Title
Development and Applications of N-terminal Protein Bioconjugation Reactions

Author
Palla, Kanwal

Publication Date
2016

Peer reviewed|Thesis/dissertation

Development and Applications of N-terminal Protein Bioconjugation Reactions

by

Kanwal Siddique Palla

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 Jamie Cate
Professor Wenjun Zhang
Summer 2016

Development and Applications of N-terminal Protein Bioconjugation Reactions

By: Kanwal Siddique Palla

Abstract

Development and Applications of N-terminal Protein Bioconjugation Reactions

by

Kanwal Siddique Palla

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

With highly evolved structures and function, proteins have an extraordinarily diverse range of capabilities. In order to take advantage of their unparalleled specificity in the field of chemical biology, bioconjugation methods can be used to produce synthetically modified proteins. As applications for protein-based materials are becoming ever-increasingly complex, there is a constant need for methodologies that can covalently modify protein substrates. More specifically, there is a requirement for reliable and chemoselective reactions that result in well-defined bioconjugates that are modified in a single location. We have developed a protein transamination strategy that uses *N*-methylpyridinium-4-carboxaldehyde benzenesulfonate salt (Rapoport's salt) to oxidize the N-terminal amine to a ketone or aldehyde functionality. We have identified high-yielding conditions for this reaction, such as N-terminal sequence and pH, and shown its applicability in the modification of several protein systems. In addition, we have used N-terminal protein modification methodologies in the synthesis of protein-DNA conjugates, towards the goal of developing a generalizable DNA-directed protein immobilization platform. Overall, the work presented herein expands the toolkit of methodologies available for building protein-based materials.

Dedicated to my family.
Mummy, Daddy, and Baji- together, I think we make a pretty great team.

Table of Contents

Acknowledgments

When it comes to graduate school, the one thing I have always felt is that it really is a team effort. I could not have gone through the last five years without the support of my mentors, colleagues, friends and family. It feels impossible to recognize the efforts of everyone involved in just a couple of pages, but I am truly grateful for the wonderful people that I have had the privilege of being surrounded by throughout my time at Berkeley.

The first person I want to thank is my undergraduate advisor, Amelia Fuller. I was lucky to have several amazing professors who helped me realize my scientific interests while studying at Santa Clara University, but it wasn't until I got to work with Amelia that I found my role model and mentor. Amelia, thank you for guiding me, inspiring me and being a great friend.

Next, I would like to thank my graduate advisor, Matt Francis. The Francis lab has been a wonderful place to work because of the supportive and positive environment that Matt fosters amongst his students. Whether it came to my research pursuits, career goals, or personal endeavors, Matt has been nothing but patient, encouraging and wise.

The Francis group has always comprised of several amazing scientists, and I am a better researcher for getting to work alongside many of them. When I joined the Francis lab, there were a lot of lab members nearing the end of their graduate experience. Michel was based in the sweatshop, my first home in the lab, and was ever ready to answer all my questions, which there were a lot of while I was learning the ins and outs of the lab. Gary was the keeper of my more permanent desk and workspace, and after four years, I have to agree with his assessment that the northeast corner of 733 is definitely the best seat of the lab. Then, there was Leah, whose mentorship was a guiding force in my ultimate research pursuits. Despite being extremely busy, she helped train me and made sure I could hit the ground running. I hope I was able to do justice to the N-terminal protein modification work! While I did not get to spend much time with Chris, Kanna, or Wesley, they all set great examples of successful Francis lab graduate students for the rest of us that were jstarting out.

The rest of the graduate students are the ones that I overlapped with the most, and hence, bothered the most as well. Dan was someone who always helped lighten the mood. He was a sharp scientist and a great hallway soccer player (with cardboard boxes, of course). Troy was always fun to hang out with and we were glad we got to see more of him when his lab got moved up to the 7th floor. I got to collaborate with Kristen on some peptide library work. I appreciated the great discussions we had and her enthusiasm for the project. She was also a lot of fun to hang out with outside of lab. Amy was the keeper of all DNA knowledge, and I was able to learn a lot from her before starting on the protein immobilization work. She was incredibly supportive and took out the time to train me even as she was finishing up in lab herself. Mike was a wonderful mentor while I rotated in the Groves lab. He was patient, enthusiastic and always ready to offer helpful discussions. I owe Allie a lot of gratitude for teaching me many of the lab skills I know. She always offered a helping hand and shared her advice any time she could. I always felt comfortable approaching her for assistance, and on top of all that, she is a fellow member of bachelor nation, which made our weekly "very important meetings" all the more enjoyable. Jeff was always willing to share his protein expression

expertise, and I am still thankful for all his help with my outside proposal. Kareem was one of the more quiet ones in lab, but his work with the photoactivated OC was determined and impressive. Katherine was my mentor during my Francis lab rotation. Working alongside her and getting to learn from her for ten weeks played a large part in why I ultimately joined the lab. She was knowledgeable, insightful and kind. She showed me the ropes of the lab and I am proud of the fact that I still express any protein with the same techniques and tricks she taught me. Any time something came up, minor or major, and I needed some words of wisdom, Stacy was the one I went to. Seeing her work ethic was motivating for me. She is a great scientist and a great friend. While I have missed our impromptu staring contests since she left, I hope that she will one day move back to the bay area so we can resume our bachelor viewings. I was able to work with Jelly on the Rapoport's salt transamination project, and she was always ready to discuss ideas and enthusiastic about planning experiments. I am glad that I was able collaborate with and learn from her. Jelly and Abby were the two resident members of 733 when I moved in and they both made it an incredibly welcoming and supportive room to work in. I know that I was not easy to deal with before my qualifying exam, but they both went out of their way to encourage me.

The next set of people I need to thank is my class. These are the people I got to walk across that stage with last month and have been around throughout my tenure at Berkeley. There's a lot to be said about getting to watch a scientist go from being a first year graduate student, to sitting through one another's qual practices, to now pursuing independent careers of our own. We supported and watched each other become the scientists we are today, and became great friends along the way. From the moment Jess joined the lab, she was a welcome addition to the group. She was easy to talk to and a well-versed researcher. She also helped initiate the protein immobilization work in the lab, a project I discussed with her frequently (when we weren't chatting about concerts or restaurants to go to). I am thrilled for her upcoming career in Hong Kong. Then there is Jenna, who became the lab mom in a way. She was always involved in group activities and supportive of other lab members despite being so busy with her own work and responsibilities. She is kindhearted and compassionate and I know she is going to be amazing at her new job in San Francisco. Jim and I came into the lab at the same time and we became the new N-terminal protein modification team. The work that he has accomplished with 2PCA has been incredible and I have always appreciated his ideas and input on my research. I am excited for him as he enters this new chapter of his life, with a baby on the way and a move to Boston. Richard will always be the physical chemist that did organic chemistry, but I think that's a testament to the great scientist that he is. Whether I needed him to proofread, offer some advice, listen to me unnecessarily panic about something, or grab coffee, I never hesitated to ask Richard. He has been a great lab mate and friend, and I know he will continue to do great things at Here. Jake was my go-to when I had any synthesis questions, and I am appreciative for all the helpful conversation about research that we had. I am even more appreciative of the fact that I could always walk over to 748 and take a break anytime I needed it. Outside of the lab, I have to say thank you to Jake for introducing me to Danielle. Hanging out with the two of them was always a highlight. I cannot wait to visit them in Oregon, and continue to hear about all of their adventures. Last, but definitely not the least, is Ioana. Ioana has sat ten feet away from me in lab for the last four years, and I could not have imagined a better person to be in that seat. Ioana is a wonderful scientist, a great lab mate and an incredibly caring person. Regardless of where our careers will take us, I have no doubt that she will remain a life-long friend.

The next group of students was those younger than me (at least in grad school years) and I hope that I have been able to help and support them the way that others guided me when I started out. Am is an adorable and kind-hearted soul, and she has already had a great amount of success in her research. Joel came into 733 with 3 other girls in the room, and he handled it like a champ, even when we decorated his desk in pink. He was clearly a welcome addition to the 733 family and has already accomplished so much. I have always enjoyed our conversations, both about research and other things, especially about the time he was in Bali. Matt Smith was my rotation student and while I like to jokingly take credit for training him, his drive and passion for research is all his own. He also makes lab a very fun place to be. Adam is someone I did not get to spend much time with but I do wish him luck with his future endeavors. Rachel came into the lab and took over the peptoid library project with an eagerness and excitement, and I look forward to seeing her successes in the next few years. Emily won the lab over from the second she joined with her bubbly attitude. Her willingness and desire to foster a community environment at work have been much appreciated. And Kristin, the newest member of 733, is hard-working, driven and compassionate. She's taken on a lot in a short amount of time but is navigating through it all with skill and class. Sarah, Marco and Jing only recently joined the lab, but I wish them the best of luck in their pursuits! It's been great getting to know them thus far.

When I joined the lab, the concept of post-docs was a mysterious one, with Michelle being the only one at the time. I best remember her for her commitment to the lab softball team and definitely made it a great experience for the whole team. Over the last five years, however, our post-doc count has grown, and each of them has brought their own unique skillset to the group. Henrik, who I got to work with briefly on some 3ER things, always had great ideas that he was always willing to share. Carson taught us all about the mechanical bond, and it's been great getting to see how he thinks about problems and wants to tackle them. Rafi always has a joyful demeanor, though we don't get to see him as much because he is often at the CARA facility. Ayo has taught us all that if you don't karaoke, you're doing something wrong. Christian has been great to hang out with and always makes great suggestions. I've particularly enjoyed chatting with him about DNA immobilization work. Since joining the lab, Ariel has quickly integrated herself into the group, both scientifically and socially. I'm exiting to see how some of the DNA-OC work that she's doing goes. It's been a pleasure getting to know Meera both inside and outside the lab. She's always fun to talk to and I've asked for her advice and perspective many times. While I have not collaborated with Adel on anything, I have always felt that we've been able to keep each other updated on research progress. He is incredibly intelligent and been someone I could always bounce ideas off of. He's never hesitated to offer assistance, and for that, I am thankful.

I worked with one undergraduate during my time at Berkeley, Cecilia. Cecilia came to work in the Francis lab to broaden her interests and dove right in from the moment she got here. She is resilient and ambitious and I know she is going to be immensely successful in graduate school.

Perhaps one of my proudest graduate school accomplishments is the Francerjai collaboration. This collaboration started about two years ago and allowed me to pursue a new project at a time when I was excited to explore my own ideas and really push myself as a researcher. The success of this project would not have been possible without my partner in crime, Tyler. Tyler is smart, motivated, and a great person to work with. While he tends to turn on the sass every so often, he

has become an honorary member of the Francis lab and of 733. I am extremely proud of the protein immobilization project and all that we have accomplished thus far. Looking at the work we've done, all I can really say is "isn't our story amazing?" I wish Tyler the absolute best of luck and know that he is going to thrive as he takes this project forward.

And lastly, my family. The whole Palla clan has been incredibly supportive throughout my academic endeavors. Looking back, I think that some of them realized the value of what I was doing even before I did. Mummy, Daddy, and Baji- you three have always encouraged me to pursue any path I wanted and have made sure that I had the resources and opportunity to do so. Your love, advice, and wisdom are priceless.

Chapter 1

Protein Modification Reactions and Their Applications

Abstract

The ability to create synthetically modified proteins has fundamental value in a range of fields. By carefully selecting the protein of interest as well as the synthetic moiety that is appended to it, the resultant hybrid molecules can be used as protein therapeutics, regenerable enzymes and immobilized proteins, to name a few. Given the diversity in applications that they can serve, it is crucial to have reliable methods for the synthesis of protein-based materials. Discussed herein are a variey of protein modification strategies, such as targeting native amino acid side chains, installing reactive handles at single, well-defined locations, and modification of the uniquely available N-terminus.

1.1. Utility of Protein Bioconjugation

Proteins have evolved in nature to be able to carry out a vast array of functions. Taking a specific protein that has its inherent structure and function and appending a synthetic molecule that has its own unique properties allows for the formation of a hybrid bioconjugate that can capitalize on the properties of both individual components. These types of synthetically modified proteins that have a dual functionality associated with them can serve to exploit biomolecular properties in a number of downstream applications.[1, 2] For example, protein-fluorophore conjugates can be used for cellular imaging,[3-6] the study of protein-protein interactions[7, 8] or light harvesting applications.[9, 10] Proteins modified with drugs or targeting groups can be used for therapeutic applications.[11–13] The pharmaceutical industry has been particularly interested in antibody-drug conjugates (ADCs), with multiple products clinically approved and several more currently in advanced trials.[14, 15] Additionally, polymers can be tethered onto proteins for water remediation,[16] and enzyme regeneration purposes.[17] Biocatalysts are continually being investigated for alternative biofuel generation, and rendering these enzymes recyclable will allow for the system to be more cost-competitive.[18]

One additional area that is of large interest is the immobilization of proteins.[19–21] The ability to anchor proteins onto supports in a generally applicable way is a powerful tool that has utility in a variety of applications. Protein microarrays have been used to analyze yeast protein phosphorylation, where an *in vitro* kinase assay and a chip modified with 4400 functional proteins were used to identify 1325 proteins involved in 4200 phosphorylation events.[22] Biosensors are desired for the detection of analytes, a good example of this being the antibody-antigen detection utilized in ELISA-like experimental set ups.[19] In one example of this, microcantilevers were functionalized with antibodies against prostate-specific antigen (PSA), a disease marker. Quantificaton on PSA concentration in both human serum albumin and human plasminogen was carried out using this platform.[23] Nature's catalysts, enzymes, are highly desirable targets in industrial processes. Their immobilization allows for preservation and reuse of these enzymes, rendering their large-scale use economical.[24, 25] In each of these cases, it is necessary that after immobilization, the active sites of the protein remain accessible. Given the diverse set of applications that are associated with synthetically immobilized proteins, the generation of well-characterized and stable protein bioconjugates that covalently attach to surfaces in high density while still retaining their higher order structure, and thus, their native protein function becomes imperative.[26] In light of this, there is an ever increasing need for new and efficient protein modification methodologies.

1.2. Modification of Native Amino Acid Sidechains

One long-standing approach to developing protein bioconjugates is to take advantage of the inherent reactivity of native amino acid side chains. Certain amino acid side chains that are nucleophilic in nature, such as cysteine, tyrosine, lysine and asparate/glutamate, can be reacted with their coupling partners, given that those side chains are solvent accessible on the surface of a protein. For example, the thiol side chain of cysteines can be alkylated with derivatized maleimide reagents.[27] Cysteines residues are the most rare of the 20 canonical amino acids, constituting about 1.9% of the amino acids on a proteins.[28] Their low abundance often makes them optimal targets for modification due to the control over number of modifications it offers, and genetic engineering can help to ensure that there is at least one free cysteine residue available. The aromatic tyrosine

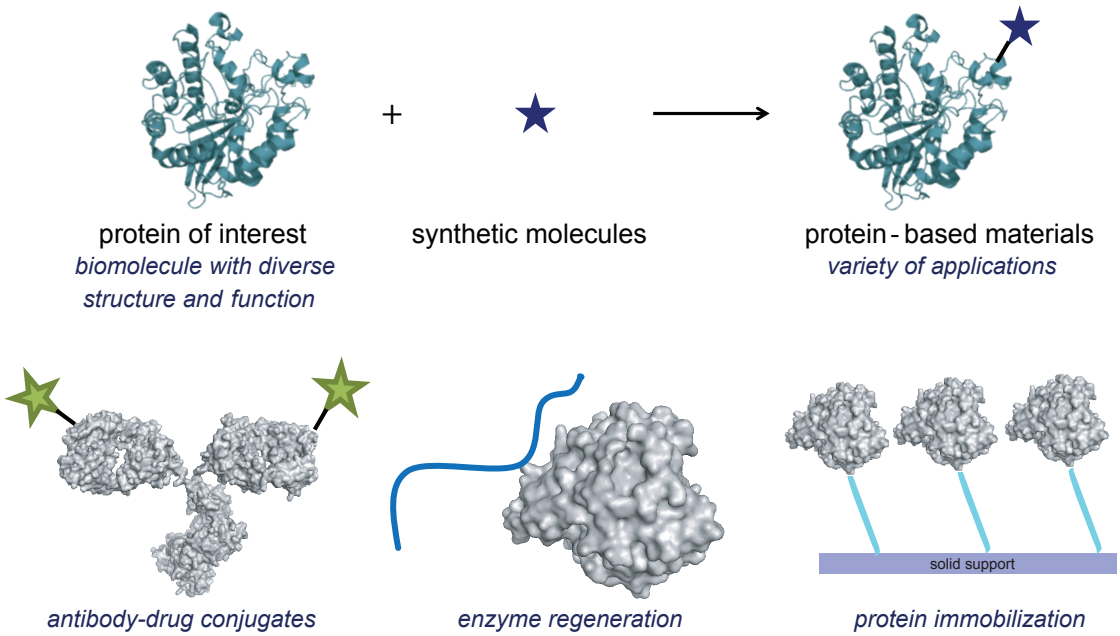


Figure 1.1. The attachment of synthetic groups to proteins is an important tool for the development of protein bioconjugates that can serve in a variety of applications. A few examples are the attachment of drugs to antibodies, recyclable polymers to enzyme, and tethers for the immobilization of proteins to solid surfaces.

residues (average abundance of 3.2%) can modified with diazonium salts.[28, 29] One additional modification methodology developed within the Francis group is the cerium ammonium nitrate mediated coupling of anisidine derivatives with tyrosine side chains.[30] The amines of lysine side chains (average abundance of 5.9%) can be coupled with *N*-hydroxysuccinimide (NHS) ester based reagents.[28, 31] Lysines can also be modified by the exposure of an aldehdye to the amine, which produces an imine that can be reduced with sodium cyanoborohydride in a reductive amination reaction.[32] Alternatively, the nucleophilic lysines will react with epoxide functionalities, but this can also lead to modification of free cysteine residues.[33] The acidic residues, aspartate and glutamate, constitute 11.6% of the amino acids on the surface of a protein, and can be modified with carbodiimides and amines to form amide bonds.[28, 34] Figure 1.2 shows the presence of these residues on the surfaces of two different proteins, the herceptin antibody (PBD: 1IGT) and the cellulase enzyme (PDB: 2ZUM). It can be seen that, while the chemistries to modify these residues are readily accessible, the distribution of multiple copies of these residues across the surface of the proteins and varying degrees of coupling efficiency offer limited control over both the numbers of modifications as well as the sites of the modifications that result. Furthermore, protein stability and activity may be affected if functionally and structurally important amino acid residues are targeted.[35] Some circumstances require the formation of heterogeneous products. As an example, in an effort to immobilize the sulfotransferase enzyme to a silicon surface, it was found that, when bound to epoxide coated or NHS-ester coated surfaces (through modification of lysine residues on the protein surface), enymatic activity of the immobililzed protein was significantly lower than its ordered counterpart (that was tethered to a His_6 tag at a single, defined site).[36] For cases such as this one, in order to be able to produce more homogenous and well-defined protein bioconjugates, there is a need for well-controlled, site-specific attachment chemistries. (Figure 1.2)

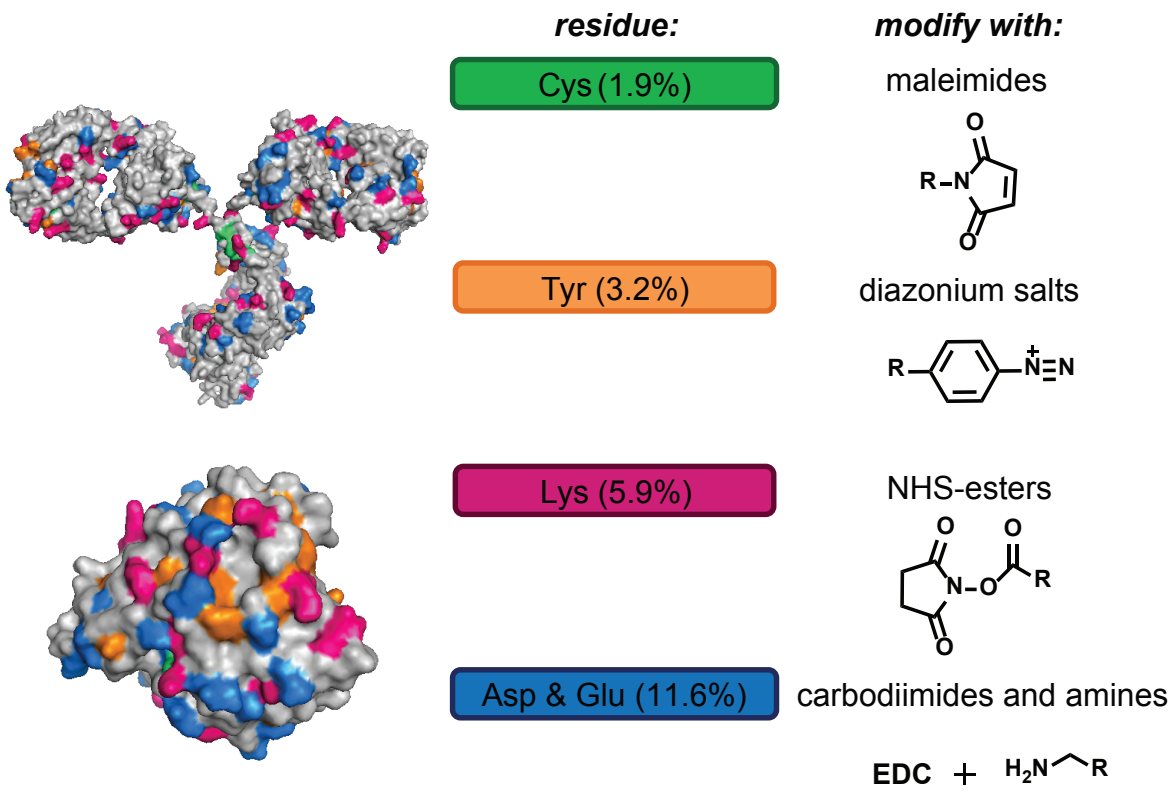


Figure 1.2. Native amino acids are present on the surface of proteins in varying abundances (percentrages listed). These amino acid side chains can be targeted for modification in various coupling reactions. Because of their varied distribution levels across a protein, targeting them results in a heterogeneous mixture of products. Proteins structures shown: Herceptin antibody (PDB: 1IGT) and cellulase enzyme (PDB: 2ZUM).

1.3. Site-specific Protein Bioconjugation Approaches

To modify a protein site-specifically, the reactive handle that the reaction targets must be present on the surface of a protein in a single copy. Additionally, the chemistry used to modify this single location must be unique to that functional group, i.e., bioorthogonal,[37] such that it does not modify any of the native functionalities present on the protein. Several reactions have been developed towards this goal. These involve incorporation of a reactive handle onto the surface of a protein, which is then modified with its coupling partner in a subsequent step. Listed in Figure 1.3 are a few of these types of reactions. The first is the copper catalyzed click chemistry between an azide and an alkyne, a biocompatible cycloaddition that can be carried out under mild reaction conditions.[38] Another example is the inverse Diels-Alder tetrazine ligation between tetrazine and cyclopropene, a alternative type of click chemistry (Figure 1.3a, b).[39]

Another class of site-specific reactions developed in the Francis group is the oxidative coupling reaction between anilines and *ortho*-aminophenols in the presence of an oxidizing agent. Two versions of this reaction have been developed: the first is mediated by sodium periodate[40] and the second is mediated by potassium ferricyanide[41] as the oxidant. In additon to being carried out in mild aqueous conditions, a particular advantage of these reactions is that they can be effected in low micromolar concentrations of both protein and aminophenol coupling partners (Figure 1.3c).

One reaction that has been widely studied and is largely used is the ligation between an aldedye or a ketone and a hydrazide or alkoxyamine.[42–45] The resulting hydrazone and oxime linkages formed are stable and allow for appendage of any alkoxyamine derivated molecule of choice. Because native amino acid side chains are nuclophilic, oximes form only at the uniquely electrophilic aldehydes and ketones that are introduced onto the surface of a protein (Figure 1.3d).

In some of the reactive handle and coupling partner pairs discussed in this section, the combination of which one is installed on the protein and which one is used to make the derivatized synthetic molecule can easily be reversed. In either case, the need for introduction of a reactive handle on the surface of a protein remains and there are a variety of ways in which this can be achieved. These include chemical, enzymatic and unnatural amino acid incorporation methods.[37]

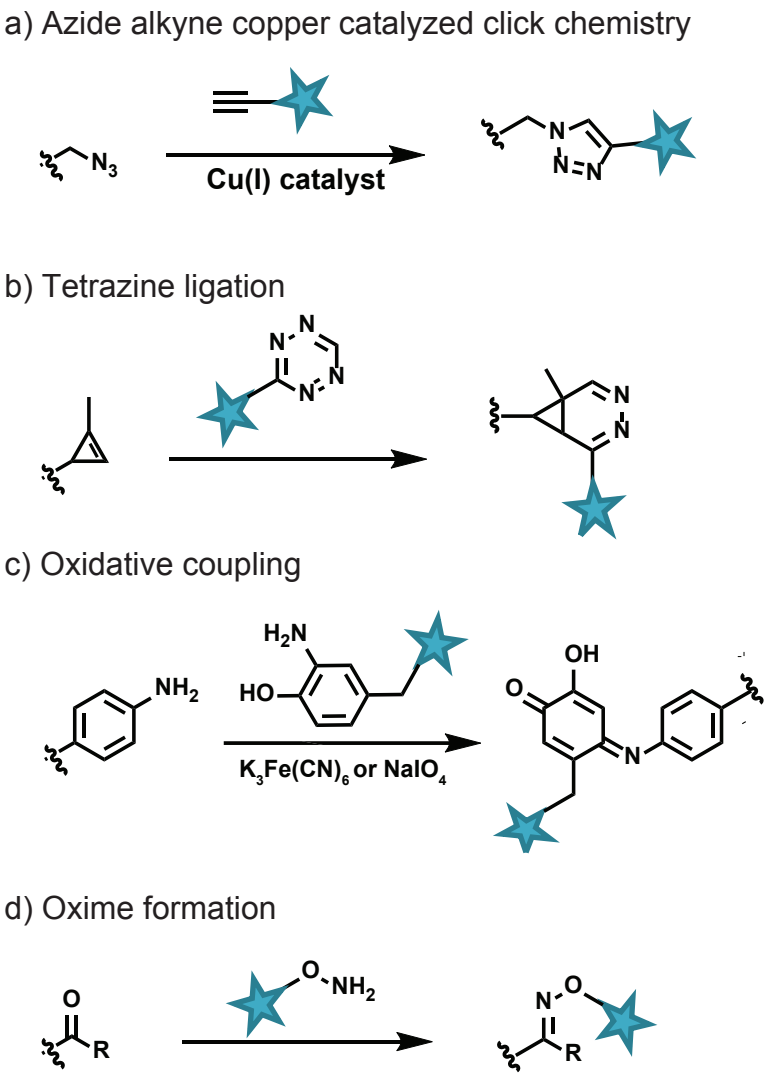


Figure 1.3. Site-specific bioconjugation can be carried out by the introduction of a bioorthogonal handle onto the surface of a protein. Once introduced, the new functional group can be modified with its coupling partner to produce desired protein bioconjugates.

1.4. Targeting the N-terminus for Site-Specific Protein Modification

One site that is present on each protein (monomer) at a single, distinct location is the N-terminus. Targeting the N-terminus would result in proteins only being modified once, leading to predictable sites of modification and minimal changes to the proteins inherent structure and function (Figure 1.4).

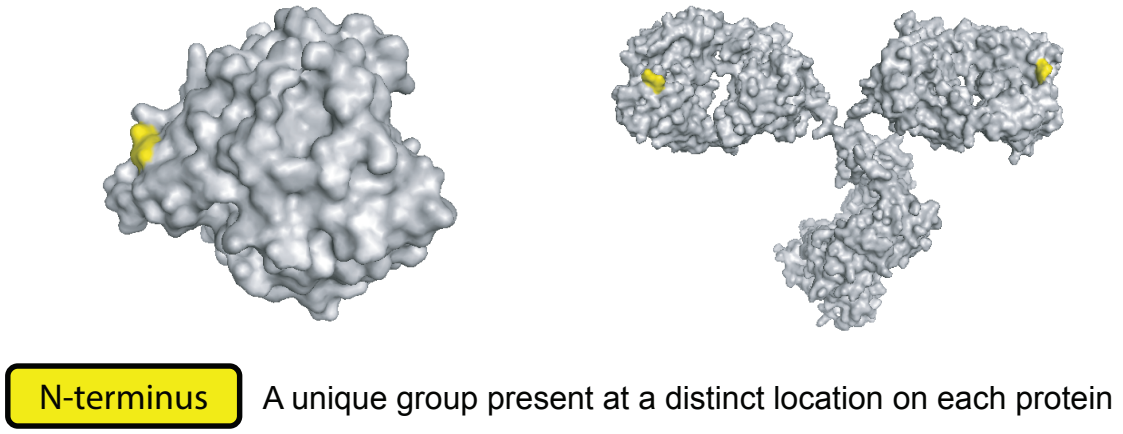


Figure 1.4. The N-terminus is a unique functional group present on every protein in a defined location that appears only a single time. This makes it an optimal native functionality to target for site-selective protein modification. (a) Crystal structure of a monomeric cellulase enzyme (PBD: 2ZUM), having one solvent exposed N-terminus shown in yellow. (b) Crystal structure of the herceptin antibody (PDB: 1IGT), made up of 2 heavy chains and 2 light chains, allowing for 4 potential sites of modification per antibody. Shown in yellow are the solvent accesible N-termini. In the case of the antibody, only the N-termini of the heavy chains can be observed in this orientation of the crystal structure.

A few approaches to modify the N-terminus of a protein are illustrated in Figure 1.5. The sodium periodate mediated cleavage of serine and threonine residues at the N-terminus results in the production of an aldehyde moiety at the N-terminus, which can then be modified in a second step with an alkoxyaine reagent.[46] Native chemical ligation can be used to modify cysteine residues with thioester reagents.[47] When tryptophan residues are present at the N-terminal position, a Pictet-Spengler reaction be used with an aldehyde reagent.[48] As effective as they are, these methods are limited by the nature of the N-terminal amino acid. Because they are restricted to the amino acid type, it becomes worthwhile to expand the types of N-terminal modification reactions available in order to broaden the amino acids they are compatible with. It becomes even more significant to develop a platform that is generalizable to the N-terminal amine, with less stringency being placed on the amino acid side chain that is present.

