while arginine and dimethyl lysine contain positive charges but are unable to form imines. These peptides were subjected to both low (100 μ M) and high (10 mM) concentrations of PLP for 18 h, and the resulting products were quantified by LC-MS (Figure 4-3). All of the positively charged residues had higher reactivity than the AAT and AET controls at low concentrations of PLP. Using the 10 mM PLP reaction conditions, all of these alanine-terminal peptides had high reactivity. There was not a significant difference observed between the functionalities that had a primary amine and those that did not, indicating the beneficial effect of lysine comes from the positive charge. Therefore, arginine may be interchangeable with lysine to fulfill this role in the reactive motif in a protein context. Arginine may have been observed less frequently during library screening due to small sample size, or because arginine-containing peptides were more difficult to sequence due to the slower deprotection kinetics of the arginine side-chain protecting group.

On peptides, the Ala-Lys motif resulted in higher reactivity at low concentrations of PLP, and maintained a superior reactivity at other reaction conditions. However, at generous reaction conditions, such as 10 mM PLP for 18 h, other sequences resulted in high yields as well. To consider if the consensus motif could have enhanced reactivity that would benefit the yield achieved on protein substrates, the proposed mechanism was considered before actual protein substrates were tested. The mechanism of the transamination

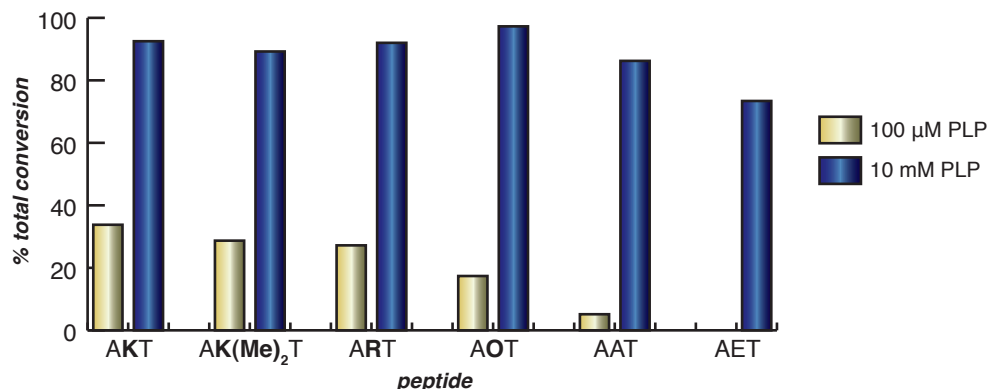


Figure 4-3. The role of the lysine in the reactive Ala-Lys motif was probed using peptides that contained a variety of positively charged functional groups in the second position, and alanine and glutamate as controls. There was not a significant difference in reactivity between lysine, dimethyl lysine, arginine, and ornithine, indicating the presence of a positive charge near the N-terminus is more important than the lysine sidechain functionality in particular.

reaction almost certainly involves reversible imine formation as the first step, followed by tautomerization. It is likely that the positive charge of the lysine residue serves to recruit the PLP molecule through interactions with the phosphate group, thus increasing the value of K_1 in Figure 4-4c. The origin of the beneficial alanine effect is less clear, as this residue could influence the equilibrium position of the tautomerization step (K_2) as well as encourage hydrolysis and/or oxime formation. Protein modification reactions are typically performed at substrate concentrations that are significantly lower than those used for peptide modification. Although we have not yet determined the rate law for the transamination reaction, it is reasonable to expect it to depend on the concentration of both PLP and the N-terminal group. Correspondingly, the low concentration of a protein will be particularly sensitive to the magnitude of K_1 in Figure 4-4c.

4.3 POSITIONAL SCANNING

A positional scanning experiment was performed to determine the ideal position (second or third) of the lysine in the motif, since the sequences identified from library screening included positive charges in both positions. The positional scanning experiment could also identify any residues that might suppress the reactivity of the motif, which would be important to avoid when using this motif as a tag sequence on proteins. We synthesized 38 octapeptides, in which the three N-terminal residues were of the form AXK or AKX. The identity of X was varied to include each of the 20 natural amino acids, with the exception of cysteine. After subjecting these peptides to standard PLP transamination conditions (10 mM, 18 h), we found that all of the sequences with alanine in the first position and lysine in the second or third position gave relatively high yields of oxime product (Figure 4-5). Despite some variation in yield, the lowest observed was still above 50%, and no residues were found to quench the reactivity of the motif substantially. The most apparent trend was a slight preference for lysine in the second position over the third position. This verified our selection of alanine-lysine-threonine as a specific sequence to pursue as an optimized protein modification tag.

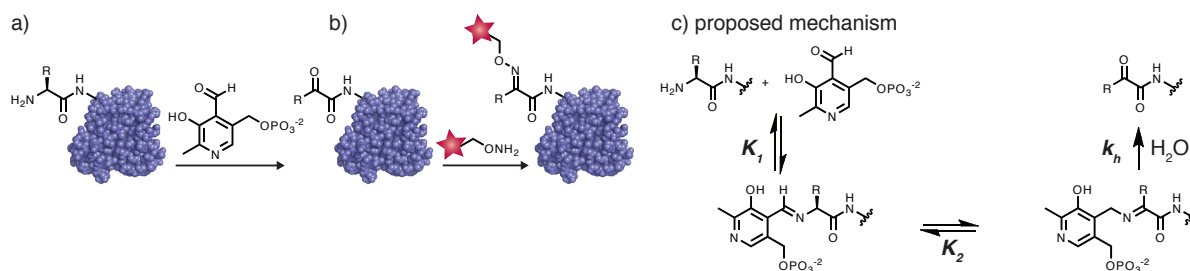


Figure 4-4. The general scheme of PLP-mediated bioconjugation. (a) and (b). c) The proposed mechanism begins with Schiff base formation between the N-terminal amine and the PLP aldehyde. Tautomerization, followed by hydrolysis, affords the keto-protein product. Based on this mechanism, the reaction rate is expected to depend on both the concentration of protein and PLP.

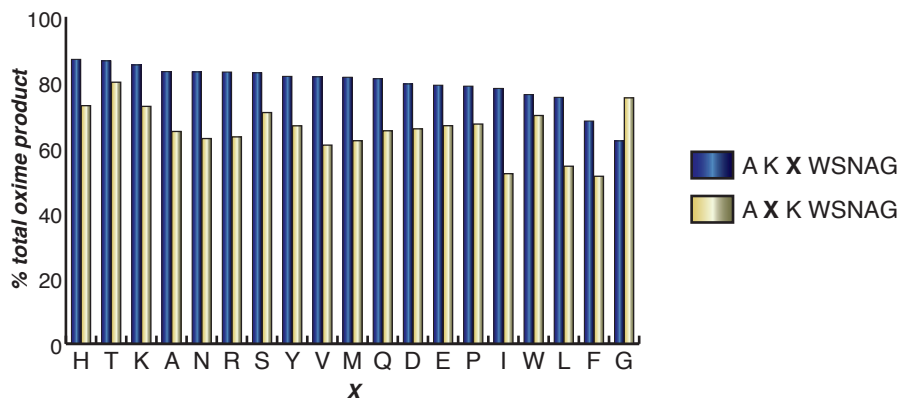


Figure 4-5. Positional scanning was performed to explore the scope of the AKX and AXK motifs. The total oxime yield after reaction with standard PLP conditions (10 mM, 18 h) was determined for 38 peptides with varied amino acids in the X positions (all natural amino acids except cysteine were included). Although high yields were seen in all cases, peptides with lysine in the second position consistently provided higher conversion.

4.4 APPLICATION TO PROTEIN SUBSTRATES: AKT-TERMINAL ANTIFREEZE PROTEIN

To determine if the alanine-lysine-threonine motif would improve the reactivity of protein substrates, we used mutagenesis to install this sequence at the N-terminus of two recombinant proteins used in the Francis group for materials purposes. The first of these proteins was the Type III “Antifreeze” Protein (AFP).⁷ In Figure 4-6a the location of the new sequence on the protein surface can be seen. The AKT motif was introduced as a two-residue extension of the wild-type N-terminus, which was GNQ-terminal. After expression and purification, the wild-type and AKT-terminal variants of these proteins were transaminated with PLP and reacted with benzylalkoxyamine for quantification of the oxime yield by mass spectrometry.

The general procedure for PLP-mediated transamination on protein substrates was performed as follows. Protein and PLP stock solutions were prepared at twice the desired final concentration and mixed in equal volumes in a 1.5 mL Eppendorf tube. The final

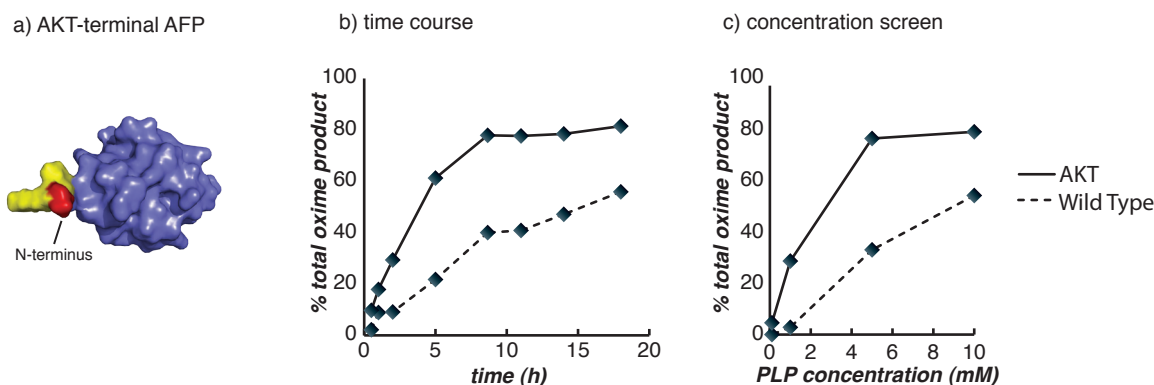


Figure 4-6. The performance of wild-type and AKT-terminal sequences were compared for the transamination of an Anti-Freeze Protein (AFP). a) The Ala residue is shown in red and the Lys-Thr portion in yellow. b) The transamination of the AFP mutant showed that the AKT sequence outperformed the wild-type terminus (GNQ) at every time point analyzed. c) The AKT-terminal AFP also outperformed the wild-type sequence at different concentrations of PLP.

volume of each reaction was 100 μL . The 2x protein stock solutions were prepared at 40-600 μM using 25 mM potassium phosphate buffer at pH 6.5 with 0.02% NaN_3 . The 2x PLP stock solutions were freshly prepared before each experiment in 25 mM phosphate buffer, pH 6.5 with 0.02% NaN_3 . Following addition of the PLP, the pH was re-adjusted to pH 6.5 using NaOH. The reaction mixture was briefly agitated to ensure mixing and then incubated without further agitation at 37 $^{\circ}\text{C}$ for 1-18 h. Following the reaction, the PLP was removed by buffer exchange using Milipore 0.5 mL spin concentrators (MWCO 3 kDa). The buffer exchange involved the dilution of each sample to 500 μL with 25 mM phosphate buffer (pH 6.5). Each sample was then concentrated to 100 μL , and the process was repeated 2-4 times, depending on the initial concentration of PLP that was used. The resulting 100 μL of reacted protein was then treated with 25 μL of 250 mM BnONH_2 (pH adjusted to 5.5) in a 1.5 mL Eppendorf tube and incubated at rt for 18 h. Buffer exchange steps were repeated as described above to stop the reaction and remove the excess alkoxyamine.

The wild-type AFP showed intermediate levels of oxime product formation, which followed the expectations of a GNQ N-terminal sequence (Figure 4-6). However, the AKT-terminal mutant outperformed the wild-type under every set of reaction conditions screened,

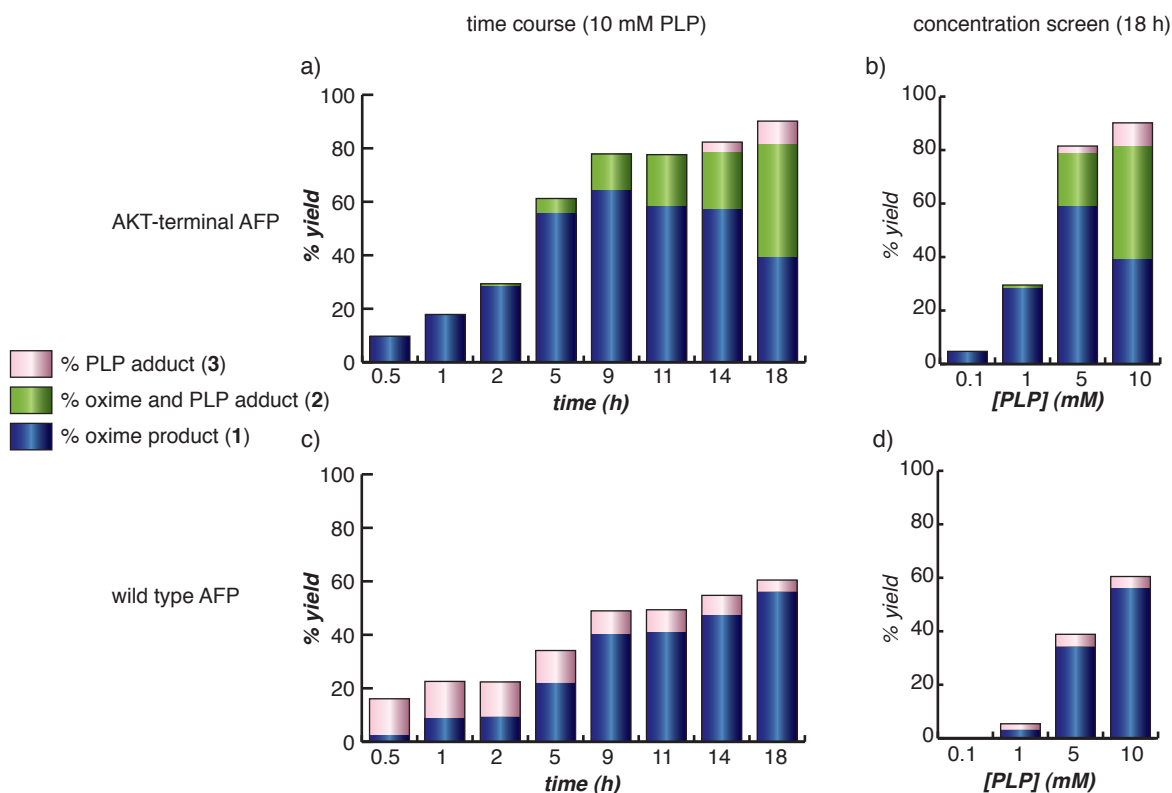


Figure 4-7. Anti-Freeze Protein (AFP) product distributions after modification with PLP and benzylalkoxyamine. a) AKT-terminated AFP reached high oxime yields after 9 hours. Longer reaction times led to increased byproduct formation. b) The AKT-mutant product distribution showed similar behavior at different concentrations of PLP, with a high yield of oxime product being formed with 5 mM PLP, and higher concentrations resulting in high amounts of byproduct. The wild type AFP (with a GNQ terminus) was less reactive than the AKT mutant, in terms of both oxime product and byproduct formation. The product distributions are shown for a time course in (c) and concentration screen in (d).

including variation of the concentration of PLP and the reaction time (Figure 4-6). This demonstrated that the oxime yield after PLP-mediated transamination could be increased for protein substrates with moderate wild-type reactivity by installing the reactive motif.

Upon inspection of the product distribution for the AKT-terminal mutant (Figure 4-7), a significant amount of covalent PLP addition to the N-terminus was observed, which corresponded by mass to an aldol addition of PLP to the introduced ketone group (**3**). This was particularly pronounced in the 10 mM PLP, 18 h sample (Figure 4-8a). Although the PLP addition byproduct had been observed previously, it was typically a small proportion of the overall product. It is possible that the AKT motif selected from library screening is more reactive towards PLP addition as well as towards transamination. The putative aldol addition of PLP to the N-terminus did not preclude oxime formation, as the green portions of the bars correspond to the reaction of the PLP adduct with the alkoxyamine agents (**2**). Thus, in many applications where the goal is simply to ligate the target protein to another material through oxime formation, the PLP addition byproduct may be innocuous. However, for cases in which a complex product mixture is problematic (such as therapeutic bioconjugates) we sought strategies to minimize the formation of the PLP adduct.

4.5 STRATEGIES TO MINIMIZE PLP ADDUCT FORMATION

As would be expected for a bimolecular reaction, we observed that the yield of PLP adduct increased with the same parameters as transamination (i.e., higher concentrations of PLP and increased reaction times). Because the PLP adduct likely forms after the transamination has taken place (Figure 4-8b), we hypothesized that a balance of these two reaction conditions might minimize PLP adduct formation without sacrificing the transamination yield (and correspondingly the total amount of oxime yield). For AFP, using a lower concentration of PLP for the full reaction time (5 mM for 18 h) or using the standard concentration for a shorter time (10 mM for 9 h) resulted in decreased amounts of PLP adduct (**2** and **3**, Figure 4-8a) while preserving the total amount of oxime yield, shown as a dotted line. Therefore, to minimize the PLP adduct byproduct, lower concentrations of PLP can be used for longer reaction times, or higher concentrations of PLP can be used for shorter incubation periods. In practice, the optimal balance will vary somewhat for each protein substrate, and, like most other organic reactions, a preliminary screen of the reaction variables is recommended.

4.6 APPLICATION TO PROTEIN SUBSTRATES: AKT-TERMINAL GFP

The AKT-terminal motif was also tested on GFP. The use of GFP as a test protein substrate was complementary to AFP in many ways. AFP is a small (7 kDa) protein while GFP is a larger protein (26 kDa). On AFP, the AKT mutant was a two-residue extension of the N-terminus, while on GFP, AKT was mutated to replace the three N-terminal residues with no extension. Additionally, while the wild-type sequence of AFP had fair-to-good reactivity towards PLP-mediated transamination, the wild-type GFP (N-terminal sequence MVS)

had minimal reactivity toward PLP from room temperature to 37 °C.³ Thus, the AKT-terminal GFP served as a test whether the Ala-Lys motif could ‘rescue’ the reactivity of proteins with unreactive native sequences.