Within the Francis group, several N-terminal modification approaches have been developed that add to the toolbox of available reactions. One such reaction is the recently reported one-step N-terminal modification of native proteins with 2-pyridinecarboxyaldehyde (2PCA) derivatives (Figure 1.6a). This reaction involves the formation of a cyclic imidazolidinone product through the addition of an immediately adjacent amide N-H group to an imine intermediate. Importantly, this cyclization cannot occur when 2PCA imines are formed with lysine side chain amines, confining the reaction to the N-terminal position.[49]

Another reaction is the pyridoxal 5'-phosphate[50–54] or *N*-methylpyridinium-4-carboxaldehdye (Rapoport's salt)[55,56] mediated transamination reaction, were the N-terminal amine is chemoselectively oxidized to a ketone functionality that can then be modified in a subseqent

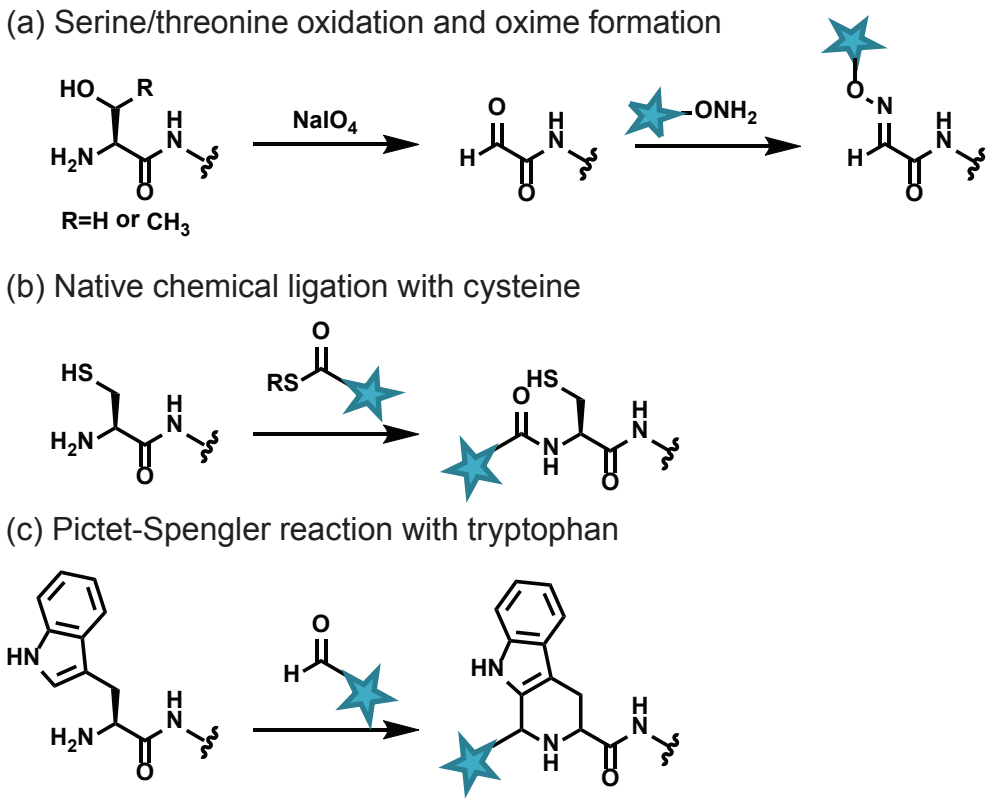


Figure 1.5. Targeting the N-terminus for site-specific protein modification. (a) Oxidation of serine or threonine with sodium periodate. (b) Native chemical ligation between cysteines and thioesters. (c) Pictet-Spengler reaction between tryptophans and aldehdyes. Each of these approaches is dependent on the nature of the amino acid present at the N-terminus.

oxime forming step (Figure 1.6b). These reactions are able to target the N-terminal amine and not the lysine amines due to their differences in pKa values, and hence their protonation states. Of the two, Rapoport's salt[55,57] is the more recently identifed one, and our progress in its development and optimization as a protein modification strategy, using both the herceptin antibody and cellulase enzyme as model protein systems, is discussed in subsequent chapters.

Lastly, the Francis group has published on a potassium ferricyanide mediate oxidative coupling reaction between the N-terminal amine of a peptide or protein and an *ortho*-aminophenol coupling partner (Figure 1.6c).[58] This reaction has been shown to be compatible with all N-terminal residues, but has a particularly efficient reactivity with proline residues. Additionally, only a few equivalents of the *ortho*-aminophenol relative to the protein are required for this reaction occur, making it particularly valuable for modification of proteins with synthetic molecules that are costly or may suffer from solubility limitations. One such synthetic molecule is DNA, where solubility levels are indirectly related to the lengths of the strands, and usage of large concentrations would prove costly and uneconomical. Protein-DNA conjugates have been used in DNA directed immobilization applications,[59] and our contributions to this field will be discussed in this work.

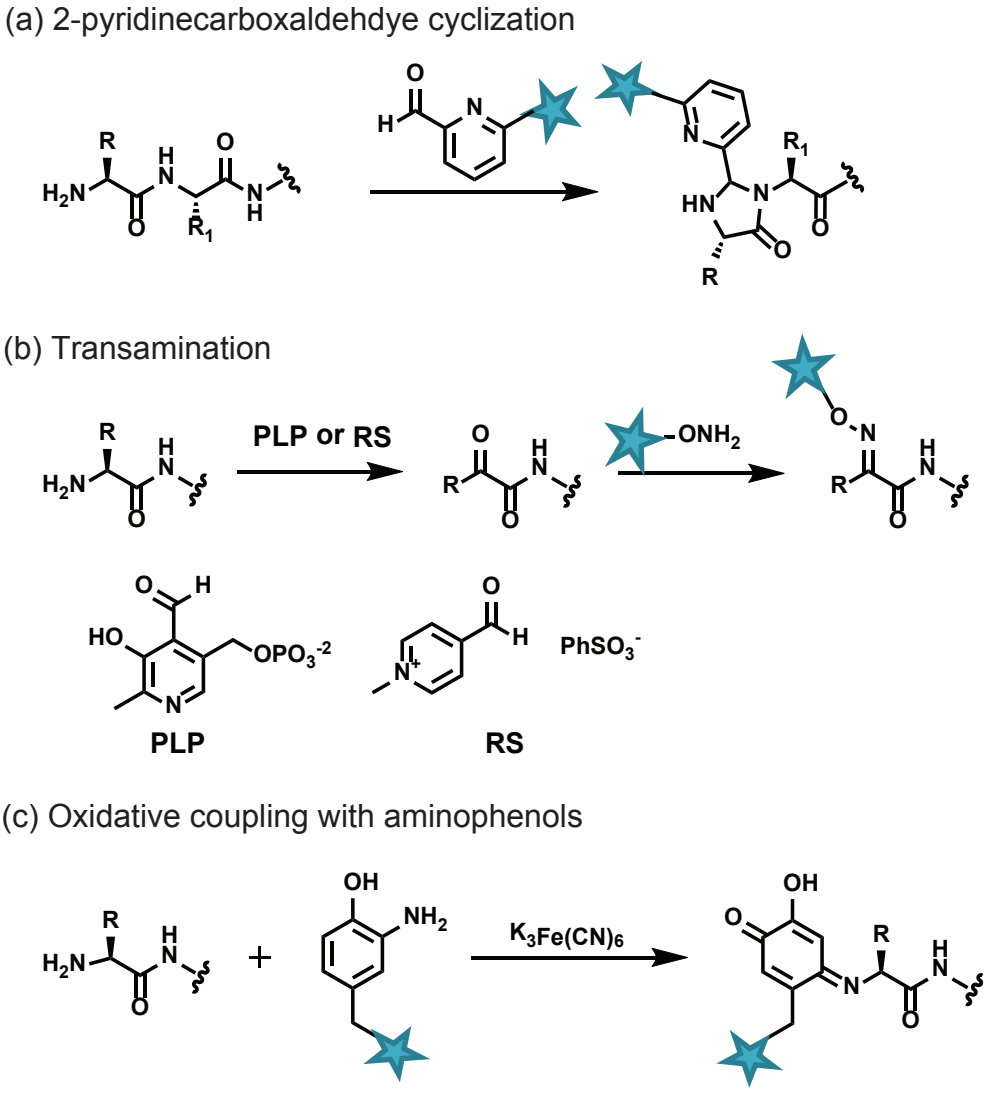


Figure 1.6. N-terminal modification strategies developed in the Francis group. (a) One-step N-terminal modification of native proteins with 2-pyridinecarboxyaldehyde (2PCA) derivatives. This reaction involves the formation of a cyclic imidazolidinone product through the addition of an immediately adjacent amide N-H group to an imine intermediate. (b) N-terminal transamination mediated by pyridoxal 5'-phosphate (PLP) or N-methylpyridinium-4-carboxaldehyde benzene sulfonate salt (Rapoport's salt, RS). After transamination, the newly introduced aldehdye or ketone can be modified with an alkoxyamine reagent of choice. (c) Potassium ferricyanide mediated oxidative coupling between ortho-aminophenols and protein N-termini.

1.5. References

1. Hermanson, G. T., Bioconjugate Techniques, 2nd ed.; Academic Press: San Diego, **2008**.

2. Witus, L. S.; Francis, M. B., Using synthetically modified proteins to make new materials. *Acc. Chem. Res*. **2011**, *44*, 774–783.

3. 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.

4. O'Hare, H. M.; Johnsson, K.; Gautier, A., Chemical probes shed light on protein function. *Curr. Opin. Struct. Biol.* **2007**, *17*, 488–494.

5. Johnsson, N.; Johnsson, K., Chemical tools for biomolecular imaging. *ACS Chem. Biol.* **2007**, *2*, 31–38.

6. Marks, K. M.; Nolan, G. P., Chemical labeling strategies for cell biology. *Nat. Meth.* ***2006***, *3*, 591–596.

7. Joo, C.; Balci, H.; Ishitsuka, Y.; Buranachai, C.; Ha, T., Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **2008**, *77*, 51–76.

8. Cecconi, C.; Shank, E. A.; Bustamante, C.; Marqusee, S., Direct observation of the three-state folding of a single protein molecule. *Science* **2005**, *309*, 2057–2060.

9. 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.

10. 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.

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. U.S.A.* **2009**, *106*, 15231–15236.

12. Zalipsky, S., Functionalized poly(ethylene glycol) for preparation of biologically relevant conjugates. *Bioconjugate Chem.* **1995**, *6*, 150–165.

13. 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 PEGylated human interferon-β-1a with improved pharmacokinetic properties and *in vivo* efficacy in a melanoma angiogenesis model. *Bioconjugate Chem.* **2005**, *17*, 179–188.

14. Wu, A. M.; Senter, P. D., Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotech.* **2005**, *23*, 1137–1146.

15. 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.

16. 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.

17. Mackenzie, K. J.; Francis, M. B., Recyclable thermoresponsive polymer–cellulase bioconjugates for biomass depolymerization. *J. Am. Chem. Soc.* **2013**, *134*, 293–300.

18. Hood, E. E.; Bauer, L., Catalysts and enzymes in biofuel production. *Biofuels Digest* **2016**.

19. Wong, L. S.; Khan, F.; Micklefield, J., Selective covalent protein immobilization: Strategies and applications. *Chem. Rev.* **2009**, *109*, 4025–2053.

20. Scouten, W. H.; Luong, J. H. T.; Brown, R. S., Enzyme or protein immobilization techniques for applications in biosensor design. *Trends in Biotechnology* **1995**, *13*, 178–185

21. Bickerstaff, G. F., Protein immobilization. Fundamentals and applications: (Bioprocess Technology Series/14) Taylor, R. F., *Journal of Pharmacy and Pharmacology* **1992**, *44*, 71.

22. Zhu, H.; Snyder, M., Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* **2001**, *5*, 40–45.

23. Wu, G.; Datar, R. H.; Hansen, K. M.; Thundat, T.; Cote, R. J.; Majumdar, A., Bioassay of prostate-specific antigen (PSA) using microcantilevers. *Nat. Biotechnol.* **2001**, *19*, 856–860.

24. Khan, A. A.; Alzohairy, M. A., Recent advances and applications of immobilized enzyme technologies: A review. *Res. J. Bio. Sci.* **2010**, *5*, 565–575.

25. DiCosimo, R.; McAuliffe, J.; Poulose, A. J.; Bohlmann, G.; Industrial use of immobilized enzymes. *Chem. Soc. Rev.* **2013**, *42*, 6437–6474.

26. Brena, B.; González-Pombo, P.; Batista-Viera, F., Immobilization of enzymes: a literature survey. *Methods Mol. Biol.* **2013**, *1051*, 15–31.

27. Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R., Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960–1965.

28. Tilley, S. D.; Joshi, N. S.; and Francis, M. B., Proteins: Chemistry and Chemical Reactivity. Wiley Encyclopedia of Chemical Biology. **2008**, 1–16.

29. Jones, M. W.; Mantovani, G.; Blindauer, C. A.; Ryan, S. M.; Wang, X.; Brayden, D. J.; Haddleton, D. M., Direct peptide bioconjugation/PEGylation at tyrosine with linear and branched polymeric diazonium salts. *J. Am. Chem. Soc.* **2012,** 134, 7406–7413.

30. 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.

31. Kalkhof, S.; Sinz, A., Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **2008**, *392*, 305–312.

32. Roy, R.; Katzenellenbogen, E.; Jennings, H. J., Chemistry for peptide and protein PEGylation. Improved procedures for the conjugation of oligosaccharides to protein by reductive amination. *Canadian Journal of Biochemistry and Cell Biology* **1984**, *62*, 270–275.

33. Chen, G.; Heim, A.; Riether, D.; Yee, D.; Milgrom, Y.; Gawinowicz, M. A.; Sames, D., Reactivity of functional groups on the protein surface: development of epoxide probes for protein labeling. *J. Am. Chem. Soc.* **2003**, *125*, 8130–8133.

34. Emmanuel Baslé, E.; Joubert, N.; Pucheault, M., Protein chemical modification on endogenous amino acids. *Chemistry and Biology*, **2010**, *17*, 213–227.

35. Doolittle, R. F., Redundancies in protein sequences. In: Prediction of protein structure and the

principles of protein conformation. Fasman GD, ed. **1989**. Plenum Press, New York.

36. Cha, T.; Guo, A.; Zhu, X. Y., Enzymatic activity on a chip: the critical role of protein orientation. *Proteomics* **2005**, *5*, 416–419.

37. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974–6998.

38. 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.

39. Yang, J.; Šečkutė, J.; Cole, C. M.; Devaraj, N. K.; Live-cell imaging of cyclopropene tags with fluorogenic tetrazine cycloadditions. *Angew. Chem. Int. Ed.* **2012**, *51*, 7476–7479.

40. Behrens, C. R.; Hooker, J. M., Obermeyer, A. C.; Romanini, D. W.; Katz, E. M.; Francis, M. B., Rapid chemoselective bioconjugation through the oxidative coupling of anilines and aminophenols. *J. Am. Chem. Soc.* **2011**, *133*, 16398–16401.

41. Obermeyer, A. C.; Jarman, J. B.; Netirojjanakul, C.; Francis, M. B., Mild bioconjugation through the oxidative coupling of ortho-aminophenols and anilines with ferricyanide. *Angew. Chem. Int. Ed.* **2014**, *53*, 1057–1061.

42. Dirksen, A.; Dawson, P. E., Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconj. Chem.* **2008**, *19*, 2543–2548.

43. Kalia, J.; Raines, R. T., Hydrolytic stability of hydrazones and oximes. *Angew. Chem. Int. Ed.* **2008**, *47*, 7523–7526.

44. Jencks, W. P., Studies on the mechanism of oxime and semicarbazone formation. *J. Am. Chem. Soc.* **1959**, *81*, 475–481.

45. Dirksen, A.; Hackeng, T. M.; Dawson, P. E., Nucleophilic catalysis of oxime ligation. *Angew. Chem. Int. Ed.* **2006**, *118*, 7743–7746.

46. 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.

47. Dawson, P. E.; Kent, S. B., Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

48. 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.

49. MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B., One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes, *Nat. Chem. Biol.* **2015**, *11*, 326–331.

50. Witus, L. S.; Francis, M. B., 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.

51. 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.

52. Scheck, R. A.; Francis, M. B., Regioselective labeling of antibodies through N-terminal transamination. *ACS Chem. Biol.* **2007**, *2*, 247–251.

53. 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.

54. 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.

55. Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C. H.; Iavarone, A. T.; Francis, M. B., Site-specific protein transamination using N-methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **2013**, *135*, 17223–17229.

56. Palla, K. S.; Witus, L. S.; Mackenzie, K. J.; Netirojjanakul, C.; Francis, M. B., Optimization and expansion of a site selective *N*-methylpyridinium-4-carboxaldehyde mediated transamination for bacterially expressed proteins. *J. Am. Chem. Soc.* **2015**, *137*, 1123–1129.

57. Buckley, T. F.; Rapoport, H., Mild and simple biomimetic conversion of amines to carbonyl compounds. *J. Am. Chem. Soc.* **1982**, *104*, 4446–4450.

58. Obermeyer, A.C.; Jarman, J.B.; Francis, M.B., *N*-Terminal modification of proteins with *o*-aminophenols. *J. Am. Chem. Soc.* **2014**, *136*, 9572–9579.

59. Niemeyer, C. M.; Boldt, L; Ceyhan, B; Blohm, D., DNA-directed immobilization: Efficient, reversible, and site-selective surface binding of proteins by means of covalent DNA- streptavidin conjugates. *Anal. Biochem.* **1999**, *268*, 54–63.

Chapter 2

Development and Optimization of a Site Selective *N*-methylpyridinium-4-carboxaldehyde Mediated Transamination for proteins

Abstract

Site-selective bioconjugation methods are valuable due to their ability to confer new properties to proteins by the chemical attachment of specific functional groups. Well-defined bioconjugates obtained through these methods have found utility for the study of protein function and the creation of protein-based materials. Here, we present a protein modification strategy to modify the N-terminus of peptides and proteins using *N*-methylpyridinium-4-carboxaldehyde benzene sulfonate (Rapoport's salt, RS) as a transamination reagent, which oxidizes the N-terminal amino group to provide a uniquely reactive aldehyde or ketone. This functional handle can subsequently be modified with an alkoxyamine reagent of choice. To expand the scope of this reaction, we have used a combinatorial peptide library screening platform as a method to explore new transamination reagents while simultaneously identifying their optimal N-terminal sequences. RS was identified as a highly effective transamination reagent when paired with glutamate-terminal peptides and proteins. Using a known therapeutic antibody, herceptin, it was demonstrated that RS can be used to modify the heavy chains of the wild type antibody, or both the heavy and the light chains after N-terminal sequence mutation to add glutamate residues.

One limitation that this finding posed was that for proteins of interest that are recombinantly expressed in *E. coli*, the expression of a glutamate-terminal protein is rendered difficult because the N-terminal methionine derived from the start codon is not cleaved when Glu is in the second position. In an effort to expand on the utility of RS mediated transamination, we describe a way to overcome this difficulty via the insertion of a Factor Xa proteolytic cleavage site to acquire the optimal glutamate residue at the N-terminus. Additionally, we present studies on alternative high yielding sequences containing N-terminal residues that can be expressed directly. We have used site-directed mutagenesis to validate these findings on a model cellulase enzyme, an endoglucanase from the thermophilic *Pyrococcus horikoshii*. Activity assays performed with these mutants show that RS transamination and subsequent modification with alkoxyamines have no negative impact on cellulolytic ability.

Portions of the work described in this chapter have been reported in separate publications.[1, 2]

2.1. Introduction

The attachment of synthetic molecules to proteins with desired structure and function allows for the formation of hybrid materials that can capitalize on the properties of both components. The utility of these bioconjugates has been realized in a variety of contexts, such as cellular imaging,[3] light harvesting,[4] drug delivery,[5] water remediation[6] and tissue engineering.[7] Traditional methods used in protein modification target the side chains of native amino acid residues, such as lysines and cysteines.[8-10] However, depending on the relative abundance of these residues on the solvent exposed protein surface, it can be difficult to target them in a controlled and site-specific manner while maintaining native protein function. As a result, many newer techniques for protein modification target uniquely reactive sites, such as C-terminal thioesters,[11,12] N-terminal groups,[13-16] artificial amino acids,[17] and specific recognition sequences for enzymatic ligations.[18-20] Each of these new methods has expanded the scope of bioconjugates that can be accessed, and the applications of well-defined bioconjugates are proceeding apace.

We have previously reported a site-specific transamination reaction mediated by pyridoxal 5'-phosphate (PLP, **1a**), which chemoselectively oxidizes the N-terminal amine of a polypeptide or protein substrate to afford a ketone or an aldehyde group.[14,21] This carbonyl functionality introduced by this reaction are not naturally occurring functionalities in proteins, and therefore can be used as a handle for appending a synthetic group of choice through oxime or hydrazone formation[22,23] (Figure 2.1). 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 acid in the N-terminal position.[24] Given this situation, we previously developed a combinatorial peptide library screening platform to identify highly reactive sequences towards PLP-mediated transamination, leading to the identification of Ala-Lys N-terminal motifs.[25] In the present work, this new bioconjugation development tool was used as a way to identify a new protein transamination reagent, *N*-methylpyridinium-4-carboxaldehyde benzenesulfonate salt (RS,[26] **1b**), while simultaneously revealing glutamate-rich sequences as particularly reactive substrates for this reagent. This finding diversifies the bioconjugation toolbox because of its sequence complementarity

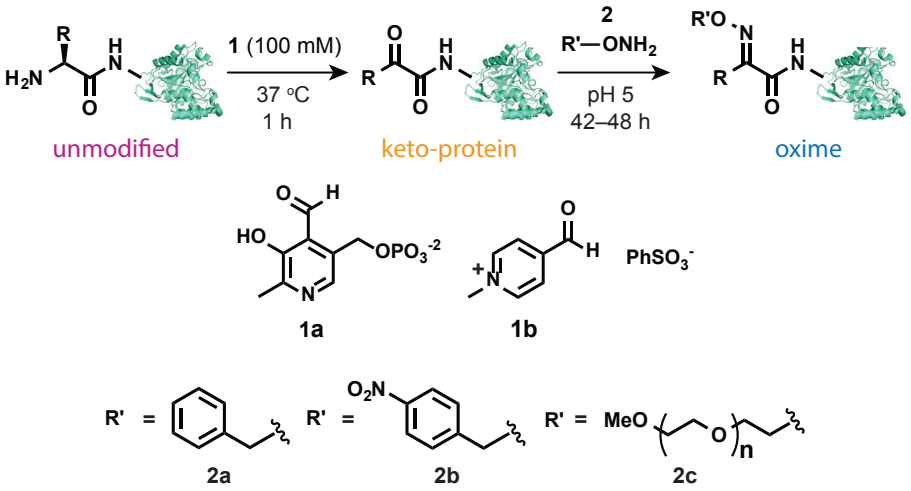


Figure 2.1. Using pyridoxal-5'-phosphate (PLP, **1a**) or Rapoport's salt (RS, **1b**), site-specific protein modification can be achieved at the N-terminus. PLP and RS oxidize the N-terminal amine into a ketone or an aldehyde group. This unique carbonyl functionality can subsequently form a stable oxime linkage with a synthetic alkoxyamine reagent. A few of the alkoxyamines used in this study are benzylalkoxyamine (**2a**), nitrobenzylalkoxyamine (**2b**) and PEG_{5kDa} alkoxyamine (**2c**).

to PLP and renders this approach particularly amenable to antibody substrates since many human IgG1 isotypes, which are promising therapuetics, contain at least one glutamate-terminal chain.[27, 28, 29] The stoichiometry of drug attachment has been identified as a critical parameter for ADC efficacy, with two to four attachments reported as being optimal in at least some cases.[30,31] Since antibodies have four N-termini (two identical heavy chains and two identical light chain units), we hypothesized that the sequence dependence of this reaction could give control over the total number of attachments. Antibodies such as anti-HER2 have one set of glutamate-terminal chains, for up to two attachment sites, but by mutating the antibody terminal sequences, the modification of both sets of chains also proved possible, enabling the attachment of up to four groups. The attachment of small molecules to the N-termini via oxime formation did not disrupt the antigen binding ability, as determined by flow cytometry. Overall, these findings establish RS as a facile and readily scalable method to obtain antibody conjugates in good yield and with control over the number of attached groups.

Using RS, the the site-selective modification of the heavy chains of anti-HER2 human IgG1 (which have N-terminal glutamate residues) could be achieved with significant efficiency. However, many proteins of interest are recombinantly expressed in bacterial cells that do not readily allow glutamate residues to be expressed at the N-termini. This is because the N-terminal methionine residues that arise from the start codons are not cleaved post-translationally when the second residue is large[24] (in contrast, the methionine residues are always removed in eukaryotic organisms). Keeping these expression parameters in mind, we sought to expand the scope of RS-mediated protein transamination to bacterially expressed proteins. We developed two different strategies for obtaining good yields of N-terminal bioconjugates through bacterial protein expression, followed by RS-mediated transamination. In the first strategy, we inserted a four-residue proteolytic site immediately before the desired N-terminal amino acid. Following expression of this construct in *E. coli*, the protein was cleaved with the protease Factor Xa to obtain the desired EES-terminal protein. Optimization of the RS transamination was then carried out using this model system. In the second strategy, we have turned to recently identified higher pH reaction conditions that increase the reactivity of non-glutamate residues. Taken together, the greater diversity of compatible N-termini and the expanded range of pH values at which this reaction can be performed make this RS-mediated transamination applicable to a significantly larger spectrum of proteins than was previously possible.

2.2. Results and Discussion

Although we have found PLP to be an effective transamination reagent in a number of different contexts, we found its reactivity can suffer from batch-to-batch variability. In addition, the preference for positive charges near the N-terminal positions to obtain optimal activity may be difficult to achieve in some cases.[21] We thus sought alternative aldehyde reagents that possessed similar reactivity, but might be more applicable to large scale protein modification reactions.

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 a uniquely effective reagent.[16] However, this screen was carried out using angiotensin as a single substrate. Given our later findings in terms of the dependence of PLP-mediated transamination on the N-terminal sequence, it seemed likely that this initial screen could have failed to identify other, more practical aldehyde reagents that could achieve similarly high levels of transamination,

but with other sequence preferences. Such combinations would most readily be identified using combinatorial peptide library screening platforms that can evaluate the transamination ability of candidate reagents against all combinations of N-terminal sequences simultaneously (Figure 2.2a).

2.2.1. Synthesis of the Combinatorial Peptide Library and Assay Calibration for Transamination Screening.

A one-bead-one-sequence combinatorial peptide library was first prepared using split-and-pool synthesis.[33] The capping of a small portion of the growing peptide chains during the synthesis of the variable positions provided a truncation ladder for sequencing beads of interest by intact ion mass spectrometry.[25] Because we sought to identify short motifs that could easily be incorporated into proteins, the diversity in the library was limited to the three N-terminal positions. The variable positions included each of the 20 natural amino acids, resulting in an 8,000 member library.

To identify the library sequences that had been transaminated to form a keto-peptide, a colorful, Disperse Red-based alkoxyamine reagent (**2d**, DispRed-ONH_2)[25] was used in a subsequent oxime-formation reaction. The resulting beads possessing the desired oxime product thus took on a bright red color. In our previous work, however, it was found that the color on the beads saturated at ~30% oxime yield, making it difficult to identify the highest yielding reaction combinations.[25] For the next generation of this screening technique, we recalibrated the detection scheme such that a visually red bead would correspond only to a high oxime yield (Figure 2.2b). To do this, reaction conditions were established that would result in beads with a low and high level of modification using a known transamination reagent (PLP) and peptide sequence. The peptide used was AOTWSNAG (where O stands for ornithine), which was found to have similar reactivity

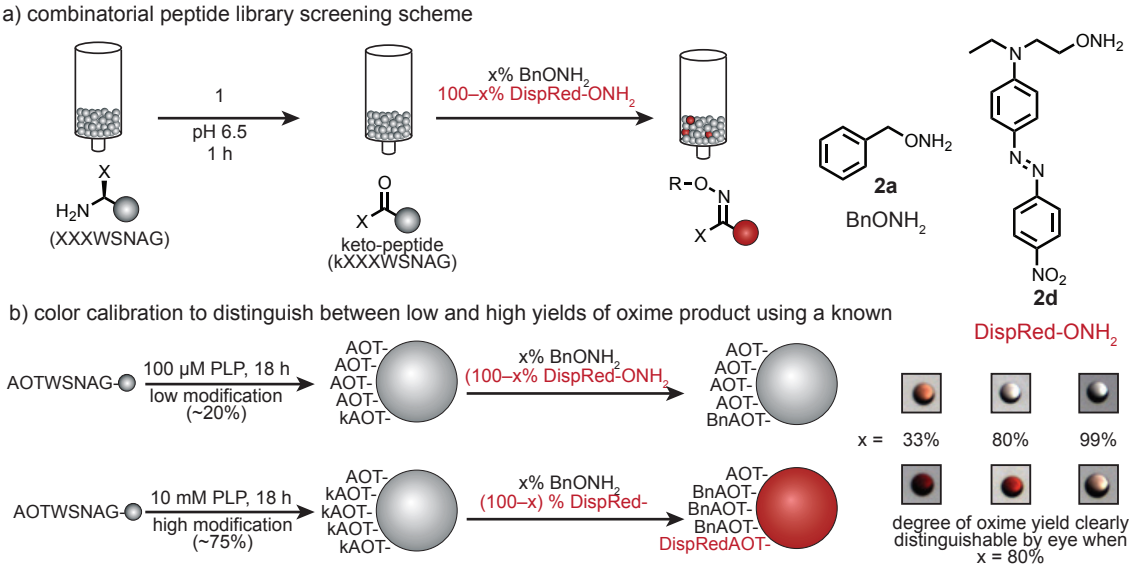


Figure 2.2. (a) A one-bead-one-sequence combinatorial peptide library in which the three N-terminal residues were varied was used to screen potential transamination reagents (1) against all possible N-terminal sequences simultaneously. After transamination, the keto-group-containing peptides were identified through the covalent attachment of a visible dye (DispRed-ONH_2, **2d**) through oxime formation. (b) To adjust the sensitivity of the colorimetric detection method, a colorless alkoxyamine ($BnONH_2$, **2a**) was combined with DispRed-ONH_2 in varying ratios. A peptide sequence (AOT-terminal, with similar reactivity to AKT-termini) and a transamination reagent with known reactivity (PLP, 1a) were used to generate 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 80:20 ratio of **2d** to **2a**.

to a previously reported AKT sequence.[25] With this sequence, incubation with 100 μM 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-ONH_2 (**2d**) was combined with different ratios of a colorless alkoxyamine (benzylalkoxyamine, $BnONH_2$, **2a**) to find a ratio that resulted in the greatest visual contrast between low and high oxime yields. With a low stringency ratio (a low proportion of $BnONH_2$, 33%), beads with both high and low levels of modification resulted in a red color, and with a high stringency detection ratio (99% $BnONH_2$) neither case had a red color. An intermediate ratio was found (80% $BnONH_2$), wherein only the highly transaminated beads were evident. Thus, this ratio of alkoxyamines was used for subsequent library screening to identify transamination conditions that resulted in maximal yields.

2.2.2. Identification of a New Transamination Reagent for Glutamate-terminal Sequences.

Portions of the library were incubated with 10 mM concentrations of a series of aldehyde reagents at pH 6.5. Substrates screened included salicylaldehyde, 2-hydroxy-5-nitrobenzaldehyde, benzaldehyde, glyoxylic acid, and *N*-methylpyridinium-4-carboxaldehyde (Rapoport's salt, **1b**). In the case of Rapoport's salt (RS), it was found that a large proportion of the beads turned red using the low stringency assay conditions, indicating that most sequences had at least some degree of reactivity (Figure 2.3a). This reagent is particularly promising for transamination reactions because it is inexpensive (particularly if synthesized in-house)[26] and the reagent can be purified before use via recrystallization from acetonitrile. Relative to PLP, this reagent is also more amenable to structural modifications, which could be explored as ways to increase reactivity in future studies. RS has been reported previously as a transamination reagent for small molecules in organic solution[29,30] suggesting that it could serve as a protein transamination reagent as well. The other aldehydes screened had lower activity and some suffered from limited aqueous solubility. In some cases initial leads were followed, but found to lack reactivity upon verification.

Since most sequences showed some reactivity towards RS using the low stringency assay conditions, the few remaining colorless beads were selected to see which sequences showed no transamination reactivity. Proline in the N-terminal position was identified as the common motif among the colorless beads. 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,[24] although the resulting product has not yet been fully characterized and was not observed with RS.

To identify the optimal N-terminal sequences for RS-mediated transamination, higher stringency detections of oxime yield were applied to the library (Figure 2.3b). After reaction with 10 mM RS for 1 h in pH 6.5 phosphate buffer followed by oxime formation with a 80:20 ratio of **2d** to **2a**, a small number of red beads were observed. These beads were collected and sequenced, and the common pattern identified was multiple glutamate residues at the N-terminus (the full list of sequences appears in Figure 2.4). The screen was next repeated using a higher ratio of the colorless alkoxyamine $BnONH_2$ (95%) to provide an even more stringent filter for high oxime yield (Figure 2.3c). 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.

	stringency:	x: (% $BnONH_2$)	beads selected:	sequence motif:
a)	low	0	colorless	PXX
b)	intermediate	80	red	EEX
c)	high	95	red	EEX

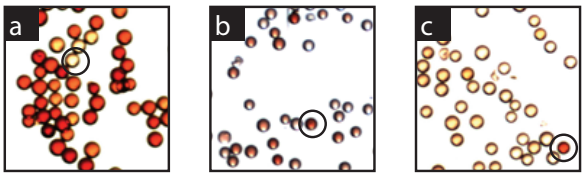


Figure 2.3. Library screening to identify reactive sequences towards Rapoport's salt-mediated transamination. The library was treated with 10 mM RS for 1 h at pH 6.5, followed by oxime formation with the specified ratios of $BnONH_2$ (x%) and DispRed-ONH_2 (100–x%). The active sequences were identified by selecting and sequencing the red beads in the library. When a low stringency screen was applied (a) many beads had a red color, indicating that many sequences form some degree of oxime product. In this case, the beads that remained colorless were selected, and proline was identified as an N-terminal residue that prevented reactivity. To identify sequences that led to high levels of oxime product, more stringent screens were used that could distinguish low from high levels of oxime yield (b, c).

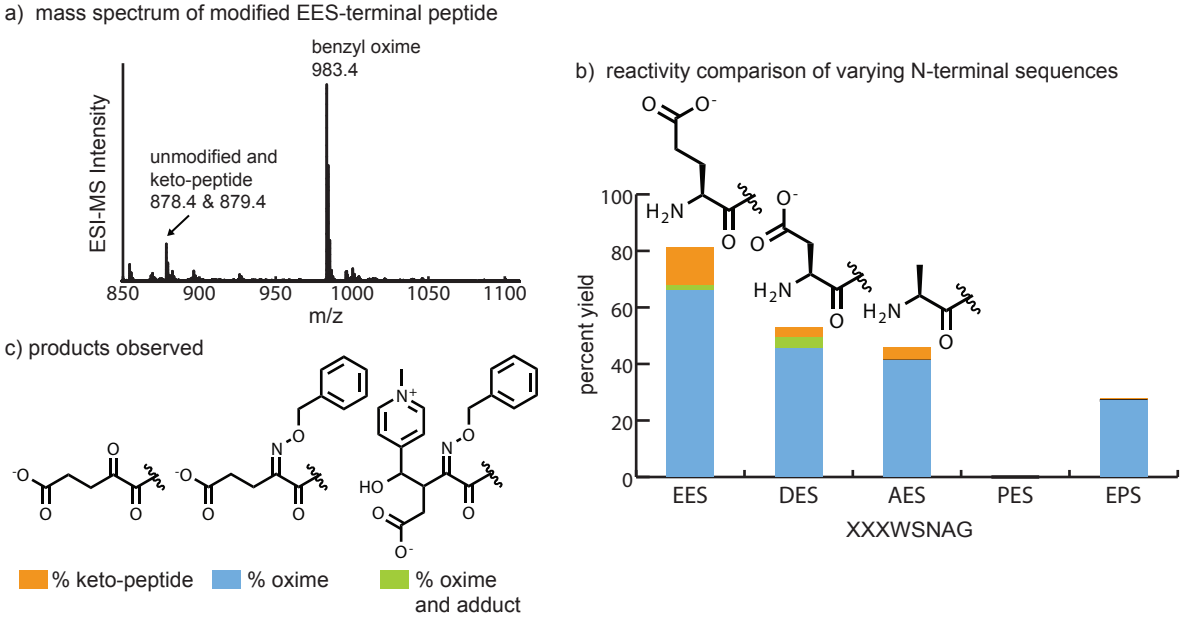
Another question about reactivity that could be answered using the library was whether residues in the second or third positions could significantly reduce the high reactivity of glutamate-terminal sequences. Knowledge of whether certain neighboring residues should be avoided is particularly important for incorporation of the reactive motif onto protein substrates. To address this question, a glutamate-terminal subset of the library, in which all peptides had Glu (E) as the N-terminal residue and the second and third positions were varied (EXX), was screened. Sequencing the colorless beads after transamination and oxime formation on this subset of the library revealed a consensus motif that had proline in the second position (Figure 2.4). 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 RS-mediated transamination. Although the third position had a lesser effect on reactivity, we selected EES as the representative highly reactive sequence and PES and EPS as the representative non-reactive sequences to explore how much the change of a single amino acid could affect the reactivity. We selected these representative sequences in order to validate our library findings on peptide and protein substrates.

library:	beads selected:	sequences identified:
entire library	red	EEQ, EED, EEI, EEM, ENE, EDD, EDW (2x), EDN, EDH, ETE, DEE, FED, FDM, FEE, WEE, WDD, YEE
E-terminal subset of library (EXX)	red	EES (3x), EEE, EEW, EET, EEN, EEY, EEA, EDE (3x), EDH, EDM, EDW, EDL, EDD, EDE, EDE, EDM, EDP, EDL, EDF EYD (2x), ENH, EYE, EFE, EFD
entire library	colorless	PTL, PAW, PSL, PFL, PTI, PTW, PIQ, ADL
E-terminal subset of library (EXX)	colorless	EPT (2x), EPP, EPS, EFV (2x), EVA

Figure 2.4. Sequences identified from library screening. The red beads indicated formation of the oxime product and corresponded to sequences that were highly reactive towards RS-mediated transamination. The colorless beads corresponded to sequences that were not reactive.

2.2.3. Verification of RS-mediated Transamination Using Peptides on Bead.

To verify the reactive and non-reactive sequence motifs identified by library screening, peptides of the form XXXWSNAG were synthesized. After synthesis, the peptide modified beads were treated with RS as described above, followed by benzylalkoxyamine for quantification of the oxime yield by LC-MS. The optimal sequence peptide, EESWSNAG, led to high modification yield, as seen in Figure 2.5a. A screen of the reaction conditions, varying the concentration of RS and the reaction time, was next performed. These screens (data shown in Figure 2.6) identified exposure to 100 mM RS for 1 h at pH 6.5 as the conditions resulting in high transamination yield with few byproducts. This reaction protocol was subsequently used to evaluate all of the peptide substrates. As seen in Figure 2.5b, the EE-terminal peptide resulted in over 80% transamination. Not all of the transaminated, keto-peptide species (shown in orange) was converted to oxime (blue) under the oxime formation conditions used. A small amount of covalent addition of RS to the N-terminus was also observed (green), which is presumed to be an aldol-type addition. Our proposed mechasim for the different product observed is show in in Figure 2.7. This byproduct was observed in higher



c) Percentage errors of XXXWSNAG studies

	% oxime	% oxime and adduct	% transaminated	% total conversion
EES	66.1 ± 8.5	1.9 ± 0.6	13.3 ± 3.4	81.4 ± 5.3
DES	45.3 ± 13.9	4.2 ± 0.9	3.5 ± 1.4	53.1 ± 13.6
AES	45.3 ± 13.9	0 ± 0	4.9 ± 1.4	46.1 ± 16.7
PES	0 ± 0	0 ± 0	0 ± 0	0 ± 0
EPS	27.1 ± 8.3	0 ± 0	0.9 ± 0.5	28 ± 8.8

Figure 2.5. 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 under the standard reaction conditions. (b) Various N-terminal sequences were subjected to the same reaction conditions and the product yields were quantified by LC-MS. (c) Three products were observed: transaminated keto-peptide that did not form an oxime, the desired oxime product, and oxime with the addition of RS to the N-terminus (proposed structure drawn). Aspartate-terminal peptides were observed by mass spectrometry to decarboxylate during transamination. A proline-terminal sequence was confirmed to have no reactivity towards RS-mediated transamination. With proline in the second position, the reactivity of the E-terminal sequence was substantially decreased. d) The reported data represent the average of three replicate experiments. Listed here are the percentage errors after the reaction (standard conditions) was performed in triplicate.

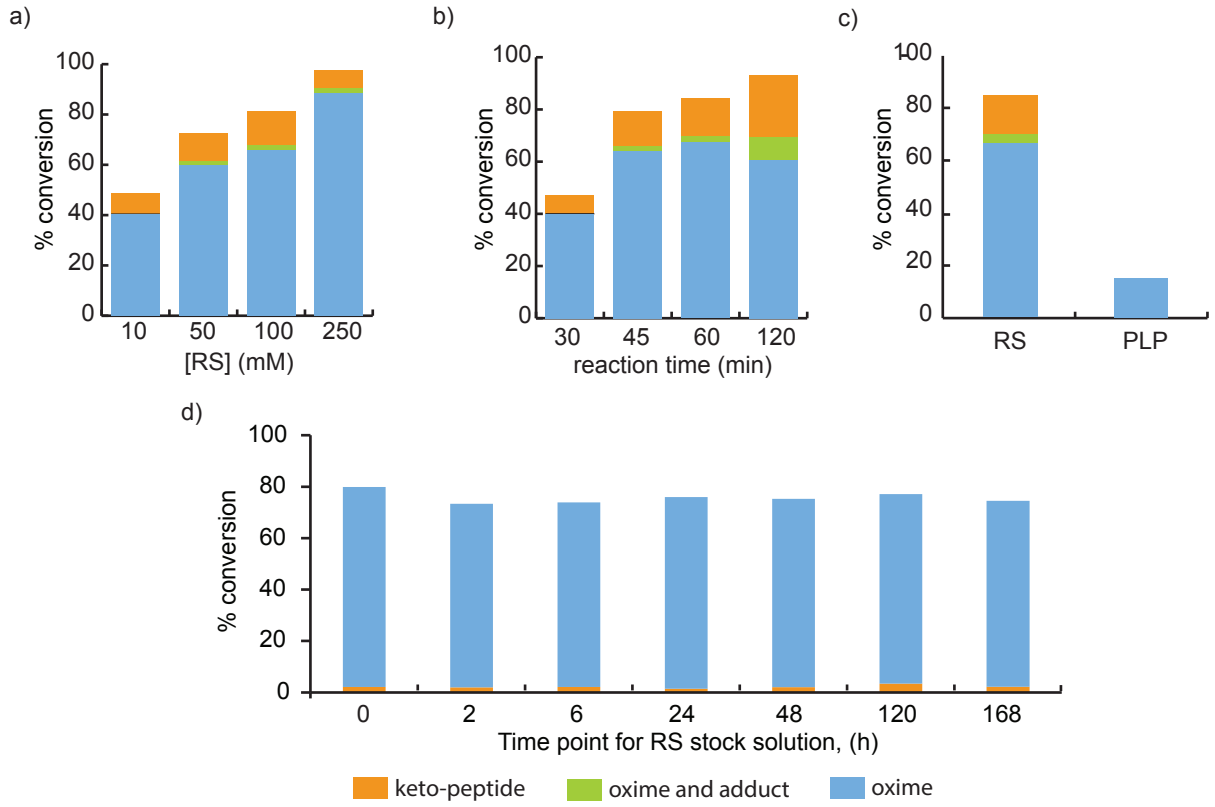


Figure 2.6. Reaction conditions screen for RS-mediated transamination of the EESWSNAG peptide on bead followed by oxime formation with benzyl alkoxyamine. (a) The concentration of RS was varied during a 1 h reaction at pH 6.5. Although library screening was done using 10 mM RS for a 1 h reaction, higher yields were found with higher concentrations. (b) Using 100 mM RS 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) To see if the EE-terminal sequence identified by library screening was an optimal sequence for RS 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 RS under the same conditions. The yield was significantly higher in the RS-mediated transamination case, indicating that the library screening had indeed identified an optimal transamination reagent/sequence pair. (d) Rapoport's salt stability screen on EESWSNAG peptide on-bead. A stock concentration of RS was made in 25 mM phosphate buffer, pH 6.5 and allowed to sit at RT. At the listed time points, 100 mM of this stock solution was added to EESWSNAG peptide beads for transamination (1 h at RT). Subsequent oximation and then cleavage off the bead was conducted prior to MS analysis of the product distribution, shown here. The results indicated that RS remains stable in solution over extended periods of time.

yields with longer reaction times, but could be minimized through the use of shorter reaction times (45 min to 1 h, Figure 2.6b).

To test whether the EE-terminal motif was optimal only for RS or was highly reactive towards transamination in general, the modification of this sequence with PLP-mediated transamination was examined as well (Figure 2.6c). Under the same reaction conditions the yield was much higher using RS, indicating the library screening protocol had identified an optimal reagent/sequence pair. The observed reactivity of the EE-terminal peptides with RS was comparable to the AKT-terminal sequence we had previously identified as optimal for PLP.[25] The N-terminal site-specificity of RS-mediated transamination was confirmed using tandem mass spectrometry for both the transaminated keto-peptide and the benzyl oxime (Figure 2.8).

Due to their structural similarity, aspartate-terminal sequences were also examined. E- and D- terminal peptides were found to have significantly different yields towards RS-mediated transamination, as shown in Figure 2.5. The mass spectra indicated that decarboxylation of the

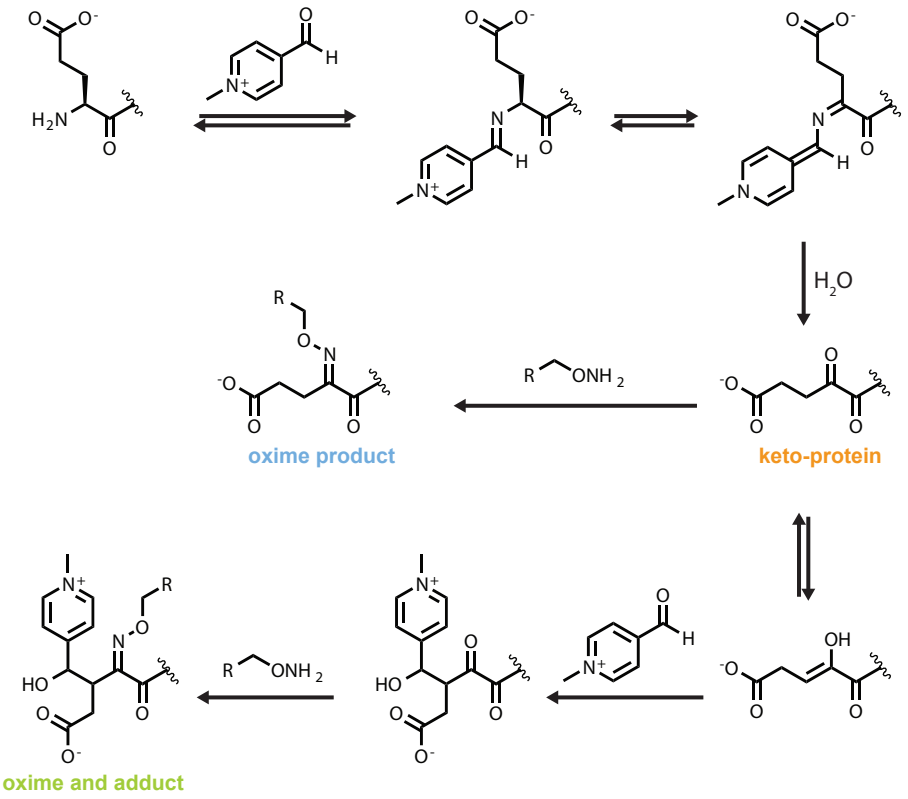


Figure 2.7. Proposed mechanism of RS-mediated transamination. The N-terminus of the protein first forms a Schiff base with the aldehyde. Alleviation of the charge on the pyridinium nitrogen helps drive the next step, which is deprotonation of the N-terminal α-hydrogen. Hydrolysis leads to the transaminated keto-protein species, which is then able to form the desired oxime product upon incubation with an alkoxyamine reagent of choice. As a mass corresponding to the addition of RS is observed, a likely reaction pathway for this is an aldol reaction with another equivalent of RS. This structure leads to the oxime and adduct species after reaction with the alkoxyamine.

aspartate-terminal peptide occured during the transamination reaction, as has been observed previously using PLP.[24] 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 RS-mediated transamination are glutamate-terminal sequences, and that Asp and Glu are not interchangable 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 neither 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 multiglutamate-terminal peptide. This result underscores the impact of a single amino acid substitution in an internal sequence position.

2.2.4. Chain-specific Modification of Anti-HER2 Human IgG1.

Given that the heavy chain of human anti-HER2 IgG1 (Herceptin)[34] has an N-terminal glutamate residue, the wild-type antibody provided a suitable substrate to test RS-mediated transamination on a protein. The wild-type anti-HER2 antibody (denoted as wild-type, WT), which

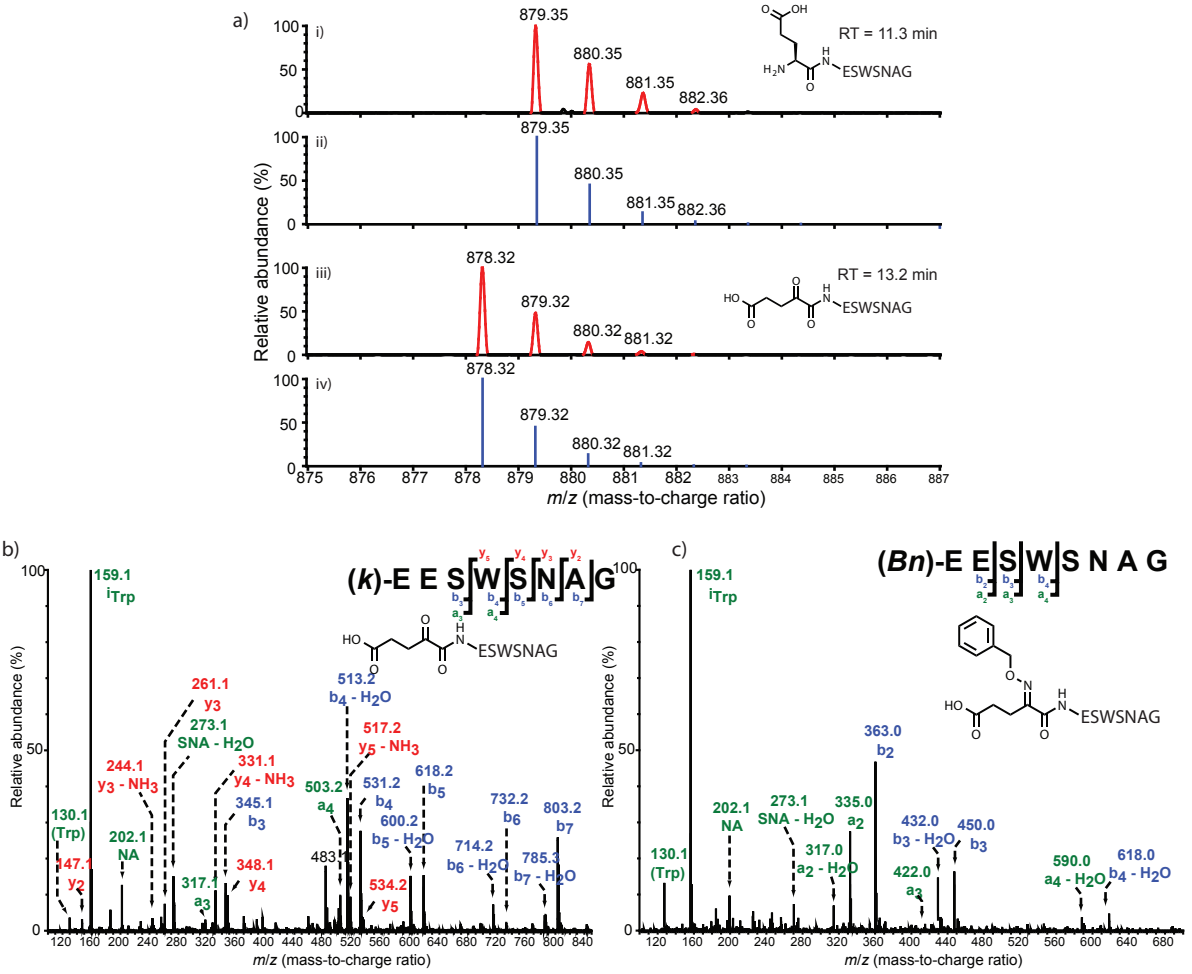


Figure 2.8. (a) Comparison between measured (red) and theoretical (blue) isotopic distributions of peptide ions measured from the 1 h reaction of EESWSNAG with 100 mM Rapoport's salt. The spectra correspond to the measured (i) and theoretical (ii) isotopic distributions for the $[M + H]^+$ ion of the 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 = $C_{36}H_{47}N_9O_{17}$; retention time = 13.2 min). 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. 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.

was commercially obtained from Eureka Therapuetics (Emeryville, CA), has a EVQ-terminal heavy chain and a DIQ-terminal light chain.[34] We also expressed an N-terminal mutant of the antibody in HEK cells to introduce additional reactivity using the EES optimal sequence on both chains. For the transamination reaction, the antibodies were incubated with a freshly prepared 100 mM solution of RS for 60 min in pH 6.5 phosphate buffer at 37 °C. After transamination and removal of the excess small molecule, the antibodies were incubated with $BnONH_2$ for 48 h at RT in pH 5.5 phosphate buffer. The modification of each chain was then analyzed and quantified using mass spectrometry.

The heavy chain wild-type sequence provided 67% conversion to the oxime product, in which 15% also included the RS adduct (Figure 2.7). In contrast, no modification was observed for the light chain (Figure 2.9). This is likely because the steric environment of the folded protein reduces the accessibility of the already less-reactive substrate (DIQ terminal sequence). The complete lack of modification also clearly demonstrated that RS does not react with lysine side chain amines or other residues. The net result of this experiment is that RS-mediated transamination of the wild-type sequence allowed the selective modification of only the heavy chain. We have not yet carried out studies to determine whether the presence of the small percentage of adduct affects the biological properties of the antibody. This reactivity pattern was also observed on another widely used human IgG1, anti-hTNFα (Invivogen, San Diego, CA), which also has a glutamate-terminal heavy chain and an aspartate-terminal light chain (Figure 2.10).[35]

To see if both antibody chains could be modified through sequence alteration, we prepared a herceptin analog in which all N-termini were extended by three residues to add the EES motif. These antibodies were modified on both the light and the heavy chains (56% and 68%, respectively). During the expression of this mutant, we observed some improper cleavage of the IL2 signal sequences, which led to the production of a small proportion of light and heavy chains that lacked

a) MS analysis of antibody modification

mutant:	chain (N-terminal sequence):	% oxime yield:	chains modified:
wild type	light (DIQ) heavy (EVQ)	0 67	
EES	light (EES) heavy (EES)	56 68	

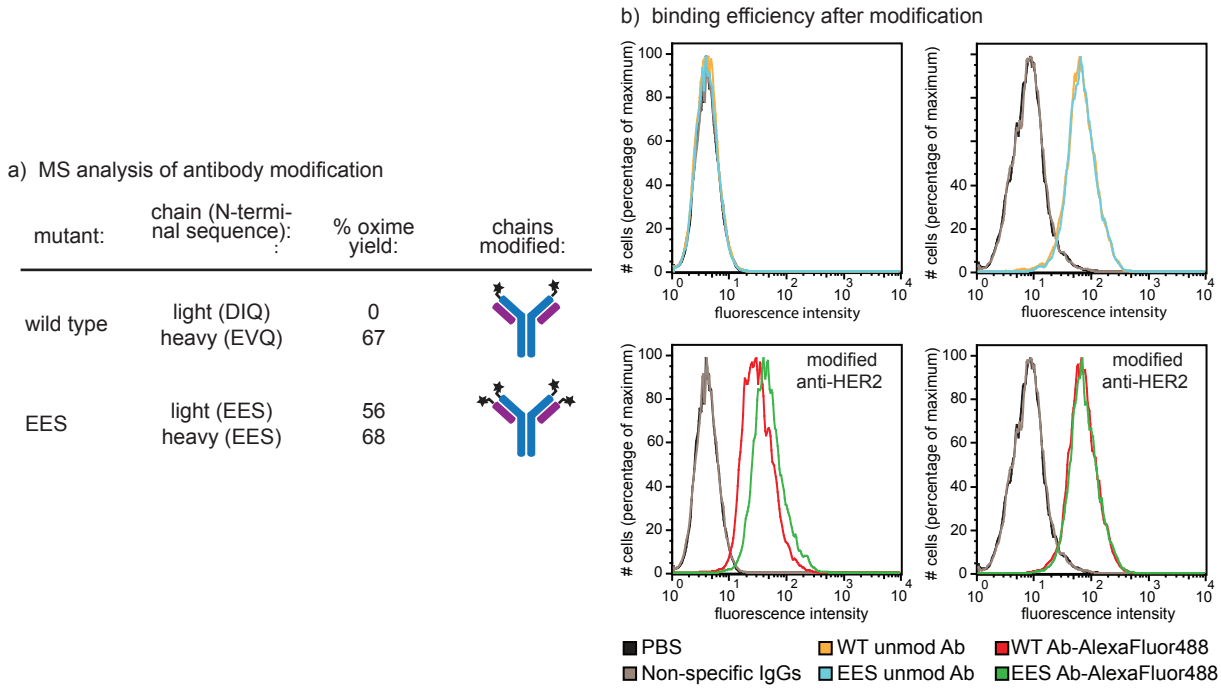


Figure 2.9. IgG modification using RS-mediated transamination. (a) Wild-type anti-HER2 human IgG1 antibodies, as well as mutants with EES N-terminal sequences, were treated with RS, followed by oxime formation with $BnONH_2$. The heavy and light chains were separated for mass spectrometry quantification of the benzyl oxime product. The wild-type antibody showed modification of the heavy chain and no modification of the light chain, while the mutant with EES on both sets of termini exhibited modification of both chains. (b) The antigen binding ability of wild-type (WT) and EES-terminal anti-HER2 antibodies post-modification (transaminated by RS, followed by oxime formation with AlexaFluor488-ONH_2) was confirmed using flow cytometry. MCF7 clone 18 cells (a HER2 positive cell line) were subjected to modified (bottom histograms) and unmodified (top) wild-type and EES-terminal antibodies. By direct detection of the AlexaFluor488, the binding of modifed antibodies was seen (bottom left), while the unmodified showed no fluorescent shift (top left). In order to compare the binding of the modified antibodies to unmodified, the unmodified species were detected using a secondary detection method: anti-human IgG PerCP-conjugated 2º $F(ab')_2$ (right histograms). Both wild-type and EES-terminal antibodies modified with the fluorophore retained similar binding affinity and specificity to the unmodified ones. Non-specific human IgG1 and PBS were used as negative control agents. Jurkat cells (a HER2 negative cell line) did not bind to either unmodified or modified anti-HER2 antibodies.

the N-terminal EE groups. The resulting serine-terminal analogs were not modified to the same extent as the EES sequence on the heavy chain, and we observed no modification for the serine-terminal light chain. Overall, these experiments demonstrate that this new RS method can achieve a previously inaccessible degree of labeling control for one or both sets of antibody chain termini. Since the N-termini of the antibody heavy and light chains flank the antigen binding domains (but are not part of the hypervariable loops),[48] it was important to confirm that the modification of these locations did not disrupt antigen binding. To evaluate this, we used AlexaFluor 488-ONH_2 to label transaminated antibodies, and we used flow cytometry (Figure 2.9b) to compare the unmodified antibodies (top histograms) to their N-terminally modified counterparts (bottom histograms). In each experiment, the antibodies were incubated with HER2 overexpressing cells (MCF7 clone 18).[37] The AlexaFlour 488 dyes were detected directly, and an anti-human IgG fluorescently labeled with peridinin chlorophyll protein (PerCP) was used as a secondary detection method. As shown in Figure 2.9b (right set of histograms), no significant disruption of binding was observed for

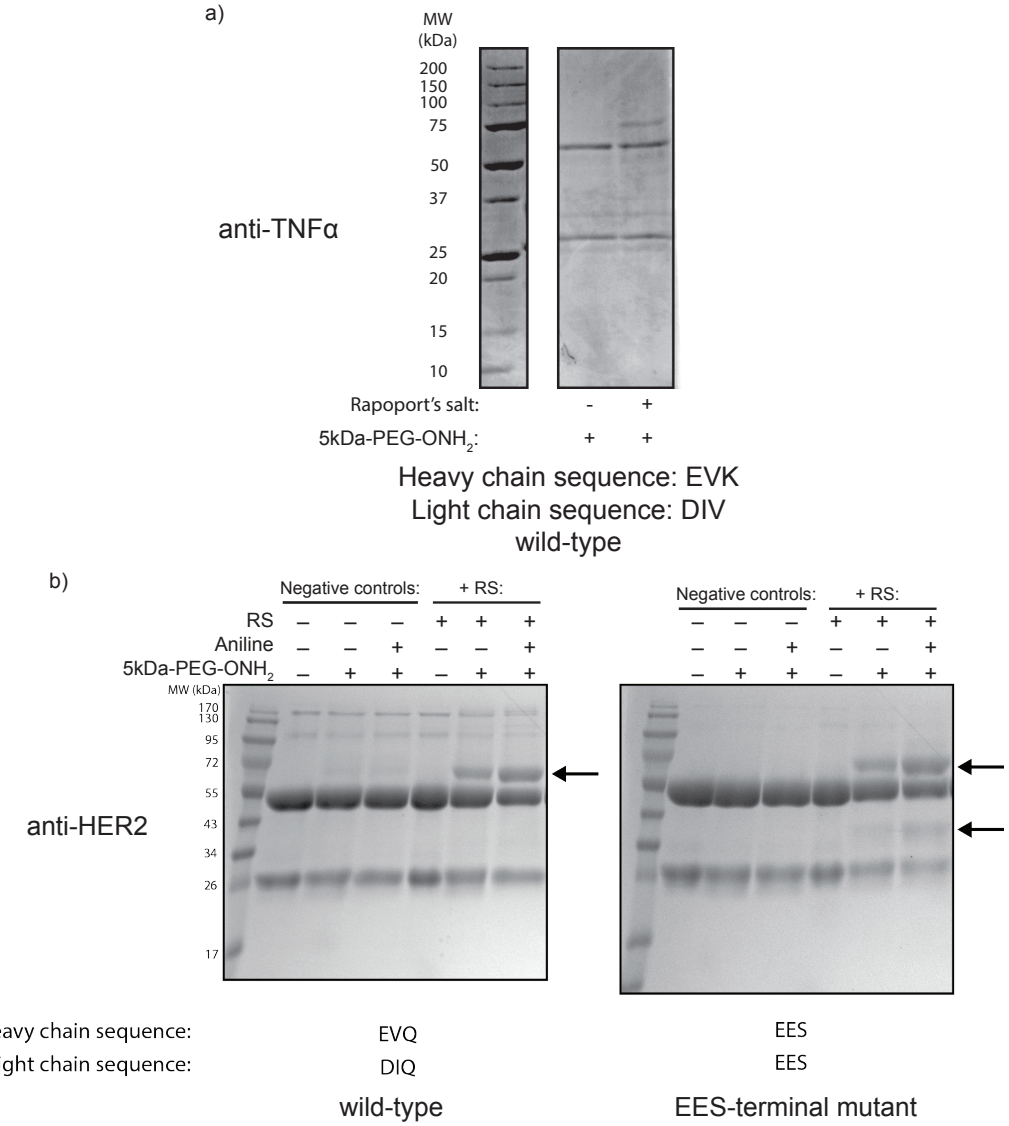


Figure 2.10. (a) SDS-PAGE gel showing attachment of a 5 kDa poly(ethylene glycol) (PEG) alkoxyamine to the light and heavy chains of the anti-TNFα antibody. (b) SDS-PAGE gel showing attachment of a 5 kDa poly(ethylene glycol) (PEG) alkoxyamine to the light and heavy chains of the anti-HER2 antibody wild-type and EES-mutant after RS-mediated transamination. The arrows point to the bands corresponding to heavy and light chain proteins with PEG attachment.

any of the AlexaFluor 488 oxime conjugates. The left histograms of Figure 2.9b indicate that the fluorophore-labeled antibodies were specifically involved in the binding, as only species modified with the AlexaFluor dye showed fluorescent shifts in this channel. Jurkat cells, which are a HER2-negative cell line, were used as a negative control. Neither modified nor unmodified antibodies were observed to bind these cells. To demonstrate that it is possible to attach larger groups using RS-mediated transamination, we also used a 5,000 Da MW poly(ethylene glycol) alkoxyamine (**2c**) reagent[38] to attach a polymer to the antibody (Figure 2.10). With these examples, we were able to validate that peptide motifs selected from a cominatorial libary screen could be succcessfully applied on the protein scale through simple amino acid mutations.