The strategies for maximizing oxime yield with minimal PLP addition byproduct were applied to AKT-terminal GFP and led to striking results. Using the strategy of a high concentration of PLP with a short reaction time (100 mM PLP for 90 min), over 90% yield of oxime product with negligible amounts of byproduct formation was achieved (Figure 4-9). Under the same conditions, the wild-type GFP resulted no conversion, highlighting the importance of the AKT sequence for achieving increased reactivity. Previously, the standard reaction conditions for peptide and protein substrates had been 10 mM PLP for 18 h. However, these data indicate that the use of a much higher concentration of PLP, 100 mM, can dramatically lower the reaction time. This concentration of PLP did not noticeably affect protein solubility or function as measured by protein concentration

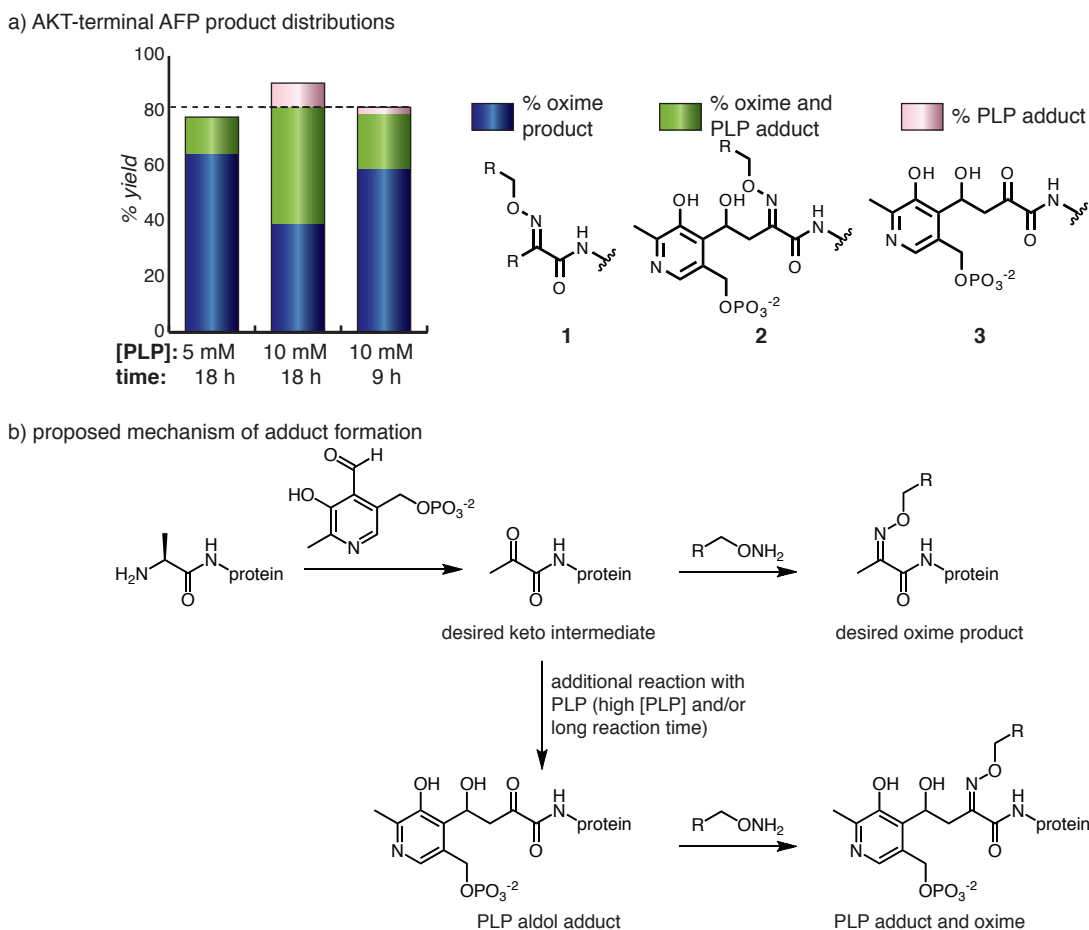


Figure 4-8. a) The product distribution of the AKT-terminal AFP mutant showed that maximum oxime yield (1 plus 2) and minimum PLP adduct (2 plus 3) could be achieved by optimizing the reaction conditions. These reaction conditions include using lower concentrations of PLP for longer reaction times, or higher concentrations of PLP with shorter reaction times. The dotted line marks the highest total oxime yield achieved, demonstrating that the strategies that decreased byproduct formation did not sacrifice overall yield. b) In the proposed mechanism of PLP adduct formation, transamination occurs first, and the PLP addition is a biomolecular reaction between the keto-protein and another molecule of PLP.

and GFP fluorescence. Using 100 mM PLP, high oxime yields were observed with 60-90 minute reactions. Although the overall yield was somewhat lower in the 60 minute reaction than in the 90 minute case, no PLP adduct was observed. These data indicate that with a small sacrifice in overall oxime yield it is possible to find reaction conditions that eliminate the byproduct if the application so requires. The product distributions of AKT-terminal GFP under other reaction conditions are shown in Figure 4-9d. When 100 mM PLP was used with long reaction times, unknown (and previously unobserved) masses resulted. Therefore, the concentration of PLP should be correlated with an inverse relationship with the reaction time. As GFP is commonly fused to protein domains to allow visualization using fluorescence, we anticipate that the availability of the fast and reliable method for the introduction of a second functional group reported herein will benefit many applications.

Since the initial report of the Ala-Lys reactive motif, the AKT terminus has been used on a

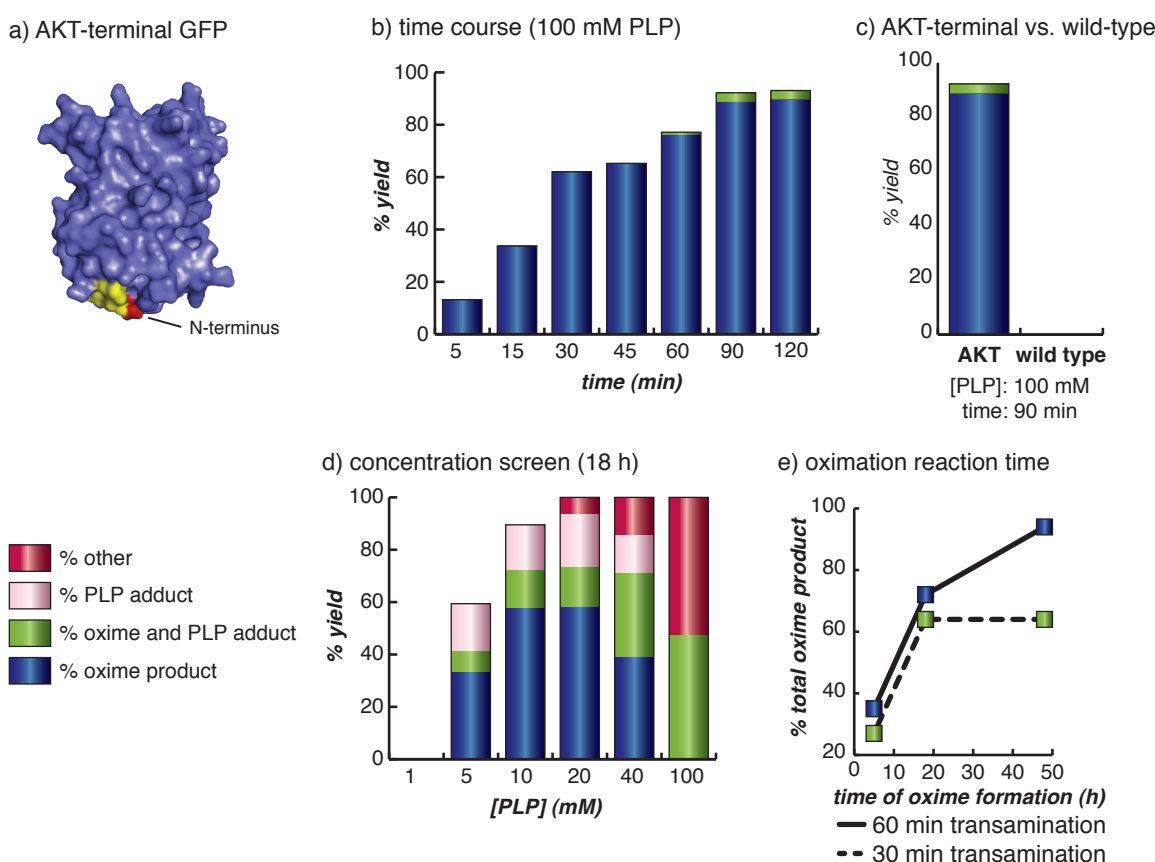


Figure 4-9. Product distributions for the modification of GFP with PLP. a) The location of the N-terminal residues on GFP are highlighted. b) When strategies to minimize byproduct formation were applied, we found that using 100 mM PLP for short reaction times (1-2 h) led to high yields of oxime product with little-to-no adduct formation. c) Wild type GFP (MVS terminal sequence) provided only trace amounts of product under similar reaction conditions, emphasizing the importance of the AKT motif for reactivity. d) AKT-terminal GFP was particularly reactive towards byproduct formation, especially at high concentrations of PLP, indicating optimal conditions are obtained for AKT-terminal sequences when the concentration of PLP varies inversely with the reaction time. The bar labeled 'other' included unknown modification products. e) Samples of AKT-terminal GFP that had been subjected to 100 mM PLP for 30 (dotted line) and 60 (solid line) minutes of transamination were compared after different lengths of oxime formation. The data indicated that incomplete oxime formation may result in under-reporting the transamination yield in some cases (18 h of oxime formation was used for all other protein samples).

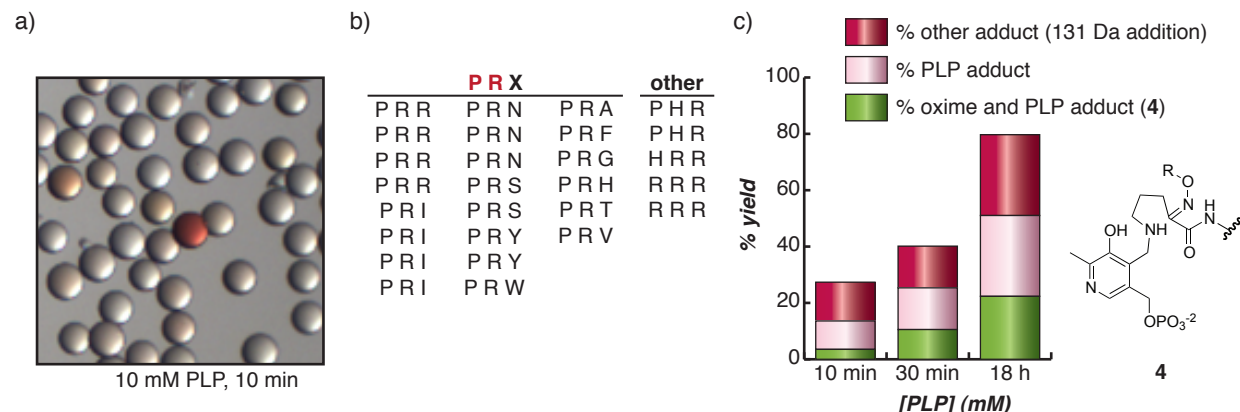


Figure 4-10. The library was also screened using a short reaction time. a) Using 10 mM PLP for just 10 minutes of transamination, a small number of beads were identified. b) The sequences from the red beads revealed a strong pattern for proline in the N-terminal position, followed by arginine. c) Upon resynthesis of a PRRWSNAG peptide and analysis of the products by LC-MS, significant amounts of oxime formation were not observed, indicating these red beads may have been false positives.

number of other proteins in the Francis group. These proteins include a human IgG1, the Fc domain of an antibody, a thermostable cellulase, and the soluble domain of a human estrogen receptor. So far, the AKT-terminal mutants have resulted in over 80% oxime yield in all of these cases. Additionally, it is worth noting that since the transamination yield cannot be measured directly for protein substrates (given the mass change of only 1 Da), the actual degree of transamination may be under-reported by incomplete oxime formation. As shown in Figure 4-9e, using 18 h of oxime formation would indicate that the transamination yield was similar for 30 min and 60 min transaminations of AKT-terminal GFP using 100 mM PLP. However, the transamination yield was actually higher for the 60 min sample, as seen when the oxime formation reaction was allowed to go to completion (48 h).

4.7 LIBRARY SCREENING WITH SHORT REACTION TIME

The AKT motif was identified from library screening, and its enhanced reactivity was subsequently verified on peptide and protein substrates. However, we also found that the library screening process can also lead to false positives, which emphasized the importance of verifying library results on resynthesized peptides. Originally, in order to access stringent transamination conditions that would reveal highly reactive sequences, we identified two parameters to vary: the concentration of PLP and the reaction time. The low concentrations of PLP resulted in the Ala-Lys motif, as described in the preceding sections. Subjecting the library to a short transamination time (10 mM PLP for 10 min) also resulted to a small number of red beads. These beads were sequenced and a strong consensus sequence emerged: Pro-Arg (Figure 4-10). We were interested in the transamination of proline-terminal sequences because we had previously identified whose mass corresponded to a ring opening addition of PLP to the N-terminus with oxime formation (4). Although small amounts of this product were observed upon LC-MS analysis of a resynthesized Pro-Arg terminal peptide, this motif did not result in high

reactivity under various reaction conditions tested, therefore this sequence appears to have been a false positive and was not pursued.

Using a combinatorial peptide library, a highly reactive sequence towards PLP-mediated transamination was identified. The proteins AFP and GFP demonstrated that mutating the N-terminal sequence of a protein to AKT can increase reactivity toward PLP-mediated bioconjugation, and thus these results serve to validate our combinatorial peptide library screening process. We have also identified reaction conditions for AKT-terminal proteins that maximize yield and purity. We anticipate that the development of this encodable motif will be of interest to those in the chemical biology and biomolecular materials communities who wish to apply this mild, site-specific transamination reaction to their own protein substrates. On the basis of the success of these studies, we are continuing to use this combinatorial library screening platform to accelerate the discovery and optimization of new bioconjugation reactions.

4.8 MATERIALS AND METHODS

General Procedures. Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H₂O) used as reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Pyridoxal 5'-phosphate monohydrate was obtained from Aldrich. All Fmoc-protected amino acids were obtained from Novabiochem (EMD, Germany). Tentagel S OH resin was obtained from Advanced ChemTech (Louisville, KY). Centrifugations were conducted with an Eppendorf Mini Spin Plus (Eppendorf, Hauppauge, NY).

Combinatorial Library Synthesis, and General Procedure for Library Screening, Sequencing, and Solid Phase Peptide Synthesis. Described in Chapter 3.

General Procedure for PLP-Mediated Transamination of Peptide Substrates. See Section 4.2.

General Procedure for PLP-mediated Transamination of Protein Substrates. See Section 4.4.

Liquid Chromatography and Mass Spectrometry Materials. Acetonitrile (Fisher Optima grade, 99.9%) and formic acid (1 mL ampules, 99+%) were used to prepare mobile phase solvents for liquid chromatography-mass spectrometry (LC-MS). Water used in these experiments was purified to a resistivity of 18.2 MΩ·cm (at 25 °C) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA).

LC-MS Analysis of Peptide Bioconjugates. Peptide bioconjugates were analyzed using a nanoAcquity ultraperformance liquid chromatograph (UPLC; Waters, Milford, MA) that was connected in-line with a quadrupole time-of-flight mass spectrometer (Q-Tof

Premier, Waters). The UPLC was equipped with C18 trapping ($5\ \mu\text{m}$, $20\ \text{mm} \times 180\ \mu\text{m}$) and analytical ($1.7\ \mu\text{m}$, $100\ \text{mm} \times 100\ \mu\text{m}$) columns and a $10\ \mu\text{L}$ sample loop. Solvent A was 0.1% formic acid/99.9% water and solvent B was 0.1% formic acid/99.9% acetonitrile (v/v). Sample solutions contained in $0.3\ \text{mL}$ polypropylene snap-top vials sealed with septa caps were loaded into the nanoAcquity autosampler compartment prior to analysis. Following sample injection ($1.5\ \mu\text{L}$, partial loop), trapping was performed for 5 min with 100% A at a flow rate of $3\ \mu\text{L}/\text{min}$. The injection needle was washed with $500\ \mu\text{L}$ each of solvents A and B after injection to avoid cross-contamination between samples. The elution program consisted of a linear gradient from 25% to 50% B over 30 min, a linear gradient to 95% B over 0.33 min, isocratic conditions at 95% B for 6.67 min, a linear gradient to 1% B over 0.33 min, and isocratic conditions at 1% B for 11.67 min, at a flow rate of $500\ \text{nL}/\text{min}$. The analytical column and sample compartment were maintained at $35\ ^\circ\text{C}$ and $8\ ^\circ\text{C}$, respectively.

The column exit was connected to a NanoEase nanoelectrospray ionization (nanoESI) emitter mounted in the nanoflow ion source of the Q-ToF. The nanoESI source parameters were as follows: nanoESI capillary voltage $2.3\ \text{kV}$, nebulizing gas (nitrogen) pressure $0.15\ \text{mbar}$, sample cone voltage $30\ \text{V}$, extraction cone and ion guide voltages $4\ \text{V}$, and source block temperature $80\ ^\circ\text{C}$. No cone gas was used. The collision cell contained argon gas at a pressure of $8 \times 10^{-3}\ \text{mbar}$. The ToF analyzer was operated in “V” mode. Under these conditions, a mass resolving power of 1.0×10^4 (measured at $m/z = 771$) was routinely achieved, which was sufficient to resolve the isotopic distributions of the singly and multiply charged peptide bioconjugate ions measured in this study. Thus, an ion’s mass and charge could be determined independently, i.e., the ion charge was determined from the reciprocal of the spacing between adjacent isotope peaks in the m/z spectrum. External mass calibration was performed immediately prior to analysis using solutions of sodium formate. Mass spectra were acquired in the positive ion mode over the range $m/z = 100\text{--}2000$, in continuum data format, using a $0.95\ \text{s}$ scan integration and a $0.05\ \text{s}$ interscan delay. Mass chromatograms and spectra were processed using MassLynx software (version 4.1, Waters).

LC-MS Analysis of Protein Bioconjugates. Protein bioconjugates were analyzed using an Agilent 1200 series liquid chromatograph (LC; Santa Clara, CA) that was connected in-line with an LTQ Orbitrap XL hybrid mass spectrometer equipped with an Ion Max electrospray ionization source (ESI; Thermo Fisher Scientific, Waltham, MA).

The LC was equipped with C8 guard (Poroshell 300SB-C8, $5\ \mu\text{m}$, $12.5 \times 2.1\ \text{mm}$, Agilent) and analytical ($75 \times 0.5\ \text{mm}$) columns and a $100\ \mu\text{L}$ sample loop. Solvent A was 0.1% formic acid/99.9% water and solvent B was 0.1% formic acid/99.9% acetonitrile (v/v). Sample solutions contained in $0.3\ \text{mL}$ polypropylene snap-top vials sealed with rubber septa caps were loaded into the Agilent 1200 autosampler compartment prior to analysis. For each sample, approximately 100 to 200 picomoles of protein analyte was injected onto the column. Following sample injection, analyte trapping was performed for 5 min with 99.5% A at a flow rate of $90\ \mu\text{L}/\text{min}$. The elution program consisted of a linear gradient from 30% to 95% B over 19.5 min, isocratic conditions at 95% B for 5 min, a linear

gradient to 0.5% B over 0.5 min, and then isocratic conditions at 0.5% B for 9.5 min, at a flow rate of 90 μ L/min. The column and sample compartments were maintained at 35 °C and 10 °C, respectively. Solvent (Milli-Q water) blanks were run between samples, and the autosampler injection needle was rinsed with Milli-Q water after each sample injection to avoid cross-contamination between samples.

The connections between the LC column exit and the mass spectrometer ion source were made using PEEK tubing (0.005" i.d. $\times$ 1/16" o.d., Western Analytical, Lake Elsinore, CA). External mass calibration was performed prior to analysis using the standard LTQ calibration mixture containing caffeine, the peptide MRFA, and Ultramark 1621 dissolved in 51% acetonitrile/25% methanol/23% water/1% acetic acid solution (v/v). The ESI source parameters were as follows: ion transfer capillary temperature 275 °C, normalized sheath gas (nitrogen) flow rate 25%, ESI voltage 2.5 kV, ion transfer capillary voltage 33 V, and tube lens voltage 125 V. Mass spectra were recorded in the positive ion mode over the range m/z = 500 to 2000 using the Orbitrap mass analyzer, in profile format, with a full MS automatic gain control target setting of 5×10^5 charges and a resolution setting of 6×10^4 (at m/z = 400, FWHM). Raw mass chromatograms and spectra were processed using Xcalibur software (version 4.1, Thermo) and measured charge state distributions were deconvoluted using ProMass software (version 2.5 SR-1, Novatia, Monmouth Junction, NJ), using default "small protein" parameters and a background subtraction factor of 1.5.

Constuction of GFP and AFP Expression Plasmids. The construction of the ISP-pTXB1 and EGFP-pET28b plasmids has been previously reported.^{3,4} The Quick Change II System Kit (Stratagene, La Jolla, CA) was used to mutate the N-termini of each plasmid to AKT. The plasmids used had previous N-terminal mutations, therefore site-directed mutagenesis was also used to make mutants containing the wild type N-terminal sequence. The AFP wild-type sequence was obtained by removing two amino acids that had been added to the N-terminus during the original construction of the plasmid using the primers:

Forward: 5'-gaaggagatatacatatgggtaaccaggcgagcgt-3'
Reverse: 5'-acgctcgctggttaccatattatatctccttc-3'

The AKT-terminal AFP was obtained on the extended N-terminus using the primers:

Forward: 5'-ggagatatacatatggcgaaaaccaaccaggcgagcgt-3'
Reverse: 5'-acgctcgctggttggtttcgccatattatatctcc-3'

The wild-type GFP plasmid was constructed using the primers:

Forward: 5'-atacatatggtgagcaagggcgaggagctgttcaccg-3'
Reverse: 5'-gaaattcttctctatatgtataccactcgttccgc- 3'

The GFP AKT mutant was constructed using the primers:

Forward: 5' -catatggccaaaacgggcgaggagctgttcaccggggtg- 3'
Reverse: 5' -cctcgcccgttttggccatatgtatatctccttcttaaag- 3'

Incorporation of the point mutations was verified by sequencing.

General Procedure for Expression and Purification of GFP and AFP. The plasmids were transformed into T7 Express LysY/lq Competent E. coli (NEB, Ipswich, MA) and plated on LB Agar Plates (ampicillin at 100 µg/mL). Cells were grown in 1 L of Luria Broth (LB) containing ampicillin at 100 µg/mL at 37 °C until an optical density (OD) of 0.7 was observed at 600 nm. Protein expression was induced by the addition of 1 mL of 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cultures were grown for an additional 16 h at 16 °C. The cells were then centrifuged for 10 min at 8,000 rcf at 4 °C. The collected cells were re-suspended in 10 mL of Lysis/Wash Buffer (0.02 M Tris, 0.5 M NaCl, 1 mM EDTA, pH 7.5) by vortexing. The cells were lysed by sonication using a Branson Digital Sonifier (VWR Scientific, West Chester, PA) for 20 min with a blunt ended tip. Debris was removed by centrifugation. The protein-containing lysate was incubated with 5-6 mL of chitin resin (NEB) in for 1 h. The resulting resin-bound protein was washed with 100 mL of Wash Buffer that was cooled to 4 °C. A 30 mL solution of 50 mM sodium 2-sulfanylethanesulfonate (MESNA) in Wash Buffer was flowed over the resin-bound protein using suction in order to cleave it from the intein-chitin binding domain. The column bed was then allowed to stand in a minimal amount of this solution at 4 °C for 15 h. The protein was eluted from the column with the addition of 15 mL of Wash Buffer. Purified protein was then buffer exchanged 3 times with Wash Buffer using Amicon Ultra 15 mL 3,000 MWCO (Millipore, Billerica, MA) and incubated with an additional 5 mL of chitin resin for 1 h. The chitin resin was removed by filtration through a Steriflip (Millipore). Purified protein was then buffer exchanged into 25 mM sodium phosphate (pH 6.5) using Amicon Ultra 15 mL 3,000 MWCO (Millipore) centrifugal ultrafiltration membranes. The entire procedure was performed in a cold room maintained at 4 °C. Purity was evaluated by SDS-PAGE with Coomassie staining.

4.9 REFERENCES

1. Witus L.S. Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., Francis, M.B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, 132, 16812-16817.
2. Christman, K.L., Broyer, R.M., Tolstyka, Z.P., Maynard, H.D. Site-specific Protein Immobilization Through N-Terminal Oxime Linkages. *J. Mater. Chem.* **2007**, 17, 2021-2027.
3. Esser-Kahn, A.P., Francis, M.B. Protein-Cross-Linked Polymeric Materials through Site-Selective Bioconjugation. *Angew. Chem. Int. Ed.* **2008**, 120, 3811-3814.
4. Dedeo, M.T., Duderstadt, K.E., Berger, J.M., Francis, M.B. Nanoscale Protein Assemblies from a Circular Permutant of the Tobacco Mosaic Virus. *Nano Lett.* **2010**, 10, 181-186.
5. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S., Francis M.B. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **2006**, 45, 5307-5311.
6. Scheck R.A., Dedeo M.T., Iavarone A.T., Francis M.B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, 130, 11762-11770.
7. Esser-Kahn A.P., Trang V., Francis M.B. Incorporation of Antifreeze Proteins into Polymer Coatings Using Site-Selective Bioconjugation. *J. Am. Chem. Soc.* **2010**, 132, 13264-13269.

CHAPTER 5:

SITE- AND SEQUENCE-SPECIFIC PROTEIN TRANSAMINATION USING AN *N*-ALKYLPYRIDINIUM CARBOXALDEHYDE

ABSTRACT

The controlled attachment of synthetic groups to proteins is important for a number of fields, including therapeutics, where antibody-drug conjugates are an emerging area of biologic medicines. We have previously reported a site-specific protein modification method using a transamination reaction that specifically oxidizes the N-terminal amine to a ketone or an aldehyde group. The newly introduced carbonyl group can be used for the conjugation of a synthetic group to one location through the formation of an oxime or a hydrazone linkage. Here we present a new transamination reagent that is particularly well-suited for antibody modification. Although the side chain of the N-terminal amino acid does not directly participate in the transamination mechanism, in previous work using pyridoxal 5'-phosphate (PLP) as the transamination reagent we found that certain combinations of two or three N-terminal residues could result in high transamination reactivity. In this chapter the use of a combinatorial peptide library screening platform as a method to explore new transamination reagents and simultaneously identify their optimal N-terminal sequences is described. *N*-methylpyridinium carboxaldehyde benzene sulfonate salt (Rapoport's salt) was identified as a highly effective transamination reagent when paired with glutamate-terminal peptides and proteins. We also used the library colorimetric detection scheme to identify non-reactive sequences, which enables for protein-specific modification based on sequence. Mutation of antibody light and heavy chain N-terminal residues allowed modification of one or both chains, thus enabling control of the number of attachments to the antibody (two or four) and demonstrating protein-specific modification based on sequence.

5.1 INTRODUCTION

The chemical modification of proteins is an important tool for a wide range of fields, including cell biology research,^{1,2,3} the construction of new biomaterials,⁴ and the development of novel therapeutics.^{5,6} The pharmaceutical industry has been particularly interested in antibody-drug conjugates (ADCs), with multiple products clinically approved and several more currently in advanced trials.^{7,8} Ideally, ADCs will need to be prepared using site-selective bioconjugation reactions that can control the stoichiometry and position of the attached drugs. However, antibodies are particularly difficult to modify in a controlled manner due to their large size, multiple polypeptide chains, glycosylation, and structurally important disulfide bonds. Traditional methods such as lysine modification⁹ are indiscriminate given the abundance of these residues (up to 100)⁸ leading to heterogeneous mixtures that complicate pharmacokinetic characterization. Current alternative methods for site-specific antibody modification include genetic mutation to alter the number of solvent-accessible cysteines^{10,11} or to introduce enzyme recognition tags.¹² Nonetheless, the growing interest in antibody drug conjugates (ADCs) as commercial treatments provides a need for a whole series of readily-scalable methods that can provide well-defined conjugates with high efficiency and control over attachment stoichiometry.

We have previously reported a site-specific transamination reaction that introduces a new carbonyl group at the N-terminus of proteins through incubation with pyridoxal 5'-phosphate (PLP, **1a**).^{13,14} The carbonyl groups introduced by this reaction are not naturally occurring functionalities on protein surfaces, and can therefore be used as a unique point of attachment for a synthetic group through oxime or hydrazone formation,¹⁵ Figure 5-1. In previous work, we have applied the PLP-mediated transamination reaction to the modification of monoclonal antibodies.¹⁶ However, the yields were not high and elevated temperatures were required, limiting the practical application of this approach. Additionally, the stoichiometry of drug attachment has been identified as a critical parameter for ADC efficacy, with 2 to 4 attachments reported as being optimal in at least some cases.^{8,17} Since antibodies have 4 N-termini (2 identical heavy chains and 2 identical light chain units), we hypothesized that sequence-selective transamination could give control over the number of attachments within the optimal range (2 or 4). Thus, we set out to identify more reliable and scalable conditions for antibody transamination that could not only achieve improved yields and greater bioconjugate purity than were possible previously, but also could modify one or both sets of chains selectively.

We hypothesized that N-terminal transamination could achieve chain-specific antibody modification based on previous evidence of the sequence dependence of PLP-mediated transamination. Although the side chain of the N-terminal residue does not participate directly in the transamination mechanism, the reaction yield was found to vary significantly depending on the amino acids in the N-terminal positions.¹⁸ Given this situation, we used a combinatorial peptide library to identify highly reactive sequences towards PLP-mediated transamination and found an Ala-Lys N-terminal motif.¹⁹ In the present work, this new bioconjugation development tool was used to identify a new protein transamination reagent, *N*-methylpyridinium carboxaldehyde benzene sulfonate salt (Rapoport's salt,²⁰

5.2 RAPOPORT'S SALT

a)

$\text{H}_2\text{N}-\text{R}$ $\xrightleftharpoons{\text{1a or 1b}}$ $\text{R}-\text{O}-\text{N}=\text{R}'$

IgG

b)

1a **1b**

c) proposed mechanism

$\text{H}_2\text{N}-\text{R} + \text{1b} \rightleftharpoons \text{Intermediate} \xrightleftharpoons{\text{H}_2\text{O}} \text{keto-protein} \xrightarrow{\text{R}'\text{OH-NH}_2} \text{oxime product} \xrightarrow{\text{R}'\text{OH-NH}_2} \text{oxime and adduct}$

57

for small molecules in organic solution,^{20,24,25} suggesting that it could serve as a protein transamination reagent. In our original report of site-specific N-terminal transamination on proteins, a number of common aldehydes were screened for transamination ability, and PLP was identified as the most effective.¹³ However, this screen was carried out using angiotensin as a single substrate. Given what we later found in terms of the dependence of PLP-mediated transamination on the N-terminal sequence, it is possible that this initial screen failed to identify aldehydes, such as Rapoport's salt, that can achieve similarly high levels of transamination with other sequences. To match more practical reagents with their proper sequences, the combinatorial peptide library screening platform serves as an important tool to evaluate the transamination ability of candidate reagents against all combinations of N-terminal sequences simultaneously (Figure 5-2a).

5.3 LIBRARY SCREENING

Incubation of the combinatorial peptide library with 10 mM Rapoport's salt for 1 h in pH 6.5 phosphate buffer, followed by oxime formation, resulted in the formation of a small number of red beads (Figure 5-2b). Oxime formation was achieved using a ratio of DispRed-ONH₂ to BnONH₂ of 20%, which was identified to result in red colored beads only for high levels of oxime yield, as described in Chapter 3. These data indicated that Rapoport's salt could effect transamination to high yields on certain N-terminal sequences.

To identify the optimal N-terminal sequences for Rapoport's salt-mediated transamination, the red beads were collected and sequenced (Figure 5-2). The common pattern identified from sequencing was multiple glutamate residues at the N-terminus. The screen was also repeated using a lower ratio of DispRed-ONH₂ to BnONH₂ (5%) to provide a more stringent filter for high oxime yield. Again the same reactive motif was identified by sequencing these red beads: glutamate in the N-terminal and second positions, with less consensus at the third position.

The identification of sequences that were non-reactive towards Rapoport's salt-mediated transamination was achieved by sequencing colorless beads. To ensure that very low levels of reactivity would be identified, 100% DispRed-ONH₂ was used during the oxime formation so that even low amounts of oxime formation would result in a red bead (the colorimetric detection of oxime yield is explained in more depth in Chapter 3). The resulting colorless beads were selected and sequenced. Proline in the N-terminal position was identified as the common motif among the colorless beads (Figure 5-2b), indicating that proline-terminal sequences would be unreactive towards Rapoport's salt-mediated transamination. This result was not unexpected given that as a secondary amine, proline is not able to undergo the standard transamination mechanism. Previous studies with PLP did show appreciable reactivity using the transamination/oximation procedure, although the resulting product has not yet been fully characterized and was not observed with Rapoport's salt.

Another question about reactivity that could be answered using the library was whether

any residues in the second or third positions could significantly alter the high reactivity of glutamate-terminal sequences. Knowledge of whether certain neighboring residues should be avoided is particularly important for incorporation of the reactive motif onto protein substrates. To address this question, the glutamate-terminal subset of the library, wherein all peptides had E as the N-terminal residue and the second and third positions were varied (EXX), was used. Sequencing the colorless beads after transamination on this subset of the library revealed a consensus motif that had proline in the second position (Figure 5-2b). Thus the library screening identified glutamate-terminal sequences as highly reactive, and sequences with proline in the N-terminal or second position as less reactive towards Rapoport's salt mediated transamination. Although the third position had a lesser effect on reactivity, we selected EES as the highly reactive sequence to test on peptide and protein substrates. PES and EPS were used as representative non-reactive sequences to explore how much the change of a single amino acid could affect the reactivity of the optimal motif.

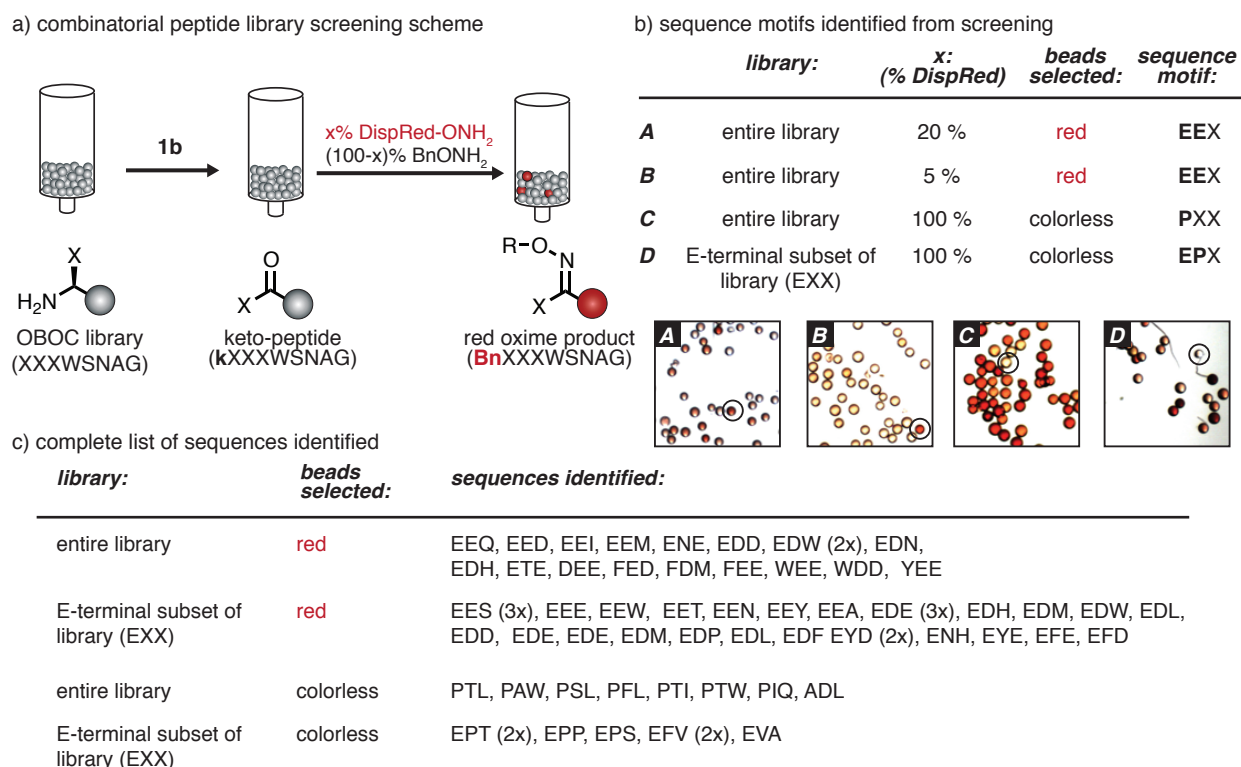


Figure 5-2. a) A one-bead-one-sequence combinatorial peptide library in which the three N-terminal residues were varied was used to screen potential transamination reagents such as Rapoport's salt (**1b**) against all possible N-terminal sequences simultaneously. After transamination, the keto-group-containing peptides were identified through the covalent attachment of a visible dye (DispRed-ONH₂) through oxime formation. b) The library was treated with 10 mM Rapoport's salt for 1 h at pH 6.5, followed by oxime formation with the specified ratios of DispRed-ONH₂ and BnONH₂. The sequences active towards Rapoport's salt-mediated transamination were identified by selecting and sequencing the red beads in the library. Glutamate-terminal sequences were identified as a common motif. N-termini that were not reactive toward Rapoport's salt-mediated transamination were also identified by sequencing the beads that remained colorless after transamination and oxime formation. In these cases, no BnONH₂ was combined with the DispRed-ONH₂. Some colorless beads still remained, and proline was identified as an N-terminal residue that prevented transamination/oxime formation. When the glutamate-terminal subset of the library was examined, proline in the second position was also found to lower the reactivity of E-terminal peptides. The complete list of sequences identified is shown in (c).