2.2.5. Obtaining a Bacterially Expressed EES Terminal Protein and Evaluating its RS Reactivity.

Given that glutamate residues prevent post-translational cleavage of methionine residues encoded for by the start codon, we sought to expand the RS-mediated transamination methodology to allow for the site-selective protein modification of bacterially expressed proteins. A cellulase was chosen as a protein of particular interest for these studies because of its relevance to the biofuels industry. This class of enzymes has the ability to hydrolyze the cellulose in biomass, releasing glucose for ultimate biofuel generation.[39] In previous studies,[40] we showed that a pyridoxal phosphate (PLP)-based N-terminal bioconjugation strategy could be used to install polymer chains with lower critical solution temperatures (LCSTs) on a thermostable endoglucanase from *Pyrococcus horikoshii* (EGPh).[41] The resulting enzyme could be recovered through modest increases in solution temperature and then reused. One hurdle to the practical use of this system, however, is the relatively high cost of the PLP reagent, limiting the scale on which the bioconjugate could be produced. In contrast, RS can be made from inexpensive precursors (pyridine carboxaldehyde and methyl benzenesulfonate) in high yields.[26] However, as this enzyme is most conveniently expressed in *E. coli*, it can not be generated with the N-terminal glutamate residue identified in the previous RS-transamination study.[1]

EGPh is a monomeric protein that is amenable to site-directed mutagenesis and can be heterologously expressed in high yield (>120 mg per liter of culture). Additionally, it has a solvent exposed N-terminus that is located on the opposite face of the enzyme active site, rendering it accessible for transamination and oxime formation.[42] To begin these studies, site directed mutagenesis was carried out to obtain an IEGREES-terminal EGPh mutant. IE(or D)GR is the preferred cleavage site for the protease Factor Xa, a commercially available enzyme derived from the blood clotting cascade.[43,44] Factor Xa will not cleave a site followed by a proline or arginine, but remains unbiased towards all other residues if they follow the recognition site. During the expression and N-terminal proteolysis process, however, we found that the cellulase had an unexpected, yet inherent Factor Xa cleavage site of SVIR 20 amino acids away from the C-terminus. Because the His_6 tag that was inserted for ease of protein purification was also directly at the C-terminus, we were able to use Ni-NTA purification following proteolytic cleavage to separate EES-EGPh that was truncated at the C-terminus from EES-EGPh that still retained the His_6 tag (Figure 2.11).

Once Factor Xa had been successfully used to obtain EES- terminal EGPh, the reactivity of the purified protein toward RS-mediated transamination was investigated. To assess the transamination yields, samples were subsequently reacted with nitrobenzylalkoxyamine (**2b**) at pH 4.5 for 42–48 h, followed by LC-MS analysis to quantify the oxime-modified protein, as seen in Figures 2.12 and 2.13.

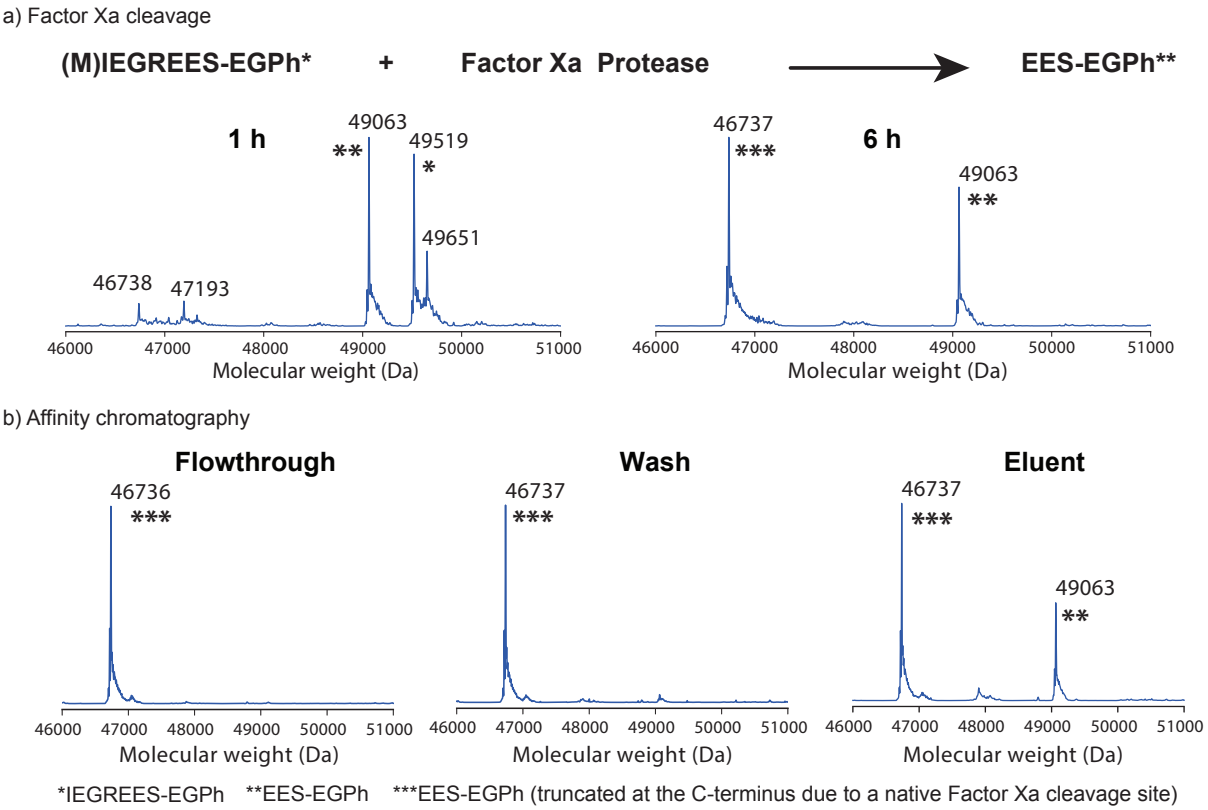


Figure 2.11. Obtaining pure EES terminal EGPh. (a) Incubation of (M)IEGREES-EGPh with Factor Xa for 6 h resulted in complete removal of the IEGR tag. Additonally, an inherent Factor Xa cleavage site after the residues SVIR, which is 20 amino acids from the C-terminus, was found. This moiety still retained the desired EES N-terminus, but did not retain the His_6 purification handle that was inserted at the C-terminus. (b) A Ni-NTA purification step separated full-length EES-EGPh from the truncated EES-EGPh. A pure sample of EES-EGPh that was truncated at the C-terminus was obtained in the flowthrough and wash steps, and this was used in subsequent studies.

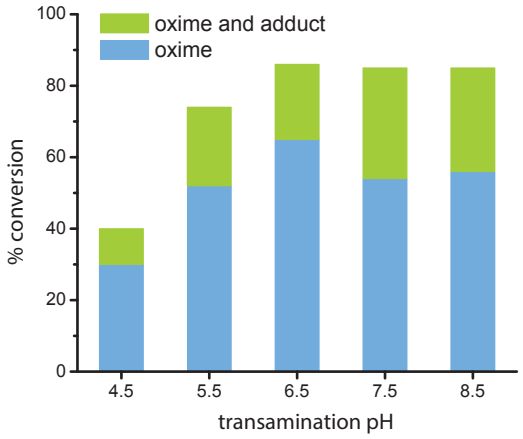


Figure 2.12. Testing the effect of pH on the EES terminal EGPh protein (after cleavage of the His_6 tag at the C-terminus). EES N-terminal EGPh was subjected to RS transamination and subsequent oximation with 2b. The pH during the transamination step was varied, but all other reaction conditions were kept the same. The levels of conversion were quantified using LC-MS. High levels of desired product were observed when the transamination step was performed at pH 6.5, 7.5, and 8.5.

The long oxime-formation times were selected to ensure complete conversion for quantitative purposes. In some cases, shorter times may suffice, but in our experience, longer incubation times are required to reach high conversion with these ketone substrates. This screen validated previously published findings, showing that at pH 6.5 the EES-terminal protein underwent high conversion to the desired oxime product (blue). Covalent addition of RS to the N-terminus was also observed, through what is presumed to be an aldol-type addition. This resulted in some amount of product that contained both the oxime and an additional equivalent of RS (green). Additionally, it was observed that the reaction was tolerant to a wider-than-expected range of pH values without sacrificing high yield, a finding that facilitates the modification of proteins that have isoelectric points (pI values) near 6.5.

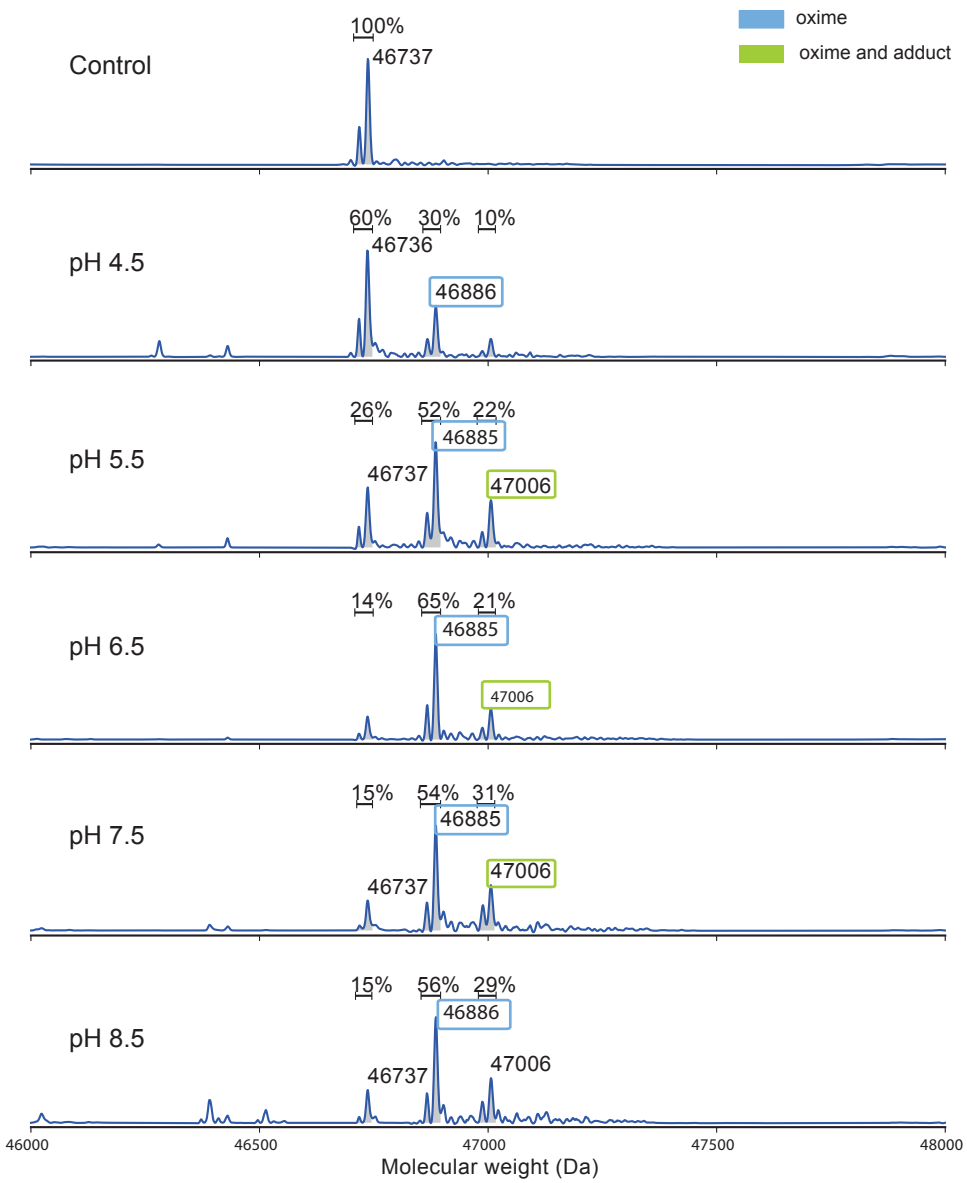


Figure 2.13. Deconvoluted LC-MS spectra of EES-EGPh after RS transamination followed by oxime formation with nitrobenzyl alkoxyamine (**2b**). The pH of RS transamination is listed next to each spectrum. Origin was used to baseline correct these data. The percentages represent the relative area underneath the highlighted peaks. Unmodified EES-EGPh (with a 20 amino acid truncation at the C-terminus due to secondary cleavage with Factor Xa) has a mass of 46737 Da.

2.2.6. Studying Reactivity of Alternate N-terminal Sequences with RS-mediated Transamination Using Purified Peptides.

While it was found that Factor Xa could be used to obtain bacterially expressed EES N-terminal proteins, it would clearly be preferable to identify termini that could be expressed directly in *E. coli* and then modified without the need for an additional protease step. In order to do this, the reactivity properties of other N-termini were also examined using different reaction conditions. More specifically, the transamination efficiency of RS was evaluated with N-terminal sequences beginning with residues that could be directly expressed bacterially. The residues for which the start codon Met will be post-translationally cleaved have been reported to be Gly, Ala, Pro, Ser and Thr.[32] Therefore, peptides of the form XEEWSNAG were synthesized, where X

represented a member of this set. In these model systems, Glu residues were incorporated in the second and third positions in order to retain the overall negative charge. Additionally, to represent those cases in which the Met residue is not cleaved off, a peptide of the sequence MEEWSNAG was included in these studies. Upon synthesis and purification, pH screens were carried out using each peptide in solution. The EESWSNAG peptide was also included so that yield comparisons to the previously identified optimal sequence could be made. Figure 3 shows the pH screen of some of these peptides with Rapoport's salt, followed by oxime formation with benzylalkoxyamine (**2a**). LC-MS was used to quantify the yields. EESWSNAG reached quantitative conversion from pH 6.5 through 8.5, which matched the protein data closely (Figure 2.14a). The slightly higher yields are likely due to the fact that the N-termini of the peptides are more accessible than the termini of large proteins folded into higher order structures.

For the glycine, serine, and threonine-terminal peptides, complex product mixtures were observed due to competing β-elimination and retro-Mannich pathways.[45,46] We therefore do not recommend using these terminal residues with this technique. These data are represented in Figure 2.15. Additionally, the proline terminal peptide showed little-to-no reactivity, and is a useful sequence if one requires reactivity to be turned "off" (Figure 2.14b). For the methionine terminal peptide, moderate yields (<70%) of desired product were still observed (Figure 2.14c). The highest yielding sequences were those that retained an alanine residue at the N-terminus. For the alanine-terminal peptide, AEEWSNAG, high oxime yields were observed at both pH 7.5 and pH 8.5, with conversion surpassing 80% in the latter case (Figure 2.14d). Additionally, no side reactivity was

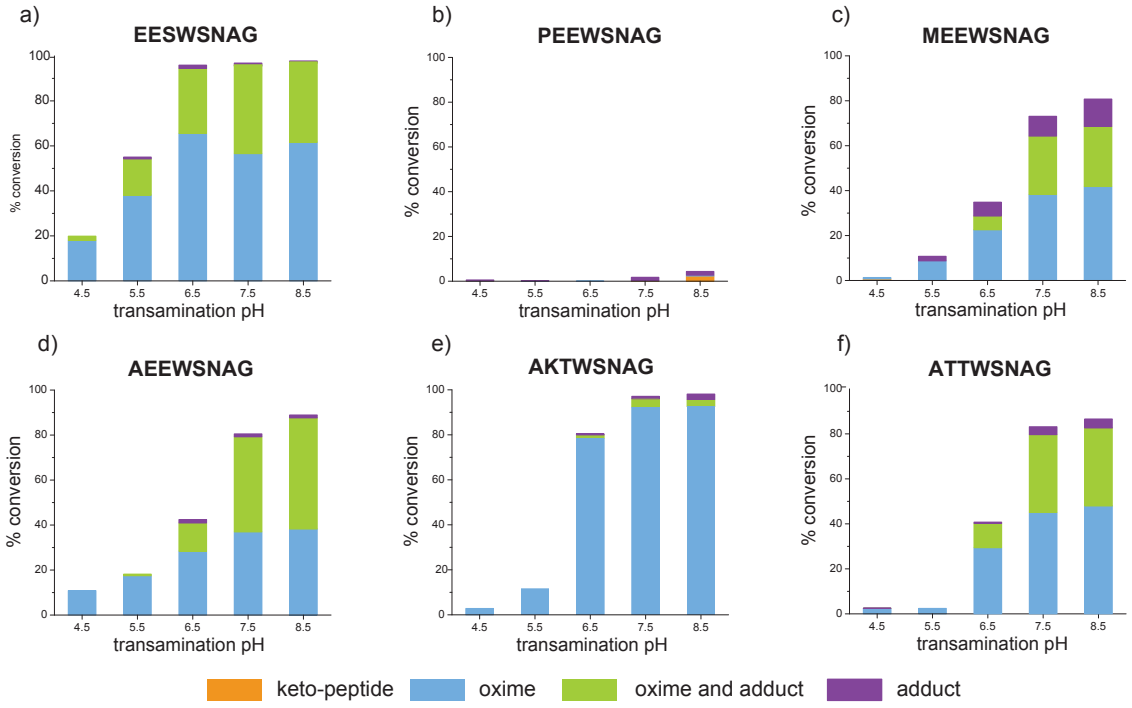


Figure 2.14. Testing the effect of pH on RS-mediated transamination with multiple N-terminal sequences. The peptides studied were of the sequence XXXWSNAG, in which the first three residues were varied as denoted on each graph. Each peptide was subjected to RS transamination at pH values ranging from 4.5–8.5, followed by oxime formation with **2a** at pH 5.0 for 40–48 h. These results indicated that the previously identified requirement for negative charges holds true at pH 6.5, but can be circumvented if the transmination reaction is performed at higher pH values. The levels of conversion were quantified using LC-MS and the data shown are an average of three replicate experiments (standard deviation is tabulated in Figure 2.15).

EESWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	80.2 ± 6.3	45 ± 4.4	4.1 ± 4.4	2.6 ± 3.7	1.7 ± 2
% transaminated	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
% oxime	18 ± 6.4	38.2 ± 5.2	65.6 ± 3.8	56.6 ± 8.2	61.7 ± 5.4
% oxime and adduct	2 ± 0.4	16.2 ± 2	29 ± 5.8	39.5 ± 10.1	35.8 ± 7.2
% adduct	0 ± 0	0.9 ± 1	1.5 ± 0.5	1.5 ± 0.6	1.1 ± 0.1

PEEWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	99.5 ± 0.9	99.8 ± 0.2	99.9 ± 0.2	98.3 ± 2.3	95.7 ± 2.4
% transaminated	0.2 ± 0.3	0 ± 0	0 ± 0	0.2 ± 0.3	2.4 ± 2.1
% oxime	0.1 ± 0.2	0.2 ± 0.2	0.2 ± 0.2	0.2 ± 0.1	0.4 ± 0.1
% oxime and adduct	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
% adduct	0.4 ± 0.7	0.2 ± 0.3	0 ± 0	1.5 ± 2.1	1.7 ± 0.4

MEEWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	98.7 ± 1.1	89.3 ± 5.4	65.2 ± 1	27 ± 1.6	19.3 ± 4.2
% transaminated	0.7 ± 1.2	0.1 ± 0.2	0 ± 0	0 ± 0	0.2 ± 0.3
% oxime	0.7 ± 0.8	8.7± 3.9	22.6 ± 2.5	38.3 ± 2.2	41.7 ± 1.6
% oxime and adduct	0 ± 0	0 ± 0	6.2 ± 0.6	26 ± 2.2	26.8 ± 3.5
% adduct	0 ± 0	2.1± 1.6	6.2 ± 2.6	8.9 ± 2.3	12.3 ± 3.5

AEEWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	89.2 ± 8.4	81.8 ± 11.7	57.6 ± 8.9	19.5 ± 4.4	11.1 ± 4.3
% transaminated	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
% oxime	10.9 ± 8.4	17.5 ± 10.8	28.5 ± 2.6	37 ± 9.1	38.3 ± 8.2
% oxime and adduct	0 ± 0	0.8 ± 1.3	12.5 ± 5	42.3 ± 6.2	49.4 ± 12.4
% adduct	0 ± 0	0 ± 0	1.6 ± 2	1.4 ± 0.6	1.3 ± 0.5

AKTWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	97.2 ± 3.9	88.5 ± 2.6	19.6 ± 2.7	2.9 ± 2.1	1.9 ± 0.8
% transaminated	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
% oxime	2.8 ± 3.9	11.5 ± 2.6	78.5 ± 2.7	92.3 ± 2.1	92.5 ± 0.9
% oxime and adduct	0±0	0 ± 0	1.5 ± 0.2	3.7 ± 0.3	3.3 ± 0.5
% adduct	0±0	0 ± 0	0.5 ± 0.4	1.1 ± 0.3	2.4 ± 1

ATTWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	97.5 ± 2.3	97.6 ± 2.7	59.3 ± 6.9	16.9 ± 8.3	13.4 ± 3.6
% transaminated	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
% oxime	2.5 ± 2.2	2.5 ±2 .7	29.4 ± 4.8	45.1 ± 4.7	48 ± 2.2
% oxime and adduct	0.1 ± 0.1	0 ± 0	10.9 ± 1.8	34.6 ± 4.5	34.7 ± 1.8
% adduct	0.1 ± 0.1	0 ± 0	0.6 ± 0.6	3.6 ± 1.5	4 ± 2

Figure 2.15. Peptides of the form XXXWSNAG were synthesized to compare the reactivity of various N-terminal sequences towards RS-mediated transamination. Listed here are the sample means and standard deviations after each reaction was performed in triplicate.

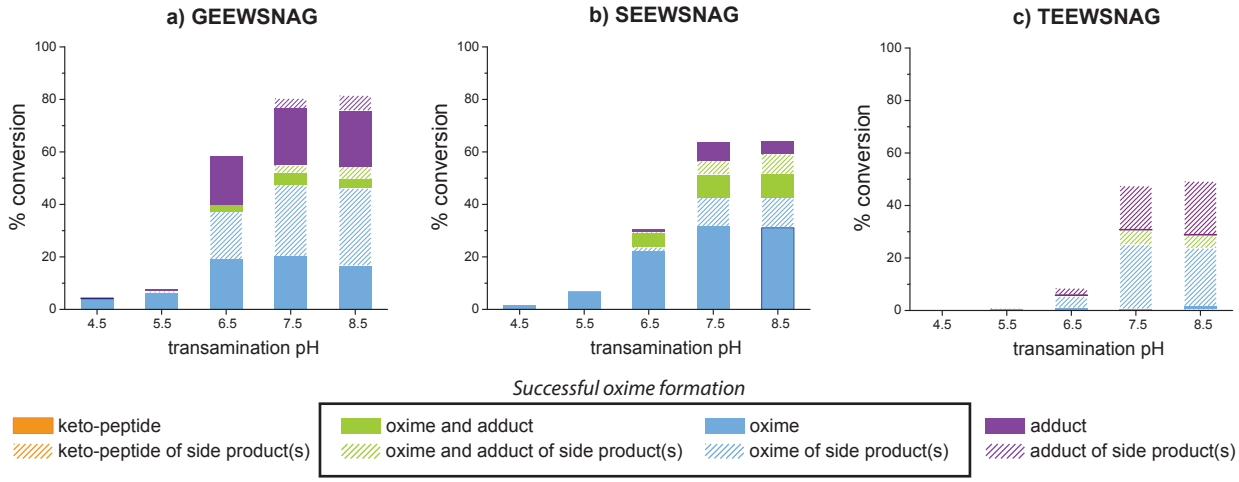


Figure 2.16. Testing the effect of pH on RS-mediated transamination with peptides of the sequence XEEWSNAG, in which the first residue was varied as denoted on each graph. (a) In the glycine terminal peptide, it was observed that conversion levels reached up to 50% at the higher pH values, but there was a large amount of adduct that also formed (purple). This is not an aldol adduct due to the lack of an R group, and is currently unidentified. Additionally, it should be noted that in the LC-MS traces, we observed large amounts of a GEEWSNAG+15 mass units product, and this molecule was able to undergo transamination and oxime formation, and is represented by the dashed bars for the side products. (b, c) Serine and threonine residues at the N-terminus underwent additional transformations in the presence of RS. As observed by LC-MS, the serine side chain experienced a beta elimination during transamination with RS, yielding an alanine terminal peptide. This resulted in a larger product distribution, but the oxime modified product, including those with RS adduct, totalled about 60% of the yield at both pH 7.5 and pH 8.5. During reaction of the threonine-terminal peptide with RS, we observed conversion of a portion of the Thr to Gly, in a retro-Mannich reaction. Of the resulting TEEWSNAG and GEEWSNAG mixture, extremeley low levels of modification of the TEEWSNAG were observed. Yields of the remaining GEEWSNAG portion are denoted by the dashed bars. The levels of conversion were quantified using LC-MS and the data shown are an average of three replicate experiments. The sample means and standard deviations are shown in Figure 2.17.

GEEWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	95.7±6	92.4±3.5	37.6±10.6	8.8±2.5	6.1±1.1
% unmodified side product	0±0	0±0	3.9±2.4	11±7.2	12.5±8.3
% transaminated	0±0	0±0	0±0	0.1±0.1	0±0
% transaminated side product	0±0	0±0	0±0	0±0	0±0
% oxime	3.9±5.1	6.3±3.1	19.1±3.5	20.2±1.4	16.7±3.7
% oxime side product	0.3±0.5	0.7±0.5	18.3±3.1	27.3±7.8	29.5±6.6
% oxime and adduct	0±0	0±0	2.8±0.7	4.5±2.9	3.5±2
% oxime and adduct side product	0±0	0±0	0.1±0.1	3±1.4	5±3.1
% adduct	0.3±0.5	0.7±0.5	18.4±3.6	21.6±2.8	21.2±3.1
% adduct side product	0±0	0±0	0.3±0.5	4±3.8	5.9±6.2

SEEWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	98.5±0.5	93.1±1.6	69.4±2.5	35.4±4.7	34.8±2
% unmodified side product	0±0	0±0	0±0	0.9±0.2	1.3±0.2
% transaminated	0±0	0±0	0±0	0±0	0±0
% transaminated side product	0±0	0±0	0±0	0±0	0±0
% oxime	1.6±0.5	6.9±1.7	22.1±3.5	31.7±2	31.2±0.5
% oxime side product	0±0	0±0	1.7±0.8	11±3.3	11.7±3
% oxime and adduct	0±0	0.1±0.2	5.3±0.6	8.8±3.3	8.8±1.2
% oxime and adduct side product	0±0	0±0	0.7±0.5	5.4±1.8	7.7±1.3
% adduct	0±0	0±0	1.1±0.6	7.2±3.8	4.9±3.7
% adduct side product	0±0	0±0	0±0	0±0	0±0

TEEWSNAG	pH 4.5	pH 5.5	pH 6.5	pH 7.5	pH 8.5
% unmodified	94.3±9.7	94.4±8.3	88.5±8.1	32.6±4.6	25.1±6.7
% unmodified side product	5.6±9.6	4.8±8.1	3.2±3.7	20±2.8	25.8-±2.1
% transaminated	0±0	0±0	0±0	0±0	0±0
% transaminated side product	0±0	0±0	0±0	0.3±0.3	0.6±0.2
% oxime	0.1±0.2	0.5±0.3	1±0.6	0.4±0.6	1.1±1.8
% oxime side product	0.2±0.2	0.5±0.5	4.5±1.5	24.5±6.5	22.2±6
% oxime and adduct	0±0	0±0	0±0	0±0	0±0
% oxime and adduct side product	0±0	0±0	0.5±0.4	5.7±4.1	5.1±2
% adduct	0±0	0±0	0.2±0.3	0.3±0.4	0.3±0.4
% adduct side product	0±0	0.1±0.1	2.6±3.6	16.6±8.2	20.3±10.1

Figure 2.17. Peptides of the form XEEWSNAG were synthesized to compare the reactivity of various N-terminal sequences towards RS-mediated transamination. Listed here are the sample means and standard deviations after each reaction was performed in triplicate.

observed other than the aldol addition pathway. With this finding, we concluded that an AEE terminus is an ideal alternative to EES, as long as transamination is carried out at the optimal higher pH values.

From these peptide data, we noticed a trend that regardless of the absolute level of transamination, Rapoport's salt reactivity was improved when the pH was increased from 6.5 to 8.5. EES was reactive at all pH values between 6.5 and 8.5, while the other sequences had significantly higher yields at higher pH values. Since the original library screen for reactive sequences was

performed at pH 6.5, this explained why EES was identified. At the higher end of pH values tested herein, the reaction appeared to be less dependent on the specific sequence at the N-terminus. To test this hypothesis, we also screened AKT and ATT terminal peptides. The sequences were chosen because they substituted the presumed beneficial negatively charged penultimate residue with a positive charge and no charge, respectively. The pH screen results of these peptides after RS transamination and oxime formation are illustrated in Figures 2.14e and 2.14f. AKTWSNAG showed high levels of modification from pH 6.5 through 8.5 and ATTWSNAG exhibited a high yield at pH 8.5. These trends validate that at a higher pH, RS is less selective towards the specific sequence at or close to the N-terminus.

2.2.7. Applying New Findings to a Model Protein System: EGPh.

From these peptide studies, we found that alanine at the N-terminus undergoes high levels of conversion with RS. To confirm these results on a protein substrate, AEE and AKT terminal EGPh mutants were made through site-directed mutagenesis and expressed. Since the start codon methionine is cleaved when alanine is in the neighboring position, the Factor Xa proteolytic cleavage site was not necessary to obtain alanine-terminal proteins after expression. The mutants were reacted with RS over a range of pH values, followed by oxime formation at pH 4.5. Yields were quantified by LC-MS. Both the AEE-EGPh and AKT-EGPh mutants resulted in high yields of modification at pH 7.5 and pH 8.5 (Figure 2.18, 2.19, and 2.20). It should be noted that some of the lack of reactivity at pH 5.5 and 6.5 for these mutants can be attributed to their theoretical isoelectric points being in that range (AEE-EGPh pI 5.77: AKT-EGPh pI: 6.01), making them less soluble during the reaction.

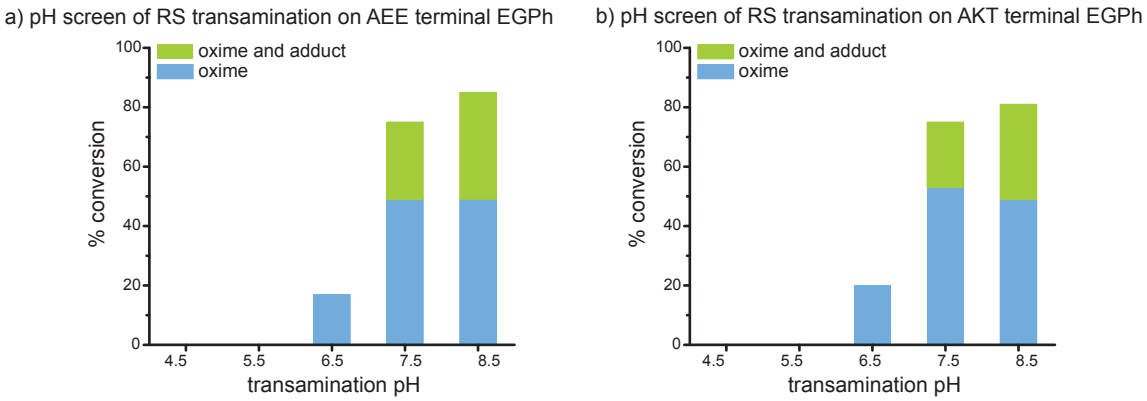


Figure 2.18. Verifying successful alanine terminal peptide trends on a protein. The proteins were subjected to RS transamination and subsequent oximation with **2b**. The pH values during the transamination steps were varied, but all other reaction conditions were kept the same. At pH 8.5, both (a) AEE-EGPh and (b) AKT-EGPh exhibited over 80% conversion to desired oxime products. The levels of conversion were quantified using LC-MS.