5.4 STUDY OF RAPOPORT'S SALT-MEDIATED TRANSAMINATION USING PEPTIDES.

To verify the reactive and non-reactive sequence motifs identified by library screening, peptides of the form XXXWSNAG were synthesized in order to quantify the reaction yields by liquid chromatography-mass spectrometry (LC-MS). After synthesis, the peptides were treated with Rapoport's salt followed by benzylalkoxyamine for quantification of the benzyl oxime yield by LC-MS. The optimal sequence peptide, EESWSNAG, led to high oxime yield, as seen in Figure 5-3. A screen of reaction conditions, varying the concentration of Rapoport's salt, the reaction time, and the pH was performed in order to identify the optimal reaction conditions for the transamination step. These screens, shown in Figure 5-4, identified 100 mM Rapoport's salt for 1 h at pH 6.5 as conditions resulting in high transamination yield with few byproducts. This protocol was subsequently used to evaluate all of the peptide substrates. As seen in Figure 5-4, the EE-terminal peptide resulted in over 80% transamination. Not all of the transaminated, keto-peptide species (shown in yellow) was converted to oxime (blue) under the oxime formation conditions used. A small amount of covalent addition of Rapoport's salt to the N-terminus was also observed (green). This byproduct was observed in higher yields with longer reaction times, but could be minimized through the use of shorter reaction times (45 min - 1 h).

To test whether the EE-terminal motif was optimal only for Rapoport's salt or was highly reactive towards transamination in general, the modification of this sequence with PLP-mediated transamination was examined as well (Figure 5-4d). Under the same reaction conditions the yield was much higher using Rapoport's salt, indicating the library did identify an optimal reagent/sequence pair. The N-terminal site-specificity of Rapoport's salt mediated transamination was confirmed using tandem mass spectrometry for both

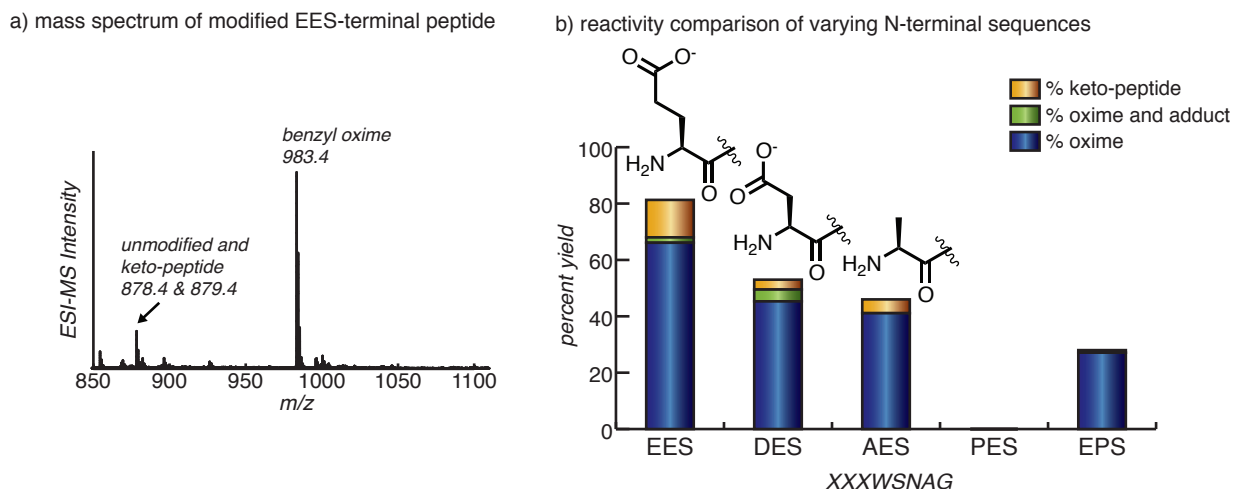


Figure 5-3. Verification of library screening results using peptides. Sequences were resynthesized in the form XXXWSNAG. a) An ESI mass spectrum of EESWSNAG showed high conversion to the oxime product. b) Various N-terminal sequences were subjected to the same reaction conditions and the product yields were quantified by LC-MS. Three products were observed: transaminated keto-peptide that did not form an oxime, the desired oxime product, and oxime with the addition of Rapoport's salt to the N-terminus. The reported data represent the average of three replicate experiments. Aspartate-terminal peptides were observed by mass spectrometry to decarboxylate during transamination, and had reactivity that was more similar to alanine-terminal sequences than to glutamate-terminal ones. A proline-terminal sequence was confirmed to have no reactivity towards Rapoport's salt mediated transamination. With proline in the second position, the reactivity of the E-terminal sequence was substantially decreased.

the transaminated keto-peptide and the benzyl oxime (Figure 5-5).

Due to their structural similarity, aspartate-terminal sequences were also examined. E- and D- terminal peptides were found to have significantly different yields towards Rapoport's salt mediated transamination, as seen in Figure 5-3b. The mass spectra indicated that decarboxylation of the aspartate-terminal peptide occurred during the transamination reaction. Comparison with an alanine-terminal peptide showed that the aspartate-terminal sequence had a more similar reactivity to alanine (which it resembles after decarboxylation) than glutamate. Therefore we concluded that optimal sequences for Rapoport's salt mediated transamination were glutamate-terminal sequences, and that D and E were not interchangeable for the N-terminal position for this reaction.

The nonreactive sequences identified by library screening were also verified on resynthesized peptides. The proline-terminal peptide (PES) resulted in no transamination nor oxime formation. The glutamate-terminal sequence with proline in the second position (EPS) did result in some oxime product, but at a yield that was significantly less than that of the multi-glutamate terminal peptide. This result underscores the impact of a single amino acid substitution in an internal sequence position.

5.5 RAPOPORT'S SALT-MEDIATED TRANSAMINATION OF ANTI-HER2 HUMAN IgG1

Given that the heavy chain of human anti-HER2 IgG1 (Herceptin)²⁶ already has an N-terminal glutamate residue, the wild type antibody provided a suitable substrate to test Rapoport's salt mediated transamination on a protein. Since both highly reactive and nonreactive sequences were identified from the library, monoclonal antibodies also presented an opportunity to test whether chain-specific modification was possible.

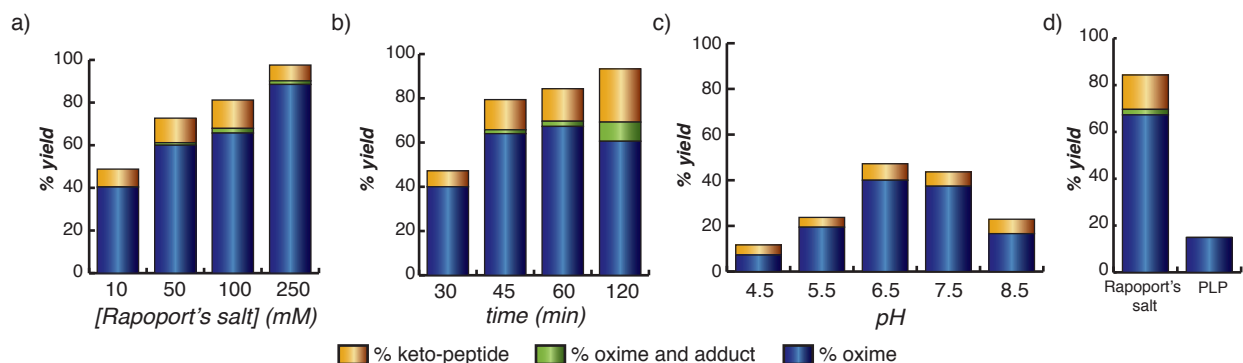


Figure 5-4. Reaction conditions screen for Rapoport's salt-mediated transamination of the EESWSNAG peptide. a) The concentration of Rapoport's salt was varied during a 1 h reaction at pH 6.5. Although library screening was done using 10 mM Rapoport's salt for a 1 h reaction, higher yields were found with higher concentrations. b) Using 100 mM Rapoport's salt at pH 6.5, the reaction time was varied. Although the total conversion was higher, the use of longer reaction times led to increased amounts of adduct. c) The pH was varied by performing the reaction in 50 mM phosphate buffers of various pHs with 100 mM Rapoport's salt for 30 min. d) To see if the EE-terminal sequence identified by library screening was an optimal sequence for Rapoport's salt in particular or a sequence that transaminated well with any reagent, conversion using 100 mM PLP for 1 h pH 6.5 was compared to that achieved using 100 mM Rapoport's salt under the same conditions. The yield was significantly higher in the Rapoport's salt-mediated transamination case, indicating that the library screening had indeed identified an optimal transamination reagent/sequence pair.

Therefore we expressed previously reported anti-HER2 antibodies (denoted as wild-type) in HEK cells, as well as a series of N-terminal mutants to introduce differential reactivity. In one case, we replaced the three N-terminal residues of the light chain with EES to introduce reactivity at this site, and the heavy chain terminus was changed to EPS to prevent reactivity (N-terminal PES antibodies were not obtained successfully). We also prepared a Herceptin analog with EES at both sets of termini to allow modification in all four locations. However, the mass spectra of the antibodies indicated that in addition to the intended mutants, other species were produced as well, which appeared to result from improper cleavage of the N-terminal leader sequence (Figure 5-6). Therefore only a portion of the antibody expressed, including the wild-type antibody, had the desired N-terminal sequence. Nevertheless the reactivity of the portion of the antibody with the

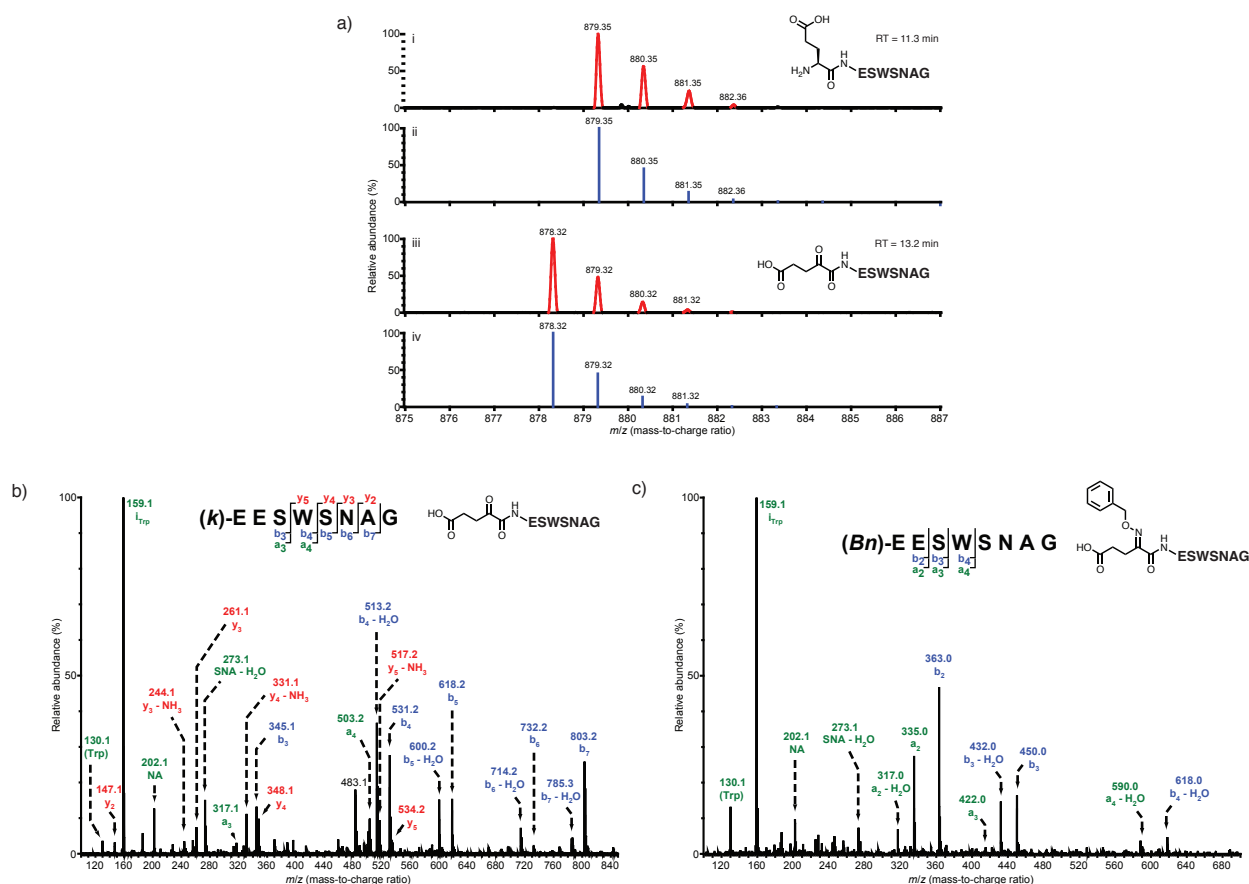
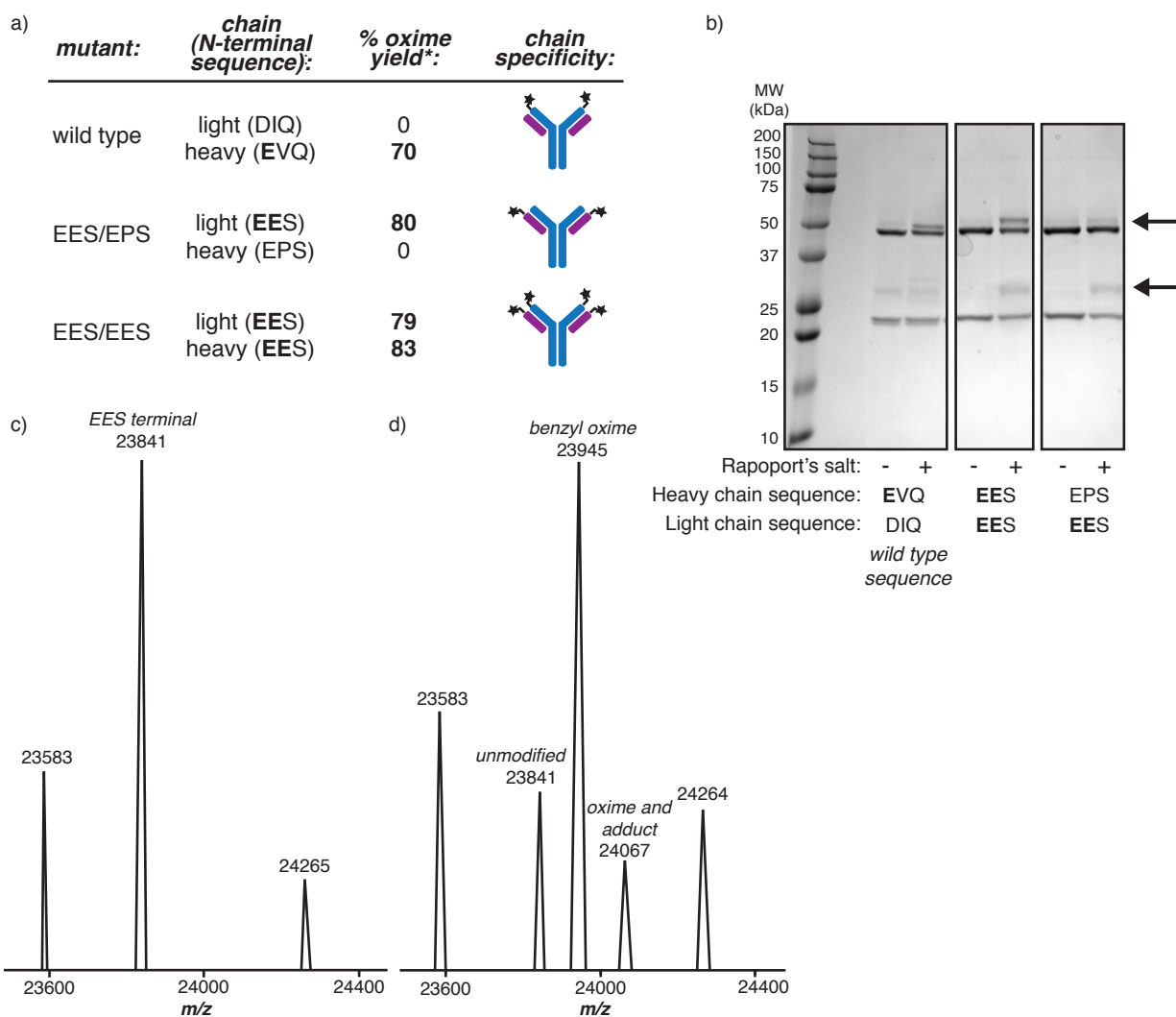


Figure 5-5. a) Comparison between measured (red) and theoretical (blue) isotopic distributions of peptide ions measured from the 1 hour reaction of EESWSNAG with 100 mM Rapoport's salt. The spectra correspond to the measured (i) and theoretical (ii) isotopic distributions for the $[M + H]^+$ ion of the unmodified peptide ($M = C_{36}H_{50}N_{10}O_{16}$; retention time = 11.3 minutes); and measured (iii) and theoretical (iv) isotopic distributions for the $[M + H]^+$ ion of the transaminated keto-peptide ($M = C_{36}H_{47}N_{9}O_{17}$; retention time = 13.2 minutes). Theoretical isotopic distributions were calculated from the natural abundances of the isotopes using MassLynx software (version 4.1, Waters, Milford, MA). b) Tandem mass (MS/MS) spectrum and corresponding sequence map resulting from collision-induced dissociation (CID) of the singly charged positive ion at $m/z = 878.3$, which was due to the $[M + H]^+$ ion of the transaminated form of the peptide EESWSNAG (precursor ion shown in a). Immonium internal fragment ions are denoted by "i" and the amino acid code. Internal cleavage fragment ions are labeled with their respective amino acid sequences. The fragment ion at $m/z = 130.1$ is due to the tryptophan side chain. Fragment ion masses were consistent with N-terminal transamination of the peptide. c) MS/MS spectrum and corresponding sequence map resulting from CID of the singly charged positive ion at $m/z = 983.5$, which was due to the $[M + H]^+$ ion of the benzyloxime modified peptide, (Bn)-EESWSNAG. Fragment ion masses were consistent with the N-terminal oxime.

intended N-terminal sequence was assessed. We are currently working on methods to produce antibodies with higher purity. These methods include the use of a different leader sequence and the insertion of a Factor Xa protease recognition site prior to the desired N-terminal residues. The Factor Xa protease site could also be used to produce glutamate-terminal proteins expressed in *E. coli*, given that otherwise the N-terminal methionine would not be clipped.²⁷

To transaminate the antibodies, the proteins were incubated with 250 mM Rapoport's salt for 45 min in pH 6.5 phosphate buffer at 37 °C. These conditions provided analogous results to the reaction conditions used for peptide substrates (100 mM Rapoport's salt at room temperature for 1 h). After transamination the antibodies were incubated with



BnONH₂ for 48 h at room temperature in pH 6.5 phosphate buffer, and the modification of each chain was then analyzed and quantified using mass spectrometry.