2.2.8. Evaluating the Activity of Modified EGPh.

The activities of the modified EGPh proteins are compared in Figure 2.21. For each mutant type, a small molecule alkoxyamine (**2b,** green) and a PEG_{5kDa} alkoxyamine (**2c**, blue) modified sample were assayed. Additionally, an unmodified, non-transaminated control of each mutant was evaluated (pink). In order to evaluate the activity of the EES-terminal mutant systematically, a second control was included with purified protein prior to proteolytic cleavage using Factor Xa. This unmodified IEGREES-terminal mutant of EGPh was studied in order to ascertain how the truncation at the C-terminus of EES-EGPh affected its activity (orange). As compared to the

controls we observed that both the small molecule modified protein and the PEG_{5kDa} modified protein exhibited similar activity levels, indicating that the tertiary structure and active site of the enzyme were not affected by the N-terminal modifications made through RS transamination. To be sure the observed activity was not merely due to remaining unmodified protein, an additional control assay was carried out using 50% less EGPh (Figure 2.24). As expected, a significant reduction in activity was observed, indicating that the modified enzyme must be responsible for a substantial portion of the activity in Figure 2.21.

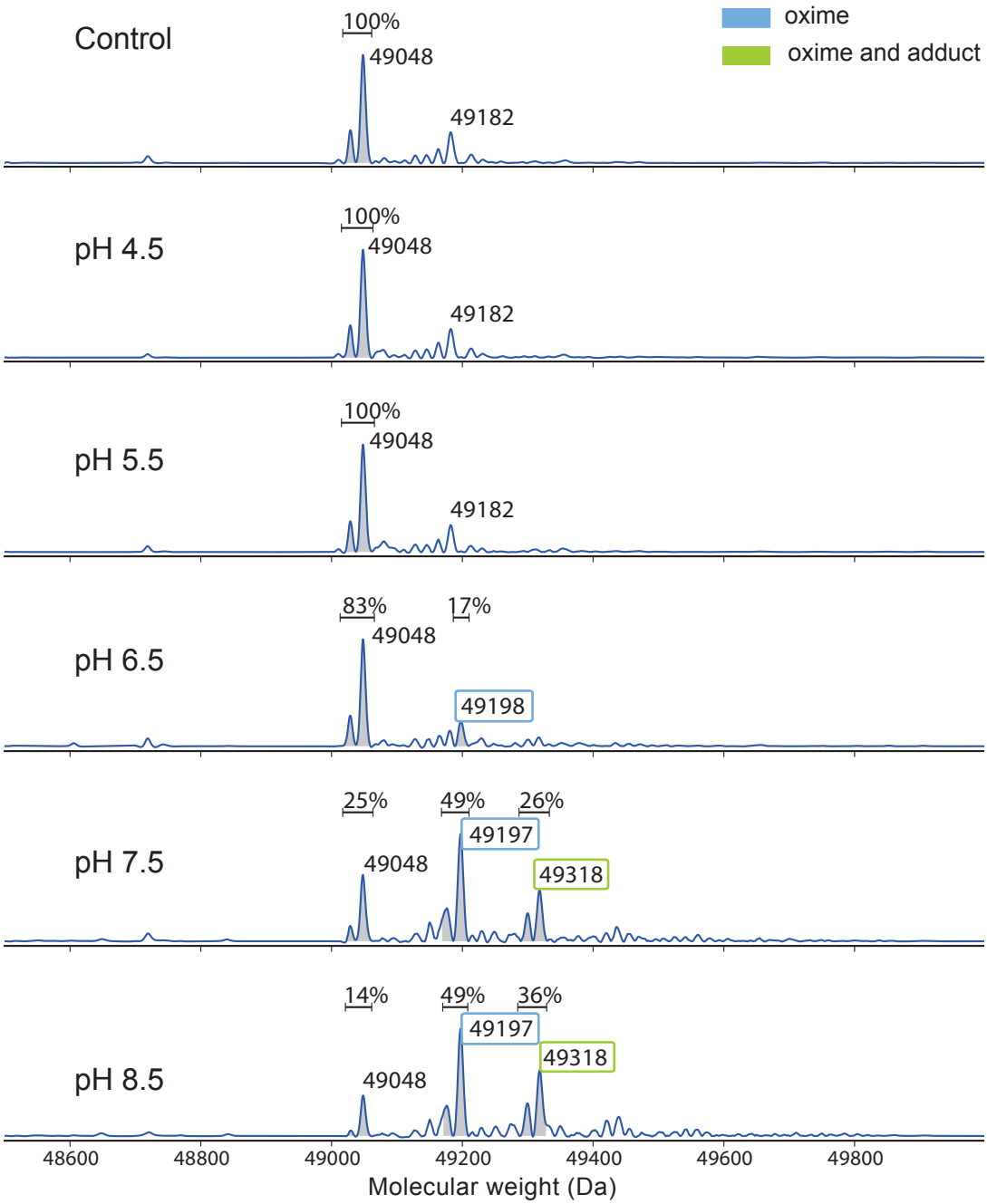


Figure 2.19. Deconvoluted LC-MS spectra of AEE-EGPh after RS transamination followed by oxime formation with nitrobenzyl alkoxyamine. The pH of RS transamination is noted next to each spectrum. Origin was used to baseline correct these data. The percentages represent the area underneath the peaks highlighted. Unmodified AEE-EGPh has a mass of 49048 Da.

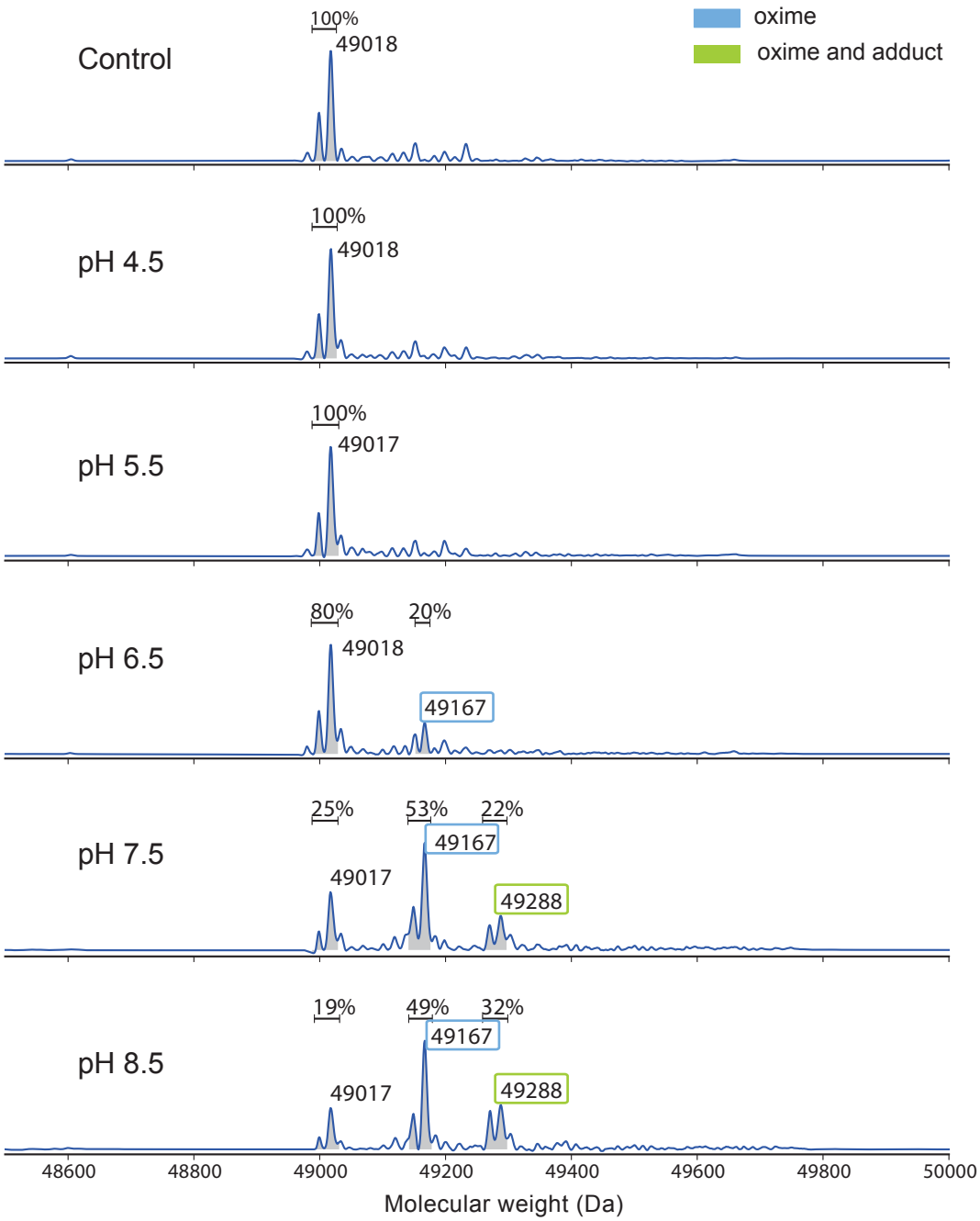


Figure 2.20. Deconvoluted LC-MS spectra of AKT-EGPh after RS transamination followed by oxime formation with nitrobenzyl alkoxyamine. The pH of RS transamination is noted next to each spectrum. Origin was used to baseline correct these data. The percentages represent the area underneath the peaks highlighted. Unmodified AKT-EGPh has a mass of 49017 Da.

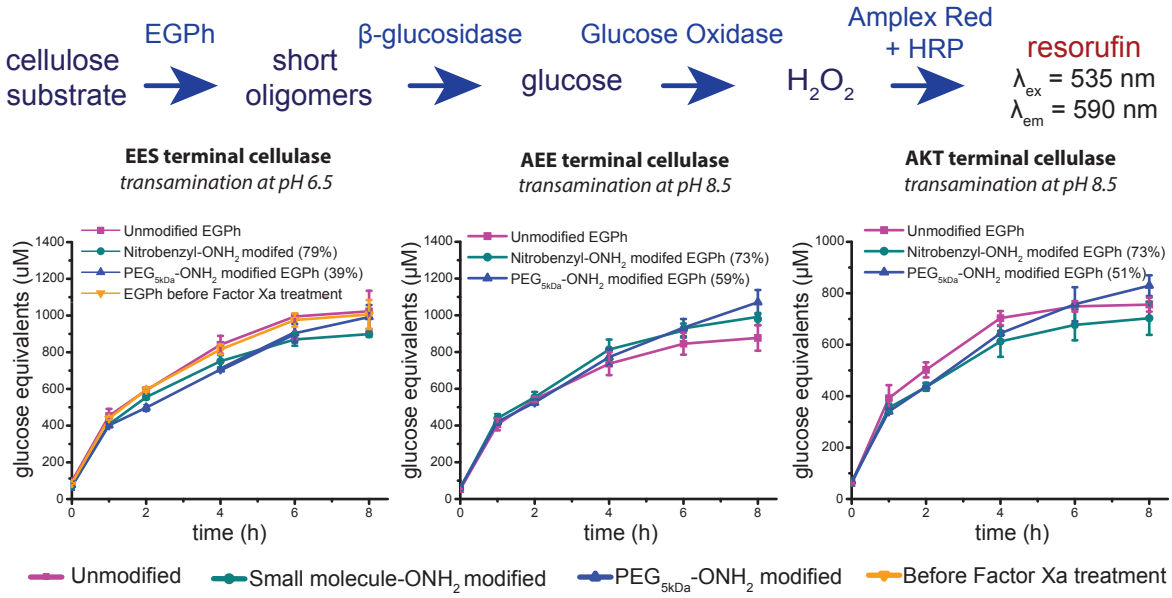


Figure 2.21. Quantification of cellulolytic activity following RS transamination and subsequent oxime formation with either a small molecule alkoxyamine or a 5 kDa PEG alkoxyamine. Samples were analyzed in triplicate and error bars indicate standard deviation. Three samples of each mutant were studied: unmodified protein that was not exposed to any modification steps (pink), RS transaminated proteins coupled to nitrobenzyl-ONH_2 (**2b**, green) and PEG_{5kDa}-ONH_2 (**2c**, blue). Additionally, the IEGREES terminal EGPh protein was used without exposure to Factor Xa. RS Transamination was carried out at pH 6.5 for (a) the EES terminal mutant and at pH 8.5 for both (b) the AEE and (c) the AKT terminal mutants. The modification levels within the samples used are denoted in the legend. All modified samples exhibited activity comparable to their unmodified controls over the span of 8 h, indicating that RS transamination, subsequent oximation and Factor Xa proteolysis all led to unaltered protein activity.

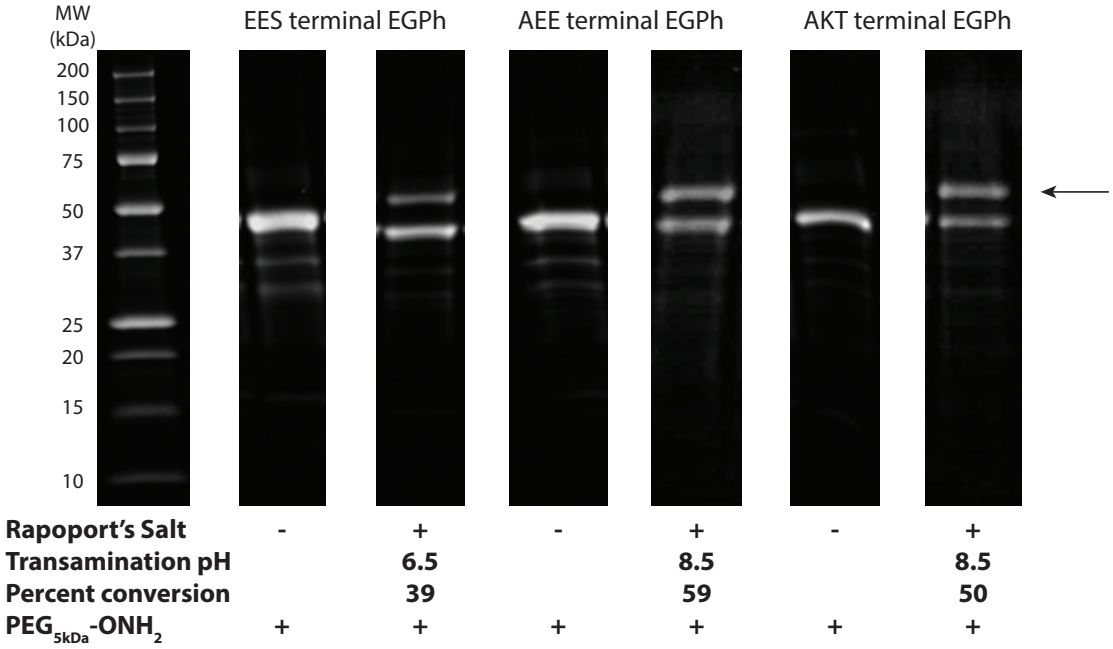


Figure 2.22. SDS-PAGE gel showing the attachment of a 5 kDa poly(ethylene glycol) (PEG) alkoxyamine to the cellulase mutants, used in the activity assays. Alkoxyamine attachment was carried out at pH 4.5 following RS mediated transamination at the pH indicated. The arrow points to the band corresponding to PEG attachment. The gel was visualized with SYPRO Ruby Protein Gel Stain, and gel densitometry was used to quantify conversion.

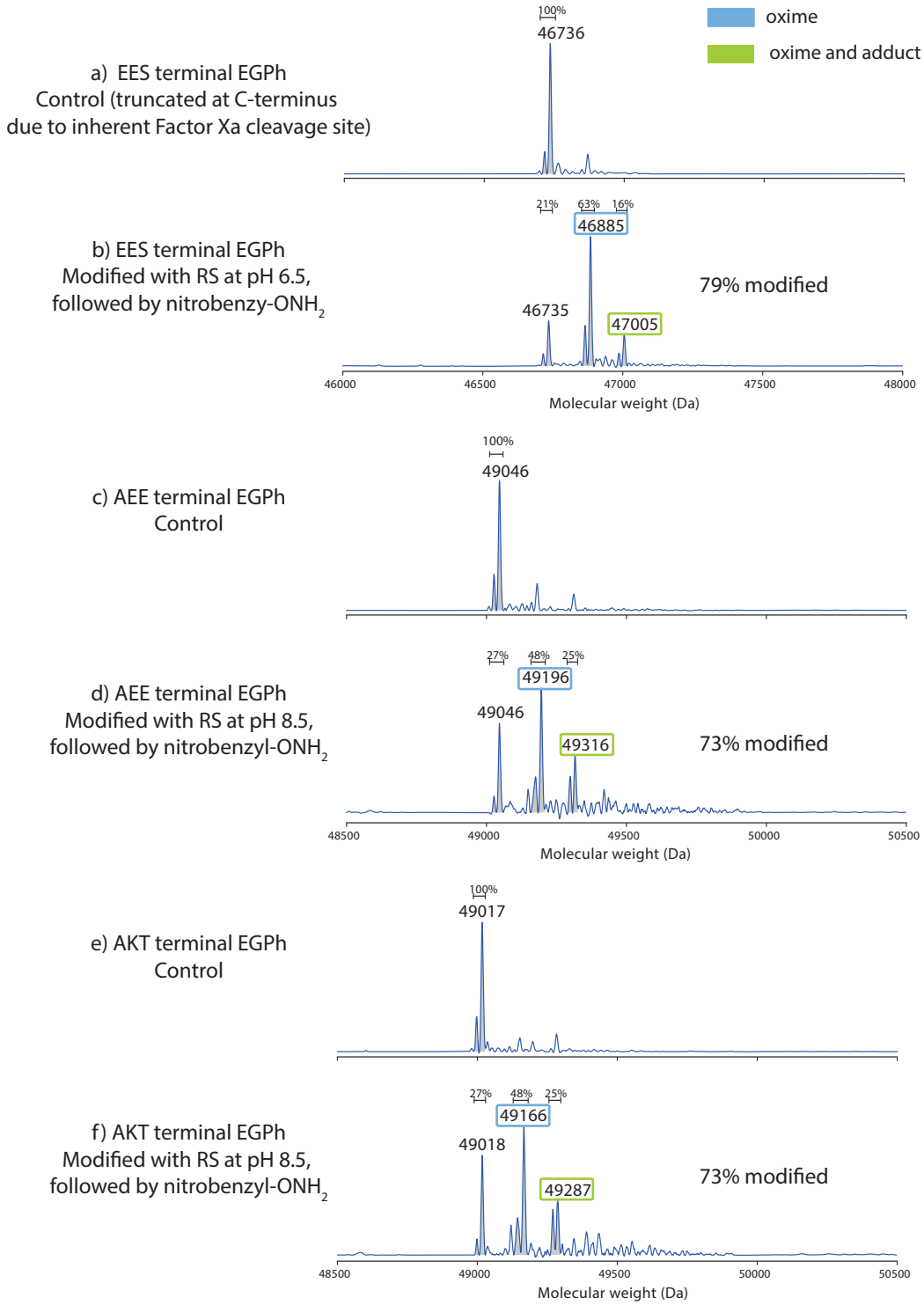


Figure 2.23. Deconvoluted LC-MS spectra of EES, AEE and AKT terminal EGPh samples modified with nitrobenzyl alkoxyamine after RS transamination, used in the activity assays. Transamination was carried out at pH 6.5 for EES terminal EGPh and at pH 8.5 for the AEE and AKT terminal proteins. Subsequent oxime formation was carried out at pH 4.5 for all three mutants. Controls were included where the reaction conditions remained the same, but the RS was omitted. Origin was used to baseline correct these data. The percentages represent the area underneath the highlighted peaks.

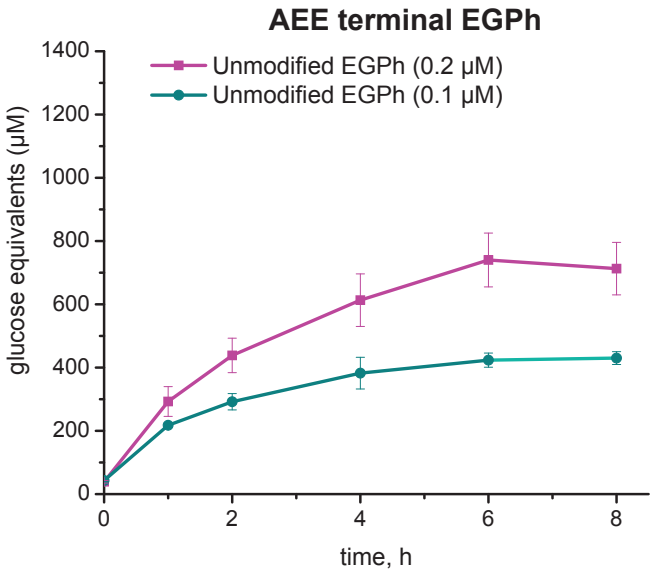


Figure 2.24. Quantification of cellulolytic activity of EGPh at 0.2 μM (same amount used in figure 5 assays), and at 0.1 μM (at 50% the original concentration) to ensure that for batches where modification was in the 50% range, activity was not solely due to the unmodified portion of protein within the sample. A decrease in rate and an approximately 50% lower turnover was observed in the 0.1 μM sample, a trend that was not observed in the modified samples of Figure 2.21.

2.3. Conclusions

The peptide and protein studies presented herein demonstrate the potential of RS to be a highly practical protein modification method. There are other effective and long standing N-terminal modification reactions, such as the oxidation of N-terminal serine residues with periodate. However, this method is incompatible with a number of protein classes. For example, the glycans on antibodies will result in periodate-based cleavage as a side reaction,[10] while the RS-mediated reaction does not suffer this competing pathway. Relative to other transamination reagents we have used, RS is inexpensive and easy to recrystallize to a high degree of purity. RS-mediated transamination enables the attachment of the synthetic group to the protein via an oxime linkage. We have found these N-terminal oximes to be stable over weeks, and even months, particularly when the protein is stored at 4 °C.[23] We also have tested the stability of oximes formed with N-terminal pyruvamide groups in previous studies.[47]

The work shown herein has validated and expanded the utility of Rapoport's salt as a site specific protein modification strategy. We have confirmed that an EES terminus is a high yielding sequence for pH values ranging from pH 6.5–8.5, and we have also observed that at the upper end of that range, transamination is less restricted to N-terminal sequences that possess negative charges and can be applied to different alanine terminal sequences with high yields (Figure 2.25a). This provides flexibility in the residues inserted at the N-terminus and lowers mutagenesis requirements. This is important because alanine residues at the N-terminus of a recombinantly expressed protein do not prevent cleavage of the starting methionine residue during post-translational processes, thus alleviating the need to insert a proteolytic cleavage site at the N-terminus of the protein.

In the proposed mechanism for transamination, the first step is the formation of a Schiff base at the N-terminus. This is followed by tautomerization due to proton abstraction at the alpha carbon. With PLP, slightly acidic conditions are required for the subsequent protonation of the nitrogen on the pyridine ring. However, RS is methylated at the analogous nitrogen, and thus

protonation is not required. This factor contributes to higher levels of RS reactivity at higher pH values.

Our combinatorial peptide library work with PLP and RS was only studied at pH 6.5. It was observed through our pH screens herein that RS reactivity for termini that did not contain a glutamate residue was enhanced at higher pH values. At low pH, the requirement for a glutamate residue in the first position could be due to its role as an internal base for the proton abstraction step. Replacing that residue at the first position with a residue that could not provide this type of anchimeric assistance results in the necessity for the reaction to be carried out at higher pH values with the buffer acting as an exogenous base in that step. (Figure 2.25b)

a)

N-terminus (XXXWSNAG)	Percent conversion at pH 6.5	Percent conversion at pH 8.5
AEE*	41.0%	87.7%
AKT*	80.0%	95.8%
ATT*	40.3%	82.7%
GEE*	28.4%	42.2%
PEE*	0.2%	0.4%
SEE*	29.8%	59.4%
TEE*	1.0%	4.6%
MEE*	28.8%	68.5%
**Termini that can be used to directly express proteins in bacteria*		
EES**	94.6%	97.5%
***Termini that do not allow post-translational methionine cleavage in bacteria and require the use of a proteolytic cleavage site for expression*		

b) *unmodified* *keto-protein*

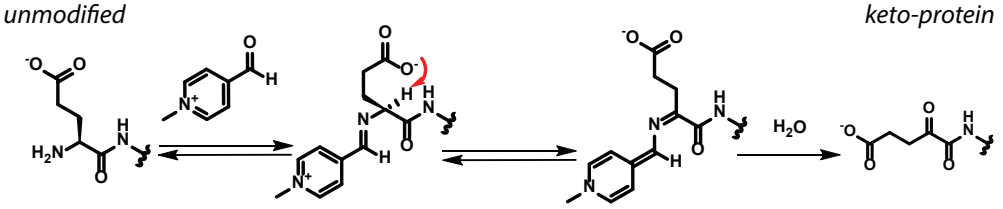


Figure 2.25. a) Summary of conversion results from the XXXWSNAG peptides studied with RS transamination at pH 6.5 and 8.5. At higher pH, peptides exhibited less dependance on nearby negative charges. Alanine terminal peptides proved most successful, as they led to high yields and do not prevent post-translational cleavage of methionine at the N-terminus in proteins expressed bacterially. The numbers are a sum of all products that retained the oxime linkage. b) Illustration of glutamate side chains serving as an endogenous base during the proposed mechanism of RS mediated transamination.

We have demonstrated that combinatorial peptide libraries can be used as a tool to accelerate 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. Using the information obtained from the library screening, we have demonstrated site-specific modification of glutamate-terminal monoclonal antibodies with good yields, and that it is possible to control whether the heavy chain, or the heavy and light chains are modified. To expand the applicability of RS-mediated transamination to bacterially expressed proteins, which do not allow for the direct expression of glutamate termini, we have also demonstrated the site-

specific modification of three mutants of a cellulase enzyme, EES-, AEE-, and AKT-EGPh in high yields. The enzymatic activity of the protein mutants remained intact following modification, thus confirming that RS-mediated transamination is a mild and simple method for the production of protein bioconjugates. The newly appreciated pH profile of RS-based transamination indicates that alanine terminal sequences give the highest yields of desired product among the residues screened, thus providing a clear design lead for expressing proteins in *E. coli* that are compatible with this chemistry. RS-mediate transamination of proteins is facile and should be readily scalable, thus providing a suitable method for the practical production of desired protein bioconjugates.

2.4. Materials and Methods

2.4.1. General Procedures and Materials.

Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H_2O) used as reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). *N*-methylpyridinium-4-carboxaldehyde benzenesulfonate hydrate (Rapoport's salt, RS) was obtained from Alfa Aesar (Ward Hill, MA). Pyridoxal 5'-phosphate monohydrate was obtained from Aldrich. Benzylalkoxyamine and nitrobenzyl alkoxyamine were obtained from Aldrich (St. Louis, MO). All Fmoc-protected amino acids were obtained from Novabiochem (EMD, Germany). TentaGel S OH resin was obtained from Advanced ChemTech (Louisville, KY). Wild-type anti-HER2 human IgG1 monoclonal antibodies were obtained from Eureka Therapeutics, Inc (Emeryville, CA). Wild-type anti-hTNFα human IgG1 monoclonal antibodies were obtained from Invivogen (San Diego, CA). Goat anti-human IgG (Fcγ specific) PerCP-conjugated $F(ab')_2$ fragments were purchased from Jackson ImmunoResearch Laboratories, Inc (West Grove, PA).Factor Xa protease was purchased from New England Biolabs (Ipswich, MA). Both Factor Xa removal resin and Ni–NTA resin were puchased from Qiagen (Valencia, CA). Samples were lyophilized using a Freezone 4.5 (LABCONCO, Kansas City, MO). Fluorescence measurements of samples in 96–well plates were obtained on a SpectraMax M2 (Molecular Devices, Sunnyvale, CA).

2.4.2. Solid-Phase Peptide Synthesis.

Peptides were synthesized using standard conditions for Fmoc-based chemistry. The side chain protecting groups used were: Asn(Trt), Asp(tBu), Arg(Pbf), Glu(tBu), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc). For the first step, the C-terminal amino acid (10 equiv.) was preactivated with 5 equivalents of diisopropylcarbodiimide (DIC), then coupled to the TentaGel S OH resin with 0.1 equivalents of *N*,*N*-dimethylaminopyridine (DMAP) as an additive. Deprotection of the Fmoc groups was accomplished by incubation with a 20% v/v piperidine/*N*,*N*-dimethylformamide (DMF) solution for 20 minutes. Subsequent 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) and 20 equivalents of *N*,*N*-diisopropylethylamine (DIPEA) in DMF for 10 minutes. Side-chain deprotection was achieved using a 1–2 h incubation with a 95:2.5:2.5 ratio of trifluoroacetic acid (TFA) to H_2O to triisopropylsilane (TIPS), followed by equilibration in 50 mM phosphate buffer (pH 6.5) using three 5 minute exposures. Cleavage was carried out in 100 mM aqueous NaOH for 30 minutes and the eluent was neutralized with 50 mM phosphate buffer

(pH 6.5) for subsequent HPLC purification.

Peptides were synthesized using standard conditions for Fmoc-based chemistry. The side chain protecting groups used were: Asn(Trt), Asp(tBU), Arg(Pbf), Cys(Trt), Gln(Trt), Glu(tBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), Tyr(tBu). For the first step, the C-terminal amino acid (10 equiv) was preactivated with 5 equivalents of diisopropylcarbodiimide (DIC), then coupled to the TentaGel S OH resin with 0.1 equivalents of *N*,*N*-dimethylaminopyridine (DMAP) as an additive. Deprotection of the Fmoc groups was accomplished by incubation with a 20% v/v piperidine/*N*,*N*-dimethylformamide (DMF) solution for 20 min. Subsequent 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) and 20 equivalents of *N*,*N*-diisopropylethylamine (DIPEA) in DMF for 10 min. Side-chain deprotection was achieved using a 1-2 h incubation with a 95:2.5:2.5 ratio of trifluoroacetic acid (TFA) to H_2O to triisopropylsilane (TIPS), followed by equilibration in 50 mM phosphate buffer (pH 6.5) using three 5 min exposures.

2.4.3. Split-Pool Library Synthesis with Partial Truncation.

The construction of the one-bead-one-sequence combinatorial peptide library with a built-in truncation ladder followed the previously reported technique.[25] The synthesis began with a 5-residue base sequence (WSNAG) prepared on 500 mg of TentaGel S OH resin. Partial capping was performed using 0.15 equivalents of bromobenzoic acid, 0.1 equivalents of 2-(6-chloro-1-*H*-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU) and 1 equivalent of *N*,*N*-diisopropylethylamine (DIPEA). The resin was then split into 20 different reaction vials, and each portion was exposed to peptide coupling conditions (10 equivalents of amino acid with 10 equivalents of HCTU and 20 equivalents of DIPEA in DMF for 10 min) using one of the twenty amino acids. After coupling and rinsing, the resins were then recombined, with the exception of the Gln and Leu samples, which were kept separate. After Fmoc deprotection, the next capping step was performed. Bromobenzoic acid was attached to the combined resin portion and methylbromobenzoic acid was used to cap the Gln and Leu samples. The Gln and Leu samples were mixed with the rest of the resin before splitting the combined library into separate vials for the next coupling reaction.

2.4.4. General Procedure for Library Screening.

Portions of resin-bound library (approx. 25 mg of resin at a time) were treated with 1 mL of freshly prepared 10 mM RS solution in 50 mM phosphate buffer, pH 6.5 with 0.02% NaN_3 and 10% DMF at rt. 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 ($BnONH_2$) in the specified ratio in a 1:1 H_2O: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 sequencing. The residual ethanol in the tubes was removed by pipetting.

2.4.5. General Procedure for Library Sequencing.

Individual beads identified in the library were incubated with 10 μL of 100 mM NaOH solution to cleave the peptide from the resin. The solution was desalted using Ziptips with 0.2 μL of C18 resin (Millipore, Billerica, MA). The peptide was eluted from the Ziptip with 2 μL of matrix solution (described below) directly onto a MALDI sample plate (Applied Biosystems, Foster City, CA). MALDI-TOF analysis was performed on a Voyager-DE instrument (Applied Biosystems) using reflector mode, and all spectra were analyzed using Data Explorer software. The truncation ladder peaks were identified from their bromine isotope pattern and the mass differences between these peaks were used to determine the sequence.[25] The matrix solution was a saturated α-cyano-4-hydroxycinnamic acid solution in 50% acetonitrile, 50% water, 0.1% TFA.

2.4.6. General Procedure for RS-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 RS solution (or the specified concentration) in 50 mM phosphate buffer, pH 6.5 with 0.02% NaN_3 and 10% DMF at rt. 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 $BnONH_2$ 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 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.

2.4.7. High Performance Liquid Chromatogaphy for the Purification of Peptides.

HPLC was performed on Agilent 1100 series HPLC systems (Agilent Technologies, USA) outfitted with an Agilent 1200 series automatic fraction collector. Sample analysis for all HPLC experiments was achieved with an inline diode array detector (DAD) and inline fluorescence detector (FLD). Semi-preparative reverse-phase HPLC of peptides was accomplished using a C18 stationary phase and a H_2O/MeCN with 0.1% TFA gradient mobile phase.

2.4.8. General Procedure for the RS-mediated Transamination of Peptides in Solution.

Peptide and RS stock solutions were prepared at twice the desired final concentrations and mixed in equal volumes in a 1.5 mL Eppendorf tube. The final volume of each reaction was 400 μL. The 2× peptide stock solutions were prepared at 50 μM in 25 mM phosphate buffer (with 0.02% NaN_3), at the desired pH. The 2× RS stock solution (200 mM) was freshly prepared before each reaction in 25 mM phosphate buffer (with 0.02% NaN_3) at the desired pH. The reaction mixture

was briefly agitated to ensure mixing and was then incubated without further agitation at 37 °C for 1 h. Following the reaction, the excess aldehyde was removed using a sep-pak C18 1 mL vac cartridge (Waters, Milford, MA). Elution was carried out with 100% acetonitrile. The resulting keto-peptide solution was concentrated through evaporation and lyophilization. The white lyophilized powder was resuspended in 40 μL of 25 mM phosphate buffer (with 0.02% NaN_3), pH 5.0, and was then treated with the alkoxyamine stock solution in a 1.5 mL Eppendorf tube and incubated at room temperature for 48 h. The alkoxyamine stock solution used was a 250 mM $BnONH_2$ solution (in water, pH adjusted to 5.0). A 10 μL portion of the stock solution was added to give a final concentration of 50 mM $BnONH_2$. Controls were conducted following the same procedure, but without RS. After oxime formation, the modified peptide sample was analyzed using mass spectrometry.

2.4.9. 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,[34] 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 ACCTTTTTTTacattgaagtgcag
F0 ctgcacttcaatgtAAAAAAAGGTCACCATGTACAGGATGCA
R24 GCAATGCAAGACAGGAGTTGCATCCTGTACATGGTG
F42 ACTCCTGTCTTGCATTGCACTAAGTCTTGCACTTGTCA
R60 TCAGTCTTAGCCGAATTCGTGACAAGTGCAAGACTTAGT
F80 CGAATTCGGCTAAGACTGACATCCAAATGACTCAGAGCC
R99 GCGCTCAGGGAACTGGGGCTCTGAGTCATTTGGATG
F119 CCAGTTCCCTGAGCGCTTCCGTAGGGGACAGG
R135 GCCCGACATGTTATTGTCACCCTGTCCCCTACGGAA
F151 GTGACAATAACATGTCGGGCTAGCCAGGATGTCAATACAG
R171 CTGGTACCAAGCGACAGCTGTATTGACATCCTGGCTA
F191 CTGTCGCTTGGTACCAGCAAAAGCCCGGAAAGGC
R208 GCTGTATATAAGAAGCTTTGGCGCCTTTCCGGGCTTTTG
F225 GCCAAAGCTTCTTATATACAGCGCCAGTTTCCTCTATTCTGG
R247 GAACCTGCTCGGCACGCCAGAATAGAGGAAACTGGC
F267 CGTGCCGAGCAGGTTCTCTGGATCTCGGTCCG
R283 TCAGTGTGAAATCGGTCCCGGACCGAGATCCAGA
F299 GGACCGATTTCACACTGACCATTAGTTCTCTGCAGCC
R317 TAGTATGTTGCAAAGTCCTCTGGCTGCAGAGAACTAATGG
F336 AGAGGACTTTGCAACATACTACTGCCAGCAGCACTAT
R357 AGGTTGGGGGTGTGGTATAGTGCTGCTGGCAG
F373 ACCACACCCCCAACCTTTGGTCAGGGCACGAA
R389 CGTACGCTTGATTTCCACCTTCGTGCCCTGACCAA
F405 GGTGGAAATCAAGCGTACGAAAAAAAcccccaactttgt
F424 acaaagttgggggTTTTTTT

An additional GCTAAAACT was added to the 5' end according to a published procedure[48] in order to create a three residue N-terminal extension (AKT, for use in other work and later mutated to EES for this work). The resulting V_L gene was inserted into a vector at BsiWI and BstEII restriction sites using standard cloning techniques. The vector used, pFUSE2-CLIg-hk from Invivogen (San Diego, CA), already contained the constant region of the kappa light chain (Figure S8) A Quikchange site-directed mutagenesis kit (Stratagene, La Jolla, CA) was used to generate the desired N-terminal mutant (EES). Incorporation of these mutations was verified by sequencing.

2.4.10. 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 light chain. In brief, the variable and constant region 1 of heavy chain (V_H and C_H1) was constructed from the following set of oligonucleotides with additional bases (CTCCAAACA) at the 5' end (corresponding to three N-terminal residues, LQT, for use in other work and later mutated to EES for this work).

R0 TTTTTTTcttagctgctttga
F0 tcaaagcagctaagAAAAAAAGAATTCGCTCCAAACAG
R21 CGACGAGTTGGACTTCTGTTTGGAGCGAATTC
F38 AAGTCCAACTCGTCGAAAGCGGAGGTGGC
R53 CCAGGCTGAACCAGGCCACCTCCGCTTT
F67 CTGGTTCAGCCTGGCGGAAGCCTGCGC
R81 GCAGCACAGCTCAAGCGCAGGCTTCCG
F94 TTGAGCTGTGCTGCCTCCGGATTTAATATCAAAGA
R108 CGAACCCAGTGTATATAAGTATCTTTGATATTAAATCCGGAG
F129 TACTTATATACACTGGGTTCGCCAGGCTCCTGGA
R150 CCACTCCAGACCCTTTCCAGGAGCCTGG
F163 AAGGGTCTGGAGTGGGTGGCGAGAATCTACC
R178 GGGTATAACCATTGGTTGGGTAGATTCTCGCCAC
F194 CAACCAATGGTTATACCCGCTATGCAGACAGCG
R212 GTAAACCGCCCTTTCACGCTGTCTGCATAGC
F227 TGAAAGGGCGGTTTACAATTAGTGCCGACACA
R243 GGTAAGCGGTATTTTTAGATGTGTCGGCACTAATT
F259 TCTAAAAATACCGCTTACCTCCAGATGAACTCTCTG
R278 TGTCCTCGGCCCTCAGAGAGTTCATCTGGA
F295 AGGGCCGAGGACACGGCTGTGTATTATTGC
R308 CACCCCACCGGCTGCAATAATACACAGCCG
F325 AGCCGGTGGGGTGGAGACGGATTCTATGCT
R338 TGACCCCAATAGTCCATAGCATAGAATCCGTCTC
F355 ATGGACTATTGGGGTCAGGGCACTCTCGTCA
R372 TGGCACTGCTTACAGTGACGAGAGTGCCC
F386 CTGTAAGCAGTGCCAGCACAAAGGGGCC
R401 CAAGGGGAAAGACACTAGGCCCCTTTGTGC
F414 TAGTGTCTTTCCCCTTGCTCCATCTAGCAAATCTAC

R431 GGTGCCCCCGCTGGTAGATTTGCTAGATGGAG
F450 CAGCGGGGGCACCGCCGCCCTGGGAT
R463 GTCCTTGACCAGGCATCCCAGGGCGGC
F476 GCCTGGTCAAGGACTATTTTCCTGAGCCAGT
R490 TCCAGGACACGGTGACTGGCTCAGGAAAATA
F507 CACCGTGTCCTGGAATAGTGGCGCCTTGA
R521 TGTGTGAACACCAGAAGTCAAGGCGCCACTAT
F536 CTTCTGGTGTTCACACATTTCCCGCCGTCC
R553 CAGCCCACTAGATTGAAGGACGGCGGGAAA
F566 TTCAATCTAGTGGGCTGTACTCTCTCTCCAGTGT
R583 TGGGTACCGTCACCACACTGGAGAGAGAGTA
F600 GGTGACGGTACCCAGTTCAAGCTTGGGCA
R614 TGCAGATATAGGTCTGTGTGCCCAAGCTTGAAC
F629 CACAGACCTATATCTGCAATGTGAACCACAAGCC
R647 CCACCTTTGTATTGCTGGGCTTGTGGTTCACAT
F663 CAGCAATACAAAGGTGGACAAAAAAGTCGAGCCT
R680 TGTCACAGCTCTTTGGAGGCTCGACTTTTTTGT
F697 CCAAAGAGCTGTGACAAAACTCACACATGCCC
R713 TACCTGGGCACGGTGGGCATGTGTGAGTTT
F729 ACCGTGCCCAGGTAAGCCAGCCCAGGC
R743 ccccattgactTTTTTTTAGGCCTGGGCTGGCT
R756 CTAAAAAAAagtcaatgggg

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

The gene encoding V_H and C_H1 was inserted into a vector comprising the crystallizable fragment (F_c) domain (i.e. C_H2 and C_H3 domains) of human IgG1 heavy chain (pinfuse-higG1-fc2 from Invivogen) at the EcoRI and BglII restriction sites (Figure S8). A Quikchange site-directed mutagenesis kit was used to generate the desired N-terminal mutant (EES). Incorporation of these mutations was verified by sequencing.

2.4.11. General Procedure for Expression and Purification of 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 5% CO_2. After two 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.

2.4.12. 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 µL of G7 reaction buffer, 4 µL PNGase, and additional PBS to a total volume of 100 µL. The mixture was then incubated at 37 °C overnight. Immediately following treatment with PNGase, buffer exchange was performed into 100 mM Tris buffer, pH 8. Dithiothrietol (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 °C for 30 min. The samples were then subjected to buffer exchange into 10 mM Tris buffer pH 8 for mass spectrometry analysis.

2.4.13. 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 × 180 µm) and analytical (1.7 µm particles, 100 mm × 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 µL/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 °C and 8 °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 °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).

2.4.14. LC-MS Analysis of Synthetic Peptide Bioconjugates.

Electrospray ionization mass spectrometry (ESI-MS) of peptides was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 Time-of-Flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a XBridge PST C18 (3.5 μm particles, 2.1 mm × 150 mm, Waters, Milford, MA) analytical column. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). For each sample, approximately 5 to 10 picomoles of analyte was injected onto the column. Following sample injection, a 20-100% B elution gradient was run at a flow rate of 0.25 mL/min for 25 min. Data were collected and analyzed using Agilent MassHunter Qualitative Analysis B.05.00.

2.4.15. General Procedure for RS-Mediated Transamination of Antibody Substrates.

Protein and RS 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 μL. The 2x protein stock solutions were prepared at 0.5 - 1 mg/mL in 25 mM phosphate buffer at pH 6.5. The 2x RS stock solution (200 mM) was freshly prepared before each reaction in 25 mM phosphate buffer (with 0.02% NaN_3), pH 6.5 from RS (recrystallized from acetonitrile). The reaction mixture was briefly agitated to ensure mixing and then incubated without further agitation at 37 °C for 1 h. Following the reaction, the excess aldehyde was removed using NAP Sephadex size exclusion columns (GE Healthcare, USA). The resulting keto-protein solution was then concentrated and buffer exchanged 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 μL with 25 mM phosphate buffer (pH 6.5). Each sample was then concentrated to 100 μL, and the process was repeated 3 times. The resulting keto-protein was then treated with the alkoxyamine stock solution of choice in a 1.5 mL Eppendorf tube and incubated at rt for 48 h. The alkoxyamine stock solutions, and their final concentrations used were: 125 mM $BnONH_2$ (in water with the pH adjusted to 5.5), 50 mM PEG(2kDa)-ONH_2[35] solution (in water), 5 mM PEG(5kDa)-ONH_2 solution (in water), and 2 mM AlexaFluor488-ONH_2 solution (in water and DMSO). For PEG and AlexaFluor488 modification, 100 mM of aniline was also used as a catalyst. After oxime formation,

the NAP column and buffer exchange steps were again repeated to remove the excess alkoxyamine to stop the reaction.

2.4.16. LC-MS Analysis of Reduced Antibody Bioconjugates.

Electrospray ionization mass spectrometry (ESI-MS) of peptides was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 Time-of-Flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a Poroshell 300SB-C18 (5 μm particles, 1.0 mm × 75 mm, Agilent, Santa Clara, CA) analytical column. 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 1260 autosampler compartment prior to analysis. For each sample, approximately 5 to 10 picomoles of analyte was injected onto the column. Following sample injection, a 20-100% B elution gradient was run at a flow rate of 0.55 mL/min for 7 min. Data was collected and analyzed using Agilent MassHunter Qualitative Analysis B.05.00.

2.4.17. Flow Cytometry Analysis of Antibody Bioconjugates.

MCF7 clone 18 (MCF7cl18), a human breast cancer cell line overexpressing HER2, cells were obtained from the Preclinical Therapeutics Core Facility, UCSF, and grown in Dulbecco's modified Eagle medium (DMEM) containing 10% fetal bovine serum (FBS) at 37 °C in 5% CO_2. Jurkat cells (negative control) were obtained from ATCC and grown in RPMI containing 10% FBS. For the binding experiment, MCF7 clone 18 cells were first trypsinized at 37 °C for 5 min, followed by the addition of complete media (DMEM + 10% FBS) to stop trypsinization. Both MCF7 clone 18 and Jurkat cells were then pelleted and resuspended in binding buffer (Dulbecco's phosphate buffered saline (DPBS) containing 1% FBS) to the density of $1x10^6$ cells/mL. Aliquots of 100 μL containing $1x10^6$ cells/mL of cells were incubated with 15 and 25 nM of the unmodified or modified antibodies for 45 min on ice. The cells were then washed twice with 150 μL and resuspended in 100 μL of binding buffer containing 1:1000 dil. goat anti-human IgG, Fcγ fragment specific PerCP-conjugated. The cells were incubated for 30 min on ice in the dark, then washed twice with 150 μL, and resuspended in 200 μL of binding buffer. The cells were analyzed by flow cytometry (FACSCalibur flow cytometer, BD Biosciences) to determine the amount of AlexaFluor 488 and PerCP fluorescence. For each sample, 10,000 cells were counted.

2.4.18. SDS-PAGE Analysis.

For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini-Protean apparatus from Bio-Rad (Hercules, CA), following the general protocol of Laemmli.[49] All protein electrophoresis samples were heated for at least 15 minutes at 100 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of any disulfide bonds. Gels were run for 75 min at 120 V to allow good resolution of bands. Commercially available markers (Bio-Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie staining or SYPRO Ruby Protein Gel (Life Technologies), following the manufacturer's protocol. Gel imaging was performed on an EpiChem3 Darkroom system (UVP, USA). ImageJ was used to determine the

level of modification by optical densitometry.

2.4.19. EGPh Plasmid Construction, Expression, and Purification.

Construction of the AKT N-terminal EGPh has been reported previously.[40] This plasmid was used to obtain the AEE-EGPh mutant, through a one-step site-directed mutagenesis using the following set of primers:

Sense: 5'- GAA GGA GAT ATA CAT ATG GCT GAA GAA CTG TTT GGT CAG G -3'
Antisense: 5'- GG AAC GAC CTG ACC AAA CAG TTC TTC AGC CAT ATG -3'

The mutated plasmid with AEE inserted at the N-terminus was transformed into XL1-blue competent *E. coli* cells (Invitrogen), and the cells were plated on LB agar plates containing kanamycin (50 μg/mL). Colonies were grown up in 4 mL overnight cultures, and the resulting plasmids were purified using a QIAprep Spin Miniprep Kit (Qiagen). The incorporation of the desired mutations was confirmed through sequencing (Sequetech, Mountain View, CA). Once confirmed, the AEE-EGPh plasmids were transformed into One Shot® BL21 (DE3) *E. coli* cells (Invitrogen) via heat shock and plated on Luria broth (LB) agar plates containing kanamycin (50 μg/mL). Cultures were grown in 1 L of LB containing kanamycin (50 μg/mL) at 37 °C until an optical density (OD) of 0.5 was observed at 600 nm. Expression of AEE- EGPh was induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were grown for 12–16 h at 37 °C and then spun down at 11000 rcf at 4 °C for 40 minutes to pellet the cells. The cells were resuspended in 8–12 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), lysed by sonication using a blunt tip for 20 min, and cell debris removed by centrifugation at 4,700 rpm at 4 °C for 40 minutes. The cleared cell lysate was incubated with 4 mL of rinsed nickel-nitrilotriacetic acid resin (Ni-NTA) in a 20 mL cartridge for 1 h at 4 °C. The resin-bound protein was washed with two 11 mL portions of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0), then eluted with four 1 mL portions of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The purified protein was buffer exchanged into 25 mM phosphate buffer (pH 7.5) through ultrafiltration (10 kDa MWCO) to yield 85–150 mg of purified protein per liter of culture.

The IEGREES-EGPh mutant was also obtained through three consecutive rounds of site-directed mutagenesis. The AEE-EGPh plasmid was used to make the first mutation, to obtain AEES-EGPh. The AEES-EGPh plasmid was then used to obtain the GREES-EGPh plasmid and the final step was to use the GREES-EGPh plasmid to obtain the desired IEGREES-EGPh plasmid. The incoporation of the desired mutations after each round was confirmed through sequencing. The primers used for each round were:

AEE to AEES:
Sense: 5'- GAA GGA GAT ATA CAT ATG GCT GAA GAA TCT CTG TTT GGT CAG G -3'
Antisense: 5'- GG AAC GAC CTG ACC AAA CAG AGA TTC TTC AGC CAT ATG -3'

AEES to GREES:
Sense: 5'- GGA GAT ATA CAT ATG GGT CGT GAA GAA TCT CTG TTT GG -3'
Antisense: 5'- GAC CTG ACC AAA CAG AGA TTC TTC ACG ACC CAT ATG -3'

GREES to IEGREES:
Sense: 5'- GGA GAT ATA CAT ATG ATC GAA GGT CGT GAA GAA TCT CTG TTT -3'
Antisense: 5'- CCT GAC CAA ACA GAG ATT CTT CAC GAC CTT CGA TCA TAT G -3'
Once the IEGREES-EGPh plasmid was obtained, the protein was expressed and purified by the same protocol outlined above for the AEE-EGPh mutant.

2.4.20. General procedure for using Factor Xa on IEGREES-EGPh to Obtain a Pure EES-EGPh Mutant.

A 1 mg/mL solution of IEGREES-EGPh was prepared in Factor Xa reaction buffer (20 mM Tris Cl pH 6.5, 50 mM NaCl, 1 mM $CaCl_2$). Factor Xa cleavage was performed at room temperature and following cleavage, was removed using Factor Xa removal resin. Both steps were carried out according the the manufacturer's protocol. Following removal of Factor Xa, the solution was buffer exchanged into Ni-NTA lysis buffer, and an affinity column was run according to the same protocol outlined in the AEE-EGPh purification protocol.

2.4.21. LC-MS Analysis of EGPh Bioconjugates.

Electrospray ionization mass spectrometry (ESI-MS) of the proteins was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 Time-of-Flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a Poroshell 300SB-C18 (5 μm particles, 1.0 mm × 75 mm, Agilent, Santa Clara, CA) analytical column. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). For each sample, approximately 5 to 10 picomoles of analyte was injected onto the column. Following sample injection, a 20-100% B elution gradient was run at a flow rate of 0.55 mL/min for 7 min. Data was collected and analyzed using Agilent MassHunter Qualitative Analysis B.05.00.

2.4.22. General Procedure for RS-mediated Transamination of EGPh.

Protein and RS stock solutions were prepared at twice the desired final concentrations and mixed in equal volumes in a 1.5 mL Eppendorf tube. The final volume of each reaction was 200 μL. The 2× protein stock solutions were prepared at 50 μM in 25 mM phosphate buffer (with 0.02% NaN_3) at the desired pH. The 2× RS stock solution (200 mM) was freshly prepared before each reaction in 25 mM phosphate buffer (with 0.02% NaN_3) at the desired pH (using RS recrystallized from acetonitrile). The reaction mixture was briefly agitated to ensure mixing and then incubated without further agitation at 37 °C for 1 h. Following the reaction, the excess aldehyde was removed and the resulting keto-protein solution was concentrated and buffer exchanged 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 μL with 25 mM phosphate buffer (pH 4.5). Each sample was then concentrated to 100 μL, and the process was repeated 5 times. The resulting keto-protein was then treated with an equal volume of the alkoxyamine stock solution of choice in a 1.5 mL Eppendorf tube and incubated at room temperature for 48 h. The alkoxyamine stock solution concentrations were the following: 62.5 μM nitrobenzyl-ONH_2 (in 25 mM phosphate buffer with the pH adjusted to 5.0), and 5 mM PEG_{5kDa}-ONH_2 solution (adjusted to pH 5.0 in water). After oxime formation, the protein concentration and buffer exchange steps were again repeated

to remove the excess alkoxyamine. Controls were conducted following the same procedure but without RS. Analysis of the protein modification was carried out with mass spectrometry for the nitrobenzyl-ONH_2 modified samples, and with an SDS-PAGE gel for the PEG_{5kDa}-ONH_2 modified samples (Figures 2.22 and 2.23).

2.4.23. EGPh Quantification.

EGPh concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific), with an extinction coefficient at 280 nm of 139,020 $M^{-1}cm^{-1}$. The molecular weights of the different mutants were: IEGREES-EGPh: 49,524 Da, EES-EGPh: 46,742 Da, AEE-EGPh: 49,052 Da, AKT-EGPh: 49,023 Da.

2.4.24. Activity of Modified EGPh.

To measure enzymatic activity, each protein sample was assayed in triplicate. Portions (1.3 mL) of a 1% (w/v) suspension of Sigmacell cellulose powder (Sigma-Aldrich) in 50 mM NaOAc buffer (pH 4.5) was added to 1.5 mL Eppendorf tubes containing magnetic stir bars. Stock solutions of each protein sample were prepared ranging from 21.4 to 26.6 μM, and 9.85 to 12.23 μL of these stocks were added to the appropriate Eppendorf tube to achieve a final protein concentration of 0.2 μM. A t = 0 h sample was taken from each reaction tube, after which they were placed in a 40 °C water bath on a stir plate. Samples were taken after 0, 2, 4, 6, and 8 h for analysis. To sample the reactions, each tube was shaken vigorously to ensure even distribution of the substrate and protein, and a 100 μL aliquot was immediately removed and transferred to a clean, empty Eppendorf tube. This aliquot was centrifuged for 1 min at 13.3k rpm, and then the clarified supernatant was transferred to a 0.6 mL Eppendorf tube and immediately frozen on dry ice. The samples were stored at –20 °C until analysis for the amount of soluble reducing sugar.

2.4.25. Analysis of Soluble Reducing Sugar.

This procedure was performed following a previously reported method,[50] using a paired glucose oxidase–peroxidase assay with OxiRed as the substrate. Analysis was performed in clear-bottom plastic 96-well plates, with each sample analyzed in triplicate. Internal standards of 300, 200, 100, 50, 25, and 0 μM glucose, and 150, 100, 50, 25, and 12.5 μM cellobiose in pH 4.5 buffer were included in each plate. Frozen aliquots from the activity assays were thawed on ice and then diluted 0- to 20- fold with cold buffer, and then 8 μL of the solution was incbated with 8 μL of β-glucosidase (5 mg/mL in 10 mM NaOAc pH 4.6) for 60 min at 37 °C to convert all of the cellobiose to glucose. The amount of glucose present was then measured by adding 65 μL of glucose oxidase (1.25 U/mL), horseradish peroxidase (1.25 U/mL), and OxiRed (60 μM) in 125 mM phosphate buffer (pH 7.45) and incubating at room temperature for 10 min in the dark. The amount of resorufin formed was measured on an optical plate reader with excitation at 535 nm and emission detection at 590 nm. The amount of resorufin formed corresponded to the amount of glucose present. Linear standard curves were made from the internal standards in each plate (all $R^2 > 0.97$), which were then used to calculate the amount of glucose equivalents present in each activity assay sample. The triplicate measurements of each supernatant sample were averaged, and then the measurements of the triplicate activity assay samples were averaged to calculate each data point.

2.5. References

1. Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M,; Weng, C. H.; Iavarone, A. T.; Francis, M. B., Site-specific protein transamination using N-methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **2013**, *135*, 17223–17229.

2. Palla, K. S.; Witus, L. S.; Mackenzie, K. J.; Netirojjanakul, C.; Francis, M. B., Optimization and expansion of a site selective N-methylpyridinium-4-carboxaldehyde mediated transamination for bacterially expressed proteins. *J. Am. Chem. Soc.* **2015**, *137*, 1123–1129.

3. Day, J. J.; Marquez, B. V.; Beck, H. E.; Aweda, T. A.; Gawande, P. D.; Meares, C. F., Chemically modified antibodies as diagnostic imaging agents. *Curr. Opin. Chem. Biol.* 2010, *14*, 803–809.

4. 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.

5. 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.

6. 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.

7. Vasita, R.; Shanmugam, K.; Katti, D. S., Improved biomaterials for tissue engineering applications: surface modification of polymers. *Curr. Top. Med. Chem.* **2008**, *8*, 341–353.

8. Basle, E.; Joubert, N.; Pucheault, M., Protein chemical modification on endogenous amino acids. *Chem. Biol.* **2010**, *17*, 213–227.

9. Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R., Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960–1965.

10. Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C. -F., A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed.* **2011**, *50*, 9611–9614.

11. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776–779.

12. Dawson, P. E.; Kent, S. B., Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

13. 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. *Bioconjugate Chem.* **1992**, *3*, 138–146.

14. 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.

15. Obermeyer, A. C.; Jarman, J. B.; Francis, M. B., N-terminal modification of proteins with *o*-aminophenols. *J. Am. Chem. Soc.* **2014**, *136*, 9572–9579.

16. Dixon, H. B. F., N-terminal modification of proteins—a review. *J. Protein Chem.* **1984**, *3*, 99–108.

17. Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C.H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; Lu, Y.; Tran, H.; Seller, A. J.; Biroc, S. L.; Szydlik, A.; Pinkstaff, J. K.; Tian, F.; Sinha, S. C.; Felding-Habermann, B.; Smider, V. V.; Schultz, P. G., Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 16101–16106.

18. Lin, C.-W.; Ting, A. Y., Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.

19. Rashidian, M.; Dozier, J. K.; Distefano, M. D., Enzymatic labeling of proteins: Techniques and approaches. *Bioconjugate Chem.* **2013**, *24*, 1277–1294.

20. Sletten, E. M.; Bertozzi, C. R., Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem., Int. Ed.* **2009**, *48*, 6974–6998.

21. Witus, L. S.; Francis, M. B., Site-specific protein bioconjugation via a pyridoxal 5'-phosphate-mediated N-terminal transamination reaction. *Curr. Prot. Chem. Biol.* **2010**, *3*, 1-10.

22. Jencks, W. P., Studies on the mechanism of oxime and semicarbazone formation. *J. Am. Chem. Soc.* **1959**, *81*, 475–481.

23. Kalia, J.; Raines, R. T., Hydrolytic stability of hydrazones and oximes. *Angew. Chem. Int. Ed.* **2008**, *47*, 7523–7526.

24. 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.

25. 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.

26. Buckley, T. F.; Rapoport, H., Mild and simple biomimetic conversion of amines to carbonyl compounds. *J. Am. Chem. Soc.* **1982**, *104*, 4446–4450.

27. Liu, H.; Gaza-Bulseco, G.; Faldu, D.; Chumsae, C.; Sun, J., Heterogeneity of monoclonal antibodies. *J. Pharm. Sciences* **2008**, *97*, 2426–2447.

28. Chelius, D.; Jing, K.; Lueras, A.; Rehder, D. S.; Dillon, T. M.; Vizel, 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.

29. Yu, L.; Vizel, A.; Huff, M. B.; Young, M.; Remmele Jr., 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.

30. Wu, A. M.; Senter, P. D., Arming antibodies: prospects and challenges for immunoconjugates. *Nat. Biotech.* **2005**, *23*, 1137–1146.

31. Hamblett, K. J.; Senter, P. D.; Chace, D. F.; Sun, M. M. C.; Lenox, J.; Cerveny, 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.

32. 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. U.S.A.* **1989**, *86*, 8247–8251.

33. Lam, K.S.; Lebl, M.; Krchňák, V., The "one-bead-one-compound" combinatorial library method. *Chem. Rev.* **1997**, *97*, 411–448.

34. 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.

35. Marusic, J.; Podlipnik, C.; Jevsevar, S.; Kuzman, D.; Vesnaver, G.; Lah, J., Recognition of human tumor necrosis factor α (TNF-α) by therapeutic antibody fragment. *J. Biol. Chem.* **2012**, *287*, 8613–8620.

36. Harris, L.J.; Skaletsky, E.; McPherson, A., Crystallographic structure of an intact IgG1 monoclonal antibody. *J. Molec. Biol.* **1998**, *275*, 861–872.

37. 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.

38. 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.

39. Yeoman, C. J.; Han, Y.; Dodd, D.; Schroeder, C. M.; Mackie, R. I.; Cann, I. K. O., Thermostable enzymes as biocatalysts in the biofuel industry. *Adv. Appl. Microbiol.* **2010**, *70*, 1–55.

40. Mackenzie, K. J.; Francis, M. B., Recyclable thermoresponsive polymer–cellulase bioconjugates for biomass depolymerization. *J. Am. Chem. Soc.* **2013**, *134*, 293–300.

41. Ando, S.; Ishida, H.; Kosugi, Y.; Ishikawa, K., Hyperthermostable endoglucanase from *Pyrococcus horikoshii*. *Appl Environ Microbiol.* **2002**, *68*, 430–433.

42. Kim, H.-W.; Ishikawa, K., Structure of hyperthermophilic endocellulase from *Pyrococcus*

horikoshii. *Proteins* **2010**, *78*, 496–500.

43. Nagai, K.; Thogersen, H. C., Generation of beta-globin by sequence-specific proteolysis of a hybrid protein produced in *Escherichia coli*. *Nature* **1984**, *309*, 810–812.

44. Jenny, R. J.; Mann, K. G.; Lundblad, R. L., A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression Purif.* **2003**, *31*, 1–11.

45. Eliot, A. C.; Kirsch, J. F., Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* **2004**, *73*, 383–415.

46. Kataoka, M.; Ikemi, M.; Morikawa, T.; Miyoshi, T.; Nishi, K.; Wada, M., Isolation and characterization of D-threonine aldolase, a pyridoxal-5'-phosphate-dependent enzyme from Arthrobacter sp. DK-38. *Eur. J. Biochem.* **1997**, *248*, 385–393.

47. Esser-Kahn, A. P.; Francis, M. B., Protein-cross-linked polymeric materials through site-selective bioconjugation. *Angew. Chem. Int. Ed.* **2008**, *47*, 3751–3754.

48. Rouillard, J.-M.; Lee, W.; Truan, G.; Gao, X.; Zhou, X.; Gulari, E., Gene2Oligo: oligonucleotide design for in vitro gene synthesis. *Nucl. Acids Res.* **2004**, *32*, 176–180.

49. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Science* **1970**, *227*, 680–685.