The portion of the heavy chain wild type sequence (EVQ²⁶) that expressed properly resulted in a 70% conversion to the oxime product, with 10% also including the Rapoport's salt aldol adduct. Although the peptide data would suggest that the DIQ sequence at the light chain terminus should result in some oxime yield, we did not observe any reactivity in this case (Figure 5-6). This is likely because the steric environment of the folded protein reduces the reactivity of this already less-reactive substrate. The complete lack of modification also clearly demonstrated that Rapoport's salt does not react with lysine side chain amines or other residues. The net result of this experiment is that Rapoport's salt mediated transamination of the wild type sequence allowed the selective modification of only the heavy chain. (This result was also obtained on a commercially available sample of the wild type anti-HER2 human IgG1, which was more pure than the expressed sample). In contrast, the transamination/oxime formation procedure gave the opposite selectivity to that of the EES/EPS mutant. In this case, with the multi-glutamate sequence on the light chain, we observed 80% conversion (of the properly expressed protein) to the oxime product. The heavy chain exhibited no detectable modification. Consistent with these findings, antibodies bearing the optimal EES motif on both sets of chains resulted in high levels of modification in all four N-terminal positions (79% for the light chains and 83% for the heavy chains). Thus the mass spectrometry data indicate that the Rapoport's salt modification method can achieve a high degree of labeling control for antibody substrates.

Since the N-termini of the antibody heavy and light chains flank the antigen binding domains (but are not part of the hypervariable loops), it was important to confirm that the modification of these locations did not disrupt antigen binding. To evaluate this, we used flow cytometry to compare the unmodified antibody mutants to those possessing the benzyloxime groups (Figure 5-7). In each experiment, the antibodies were incubated with HER2 overexpressing cells (HCC194),²⁹ and a secondary anti-human IgG fluorescently-labeled with fluorescein isothiocyanate (FITC) was used to detect them. Although the modified antibody represented only one portion of the total antibody population, no significant disruption of binding was observed from the benzyl oxime conjugates (compare orange and light blue curves). Since the benzyl oxime is a small modification, the binding ability will, of course, need to be evaluated for larger modifications. However, these results indicated that the modification procedure itself left the binding function unperturbed. To

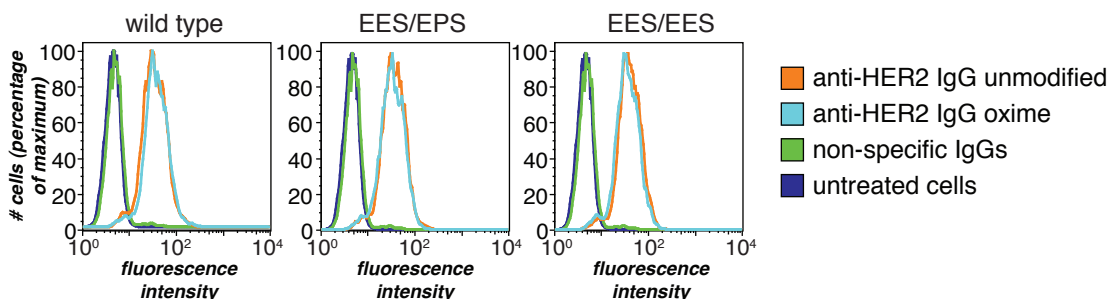


Figure 5-7. The antigen binding ability of the antibody mutants post-modification was assessed using flow cytometry. No significant difference was observed between the benzyl oxime modified (shown in light blue) and unmodified (orange)

demonstrate that it was possible to attach larger groups using Rapoport's salt-mediated transamination, we also used a 2,000 Da MW poly(ethylene glycol) alkoxyamine reagent³⁰ to attach a polymer to the antibody (Figure 5-6).

We have demonstrated that combinatorial peptide libraries can be used as a tool for the discovery of protein modification reactions. Given the importance of pairing a transamination reagent with its optimal N-terminal sequence, this method can be used to evaluate the transamination capability of candidate reagents and identify their optimal N-terminal sequences simultaneously. Currently, we are synthesizing rationally designed pyridinium derivatives that are anticipated to have enhanced reactivity. Using the information obtained from library screening in the present work, we have demonstrated that site- and chain- specific modification of monoclonal antibodies is possible with yields in excess of 80%. Rapoport's salt-mediated transamination of antibodies (and other glutamate-terminal proteins) is facile and readily scalable, thus potentially providing a suitable method for the commercial production of desired antibody conjugates.

5.6 MATERIALS AND METHODS

General Procedures. Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H₂O) used as reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). *N*-methylpyridinium carboxaldehyde benzene sulfonate salt (Rapoport's salt) was obtained from Alfa Aesar and recrystallized from acetonitrile. Pyridoxal 5'-phosphate monohydrate was obtained from Aldrich. All Fmoc-protected amino acids were obtained from Novabiochem (EMD, Germany). Tentagel S OH resin was obtained from Advanced ChemTech (Louisville, KY). Centrifugations were conducted with an Eppendorf Mini Spin Plus (Eppendorf, Hauppauge, NY).

Solid-Phase Peptide and Combinatorial Library Synthesis. See Chapter 3.

General Procedure for Library Screening. Portions of resin-bound library (appx. 25 mg resin at a time) were treated with 1 mL of freshly prepared 10 mM Rapoport's salt solution in 50 mM phosphate buffer, pH 6.5 with 10% DMF to facilitate bead swelling and 0.02% NaN₃ for preservation of the buffer. After 1 h of reaction time, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual aldehyde. The library was then incubated with 1 mL of a mixture of disperse red alkoxyamine and *O*-benzylhydroxylamine hydrochloride (BnONH₂) in the specified ratio in a 1:1 H₂O:DMF solution for 3 h at rt. The excess alkoxyamine was removed by rinsing with three portions of dichloromethane (DCM), followed by three portions of DMF. The beads were then rinsed with ethanol and transferred to a Petri dish for visual inspection. The beads were examined using a Leica S6D Microscope and L2 Light Source (Leica, Germany) equipped with a Moticam 2300 3.0 MP camera using Motic Image Plus 2.0 ML software for capturing images. Individual red beads were manually

removed using a Pipet-Lite LTS L-20 pipet (Rainin, Oakland, CA) and transferred to PCR tubes for cleavage of the peptide followed by sequencing. The residual ethanol in the tubes was removed by pipetting.

General Procedure for Library Sequencing. See Chapter 3

General Procedure for Rapoport's Salt Mediated Transamination of Resin-Bound Peptide Substrates. Portions of resin-bound peptides (appx. 10 mg of resin) were treated with 1 mL of 100 mM freshly prepared Rapoport's salt solution (or the specified concentration) in 50 mM phosphate buffer, pH 6.5 with 10% DMF to facilitate bead swelling and 0.02% NaN₃ for preservation of the buffer. After 1 h of reaction time, the resin was washed with three portions of deionized water, followed by three portions of DMF to remove residual aldehyde. The peptides were then incubated with 1 mL of a 250 mM BnONH₂ solution in water for 3 h at rt. The resin was then rinsed with three portions of deionized water, followed by three portions of DMF. The peptides were then cleaved from the resin via incubation with 300 μ L of a 100 mM sodium hydroxide solution for 30 min. The resulting peptide solution was added to 700 μ L of 50 mM phosphate buffer (pH 6.5). The resulting solution was diluted 20-fold into dd-H₂O for mass spectrometry analysis.

Construction of Light Chain Anti-HER2 Human IgG1 Expression Plasmids. To clone a plasmid for the expression of the anti-HER2 human IgG1 light chain, the sequence for the variable domain of the light chain (V_L) was obtained from the literature,²⁶ assembled into a gene, then cloned into a plasmid containing the light chain constant region. Gene2Oligo was used to generate the following set of oligonucleotides for gene assembly from the V_L sequence. An IL2 signaling sequence was also included in the N-terminal region: (The bases in lower case were added by the Gene2Oligo program and did not belong to the input sequence)

```
R0    ACCTTTTTCacattgaagtgcag
F0    ctgcacttcaatgtAAAAAAGGTCACCATGTACAGGATGCA
R24   GCAATGCAAGACAGGAGTTGCATCCTGTACATGGTG
F42   ACTCCTGTCTTGCATTGCACTAAGTCTTGCACCTTGTC
R60   TCAGTCTTAGCCGAATTCGTGACAAGTGCAAGACTTAGT
F80   CGAATTCGGCTAAGACTGACATCCAAATGACTCAGAGCC
R99   GCGCTCAGGGAACTGGGGCTCTGAGTCATTTGGATG
F119  CCAGTTCCCTGAGCGCTTCCGTAGGGGACAGG
R135  GCCCGACATGTTATTGTCAACCCTGTCCCCTACGGAA
F151  GTGACAATAACATGTCTGGGCTAGCCAGGATGTCAATACAG
R171  CTGGTACCAAGCGACAGCTGTATTGACATCCTGGCTA
F191  CTGTCGCTTGGTACCAGCAAAAGCCCGGAAAGGC
R208  GCTGTATATAAGAAGCTTTGGCGCCTTTCCGGGCTTTTG
F225  GCCAAAGCTTCTTATATACAGCGCCAGTTTCCTCTATTCTGG
R247  GAACCTGCTCGGCACGCCAGAATAGAGGAACTGGC
F267  CGTGCCGAGCAGGTTCTCTGGATCTCGGTCCG
R283  TCAGTGTGAAATCGGTCCCGGACCGAGATCCAGA
F299  GGACCGATTTCACACTGACCATTAGTTCTCTGCAGCC
```

R317 TAGTATGTTGCAAAGTCCTCTGGCTGCAGAGAACTAATGG
 F336 AGAGGACTTTGCAACATACTACTGCCAGCAGCACTAT
 R357 AGGTTGGGGGTGTGGTATAGTGCTGCTGGCAG
 F373 ACCACACCCCCAACCTTTGGTCAGGGCACGAA
 R389 CGTACGCTTGATTTCCACCTTCGTGCCCTGACCAA
 F405 GGTGGAAATCAAGCGTACGAAAAAAAccccaactttgt
 F424 acaaagttgggggTTTTTTT

An additional GCTAAACT was added to the 5' end according to a published procedure³ in order to create an alanine-lysine-threonine N-terminal mutant. The resulting VL gene was inserted into a vector at BsiWI and BstEI restriction sites using standard cloning techniques. The vector used, pFUSE2-CLlg-hk from Invivogen (San Diego, CA), already contained the constant region of the kappa light chain. A Quikchange site-directed mutagenesis kit (Stratagene, La Jolla, CA) was used to generate the desired N-terminal mutants (EES and the wild-type N-terminus, DIQ). Incorporation of these mutations was verified by sequencing.

Construction of Heavy Chain Anti-HER2 Human IgG1 Expression Plasmids. A plasmid for the expression of the anti-HER2 heavy chain was cloned in a similar fashion to that of the light chain. In brief, the variable and constant region of heavy chain (V_H and CH1) was constructed from the following set of oligonucleotides with additional bases (CTCCAAACA) at the 5' end (corresponding to three N-terminal residues, LQT).

R0 TTTTTTcttagctgcttga
 F0 tcaaagcagctaagAAAAAAGAATTCGCTCCAAACAG
 R21 CGACGAGTTGGACTTCTGTTTGGAGCGAATTC
 F38 AAGTCCAACCTCGTCGAAAGCGGAGGTGGC
 R53 CCAGGCTGAACCAGGCCACCTCCGCTTT
 F67 CTGGTTCAGCCTGGCGGAAGCCTGCGC
 R81 GCAGCACAGCTCAAGCGCAGGCTTCCG
 F94 TTGAGCTGTGCTGCCTCCGGATTTAATATCAAAGA
 R108 CGAACCCAGTGTATATAAGTATCTTTGATATTAAATCCGGAG
 F129 TACTTATATACACTGGGTTCCGCCAGGCTCCTGGA
 R150 CCACTCCAGACCCTTTCCAGGAGCCTGG
 F163 AAGGGTCTGGAGTGGGTGGCGAGAATCTACC
 R178 GGGTATAACCATTGGTTGGGTAGATTCTCGCCAC
 F194 CAACCAATGGTTATACCCGCTATGCAGACAGCG
 R212 GTAAACCGCCCTTTCACGCTGTCTGCATAGC
 F227 TGAAAGGGCGGTTTACAATTAGTGCCGACACA
 R243 GGTAAAGCGGTATTTTTAGATGTGTGCGCACTAATT
 F259 TCTAAAAATACCGCTTACCTCCAGATGAACTCTCTG
 R278 TGTCTCGGCCCTCAGAGAGTTTCATCTGGA
 F295 AGGGCCGAGGACACGGCTGTGTATTATTGC
 R308 CACCCACCGGCTGCAATAATACACAGCCG
 F325 AGCCGGTGGGGTGGAGACGGATTCTATGCT
 R338 TGACCCCAATAGTCCATAGCATAGAATCCGTCTC

F355 ATGGACTATTGGGGTCAGGGCACTCTCGTCA
 R372 TGGCACTGCTTACAGTGACGAGAGTGCCC
 F386 CTGTAAGCAGTGCCAGCACAAAGGGGCC
 R401 CAAGGGGAAAGACACTAGGCCCTTTGTGC
 F414 TAGTGTCTTTCCCCTTGCTCCATCTAGCAAATCTAC
 R431 GGTGCCCCCGCTGGTAGATTTGCTAGATGGAG
 F450 CAGCGGGGGCACCGCCGCCCTGGGAT
 R463 GTCCTTGACCAGGCATCCCAGGGCGGC
 F476 GCCTGGTCAAGGACTATTTTCCTGAGCCAGT
 R490 TCCAGGACACGGTGACTGGCTCAGGAAAATA
 F507 CACCGTGTCTTGGAATAGTGGCGCCTTGA
 R521 TGTGTGAACACCAGAAGTCAAGGCGCCACTAT
 F536 CTTCTGGTGTTACACATTTCCCGCCGTCC
 R553 CAGCCCCTAGATTGAAGGACGGCGGGAAA
 F566 TTCAATCTAGTGGGCTGTACTCTCTCTCCAGTGT
 R583 TGGGTACCGTCACCACACTGGAGAGAGAGTA
 F600 GGTGACGGTACCCAGTTCAAGCTTGGGCA
 R614 TGCAGATATAGGTCTGTGTGCCCAAGCTTGAAC
 F629 CACAGACCTATATCTGCAATGTGAACCACAAGCC
 R647 CCACCTTTGTATTGCTGGGCTTGTGGTTCACAT
 F663 CAGCAATACAAAGGTGGACAAAAAAGTCGAGCCT
 R680 TGTCACAGCTCTTTGGAGGCTCGACTTTTTTTGT
 F697 CCAAAGAGCTGTGACAAAACCTCACACATGCCC
 R713 TACCTGGGCACGGTGGGCATGTGTGAGTTT
 F729 ACCGTGCCCAGGTAAGCCAGCCCAGGC
 R743 cccattgactTTTTTTTAGGCCTGGGCTGGCT
 R756 CTAAAAAAAgtaaatgggg

The BglII site was introduced using PCR with forward primer F0 and a reverse primer containing a BglII restriction site (sequences shown below).

Forward: tcaaagcagtaagAAAAAAGAATTCGCTCCAAACAG

Reverse: ttttttAGATCTCTTTGGAGGCTCGACTTTTTTTGT

The gene encoding V_H and CH1 was inserted into a vector comprising the crystallizable fragment (Fc) domain (i.e. CH2 and CH3 domains) of human IgG1 heavy chain at the EcoRI and BglII restriction sites. A Quikchange site-directed mutagenesis kit was used to generate the desired N-terminal mutants (EES, EPS, and the wildtype N-terminus EVQ). The incorporation of these mutations was verified by sequencing.