50. Kim,T. W.; Chokhawal, H. A.; Nadler, D.; Blanch, H. W.; Clark, D.S., Binding modules alter the activity of chimeric cellulases: Effects of biomass pretreatment and enzyme source. B*iotechnol. Bioeng.* **2010**, *107*, 601–610.

Chapter 3

Immobilization of Transamination Reagents

Abstract

Protein N-terminal modification via transamination reagents proceeds in a two-step process. First, the N-terminal amine is chemoselectively oxidized to a pyruvamide. A subsequent oxime-forming reaction ultimately allows incorporation of any alkoxyamine reagent. The primary product of this reaction is the desired oxime product, although some "adduct" side-product forms when the enol form of pyruvamide reacts with Rapoport's salt (RS) or pyridoxal 5'-phosphate (PLP). Additionally, as both RS and PLP contain an aldehyde capable of undergoing an oxime formation with the alkoxyamine reagent, both reagents must be removed before the secondary conjugation. Currently, these reagents are removed using techniques that are both expensive and time-consuming. Here, we work to immobilize these transamination reagents on a support, allowing simple purification without interfering with the desired reaction. We can use immobilization to achieve rapid purification of the modified protein away from the excess transamination reagent and the aldol "adduct" side-product, leading to a more homogenous product. We synthesized a derivative of RS on polystyrene beads and carried out transamination on the peptide and small protein scale. With this novel purification technique, after oxime formation, no adduct moeities were observed in the final product mixture.

3.1. Introduction

Chemical tools to modify proteins are essential to a wide range of research fields and enable studies into cellular localization,[1,2,3] biomaterial construction,[4] and therapeutic development.[5,6] Protein modification can confer new properties to a protein through the chemical attachment of specific funtional groups. Some traditional protein modification methods capitalize on the inherent reactivity of native amino acid side chains.[7] In contrast, site-specific protein modification methods are increasingly studied, as these methods are location-specific and site-selective, which in turn produces well-defined and homogenous bioconjugates.[8] We have previously reported a site-specific transamination reaction that introduces a new ketone group at the N-terminus of proteins through incubation with pyridoxal 5'-phosphate (PLP)[9,10] or *N*-methylpyridinium-4-carboxaldehyde benzenesulfonate salt (Rapoport's salt, RS).[11,12] The resulting carbonyl groups are not naturally-occurring functionalities in proteins and can therefore be used as unique reactive synthetic handles, forming hydrazone or stable oxime bonds.[13,14] However, the pyruvamide group—formed after transamination—can tautomerize to its enol form. This enol can then undergo an aldol addition reaction, resulting in an aldol "adduct" species. This adduct, which is still transaminated, can undergo subsequent oxime formation, resulting in the "oxime + adduct" product. While both pathways lead to the formation of useful bioconjugation products, they are slightly different in chemical composition (Figure 3.1), and thus there is utility in being able to purify the two away from one another. One approach to doing so takes advantage of solid phase chemistry.

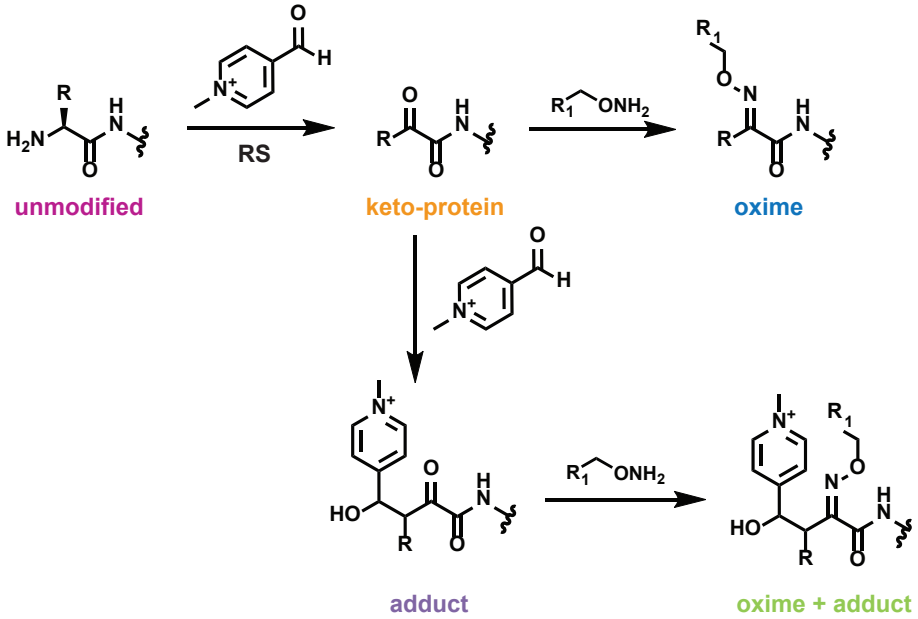


Figure 3.1. Product distribution observed as a result of Rapoport's salt (RS) mediated transamination. The N-terminal amine undergoes transamination to the keto-protein. This pyruvamide undergoes oxime formation in a second step with an alkoxyamine reagent. The keto-protein can tautomerize to its enol from and undergo an aldol addition with a second molecule of RS. This adduct species also undergoes oxime formation in a second step.

Immobilizing reagents is a long-standing strategy used to reduce side products and improve purification in the chemical industry. Solid phase synthesis enables rapid purification away from excess reagents and has propelled the usage of combinatorial libraries for screening applications.[15] One example of an immobilized reagent being used to carry out chemical reactions is polymer-

bound HOBT, which has been used to prepare amides from amines.[16] Immobilizing reagents allows for easy handling and purification while retaining yields comparable to in-solution reactions.[17] Solid phase reactions are also used frequently in molecular biology, for purposes such as protein immobilization[18], purification[19] and synthetic modification.[20] We synthesized a Rapoport's salt derivatized polymer that, when used to transaminate a peptide or protein, allowed us to separate unmodified or transaminated species away from the adduct species that remained tethered to the solid support (Figure 3.2). Additionally, in-solution RS transamination requires a series of centrifugal concentrations to remove free RS. In contrast, immobilizing the protein modification reagent allows for a more simplistic and cost-effective purification process. This technique produces a more homogenous product mixture, and the modified peptide or protein in solution can then be subjected to oxime formation—where only the oxime product was observed, and no amount of aldol "adduct" containing species was identified.

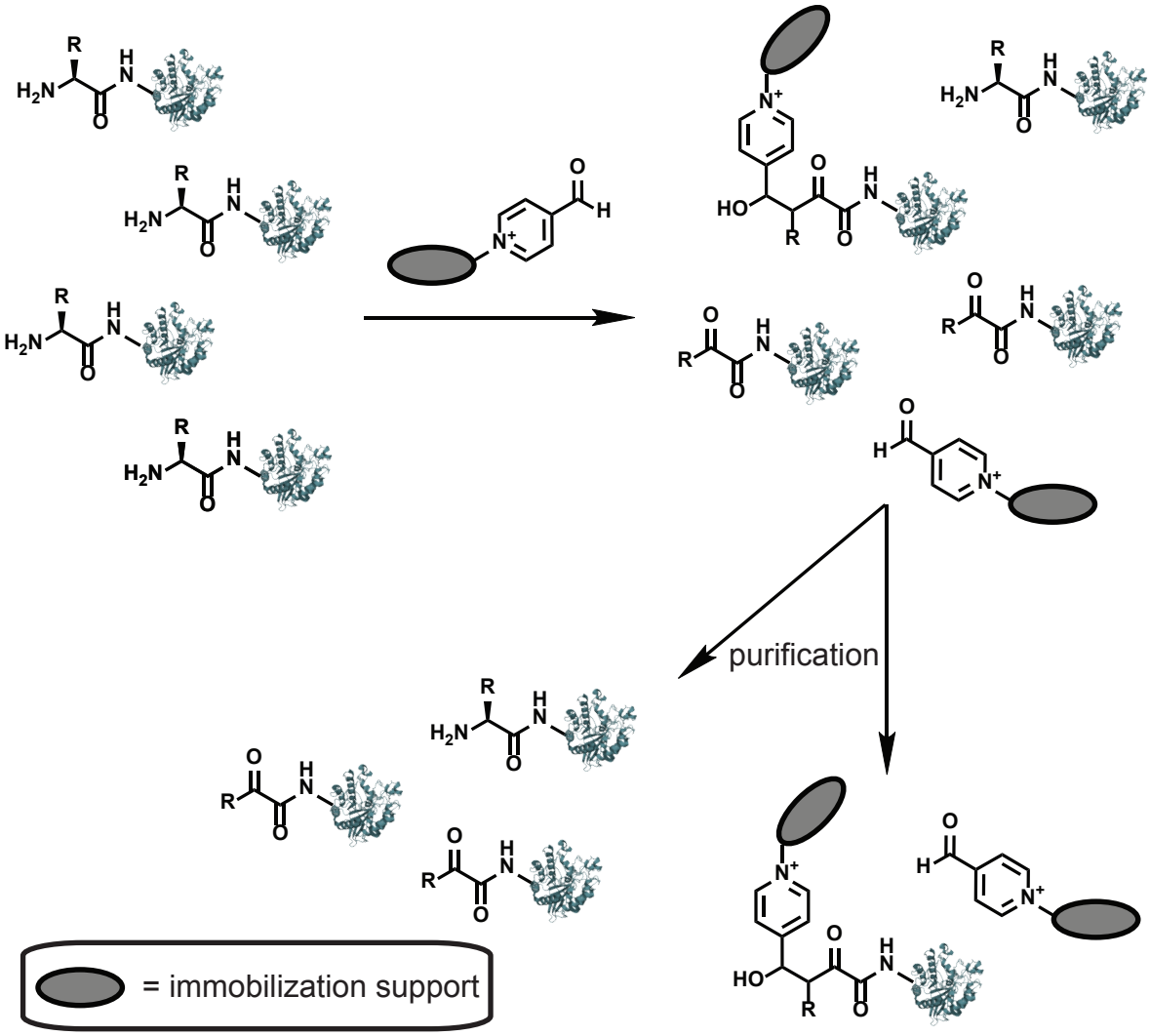


Figure 3.2. Schematic of immobilized transamination reagents. Once immobiized on any type of support, the transamination reagent can be used with peptides or proteins, resulting in a mixture of unmodified, ketone and adduct moieties. The support handle can then be used to purify out any excess reagent as well as adduct species away from the unmodified and ketone moieties. The resulting transaminated species can then be subjected to oxime formation.

3.2. Results and Discussion

3.2.1. Testing the Stability of the Rapoport's Salt Adduct

To improve product purity, we first attempted to reverse formation of the aldol adduct. In our previous work, where RS salt was kept in solution, different levels of "oxime + adduct" were observed between reactions carried out when the peptides were on bead vs. when they were also in solution. While the absolute levels of conversion to useful bioconjugation products remained consistent between the two scenarios, the ratios of oxime : oxime + adduct differed between them. In order to evaluate why the peptide on-bead reactions consistently resulted in reduced adduct formation, we looked into factors that could impact the stability of the RS adduct. After transamination and oxime formation on bead, peptides were cleaved off the TentaGel resin for mass spectrometry analysis using strong base. When the peptides were kept in solution throughout the reaction conditions, the need for this final base cleavage step was obviated. Because this was the primary difference between reaction conditions of peptides modified on bead vs. in solution, we hypothesized that under these conditions, the aldol adduct was reversible. In order to test our hypothesis, we reacted the eight residue peptide EESWSNAG (Figure 3.3a) with RS in solution,

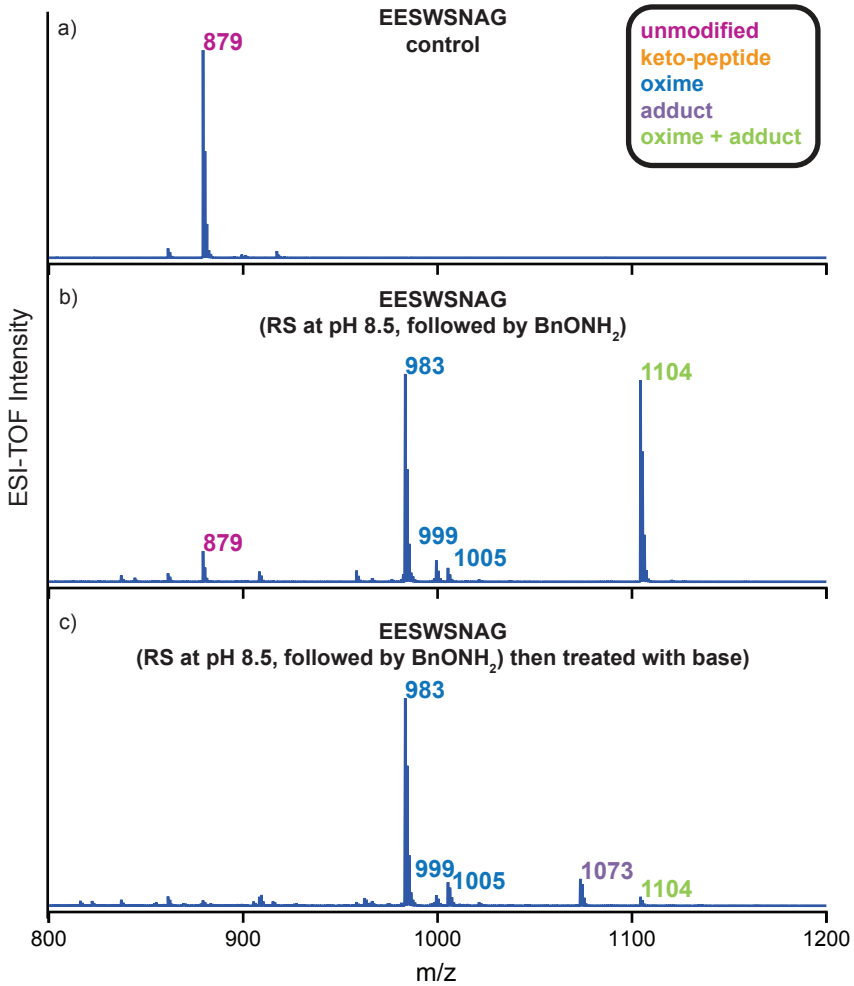


Figure 3.3. To test the stability of the Rapoport's salt adduct, EESWSNAG (a) peptide was transaminated at pH 8.5, then modified with benzyl alkoxyamine (b), and then incubated with strong base (c). Mass spectrometry analysis showed the reversibility of the adduct species in these strongly basic conditions.

purified and then incubated with benzyl alkoxyamine (Figure 3.3b). Then, this batch was incubated with 100 mM NaOH (Figure 3.3c). We found that, prior to treatment with the strong base, an oxime + adduct peak was observed with intensities comparable to the oxime peak. After base treatment, however, this peak significantly diminished, and we primarily observed a single oxime peak. Under these conditions, we did not observe complete reversion of the aldol adduct, as small peaks corresponding to "adduct" and "oxime + adduct" were detected. This result suggested that longer reaction times or stronger bases would be needed to push adduct reversion to completion.

These results confirmed that strongly basic conditions can reverse aldol adduct formation, but the necessary pH values were far higher than can be tolerated by most biological molecules. Specifically, while peptides can withstand such conditions, proteins denature and even degrade when subjected to strong bases. As a result, we concluded that improved purification techniques to separate the oxime from the adduct would be more feasible than reversing adduct formation.

3.2.2. Synthesis of a Rapoport's Salt Derivative on a Solid Phase Resin.

We next worked to immobilize transaminating reagents to improve reaction ease and purity. When selecting a support for immobilizing these reagents, we considered several parameters. An ideal support would have a large surface area to enable high loading capacities and maximal transamination yields. In addition, steric considerations—particularly given the size and steric bulk of most proteins—were crucial to ensure accessibility of the protein N-terminus to the immobilized reagent. The support should be biocompatible, such that it can be used under the aqueous conditions. And lastly, inexpensive supports that provide a simple purification handle are essential for future scale-up of these chemistries.

With these constraints in mind, we selected polystrene resins with terminal hydroxyl groups because of their compatibility with aqueous and organic solvents. Beads must swell in both conditions because the organic solvents prove useful for the attachment chemistry and the aqueous conditions are required for the subsequent application of the resin with proteins.

We then used solid phase synthesis to attach 4-pyridine carboxaldheyde—a derivative of Rapoport's salt—onto the beads to create an immobilized transaminating reagent on a polymeric support (Figure 3.4a and 3.5b).[21] The solid supports selected were TentaGel resin and Nova PEG Wang resin, and after attachment of the RS derivative, were notated as RS-TentaGel and RS-Wang, respectively. By attaching the 4-pyridine carboxaldheyde to the resin via the pyridine nitrogen, the resulting derivative produced a pyridinium carboxaldehyde, synonymous with that of RS, the primary difference being that triflate ($^{-}$OTf) served as the counter-ion instead of benzene sulfonate ($PhSO_3^-$). To confirm qualitatively that the attachment chemistry was successful, both the unmodified beads and RS modified beads were incubated with a colored alkoxyamine, called Disperse Red alkoxyamine. This reagent only reacts with aldehydes, resulting in coloration of beads containing free aldehydes moieties (Figure 3.4b and 3.5b). We observed that coloration occurred when Disperse Red alkoxyamine was incubated with RS-TentaGel or RS-Wang beads. In contrast, we saw no color change when unmodified beads were incubated with the same dye. This confirmed the presence of an aldehyde functionality, which we attributed to the successful attachment of the RS derivative.

Polystyrene resin beads are readily available, and previous studies have evaluated a large range of chemistries for compatibility. This includes the cleavage chemistries, used to release an

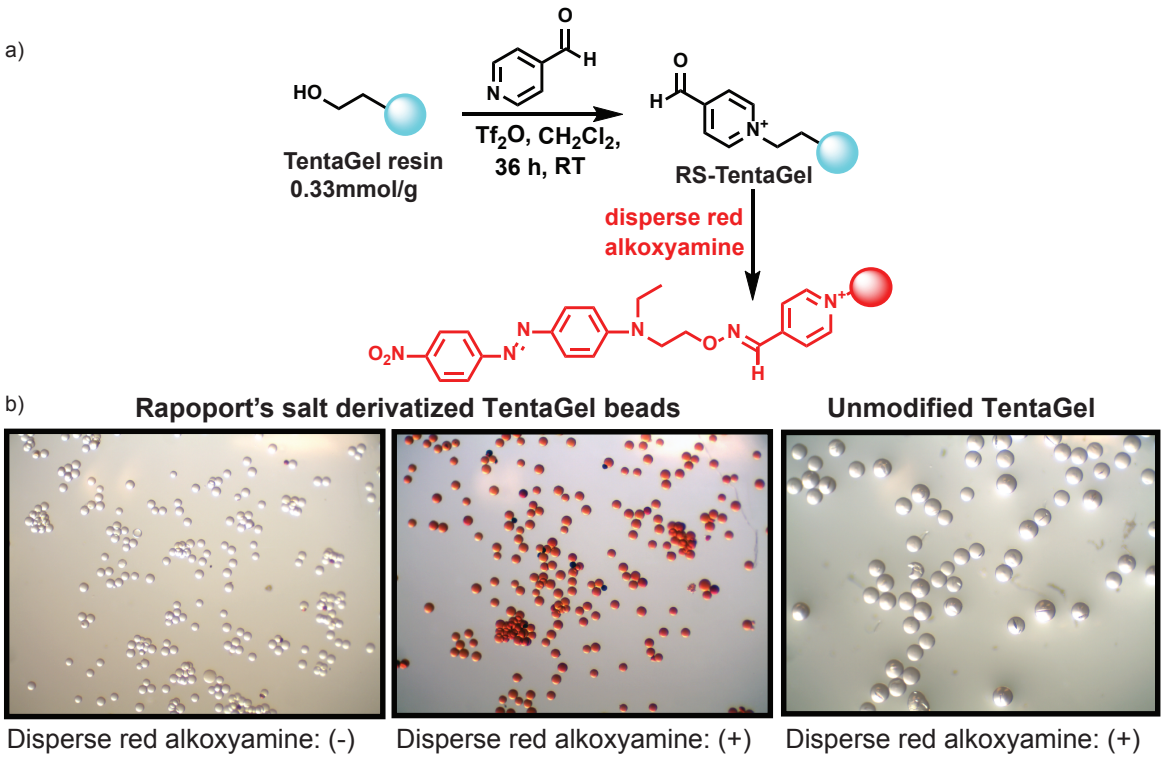


Figure 3.4. (a) Schematic of RS-TentaGel synthesis. (b) To verify the attachment of RS to TentaGel resin, the derivatized beads were incubated with disperse red alkoxyamine, a colored dye. Conversion of beads to a red color indicated the presence of an aldehdye moiety, as expected with the attachment of RS.

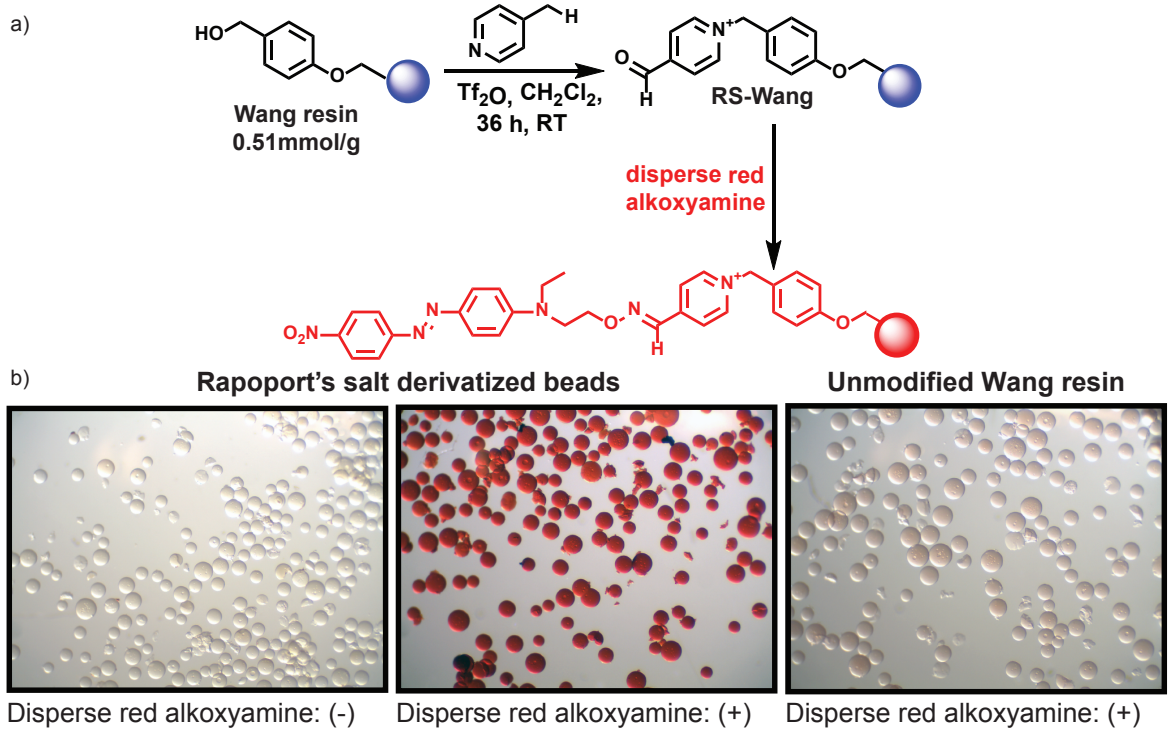


Figure 3.5. (a) Schematic of RS-Wang synthesis. (b) To verify the attachment of RS to NovaPEG Wang resin, the derivatized beads were incubated with disperse red alkoxyamine, a colored dye. Conversion of beads to a red color indicated the presence of an aldehdye moiety, as expected with the attachment of RS.

attached molecule from the bead. We used this approach to analyze our RS attachment chemistry. We first incubated RS-Wang beads with a benzyl alkoxyamine, which reacts with the aldehyde on RS. We then used TFA to cleave this molecule off the bead and analyzed the supernatant by LC-MS. We were able to confirm the expected molecular weight in the cleaved sample, further validating the synthesis of our immobilized RS derivative (Figure 3.6).

3.2.3. Quantification of Rapoport's Salt Loading Levels On-Bead.

In order to quantify the loading levels of RS on the polystyrene beads, we used a pull down assay, where the RS (and therefore aldehyde)-containing beads were reacted with a solution of Disperse Red alkoxyamine. Using the extinction coefficient of Disperse Red, we quantified the loss of signal in the supernatant above the beads. We hypothesized that this signal loss would correlate

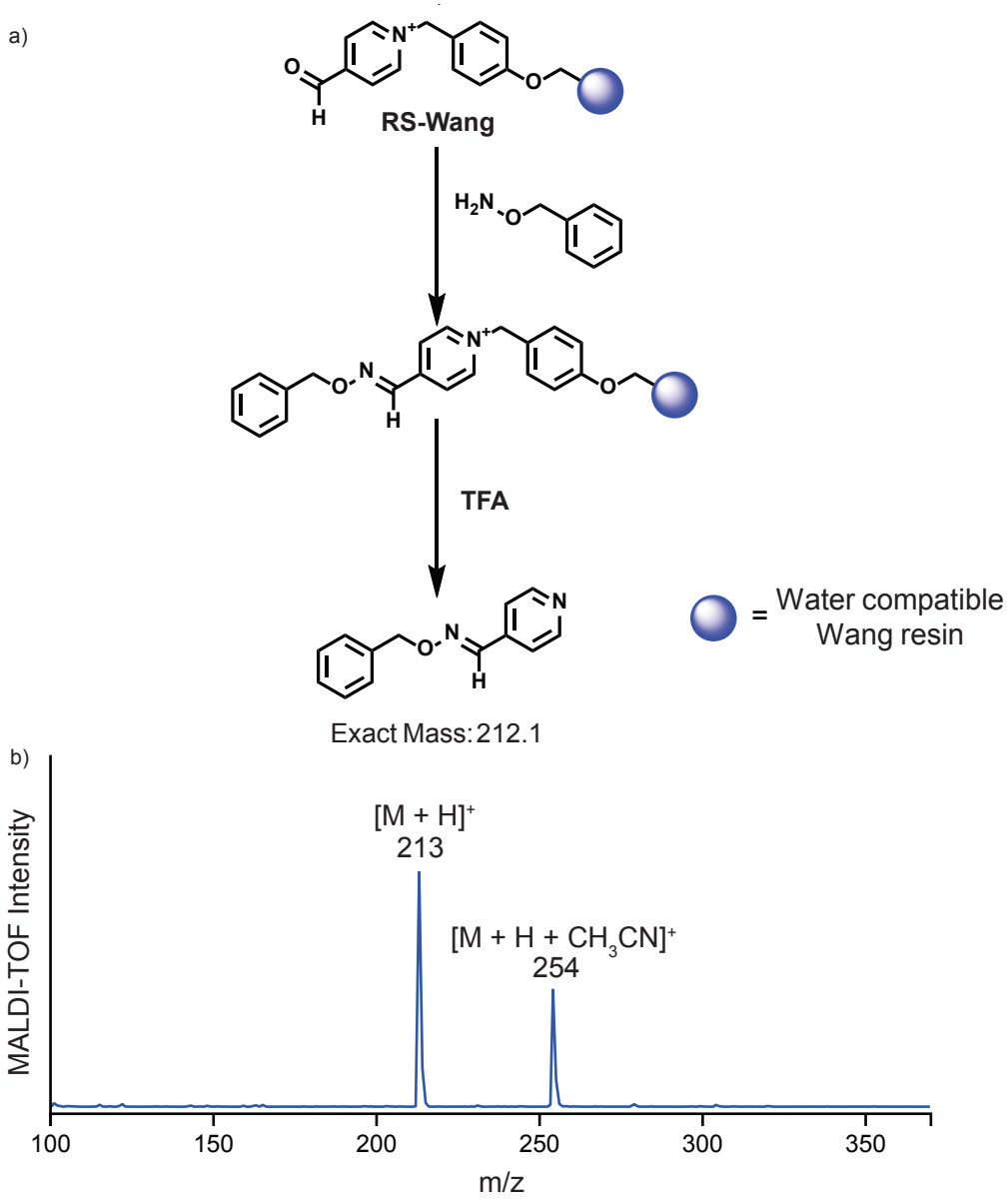


Figure 3.6. (a) Synthesis of immobilized Rapoport's salt on NovaPEG Wang resin (RS-Wang). (b) In order to confirm attachment of RS, the free aldehyde was reacted with benzyl alkoxyamine and then cleaved off the bead with 95% TFA before mass specrometry analysis. MS data confirmed the attachment of RS to the bead.

with the amount of aldehyde available to react on the bead. However, we found that changes in absorbance in the supernatant were too small to effectively quantify loading, particularly compared to the absorbance levels of the stock solution. As a second attempt, dansylhydrazine was used as the reactive dye and loss of fluorescence was quantified.[22] Again, we found that fluorescence changes were too small to reliably quantify loading levels. Together, these results indicated that the loading levels of RS on the beads were very low, and further optimization was required.

3.2.4. Testing Solid Phase Transamination on a Peptide.

After verifying RS attachment to the polystyrene bead, we proceeded to transaminate a peptide (AEEWSNAG). We incubated AEEWSNAG with 15 mg of RS-Wang beads. Following incubation, centrifugal filtration was used to separate the beads from the supernatant. We added benzyl alkoxyamine to the solution peptide, and the resulting product was evaluated by MALDI-TOF. The spectra showed that the supernatant contained both unmodified peptide and the oxime modified peptide (Figure 3.7), confirming successful transamination of the peptide with the on-bead RS derivative. Interestingly, no adduct-containing species were observed. In short, this immobilization strategy successfully purified the adduct species away from the desired oxime product. As a comparison, Figure 2.14 in Chapter 2 shows the results of in-solution RS transamination of AEEWSNAG, where about 50% of the adduct + oxime species is observed.

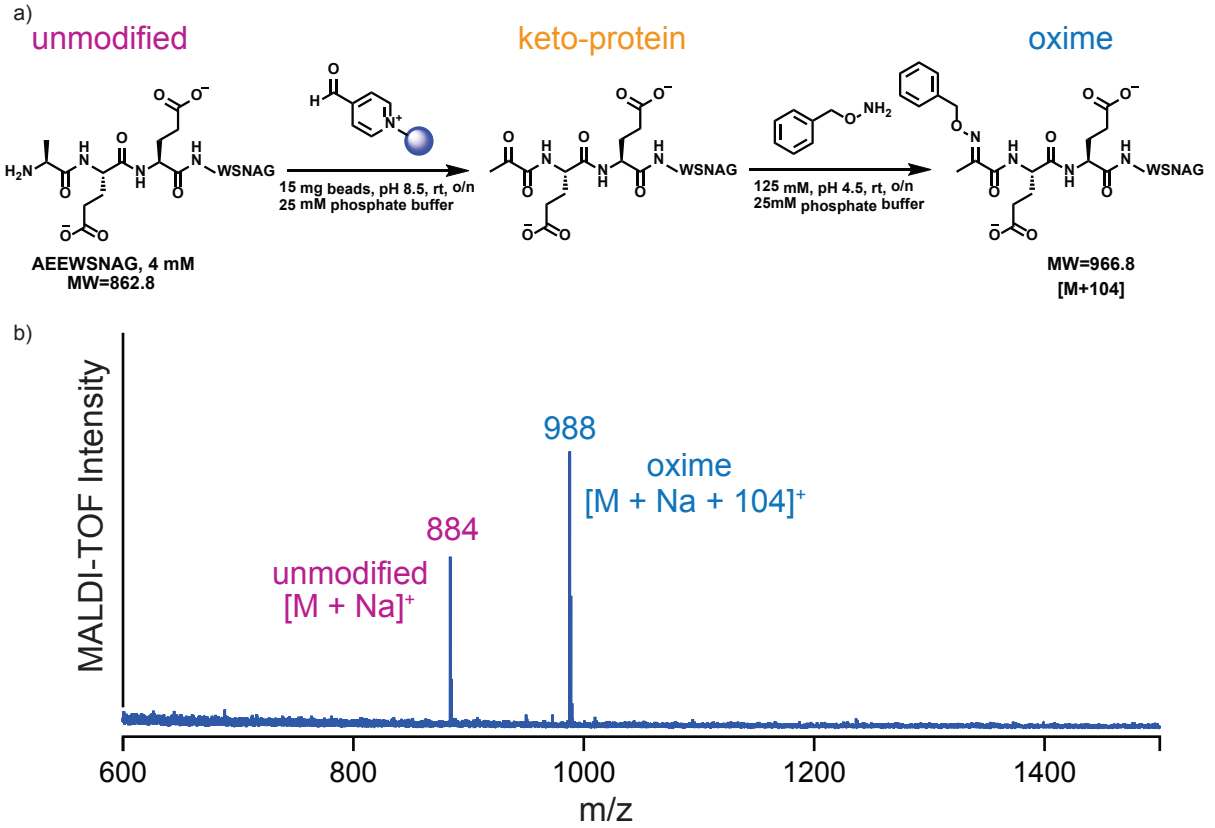


Figure 3.7. Using the RS-Wang, peptide AEEWSNAG was transaminated. Following transamination, peptide in solution was purified from the resin via filtration, and then subjected to oxime formation (with benzyl alkoxyamine) before mass spectrometry analysis. This reaction schematic is shown in (a). (b) MALDI-TOF spectra of AEEWSNAG peptide after RS-Wang transamination and oximation. Molecular weight of AEEWSNAG is 863 g/mol. Oxime attachment leads to a mass increase of 104 g/mol, which is observed. More significantly, it can be see that no adduct + oxime peak is observed in this approach (expected molecular weight addition: +225). For comparison, when using RS in solution, adduct + oxime is observed (Figure 2.14).