General procedure for expression and purification of wild-type and mutant antibodies. The plasmids for the light and heavy chains of the anti-HER2 antibody were transiently co-transfected into human embryonic kidney 293T cells in a 3:2 ratio using Lipofectamine 2000 (Invitrogen, Grand Island, NY) in Opti-MEM medium following the protocol from Invitrogen. The cells were incubated at 37 °C in a 5% CO₂ atmosphere. After 2 days, the media was collected and the secreted antibodies were purified using

protein G affinity chromatography, according to the procedure from the manufacturer (Pierce, Rockford, IL). The media was replaced and cultures were grown for an additional 3 days, after which the additional antibodies were harvested and purified as above. Purified protein was buffer exchanged into PBS using Amicon Ultra 4 mL 10,000 MWCO (Millipore) centrifugal ultrafiltration membranes. Purity was evaluated by SDS-PAGE with Coomassie staining.

General Procedure for Rapoport's Salt Mediated Transamination of Protein Substrates. Protein and Rapoport's salt stock solutions were prepared at twice the desired final concentration and mixed in equal volumes in a 1.5 mL Eppendorf tube. The final volume of each reaction was 100 μ L. The 2x protein stock solutions were prepared at 0.5 - 1 mg/mL in 25 mM phosphate buffer at pH 6.5. The 2x Rapoport's salt stock solutions (200 or 500 mM) were freshly prepared in 25 mM phosphate buffer (with 0.02% NaN_3), pH 6.5 from Rapoport's salt recrystallized from acetonitrile. The reaction mixture was briefly agitated to ensure mixing and then incubated without further agitation at 37 $^{\circ}\text{C}$ for 1 h. Following the reaction, the excess aldehyde was removed using 0.5 mL spin concentrators with a MWCO of 10 kDa (Millipore, Billerica, MA). The buffer exchange first involved the dilution of each sample to 500 μ L with 25 mM phosphate buffer (pH 6.5). Each sample was then concentrated to 100 μ L, and the process was repeated 3 times. The resulting keto-protein was then treated with an equal volume of 250 mM BnONH_2 (pH adjusted to 5.5), or 100 mM PEG(2kDa)- ONH_2 solution in a 1.5 mL Eppendorf tube and incubated at rt for 48 h. Buffer exchange steps were again repeated to stop the reaction and remove the excess alkoxyamine.

General Procedure for Antibody Disulfide Reduction and Cysteine Capping for Mass Spectrometry Analysis. To prepare the antibody mutants for mass spectrometry analysis, first the oligosaccharides were removed via treatment with *N*-Glycosidase F (PNGase F) following the protocol from the manufacturer (New England Biolabs, Ipswich, MA). Briefly, a buffer exchange into PBS was performed on the antibody samples. In a 1.5 mL Eppendorf tube, the protein was mixed with 10 μ L of G7 reaction buffer, 4 μ L PNGase, and additional PBS to a total volume of 100 μ L. The mixture was then incubated at 37 $^{\circ}\text{C}$ overnight. Immediately following treatment with PNGase, buffer exchange was performed into 100 mM Tris buffer, pH 8. Dithiothreitol (DTT) and ethylenediaminetetraacetic acid (EDTA) were then added to a final concentration of 10 mM each, and the reaction was incubated at room temperature for 20 min. After the reduction, iodoacetamide was added to a final concentration of 50 mM and the mixture was incubated at 37 $^{\circ}\text{C}$ for 30 min. The samples were then subjected to buffer exchange into 100 mM Tris buffer pH 8 for mass spectrometry analysis.

Liquid Chromatography and Mass Spectrometry Materials. Acetonitrile (Fisher Optima grade, 99.9%), formic acid (Pierce, 1 mL ampules, 99+%), and water purified to a resistivity of 18.2 $\text{M}\Omega\cdot\text{cm}$ (at 25 $^{\circ}\text{C}$) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile phase solvents for liquid chromatography-mass spectrometry (LC-MS).

LC-MS and -MS/MS Analysis of Synthetic Peptide Bioconjugates. Peptide bioconjugates were analyzed using a nanoAcquity ultraperformance liquid chromatograph (UPLC; Waters, Milford, MA) that was connected in-line with a quadrupole time-of-flight mass spectrometer (Q-ToF Premier, Waters).

The UPLC was equipped with C18 trapping (5 μm particles, 20 mm $\times$ 180 μm) and analytical (1.7 μm particles, 100 mm $\times$ 100 μm) columns and a 10 μL sample loop. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). Sample solutions contained in 0.3 mL polypropylene snap-top vials sealed with septa caps were loaded into the nanoAcquity autosampler compartment prior to analysis. Following sample injection (1 μL , partial loop), trapping was performed for 2 min with 100% A at a flow rate of 15 $\mu\text{L}/\text{min}$. The injection needle was washed with 500 μL of A and 200 μL of B after injection to avoid cross-contamination between samples. The elution program consisted of a linear gradient from 12% to 75% B over 15 min, a linear gradient to 95% B over 0.33 min, isocratic conditions at 95% B for 2.67 min, a linear gradient to 1% B over 0.33 min, and isocratic conditions at 1% B for 10.67 min, at a flow rate of 500 nL/min. The analytical column and sample compartment were maintained at 35 $^{\circ}\text{C}$ and 8 $^{\circ}\text{C}$, respectively.

The column exit was connected to a Universal Nanoflow Sprayer nanoelectrospray ionization (nanoESI) emitter mounted in the nanoflow ion source of the Q-ToF mass spectrometer. The nanoESI source parameters were as follows: nanoESI capillary voltage 2.4 kV, nebulizing gas (nitrogen) pressure 0.15 mbar, sample cone voltage 35 V, extraction cone and ion guide voltages both 4 V, and source block temperature 80 $^{\circ}\text{C}$. Cone gas was not used. The collision cell contained argon gas at a pressure of 8×10^{-3} mbar. The ToF analyzer was operated in "V" mode and routinely achieved a mass resolving power of 1.0×10^4 (measured at $m/z = 498$, full width at half maximum peak height), which was sufficient to resolve the isotopic distributions of singly and multiply charged peptide ions. Thus, a peptide ion's mass and charge could be determined independently (i.e., the ion charge was determined from the reciprocal of the spacing between adjacent isotope peaks in the m/z spectrum). External mass calibration of the ToF analyzer was performed immediately prior to analysis using a sodium formate solution. Mass spectra were acquired in the positive ion mode over the range $m/z = 300$ -1500, in continuum data format, using a 0.95 s scan integration and a 0.05 s interscan delay. In the data-dependent mode, up to three precursor ions exceeding an intensity threshold of 30 counts/second (cps) were selected from each survey scan for tandem mass spectrometry (MS/MS) analysis. MS/MS spectra were acquired over the range $m/z = 100$ -1500 using a 0.45 s scan integration, a 0.05 s interscan delay, and a collision energy of 30 eV. Ions were fragmented to achieve a minimum total ion current (TIC) of 250,000 cps in the cumulative MS/MS spectrum for a maximum of 10 s. An include list was used to select precursor ions of interest for MS/MS preferentially. To avoid the occurrence of redundant MS/MS measurements, real-time dynamic exclusion was used to preclude re-selection of previously analyzed precursor ions over an exclusion width of ± 0.2 m/z unit for a period of 120 s. Mass spectra and MS/MS spectra were processed using MassLynx software (version 4.1, Waters).

LC-MS Analysis of Reduced Antibody Bioconjugates. Reduced antibody bioconjugates were analyzed using an Agilent 1200 liquid chromatograph (LC; Santa Clara, CA) that was connected in-line with an LTQ Orbitrap XL hybrid mass spectrometer equipped with an Ion Max electrospray ionization source (ESI; Thermo Fisher Scientific, Waltham, MA). The LC was equipped with a Viva C8 (5 μ m particles, 100 mm $\times$ 1.0 mm, Restek, Bellefonte, PA) analytical column and a 100 μ L sample loop. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). Sample solutions contained in 0.3 mL polypropylene snap-top vials sealed with septa caps were loaded into the Agilent 1200 autosampler compartment prior to analysis. For each sample, approximately 100 to 300 picomoles of analyte was injected onto the column. Following sample injection, analyte trapping was performed for 5 min with 95% A at a flow rate of 150 μ L/min. The elution program consisted of a linear gradient from 30% to 95% B over 32.5 min, isocratic conditions at 95% B for 4 min, a linear gradient to 5% B over 0.5 min, and then isocratic conditions at 5% B for 9.5 min, at a flow rate of 150 μ L/min. The column and sample compartments were maintained at 35 $^{\circ}$ C and 4 $^{\circ}$ C, respectively. Solvent (Milli-Q water) blanks were run between samples, and the autosampler injection needle was rinsed with Milli-Q water after each sample injection, to avoid cross-contamination between samples.

The connections between the LC column exit and the mass spectrometer ion source were made using PEEK tubing (0.005" inner diameter, 1/16" outer diameter, Agilent). External mass calibration of the Orbitrap analyzer was performed prior to analysis using the standard Thermo LTQ calibration mixture containing caffeine, the peptide MRFA, and Ultramark 1621 dissolved in 51% acetonitrile/25% methanol/23% water/1% acetic acid (v/v). The ESI source parameters were as follows: ion transfer capillary temperature 250 $^{\circ}$ C, normalized sheath gas (nitrogen) flow rate 20%, ESI voltage 2.8 kV, ion transfer capillary voltage 35 V, and tube lens voltage 110 V. Mass spectra were recorded in the positive ion mode over the range m/z = 700 to 2000 using the Orbitrap mass analyzer, in profile format, with a full MS automatic gain control target setting of 5×10^5 charges and a resolution setting of 6×10^4 (measured at m/z = 400, full width at half maximum peak height). Raw mass chromatograms and spectra were processed using Xcalibur software (version 2.0.7, Thermo) and measured charge state distributions were deconvoluted using ProMass software (version 2.5 SR-1, Novatia, Monmouth Junction, NJ).

Gel Analyses. Sodium dodecyl sulfate-poly(acrylamide) gel electrophoresis (SDS-PAGE) was performed on a Mini-Protean apparatus (Bio-Rad) with 10-20% gradient polyacrylamide gels (BioRad). The protein samples were mixed with SDS loading buffer in the presence of dithiothreitol (DTT) and heated to 95 $^{\circ}$ C for 5 min to reduce disulfide bonds and denature the protein. Commercially available molecular mass markers (Bio-Rad) were applied for approximation of the apparent molecular masses.

Flow Cytometry of Antibody Mutants. Flow cytometry was used to determine the HER2 binding ability of the unmodified and modified anti-HER2 IgG1 mutants. For each experiment, 100 μ L containing 3×10^6 cells/mL of HCC1954, a human breast cancer cell line overexpressing HER2, was suspended in PBS containing 1% BSA. These cells were

incubated with 10 $\mu\text{g/mL}$ of the unmodified or modified antibodies for 30 min on ice. The cells were then washed with 500 μL of PBS containing 1% BSA and resuspended in 100 μL of PBS containing 1% BSA. Anti-human IgG1 antibody with FITC conjugated (Sigma) was then added to a final concentration of 22.5 $\mu\text{g/mL}$. The cells were incubated for 30 min on ice in the dark, then washed with 500 μL , and resuspended in 200 μL of PBS containing 1% BSA. The cells were analyzed by flow cytometry to determine the amount of FITC fluorescence. For each sample, 10,000 cells were counted.

5.7 REFERENCES

1. O'Hare, H.M., Johnsson, K., Gautier, A. Chemical probes Shed Light on Protein Function. *Curr. Opin. Struct. Biol.* **2007**, 17, 488–494.
2. Griffin, B.A., Adams, S.R., Tsien, R.Y. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science* **1998**, 281, 269–272.
3. Cravatt, B.F., Wright, A.T., Kozarich, J.W. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. *Annu. Rev. Biochem.* **2008**, 77, 383–414.
4. Witus, L.S., Francis, M.B. Using Synthetically Modified Proteins to Make New Materials. *Acc. Chem. Res.* **2011**, 44, 774–783.
5. Zalipsky, S. Functionalized Poly(ethylene glycols) for Preparation of Biologically Relevant Conjugates. *Bioconj. Chem.* **1995**, 6, 150–165.
6. Baker, D.P., Lin, E.Y., Lin, K., Pellegrini, M., Petter, R.C., Chen, L.L., Arduini, R.M., Brickelmaier, M., Wen, D., Hess, D.M., Chen, L., Grant, D., Whitty, A., Gill, A., Lindner, D.J., Pepinsky, R.B. N-Terminally PE-Glylated Human Interferon- β -1a with Improved Pharmacokinetic Properties and in Vivo Efficacy in a Melanoma Angiogenesis Model. *Bioconj. Chem.* **2005**, 17, 179–188.
7. Wu, A.M., Senter, P.D. Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotech.* **2005**, 23, 1137–1146.
8. Alley, S.C., Okeley, N.M., Senter, P.D. Antibody-Drug Conjugates: Targeted Drug Delivery for Cancer. *Curr. Opin. Chem. Biol.* **2010**, 14, 529–537.
9. Hermanson G.T. Bioconjugate Techniques. Academic Press. **1996**, 1st Ed.
10. Junutula, J.R., Raab, H., Clark, S., Bhakta, S., Leipold, D.D., Weir, S., Chen, Y., Simpson, M., Tsai, S.P., Dennis, M.S., Lu, Y., Meng, Y.G., Ng, C., Yang, J., Lee, C.C., Duenas, E., Gorrell, J., Katta, V., Kim, A., McDorman, K., Flagella, K., Venook, R., Ross, S., Spencer, S.D., Lee Wong, W., Lowman, H.B., Vandlen, R., Sliwkowski, M.X., Scheller, R.H., Polakis, P., Mallet, W. Site-Specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index. *Nat. Biotech.* **2008**, 26, 925–932.
11. McDonagh, C.F., Turcott, E., Westendorf, L., Webster, J.B., Alley, S.C., Kim, K., Andreyka, J., Stone, I., Hamblett, K.J., Francisco, J.A., Carter, P. Engineered Antibody–Drug Conjugates with Defined Sites and Stoichiometries of Drug Attachment. *Prot. Engineer. Design Selec.* 2006, 19, 299–307.
12. Wu, P., Shui, W., Carlson, B.L., Hu, N., Rabuka, D., Lee, J., Bertozzi, C.R. Site-Specific Chemical Modification of Recombinant Proteins Produced in Mammalian Cells by Using the Genetically Encoded Aldehyde Tag. *Proc. Natl. Acad. Sci.* **2009**, 106, 3000–3005.
13. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S., Francis M.B. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **2006**, 45, 5307–5311.
14. Witus L.S., Francis M. Site-Specific Protein Bioconjugation via a Pyridoxal 5'-Phosphate-Mediated

- N-Terminal Transamination Reaction. *Current Protocols in Chemical Biology*. John Wiley & Sons, Inc. **2010**. 125-134.
15. Jencks, W.P. Studies on the Mechanism of Oxime and Semicarbazone Formation. *J. Am. Chem. Soc.* **1959**, *81*, 475–481.
 16. Scheck R.A., Francis M.B. Regioselective Labeling of Antibodies Through N-Terminal Transamination. *ACS Chem. Biol.* **2007**, *2*, 247-251.
 17. Hamblett K.J., Senter P.D., Chace, D.F., Sun, M.M.C., Lenox, J., Cervený, C.G., Kissler, K.M., Bernhardt, S.X., Kopcha, A.K., Zabinski, R.F., Meyer, D.L., Francisco, J.A. Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. *Clinical Cancer Res.* **2004**, *10*, 7063 –7070.
 18. Scheck R.A., Dedeo M.T., Iavarone A.T., Francis M.B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, *130*, 11762-11770.
 19. Witus L.S. Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., Francis, M.B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, *132*, 16812-16817.
 20. Buckley T.F., Rapoport, H. Mild and Simple Biomimetic Conversion of Amines to Carbonyl Compounds. *J. Am. Chem. Soc.* **1982**, *104*, 4446–4450.
 21. Liu H., Gaza-Bulseco G., Faldu D., Chumsae C., Sun J. Heterogeneity of monoclonal antibodies. *J. Pharm. Sciences* **2008**, *97*, 2426-2447.
 22. Chelius D., Jing K., Lueras A., Rehder D.S., Dillon T.M., Vizek A., Rajan R.S., Li T., Treuheit M.J., Bondarenko P.V. Formation Of Pyroglutamic Acid from N-Terminal Glutamic Acid in Immunoglobulin Gamma Antibodies. *Anal. Chem.* **2006**, *78*, 2370-2376.
 23. Yu L., Vizek A., Huff M.B., Young M., Remmele R.L., He, B. Investigation of N-Terminal Glutamate Cyclization of Recombinant Monoclonal Antibody in Formulation Development. *J. Pharm. Biomed. Anal.* **2006**, *42*, 455-463.
 24. Giustiniano, M., Pirali, T., Massarotti, A., Biletta, B., Novellino, E., Campiglia, P., Sorba, G., Tron, G. A Practical Synthesis of 5-Aroyl-1-aryltetrazoles Using an Ugi-Like 4-Component Reaction Followed by a Biomimetic Transamination. *Synthesis* **2010**, *2010*, 4107–4118.
 25. Shi Q, Verdier-Pinard P, Brossi A., Hamel E., McPhail A.T., Lee K.H. Synthesis and Biological Evaluation of Novel Decetamidothiocolchicin-7-OLs and Ester Analogs as Antitubulin Agents. *J. Med. Chem.* **1997**, *40*, 961-966.
 26. Cho H.-S., Mason K., Ramyar K.X., Stanley A.M., Gabelli S.B., Denney D.W., Leahy D.J. Structure of the Extracellular Region of Her2 Alone and in Complex with the Herceptin Fab. *Nature* **2003**, *421*, 756-760.
 27. Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G., Blanquet, S. Extent of N-Terminal Methionine Excision from Escherichia Coli Proteins Is Governed by the Side-Chain Length of the Penultimate Amino Acid. *Proc. Natl. Acad. Sci.* **1989**, *86*, 8247–8251.
 28. Harris L.J., Skaletsky E., McPherson, A. Crystallographic Structure Of an Intact IgG1 Monoclonal Antibody. *J. Molec. Biol.* **1998**, *275*, 861-872.
 29. Neve R.M., Chin K., Fridlyand J., Yeh J., Baehner F.L., Fevr T., Clark L., Bayani N., Coppe J.P., Tong F., Speed T., Spellman P.T., DeVries S., Lapuk A., Wang N.J., Kuo W.L., Stilwell J.L., Pinkel D., Albertson D.G., Waldman F.M., McCormick F., Dickson R.B., Johnson M.D., Lippman M., Ethier S., Gazdar A., Gray J.W. A Collection of Breast Cancer Cell Lines for the Study of Functionally Distinct Cancer Subtypes. *Cancer Cell* **2006**, *10*, 515-527.
 30. Schlick, T.L., Ding, Z., Kovacs, E.W., Francis, M.B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.

CHAPTER 6:

NEW AVENUES FOR N-TERMINAL REACTIVITY

ABSTRACT

The site-specificity of pyridoxal 5'-phosphate (PLP)-mediated transamination is due to the unique reactivity of the N-terminal position of a protein. This position can be used to develop new types of bioconjugation reactions. Described herein are some attempts towards this goal, although in some cases much investigation remains. These avenues include new ways to improve and expand N-terminal transamination and to form attachments to the protein using methods other than oxime formation. In one project, the library screening platform was used as a tool to evaluate aldehydes as candidate transamination reagents against all N-terminal sequences simultaneously. Another investigation was on the effect of 415 nm light on PLP-mediated transamination. Finally, we also considered how to turn an unwanted byproduct into a useful form of reactivity. The addition of PLP to the N-terminus during transamination has previously been seen as a byproduct we have sought to minimize. However, if PLP was replaced by another aldehyde with a useful functionality attached, the adduct could become the desired product, thus eliminating the need for oxime formation. This was investigated using an alkyne-containing glyoxamide compound.

6.1 LIBRARY SCREENING TO EVALUATE POTENTIAL TRANSAMINATION REAGENTS

The combinatorial library screening platform discussed in Chapter 3 is a powerful tool for the evaluation of potential transamination candidates because the transamination ability of a compound can be assessed for all N-terminal sequences at once.¹ When the pyridoxal 5'-phosphate (PLP)-mediated transamination was being developed, a number of other aldehydes were investigated as well.² PLP was found to result in the highest transamination yield, thus it was selected for further development. However, this screen was performed on a single substrate, the angiotensin peptide. Subsequent studies revealed the influence of sequence on reactivity; therefore, it is possible that although PLP resulted in the highest transamination for the particular N-terminal sequence examined, other aldehydes may be successful transamination reagents for other sequences. For instance, in Chapter 5 the use of the combinatorial library to discover *N*-methyl pyridinium carboxaldehyde benzene sulfonate salt as an efficient transamination reagent for glutamate-terminal sequences was described. The library screening process has also been used to evaluate other potential transamination reagents.

The one-bead-one-compound (OBOC) peptide library, described in depth in Chapter 3, was used to evaluate the transamination ability of a series of common aldehydes (Figure 6-1). Portions of the library were incubated with candidate compounds at a 10 mM concentration for 18 h in pH 6.5 phosphate buffer. After incubation with the aldehydes, any beads that bore transaminated sequences were identified through the subsequent formation of a red colored oxime using Disperse Red alkoxyamine (DispRed-ONH₂). This screen was performed prior to the calibration of the colorimetric detection scheme described in Chapter 3, and thus a red colored bead did not necessarily correspond to a high yield of oxime formation but did indicate some degree of oxime product. The compounds screened included salicylaldehyde (**1a**), 2-hydroxy-5-nitrobenzaldehyde (**1b**), glyoxylic acid (**1c**), benzaldehyde (**1d**), tetrafluoroquinone, and 2,3-dimethoxy-5-methyl-*p*-benzoquinone. Images of the library resulting from incubation with the first four of these reagents is shown in Figure 6-1b. Under the same reaction conditions, PLP resulted in a large proportion of red beads (Chapter 4), which confirmed PLP to be a very general transamination reagent for many N-terminal sequences. The presence of a few red beads after incubation with the candidate aldehydes indicated that, although they may not be general transamination reagents, they may be capable of transamination for particular N-terminal sequences. This may be useful, as some proteins may tolerate only a few different N-terminal sequences. Therefore it would be helpful to have an arsenal of transamination reagent/optimal sequence pairs from which to draw on for various applications. Additionally, reagents that are highly specific for a certain N-terminal sequence can be used to modify one protein selectively out of a mixture. The sequencing of these red beads remains to be done, as well as the resynthesis of the motifs identified from screening to confirm the transamination by liquid chromatography-mass spectrometry (LC-MS) analysis.

Although the optimal N-terminal sequence had not yet been identified through screening, glyoxylic acid-mediated transamination was attempted on an AK-terminal peptide,

AKSWSNAG. Shown in Figure 6-1c are the resulting yields after transamination, benzyl oxime formation, and LC-MS quantification. The reaction time was varied using 10 mM glyoxylic acid in pH 6.5 phosphate buffer. The formation of oxime product seen in these data indicated that, with more sequencing of red beads to identify optimal N-terminal sequences, glyoxylic acid-mediated transamination may be a successful bioconjugation strategy. Initial attempts at sequencing revealed lysine as the preferred N-terminal residue (data not shown), but since lysine-terminal sequences have led to complex product distributions in the past³ (described in Chapter 2), more screening is necessary to evaluate whether these or other N-terminal sequences will lead to high yields of a single product. The sequencing of the red beads resulting from the other aldehydes remains to be done as well. Additionally, it may be interesting to use this platform as a method to investigate the structural properties important for transamination, by evaluating a number of PLP derivatives and other related aldehydes.

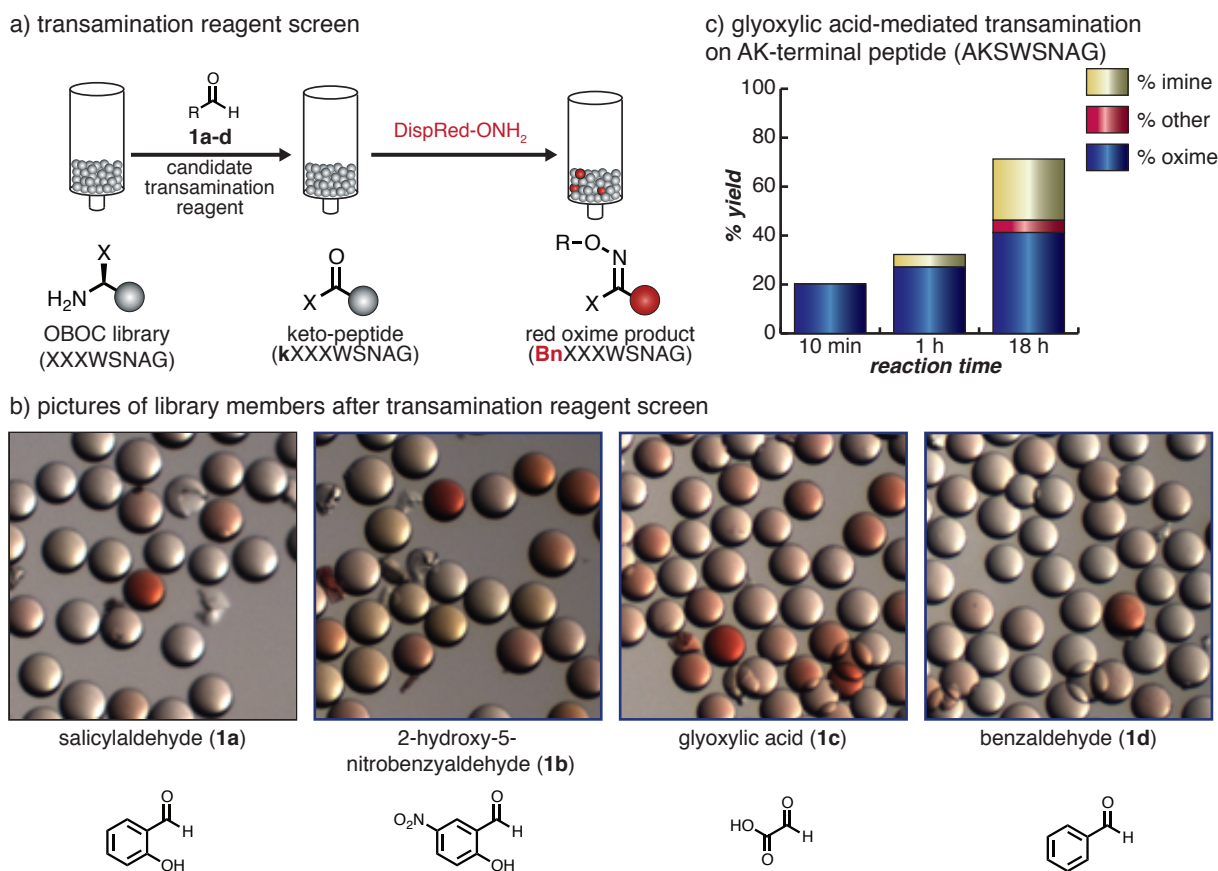


Figure 6-1. The combinatorial library was used to evaluate other aldehydes as transamination reagents. a) By screening candidates against the library, their transamination ability for all N-terminal sequences could be evaluated simultaneously. b) After incubation with common aldehydes (10 mM for 18 h, pH 6.5) followed by oxime formation with Disperse Red alkoxyamine, some red beads were observed. The calibration of the colorimetric detection was not used however, so these beads may not correspond to a high oxime yield (see Chapter 3 for details). c) Although sequencing of the red beads remains to be done, glyoxylic acid-mediated transamination was quantified on an AK-terminal peptide after reaction with 10 mM glyoxylic acid for the specified time.

6.2 LIGHT-MEDIATED ENHANCEMENT OF PLP TRANSAMINATION

Since PLP is a common biological cofactor, there have been many studies on its enzymatic role. One such study with potential relevance to the nonenzymatic PLP-mediated N-terminal transamination was a report by Hill and coworkers. They indicated that 440 nm light could enhance the rate of PLP-mediated transformations. In this paper it was proposed that the PLP-Schiff base absorbed light between 410 and 430 nm to form a highly reactive triplet state.⁴ They proposed that the pK_a of the excited state imine α -proton was lowered by 11-19 units relative to the ground state, thus facilitating transamination. The authors found the rate enhancement in the enzyme-mediated decarboxylation of aspartate, but also in the non-enzymatic transamination of amino acids in solution. Therefore we sought to investigate whether irradiation could enhance the rate of PLP-mediated transamination of proteins.

To assess the effect of blue light on PLP-mediated transamination, a gel assay was used. The protein myoglobin was incubated with various concentrations of PLP (1, 10 or 100 mM) for either 10 or 60 minutes. For each set of reaction conditions, two samples were prepared: one that was placed in a dark cabinet, and one that was irradiated with 415 nm light using a 70 LED light source capable of illumination up to 100 mW/cm² (All Blue Single Head Acnelamp, Dima Tech). The samples were prepared in glass vials, which were suspended approximately one inch above the lamp head. After transamination, a 2 kDa poly(ethylene glycol) (PEG) alkoxyamine reagent was added for analysis of oxime product formation by a gel shift assay.⁵ The protein that formed an oxime ran with an apparent molecular weight higher than the unmodified sample, thus providing a simple means to compare the degree of conversion. As shown in Figure 6-2, all of the samples that were subjected to irradiation showed higher conversion to the PEG oxime than the corresponding sample kept in the dark. Thus, these data indicated that light-enhancement of PLP-mediated transamination was possible. A second, higher molecular weight band

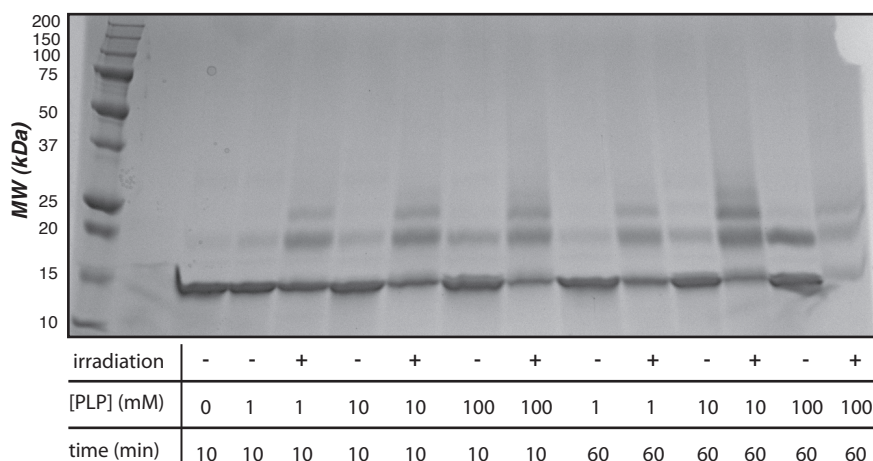


Figure 6-2. The effect of irradiation with 415 nm light on PLP-mediated transamination was investigated. Shown is an SDS-PAGE gel of myoglobin samples after reaction with PLP and a 2 kDa molecular weight poly(ethylene glycol) (PEG) alkoxyamine reagent. The higher molecular weight band indicated formation of oxime between the transaminated protein and the PEG. Enhanced yields were seen in the samples that were irradiated. The presence of two higher molecular weight bands in those samples indicated that irradiation may cause over-modification as well.

was observed in some of the samples subjected to irradiation, indicating additional PEG attachment. We hypothesized that if the imine α proton pK_a was lowered in the triplet state, lysine side chain amines may have been able to undergo transamination even though under normal, non-irradiated conditions they did not.

However, these data were not reproducible in subsequent experiments and contradictory results were also obtained (data not shown). Therefore, we decided not to pursue blue light irradiation as a component of PLP-mediated bioconjugation. It is worth noting that these experiments serve to underscore the importance of preparing fresh solutions of PLP, as visible light contains the wavelengths that may alter its properties.

6.3 PROLINE-TERMINAL BIOCONJUGATION

In addition to the aforementioned attempts at improving N-terminal transamination, alternate modes of N-terminal reactivity were also investigated. One such avenue was the formation of adducts to proline-terminal sequences. Previous studies on the PLP-mediated transamination of proline-terminal peptides had resulted in stable additions of PLP to the N-terminus with a mass that corresponded to a ring opening of the proline (proposed structure **4**).³ Although the structure of this adduct remains to be fully characterized, there are few methods for bioconjugation on proline residues, so these adducts were of high interest. Although the covalent addition of PLP to the N-terminus does not result in a

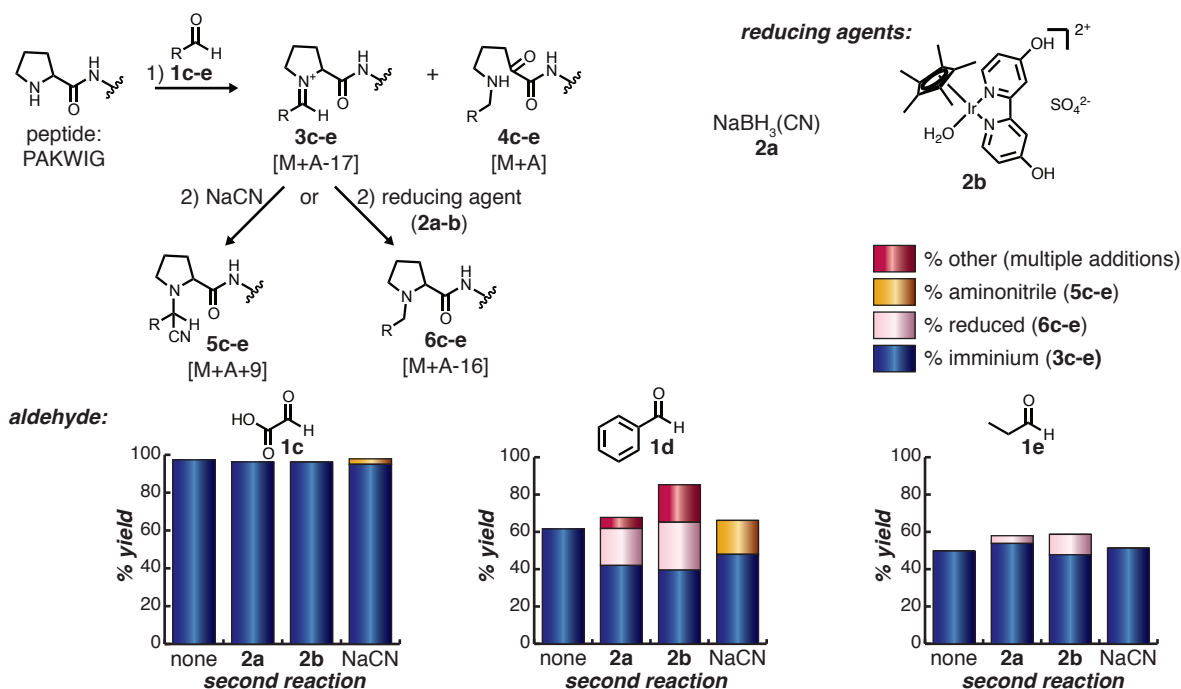


Figure 6-3. Proline-terminal reactivity was investigated by LC-MS analysis of the products resulting from the incubation of a PAKWIG peptide with aldehydes **1c-e**. Proposed structures for the resulting products based on the mass change listed (M= mass of peptide, A=mass of aldehyde) are shown. To trap the putative iminium product (**3c-e**), secondary reactions were applied as well. Although the product was not reactive toward reduction (**6c-e**) or aminonitrile (**5c-e**) formation, the incubation of the proline-terminal peptide with glyoxylic acids led to high yields of a single addition product (**3c**).

useful bioconjugate, we hypothesized that if other classes of aldehydes exhibited similar types of reactivity, the formation of a stable adduct could be a method of attaching an aldehyde containing a desired functionality to a proline-terminal protein.

To test whether aldehydes besides PLP would result in stable covalent additions to N-terminal proline residues, a peptide, PAKWIG, was used as a test substrate. The products resulting from incubation with aldehydes **1c-e** were analyzed by LC-MS. The three aldehydes examined were glyoxylic acid (**1c**), benzaldehyde (**1d**), and an aliphatic aldehyde, propionaldehyde (**1e**). After incubation of 10 mM aldehyde for 18 h at pH 7.5 masses corresponding to additions were observed. These adducts are presumed to be stable since they persisted through the rinsing steps, exposure to basic conditions during the cleavage of the peptide from the resin, and the acidic conditions of the LC-MS. None of the aldehydes examined resulted in masses that corresponded to the putative ring-opening addition observed for PLP (the M+A mass where M is the mass of the peptide and A is the mass of the aldehyde, Figure 6-3). However, large amounts of a different addition product were observed, **3c-e**, which by mass corresponded to imminium formation (Figure 6-3). Since an imminium product would not be assumed to be stable, it is possible that the actual structure is an isomer of the proposed structure drawn.

We hypothesized that if the observed adduct was an imminium ion, it could be trapped as a more stable product through a second reaction. Therefore after incubation with the aldehyde, we applied various secondary reactions. These included attempts at reduction of the imminium species with sodium cyanoborohydride and an iridium hydride transfer catalyst.⁶ Sodium cyanide was also used to form a Strecker-reaction type aminonitrile product. However, these secondary reactions did not result in large amounts of the intended products, perhaps providing evidence that the adduct structure was not an imminium ion. Although the structural characterization of the **3c-e** adducts is pending, this addition represents a promising lead for the modification of proline-terminal sequences. Using glyoxylic acid, the adduct **3c** formed in a very high yield of a single product. It remains to be seen whether a similar adduct will form for other types of glyoxamide compounds that contain useful functionalities. A proline terminal GFP mutant was expressed in order to evaluate this reactivity on a protein substrate. These initial results indicate that the addition of certain aldehydes to the proline-terminal sequences may be a promising and novel bioconjugation strategy.

6.4 TURNING THE ADDUCT INTO THE DESIRED PRODUCT

In addition to PLP-mediated transamination, the covalent addition of PLP to the N-terminus of proteins has been observed for many N-terminal sequences. Previously this has been considered to be an unwanted byproduct. Although the structure obtained varied for certain N-terminal residues (for instance N-terminal tryptophan residues form Pictet-Spengler products), for most N-terminal residues the observed adduct corresponded by mass to an aldol-type addition (**8**, Figure 6-4a). In most cases the PLP addition product was not observed in amounts above 10% of the total product. Additionally, the PLP adduct

contains a ketone, and so does not preclude oxime formation. On proteins bearing the highly reactive Ala-Lys terminal motif, particularly large amounts of adduct were observed.¹ To minimize the formation of this byproduct, reaction conditions were screened in hopes of eliminating its formation. This was done to be able to obtain a single bioconjugate, the oxime product, in high yield. However, we realized that if a similar adduct could be formed in high yields with aldehydes containing useful functionalities or reactive handles, the adduct itself could become the desired bioconjugate product.

Although structural characterization is needed, the observed mass corresponded to the covalent addition of PLP to the N-terminus via to an aldol addition without dehydration. The use of aldol reactions as a bioconjugation strategy has been reported using Mukaiyama aldol reagents. As reported by Alam and coworkers, proteins with an N-terminal glycine were transaminated using PLP and the resulting aldehyde served as an aldol acceptor.⁷ Therefore, the proposed aldol adduct (**8**) represents a novel mode of bioconjugation reactivity because in this case the protein serves as the aldol donor and the exogenous aldehyde as the aldol acceptor. Since the PLP adduct has been observed on various N-terminal residues, this type of reactivity is potentially general for a number of N-termini. The addition can be achieved in a single step since the aldehyde that performs the transamination also adds to the protein in the same pot. There is also the potential for dual functionalization since the product contains a ketone that could subsequently be used for

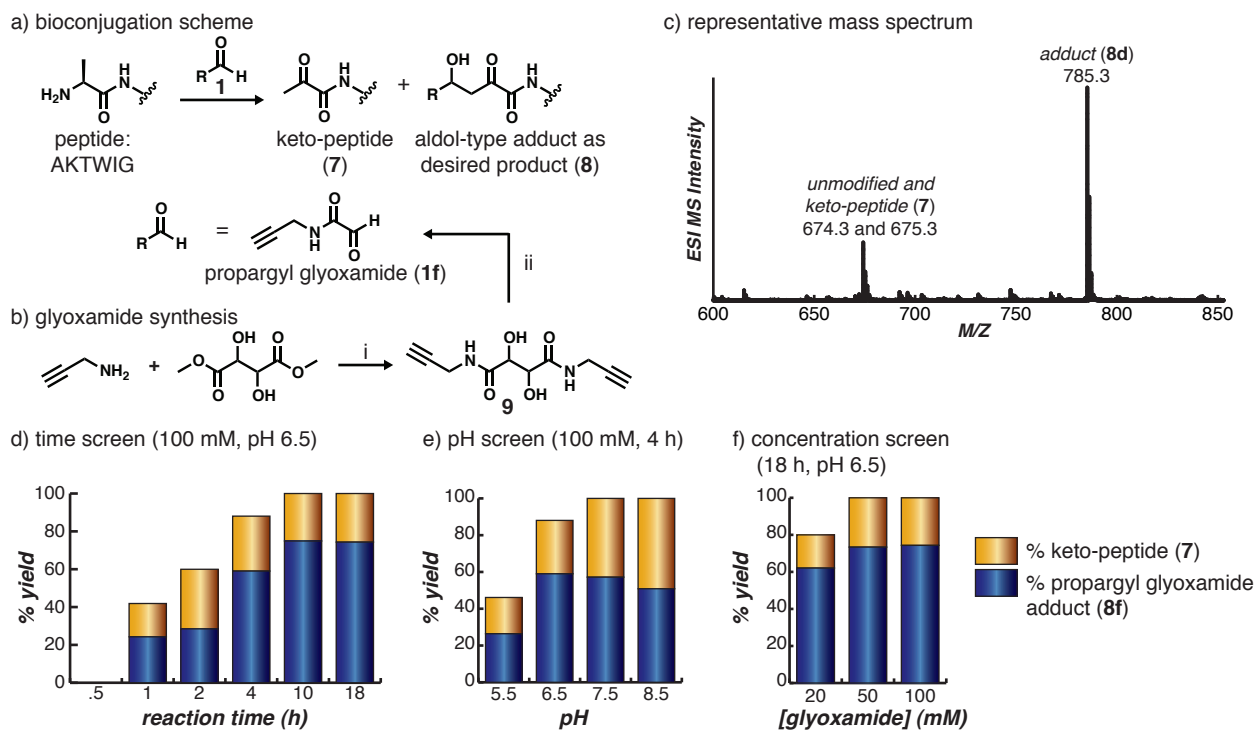


Figure 6-4. Investigation of an aldol addition to the N-terminus as a desired product. a) When the transamination reagent (**1**) is pyridoxal 5'-phosphate (PLP), the formation of a product (**8**) was observed that corresponded by mass to an aldol addition with the keto-peptide (**7**). This was initially considered to be an unwanted byproduct. However, if the transamination reagent included an interesting functionality or reactive handle (such as an alkyne as in **1f**) then the putative aldol addition product could be the desired product. b) A propargyl glyoxamide (**1f**) was synthesized to demonstrate this type of reactivity. The synthesis was accomplished by *i.* stirring overnight at room

oxime formation. Given these promising attributes, what remained to transform the PLP adduct into a viable bioconjugation strategy was the discovery of aldehydes other than PLP which could perform similar reactivity but contain more useful functionalities.

Candidates for aldol addition reagents needed to be aldehydes capable of transamination and with ready access to a number of functionalities. Alkylated pyridinium carboxaldehyde reagents meet these criteria, as Rapoport's salt is capable of transamination, has been found to form some aldol adduct, and a wide number of derivatives could be synthetically accessed through the group attached to the pyridine nitrogen. Other promising candidates were functionalized glyoxamides. Glyoxylic acid had been used for transamination,² and can be readily functionalized through selection of any primary amine for the amide linkage. The two-step synthetic route we used to access a glyoxamide compound is shown in Figure 6-4b. An amine of choice was stirred with dimethyltartrate to form a symmetric tartaric amide (**9**). This was cleaved with sodium periodate to form the glyoxamide reagent.⁸ Although a number of derivatives could be accessed by using various synthetic or commercially available amines, propargylamine was used to create propargyl glyoxamide (**1f**) as an initial test compound. The benefits of using propargyl glyoxamide include its water solubility and the potential of the alkyne group for further secondary bioconjugation reactions using copper-catalyzed alkyne azide cycloaddition (CuAAC) with any commercially available azide-containing reagent. There are currently few simple methods for installing an alkyne in a single location on a protein, so the site-specific aldol addition of propargyl glyoxamide to the N-terminus may be of great interest.

Since the Ala-Lys terminal motif was found to result in higher than average proportions of PLP-adduct, an AKT-terminal peptide was used as a test substrate for the glyoxamide reagent. Incubation of the peptide, AKTWIG, with propargyl glyoxamide resulted in transamination and a product which by mass corresponded to the aldol addition product. A representative mass spectrum is shown in Figure 6-4c. Not all of the keto-peptide produced from transamination formed the aldol addition, but high yields of the adduct were observed. The reaction time, pH, and concentration of glyoxamide were screened. A yield of 80% glyoxamide addition was achieved after 10 h of incubation with 100 mM propargyl glyoxamide. The transamination reached even higher conversions, so investigation into how to increase the efficiency of the aldol addition step may be beneficial. Additionally, further reaction screens may be necessary to find conditions that achieve high yields with a lower concentration of glyoxamide. If this is not possible, than the use of a glyoximide that installs a reactive handle on the protein (such as the propargyl glyoxamide) may be more economical than including the desired exogenous functionality in the glyoxamide compound. Other initial experiments that were performed included adding the glyoxamide as an aldol acceptor subsequent to transamination of the peptide using PLP. This resulted in glyoxamide addition, but did not enhance the yield over the one-pot incubation with the glyoxamide (data not shown). Additionally, a catalyst for aqueous aldol reactions, norcotine,⁹ was added but was not found to affect the reaction rate significantly.

Given the influence of the N-terminal sequence on transamination, it is likely that there is a strong sequence effect on reactivity towards aldol addition as well. In the future this

can be explored using colorful aldehyde reagents on the library to identify beads bearing sequences that formed high levels of addition products. Other pending experiments include the structural characterization of the products, the demonstration of this reaction on protein substrates, the confirmation of its site-specificity for the N-terminus using tandem mass spectrometry, and the demonstration that the adduct alkyne can be used to attach useful reagents, such as dyes or polymers, through CuAAC.

The site-specificity of N-terminal transamination has demonstrated that the N-terminus has unique reactivity relative to other functional groups on a protein. Our studies have revealed many avenues for the development of new site-specific bioconjugation reactions using the N-terminus. Exciting!

6.5 MATERIALS AND METHODS

General Procedures. Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H₂O) used as reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Pyridoxal 5'-phosphate monohydrate was obtained from Aldrich.

Combinatorial Library Synthesis, and General Procedure for Library Screening, Sequencing, and Solid Phase Peptide Synthesis. Procedures described in Chapter 3 were used.

General Procedure for Glyoxylic Acid-Mediated Transamination of Resin-Bound Peptide Substrates. Portions of resin-bound AKWSNAG peptide (appx. 10 mg each) were treated with 1 mL of 10 mM glyoxylic acid solution in 50 mM potassium phosphate buffer, pH 6.5 with 10% DMF to facilitate bead swelling and 0.02% NaN₃ for preservation of the buffer. After the specified time of gentle rotation at room temperature, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual aldehyde. The peptides were then incubated with 1 mL of a 250 mM *O*-benzylhydroxylamine hydrochloride (BnONH₂) solution in water for 3 h at rt. Excess alkoxyamine was removed by rinsing with three portions of deionized water, followed by three portions of DMF. The peptides were then cleaved from the resin through incubation with 300 μ L of a 100 mM sodium hydroxide solution for 30 min. The resulting peptide solution was added to 700 μ L of 50 mM phosphate buffer (pH 6.5). The resulting solution was diluted 20-fold into dd-H₂O for mass spectrometry analysis.

Liquid Chromatography and Mass Spectrometry of Peptide Bioconjugates. Procedure described in Chapter 4 was used.

Gel Analyses. Procedure described in Chapter 5 was used.

General Procedure for Primary Modification of Resin-Bound Proline-Terminal

Peptide Substrates. Portions of resin-bound PAKWIG peptide (appx. 10 mg each) were treated with 1 mL of 10 mM aldehyde solution in 50 mM potassium phosphate buffer, pH 7.5 with 10% DMF to facilitate bead swelling and 0.02% NaN₃ for preservation of the buffer. After 18 h of gentle rotation at room temperature, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual aldehyde. The peptides were then subjected to a secondary reaction (described below) or cleaved from the resin through incubation with 300 μ L of a 100 mM sodium hydroxide solution for 30 min. The resulting peptide solution was added to 700 μ L of 50 mM phosphate buffer (pH 6.5). The resulting solution was diluted 20-fold into dd-H₂O for mass spectrometry analysis.

General Procedure for Secondary Modification of Resin-Bound Proline-Terminal Peptide Substrates. The proline-terminal peptide samples that were subjected to a secondary reaction were exposed to the secondary reagents after incubation with the aldehyde and rinsing. One mL of the following solutions was added to the resin: 20 mM NaBH₃(CN) in ethanol, 0.3 mM [Cp*Ir(4,4'-dihydroxy-2,2'-bipyridine)(H₂O)]SO₄ catalyst⁶ in 200 mM phosphate buffer pH 7.4, 100 mM potassium formate, or 20 mM NaCN in phosphate buffer pH 7.5. After 3 h of gentle rotation at room temperature, the resin was washed with three portions of deionized water, followed by three portions of DMF and the peptide was subsequently cleaved from the resin for mass spectrometry analysis.

Synthesis of Propargyl Glyoxamide.

NMR. ¹H and ¹³C spectra were measured with a Bruker AVQ-400 (400 MHz) or AV-300 (300 MHz) spectrometer. ¹H NMR chemical shifts are reported as δ in units of parts per million (ppm). Multiplicities are reported as follows: s (singlet), d (doublet), t (triplet), q (quartet), 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 nH, and is based on spectral integration values. ¹³C NMR spectra are reported as δ in units of part per million (ppm).

Propargyl tartaric amide (9). To a 20 mL scintillation vial was added dimethyl L-tartrate (0.25 g, 1.4 mmol, 1.0 equiv) and propargyl amine (0.9 mL, 14.0 mmol, 10 equiv). The mixture was stirred at room temperature for 24 h. The mixture was concentrated under reduced pressure to remove excess amine. The resulting residue was dissolved in methanol and concentrated under reduced pressure again. The resulting off-white solid was used in subsequent experiments without further purification. ¹H-NMR (400 MHz, MeOD) δ 2.89 (s, 2H) 2.42 (m, 4H) 1.73 (t, 2H, J=1.6) 1.00 (t, 2H, J=2.4) ¹³C-NMR (400 MHz, MeOD) δ 172.9, 79.0, 72.1, 70.7, 27.8.

Propargyl glyoxamide (1f). To a 500 mL round bottom flask was added propargyl tartaric amide (0.283 g, 1.27 mmol, 1.0 equiv). The white solid was dissolved in 50 mL dichloromethane with 5 mL methanol. Sodium periodate adsorbed on silica gel was prepared according to the procedure by Zhong *et al.*,⁸ and 3.16 g was added to the reaction. The resulting slurry was stirred for 3 h at room temperature. The mixture was then filtered and the filtrate was collected and concentrated under reduced pressure. The resulting colorless oil was used without further purification.

General Procedure for Glyoxamide Addition to Resin-Bound Peptide Substrates.

Portions of resin-bound AKTWIG peptide (appx. 10 mg each) were treated with 1 mL of the specified concentration of propargyl glyoxamide solution in 50 mM potassium phosphate buffer, at the specified pH with 10% DMF to facilitate bead swelling and 0.02% NaN_3 for preservation of the buffer. After the specified reaction time of gentle rotation on a LabQuake shaker at room temperature, the resin samples were washed with three portions of deionized water, followed by three portions of DMF to remove residual aldehyde. The peptides were then cleaved from the resin through incubation with 300 μL of a 100 mM sodium hydroxide solution for 30 min. The resulting peptide solution was added to 700 μL of 50 mM phosphate buffer (pH 6.5). The resulting solution was diluted 20-fold into dd- H_2O for mass spectrometry analysis.

6.6 REFERENCES

1. Witus L.S. Moore, T., Thuronyi, B.W., Esser-Kahn, A.P., Scheck, R.A., Iavarone, A.T., Francis, M.B. Identification of Highly Reactive Sequences for PLP-Mediated Bioconjugation Using a Combinatorial Peptide Library. *J. Am. Chem. Soc.* **2010**, 132, 16812-16817.
2. Gilmore J.M., Scheck R.A., Esser-Kahn A.P., Joshi N.S., Francis M.B. N-Terminal Protein Modification Through A Biomimetic Transamination Reaction. *Angew. Chem. Int. Ed.* **2006**, 45, 5307-5311.
3. Scheck R.A., Dedeo M.T., Iavarone A.T., Francis M.B. Optimization of a Biomimetic Transamination Reaction. *J. Am. Chem. Soc.* **2008**, 130, 11762-11770.
4. Hill, M.P., Carroll, E.C., Vang, M.C., Addington, T.A., Toney, M.D., Larsen, D.S. Light-Enhanced Catalysis by Pyridoxal Phosphate-Dependent Aspartate Aminotransferase. *J. Am. Chem. Soc.*, **2010**, 132, 16953–16961.
5. Schlick, T.L., Ding, Z., Kovacs, E.W., Francis, M.B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, 127, 3718–3723.
6. McFarland, J.M., Francis, M.B. Reductive Alkylation of Proteins Using Iridium Catalyzed Transfer Hydrogenation. *J. Am. Chem. Soc.* **2005**, 127, 13490–13491.
7. Alam, J., Keller, T.H., Loh, T.-P. Functionalization of Peptides and Proteins by Mukaiyama Aldol Reaction. *J. Am. Chem. Soc.* **2010**, 132, 9546–9548.
8. Zhong Y.-L., Shing T.K.M. Efficient and Facile Glycol Cleavage Oxidation Using Improved Silica Gel-Supported Sodium Metaperiodate. *J. Org. Chem.* **1997**, 62, 2622–2624
9. Dickerson, T.J., Janda, K.D. Aqueous Aldol Catalysis by a Nicotine Metabolite. *J. Am. Chem. Soc.* **2002**, 124, 3220–3221.