3.2.5. Testing Solid Phase Transamination on a Protein.

To test the compatibility of the immobilized RS reagent, RS-Wang was incubated with the Epidermal Growth Factor protein (EGF). EGF is a small protein, approximately 6 kDa in molecular weight. We used a similar protocol to the one described above for peptide modification. Oxime-modified EGF was observed via ESI-TOF (Figure 3.8). For comparison, transamination of EGF with in-solution RS is shown in Figure 3.9, and approximately 20% (via peak integration) of the product observed retains the aldol adduct species. These species are not observed when using the RS-Wang resin to carry out the transamination.

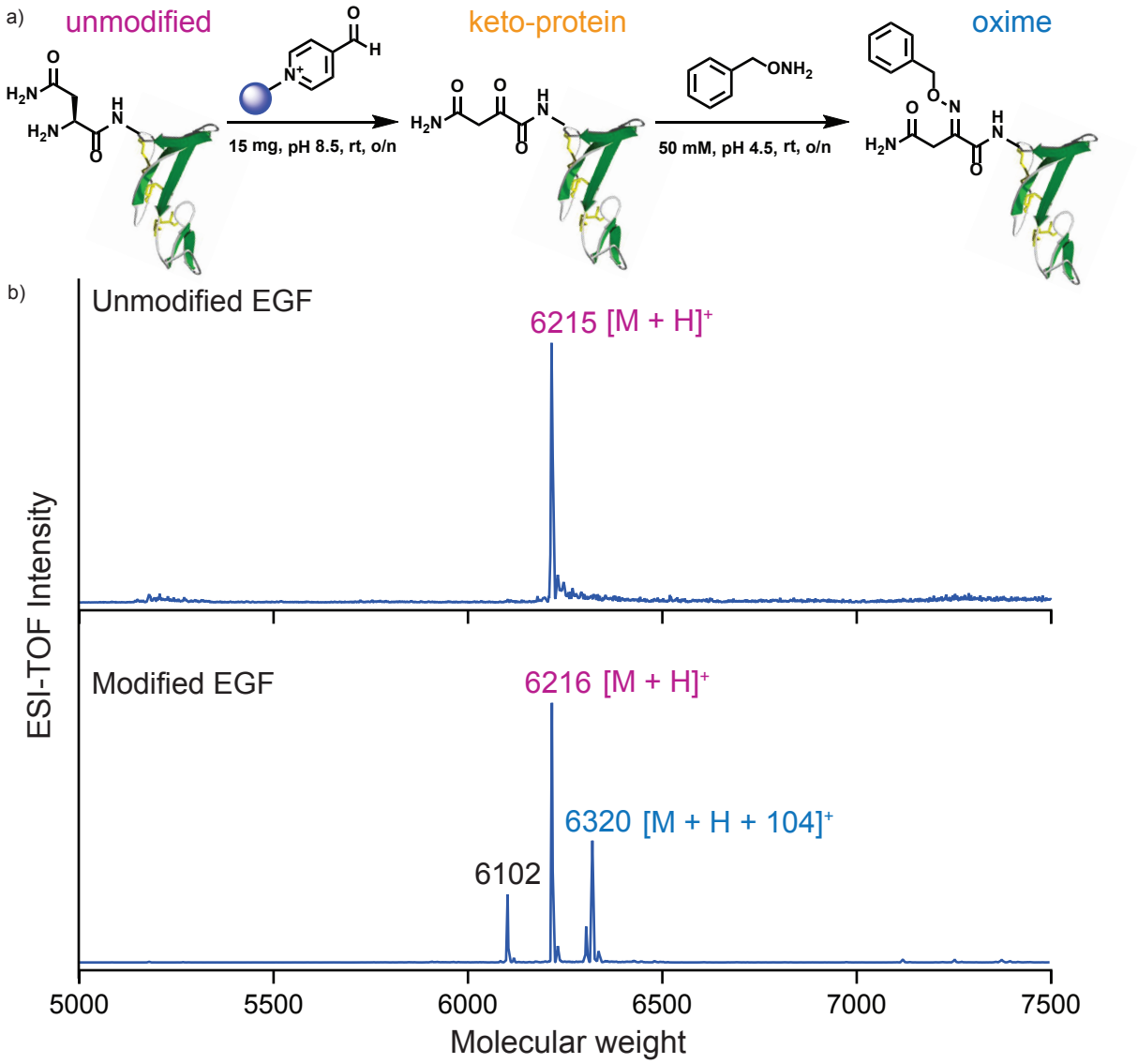


Figure 3.8. Using the RS-Wang resin, the Epidermal Growth Factor protein (EGF) was transaminated (N-terminal asparagine residue). Following transamination, EGF in solution was purified from the resin via filtration, and then subjected to oxime formation (with benzyl alkoxyamine) before mass spectrometry analysis. This reaction schematic is shown in (a). (b) MALDI-TOF spectra of EGF after RS-Wang transamination and oximation. Molecular weight of EGF is 6214 g/mol. Oxime attachment leads to a mass increase of 104 g/mol, which is observed. More significantly, it can be see that no adduct + oxime peak is observed in this approach (expected molecular weight addition: +225). The peak at 6102 is indicative of a loss of the N-terminal asparagine residue.

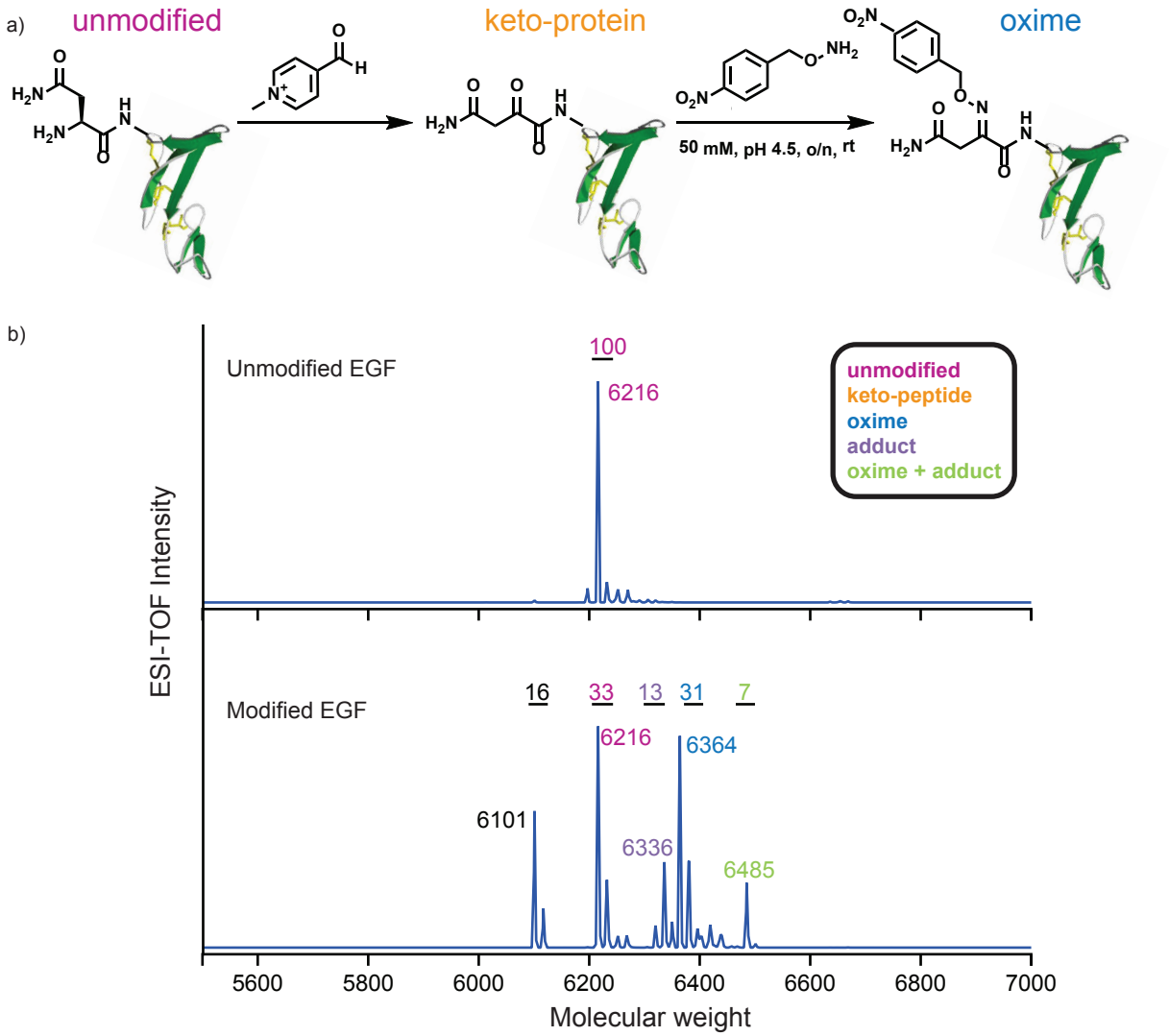


Figure 3.9. Using in-solution RS transamination, the Epidermal Growth Factor protein (EGF) was transaminated (N-terminal asparagine residue). Following transamination, EGF was subjected to oxime formation (with nitrobenzyl alkoxyamine) before mass spectrometry analysis. This reaction schematic is shown in (a). (b) ESI-TOF spectra of EGF after RS transamination and oximation. Molecular weight of EGF is 6214 g/mol. Oxime attachment leads to mass increase of 150 g/mol, which is observed. In addition, when carried out in solution, an aldol addition of RS is observed (purple and green). Peak integrations are listed above each peak. The peak at 6102 is indicative of a loss of the N-terminal asparagine residue.

3.3. Conclusions

Here, we have prepared and characterized solid phase polystyrene beads that are functionalized with the protein transamination reagent, Rapoport's salt. These polymers are also available in varying loading capacity and pore sizes to allow for larger proteins to fit between the pores of the resin. The porosity of the resin allows for maximal surface area to volume, and thus high loading levels can be achieved. Finally, cleavable linkers on these resins allow easy verification of attachment chemistry by mass spectrometry.

We have successfully used the immobilized RS to transaminate the N-terminus of a peptide and protein. This strategy allows protein modification to occur with an on-bead reagent, allowing for reagent removal and product purification. This chemistry results in a purer product for subsequent oxime formation. While EGF was used in this work as a model protein system, very

clear differences were observed between transamination with RS, both in-solution and on-bead. These preliminary data validate the utility of immobilized transamination reagents as a protein modification platform that can be generalized to different proteins.

We believe that further investigation into the types of polymer supports that can be employed in this context, in tandem with the optimized conditions for Rapoport's salt transamination, can be used to develop pure, facile and cost-effective protein-based materials in high yields. Alternatively, we also envision capitalizing on the adduct species as a way to immobilize proteins on resin, though additional work will be necessary to find conditions that yield high adduct formation.

3.4. Materials and Methods

3.4.1. General Procedures and Materials

Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H_2O) used as a reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). N-methylpyridinium-4-carboxaldehyde benzenesulfonate hydrate (Rapoport's salt, RS) was obtained from Alfa Aesar (Ward Hill, MA). 4-pyridine carboxaldehhdye was obtained from Aldrich. Benzylalkoxyamine, nitrobenzyl alkoxyamine and dansylhydrazine were obtained from Aldrich (St. Louis, MO). EGF was purchased from Life Technologies-Invitrogen. All Fmoc-protected amino acids were obtained from Novabiochem (EMD, Germany). TentaGel S OH and NovaPEG Wang resin was obtained from Advanced ChemTech (Louisville, KY).

3.4.2. Instrumentation and Sample Analysis

MALDI-TOF Analyis of Peptides

Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DE system (PerSeptive Biosystems, USA), and data were analyzed using Data Explorer software.

LC-MS Analysis of Synthetic Peptide Bioconjugates.

Electrospray ionization mass spectrometry (ESI-MS) of peptides was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 Time-of-Flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a XBridge PST C18 (3.5 μm particles, 2.1 mm × 150 mm, Waters, Milford, MA) analytical column. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). For each sample, approximately 5 to 10 picomoles of analyte was injected onto the column. Following sample injection, a 20-100% B elution gradient was run at a flow rate of 0.25 mL/min for 25 min. Data were collected and analyzed using Agilent MassHunter Qualitative Analysis B.05.00.

LC-MS Analysis of EGF Bioconjugates.

Electrospray ionization mass spectrometry (ESI-MS) of the peptides was performed using an Agilent 1260 series liquid chromatograph outfitted with an Agilent 6224 Time-of-Flight (TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a Poroshell 300SB-C18 (5 μm particles, 1.0 mm × 75 mm, Agilent, Santa Clara, CA) analytical column. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). For each

sample, approximately 5 to 10 picomoles of analyte was injected onto the column. Following sample injection, a 20-100% B elution gradient was run at a flow rate of 0.55 mL/min for 7 min. Data was collected and analyzed using Agilent MassHunter Qualitative Analysis B.05.00.

Chromatography was performed using a Proswift RP-4H (Thermo Scientific, USA) column with a H_2O/MeCN gradient mobile phase containing 0.1% formic acid. Mass spectra of proteins and protein conjugates were deconvoluted with the MassHunter Qualitative Analysis Suite B.05 (Agilent Technologies, USA).

3.4.3. Rapoport's Salt Adduct Stability.

Peptide and RS stock solutions were prepared at twice the desired final concentrations and mixed in equal volumes in a 1.5 mL Eppendorf tube. The final volume of each reaction was 400 µL. The 2× peptide stock solutions were prepared at 50 µM in 25 mM phosphate buffer (with 0.02% NaN_3), at pH 8.5. The 2× RS stock solution (200 mM) was freshly prepared before each reaction in 25 mM phosphate buffer (with 0.02% NaN_3) at the desired pH. The reaction mixture was briefly agitated to ensure mixing and was then incubated without further agitation at 37 °C for 1 h. Following the reaction, the excess aldehyde was removed using a sep-pak C18 1 mL vac cartridge (Waters, Milford, MA). Elution was carried out with 100% acetonitrile. The resulting keto-peptide solution was concentrated through evaporation and lyophilization. The white lyophilized powder was resuspended in 25 µL of 25 mM phosphate buffer (with 0.02% NaN_3), pH 5.0, and was subsequently treated with the alkoxyamine stock solution in a 1.5 mL Eppendorf tube and incubated at room temperature for 45 min. The alkoxyamine stock solution used was a 250 mM benzyl alkoxyamine solution (in water, pH adjusted to 5.0). A 25 µL portion of the stock solution was added to give a final concentration of 125 mM benzyl alkoxyamine. After oxime formation, 1 µL of 5M NaOH (final concentration of 100 mM) was added and allowed to incubate for 30 min. After the incubation period, the sample was neutralized with HCl and analyzed using mass spectrometry. Two controls were run following the same procedure. The first was without RS and the second was without the base treatment step. The control samples were also analyzed using mass spectrometry for comparison.

3.4.4. General Procedure for Solid Phase Synthesis of Immobilized RS.

Solid phase synthesis of RS derivatized Wang or TentaGel resin was prepared following the general protocol of Swinnen.[21] Briefly, to an oven dried round bottom flask was added 250 mg of alcohol terminated resin (NovaPEG Wang resin or TentaGel resin). Under argon, 5 mL of dry DCM was added and allowed to stir for 30 min while the beads swelled. Stir bars were avoided to keep beads intact. Then, 200 µL of 4-pyridine carboxyaldehyde (2.1 mmol) was added, and the flask was kept over ice to allow for cooling. To the solution was added 100 µL of trifluorometane sulfonic anhydride (0.6 mmol) dropwise. After five minutes, the ice bath was removed and after 30 min, 2 mL of DCM were added to rinse off beads on the sides of the flask. Reaction was allowed to sit overnight and then beads were collected and washed with DCM, DMF, and DCM again. After drying the beads, the sample was stored at 4 °C until further use. The beads appeared light brown in color after modification.

3.4.5. Pull-down Assays to Quantify Bead Loading Levels.

Control beads and RS modified beads were dried and weighed to obtain an exact mass. DMF was added to the beads to allow for the resin to swell. Then, aldehyde reactive dye (Disperse Red alkoxyamine or dansylhydrazine) was added in a 2:1 molar ration relative to the theoretical maximum loading on the resin (in a total volume of 0.2 mL). Synthesis of Disperse Red alkoxyamine is outlined in Chapter 2. Aliquots of the supernatant were collected at varying time points of dye addition (e.g. 20 μL at 0, 15 and 60 min). These were diluted into DMF to give enough volume for absorbance and fluorescence measurements. For Disperse Red alkoxyamine, absorbance scans were collected and for dansylhdyrazine, fluorescence emission scans from 380-700 nm were collected with an excitation wavelength of 345 nm. Using the extinction coefficients of each dye and the weight of beads used in each sample, total concentration of dye in the supernatant was calculated. The loss of dye over time corresponded to the loading of RS on the beads.

3.4.6. General Procedure for Peptide Synthesis and Purification.

This procedure has been described in Chapter 2.

3.4.7. General Procedure for Peptide Modification Using Immobilized RS.

The peptide stock solution was prepared at 4 mM in 25 mM phosphate buffer (with 0.02% NaN_3) at pH 8.5. To 300 μL of this sample, 15 mg of Wang resin derivatized with RS (RS-Wang) was added. The reaction mixture was incubated with rotation at room temperature overnight. Following the reaction, centrifugal filtration was used to separate the beads away from the protein. The resulting keto-protein was then treated with an equal volume of the alkoxyamine stock solution of choice in a 1.5 mL Eppendorf tube and incubated at room temperature overnight. The alkoxyamine stock solution concentration was 250 μM benzyl alkoxyamine (in 25 mM phosphate buffer with the pH adjusted to 5.0). After oxime formation, the protein concentration and buffer exchange steps were again repeated to remove the excess alkoxyamine. Controls were conducted following the same procedure but without RS. Analysis of the protein modification was carried out with MALDI-TOF mass spectrometry.

3.4.8. General Procedure for EGF Modification Using Immobilized RS.

The protein stock solution was prepared at 25 μM in 25 mM phosphate buffer (with 0.02% NaN_3) at pH 8.5. To 300 μL of this sample, 15 mg of Wang resin derivatized with RS (RS-Wang) was added. The reaction mixture was incubated with rotation at room temperature overnight. Following the reaction, centrifugal filtration was used to separate the beads away from the protein. The resulting keto-protein was then treated with an equal volume of the alkoxyamine stock solution of choice in a 1.5 mL Eppendorf tube and incubated at room temperature overnight. The alkoxyamine stock solution concentration was 250 μM benzyl alkoxyamine (in 25 mM phosphate buffer with the pH adjusted to 5.0). After oxime formation, the protein concentration and buffer exchange steps were again repeated to remove the excess alkoxyamine. Controls were conducted following the same procedure but without RS. Analysis of the protein modification was carried out with ESI-TOF mass spectrometry.

3.4.9. General Procedure for EGF Modification Using In-solution RS.

Protein and RS stock solutions were prepared at twice the desired final concentrations and mixed in equal volumes in a 1.5 mL Eppendorf tube. The final volume of each reaction was 200 μL. The 2× protein stock solutions were prepared at 50 μM in 25 mM phosphate buffer (with 0.02% NaN_3) at the desired pH. The 2× RS stock solution (200 mM) was freshly prepared before each reaction in 25 mM phosphate buffer (with 0.02% NaN_3) at the desired pH (using RS recrystallized from acetonitrile). The reaction mixture was briefly agitated to ensure mixing and then incubated without further agitation at 37 °C for 1 h. Following the reaction, the excess aldehyde was removed and the resulting keto-protein solution was concentrated and buffer exchanged 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 μL with 25 mM phosphate buffer (pH 4.5). Each sample was then concentrated to 100 μL, and the process was repeated 5 times. The resulting keto-protein was then treated with an equal volume of the alkoxyamine stock solution of choice in a 1.5 mL Eppendorf tube and incubated at room temperature for 48 h. The alkoxyamine stock solution concentration was 62.5 μM nitrobenzyl alkoxyamine (in 25 mM phosphate buffer with the pH adjusted to 5.0). After oxime formation, the protein concentration and buffer exchange steps were again repeated to remove the excess alkoxyamine. Controls were conducted following the same procedure but without RS. Analysis of the protein modification was carried out with mass spectrometry.

3.5. 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., New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. *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. 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. *Bioconjugate Chem.* **1992**, *3*, 138–146.

6. Zalipsky, S., Functionalized poly(ethylene glycol) for preparation of biologically relevant conjugates. *Bioconjugate Chem.* **1995**, *6*, 150–165.

7. Means, G. E.; Feeney, R. E., Chemical modifications of proteins: History and applications. *Bioconjugate Chem.* **1990**, *1*, 2–12.

8. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L., Site-selective protein-modification chemistry for basic biology and drug development. *Nature Chemistry* **2016**, *8*, 103–113.

9. 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.

10. Witus, L. S.; Francis, M. B., Site-specific protein bioconjugation via a pyridoxal 5'-phosphate-mediated N-terminal transamination reaction. *Curr. Prot. Chem. Biol.* **2010**, *3*, 1–10.

11. Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M,; Weng, C. H.; Iavarone, A. T.; Francis, M. B., Site-specific protein transamination using N-methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **2013**, *135*, 17223–17229.

12. Palla, K. S.; Witus, L. S.; Mackenzie, K. J.; Netirojjanakul, C.; Francis, M. B., Optimization and expansion of a site selective N-methylpyridinium-4-carboxaldehyde mediated transamination for bacterially expressed proteins. *J. Am. Chem. Soc.* **2015**, *137*, 1123–1129.

13. Jencks W. P., Studies on the mechanism of oxime and semicarbazone formation. *J. Am. Chem. Soc.* **1959**, *81*, 475–481.

14. Kalia, J.; Raines, R. T., Hydrolytic stability of hydrazones and oximes. *Angew. Chem. Int. Ed.* ***2008****, 47, 7523–7526.*

15. Booth, R. J.; Hodges, J. C., Solid-supported reagent strategies for rapid purification of combinatorial synthesis products. *Acc. Chem. Res.* **1999**, *32*, 18–26.

16. Dendrinos, K.; Jeong, J.; Kalivretenos, A. G., HOBT immobilized on macroporous polystyrene beads: a useful reagent for the synthesis of amides. *Chem. Commun.* **1998**, *4*, 499–500.

17. Broun, G. B.; Manecke, G.; Wingard Jr., L. B., Enzyme engineering: Immobilized protein modification reagents; Springer, U.S. **1978**; pp 39–392.

18. Schmidt, T. G. M.; Skerra, A., The *Strep*-tag system for one-step purification and high-affinity detection or capturing of proteins. *Nature Protocols* **2007**, *2*, 1528–1535.

19. Rosen, C. B.; Kwant, R. L.; MacDonald, J. I.; Rao, M.; Francis, M. B., Capture and recycling of Sortase A through site-specific labeling with lithocholic acid. *Angew. Chem. Int. Ed.* **2016**, doi:10.1002/anie.201602353.

20. Lewis, C.; Scouten, W. H., Photooxidation of methionine with immobilized methylene blue as photooxidizer. *Biochim Biophys Acta.* **1976**, *444*, 326–330.

21. Crosignani, S.; Swinnen, D., A straightforward, one-pot protocol for the preparation of libraries of 2-oxazolines. *J. Comb. Chem.* **2005**, *7*, 688–696.

22. Yan, B.; Li, W., Rapid fluorescence determination of the absolute amount of aldehyde and ketone groups on resin supports. *J. Org. Chem.* **1997**, *62*, 9354–9357.

Chapter 4

Site-selective Attachment of Enzymes to Glass Surfaces Through DNA Directed Immobilization

Abstract

Enzymes are able to maintain remarkably high selectivity towards their substrates while still retaining high catalytic rates. By immobilizing enzymes onto surfaces we can heterogenize these biological catalysts, making it practical to study, use, and combine them in an easily controlled system. In this work, we develop a platform that allows for the simple and oriented immobilization of proteins through DNA directed immobilization (DDI). First, we modified a glass surface with single stranded DNA. We then site-selectively attached the complementary DNA strand to the N-terminus of a protein. Both DNA modifications were carried out using an oxidative coupling strategy, and the DNA strands served as easily tunable and reversible chemical handles to hybridize the protein-DNA conjugates onto the surface. We have used the aldolase enzyme as a model protein to conduct our studies. We characterized each step of the protein immobilization process using fluorescent reporters as well as atomic force microscopy. We also conducted activity assays on the surfaces with DNA linked aldolase to validate that, despite being modified with DNA and undergoing subsequent immobilization, the enzyme was still able to retain its catalytic activity and the surfaces were reusable in subsequent cycles.

Portions of this work were performed in collaboration with the following persons:

Platform development was done with Tyler J. Hurlburt

Atomic force microscropy was conducted by Alexander M. Buyanin

4.1. Introduction

The utility of immobilizing proteins onto a surface spans a variety of applications, including the study of protein-protein interactions, enzyme kinetic studies, biosensors, bioanalytics, and even industrial biocatalytic processes.[1,2] These studies create a constant need for effective and facile ways to assemble protein microarrays. The requirements for a surface based protein technology are dependent on the specific application at hand, but general considerations include choosing a suitable solid surface and a correspondingly appropriate chemistry.[3] In addition, the chemistry used to attach the protein of interest should be mild and compatible towards amino acid side chains, such that the protein is able to maintain its higher order structure and retain its native function.

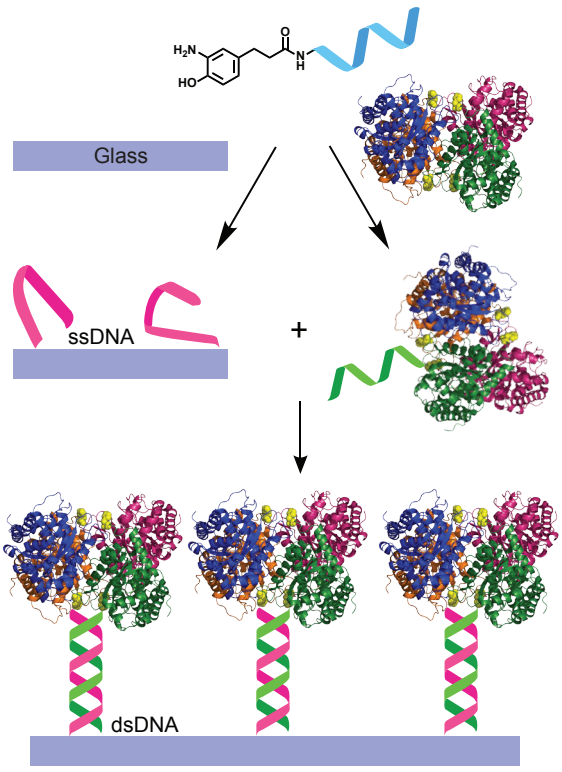
Many protein immobilization chemistries involve the direct attachment of proteins to surfaces through short linkers and reactive handles. Common approaches include nonspecific covalent modification of native amino acid side chains on the surface of a protein, such as lysine acylation with NHS esters. However, it has been found that randomly oriented proteins can exhibit reduced accessibility of active sites and display lower activities than their ordered counterparts.[3–5]

Because an ordered display of proteins is often more favored, both covalent and non-covalent strategies to orient proteins uniformly on surfaces have been developed. Covalent approaches have taken advantage of maleimide reactivity with thiols,[5] native chemical ligation,[6] photochemical thiol-ene chemistry,[7] and enzymatic tags,[8,9] to name a few. Representative non-covalent systems are exemplified by polyhistidine tag incorporation via genetic engineering to bind to Ni-NTA functionalized surfaces, as well as biotin-streptavidin complexation.[10–12] Another non-covalent protein immobilization approach is through the use of DNA, taking advantage of complementary strand hybridization. This type of DNA directed immobilization (DDI) requires that the surface be functionalized with a short oligonucleotide and that its complementary strand be conjugated to the target protein such that the hybridization of the two strands will lead to the controlled immobilization of the proteins under chemically mild and biocompatible conditions. DDI has shown reliability and has been used in tandem with a variety of other assembly processes. It has also been reported that the DDI strategy is an efficient method to immobilize proteins because of the easily adjustable linker it provides, thereby helping to prevent protein denaturation.[13–15]

Previously, protein surfaces have been created with this strategy via the complexation of biotinylated antibodies and biotinylated DNA, brought together with streptavidin, that were then immobilized on streptavidin-biotin-DNA modified surfaces via DNA hybrdization.[16] Because DNA molecules are highly stable, they can easily undergo chemical modification in preparation for DDI. In another example, a DNA-heme was generated and used to reconsitute two separate heme binding proteins: apo myoglobin and apo horseradish peroxidase. These were tethered onto microplates that were coated with the complementary DNA strands, and it was shown that enzymatic activity of both proteins was retained.[17] Even more recently, unnatural amino acid incorporation was used to insert a *p*-acetylphenylalanine residue into a monoclonal antibody, which was then used as a handle for ligation with an aminoxy-functionalized single stranded DNA.[18] These approaches illustrate both the benefits and the complexities involved in generating protein-DNA bioconjugates.

We have previously developed several site-selective protein modification reactions, and we have shown the applicability of some of them in the synthesis of DNA surfaces and DNA-protein bioconjugates.[19–27] More recently, we reported on an oxidative strategy that is able to

couple an *o*-aminophenol (AP) to either an aniline moiety or the N-terminal amino group of peptides and proteins using potassium ferricyanide as the oxidant.[28,29] Herein, we take advantage of this positional selectivity and functional group tolerance and apply it toward the development of a DDI based platform as shown in Scheme 4.1. We first coupled an *o*-aminophenol modified DNA strand to an aniline modified glass surface. Separately, we modified our protein of interest at the N-terminus with a complementary *o*-aminophenol substituted DNA strand in a single step with low concentrations of reagents. The subsequent hybridization of the surface oligo with the complementary oligo-protein conjugate allowed for the controlled attachment of the protein to surfaces in an oriented and versatile manner. Herein, we apply DNA hybridization based protein immobilization using aldolase and evaluate its catalytic activity after attachment to glass slides. We also study the reusability and regenerability of these surfaces.



Scheme 4.1. DNA directed immobilization of a site-selectively modified DNA-protein conjugate onto a glass surface displaying complementary single stranded DNA

4.2. Results and Discussion

4.2.1. Modifying Glass Slides with Single Stranded DNA and Hybridizing to Complementary DNA.

Most previous studies of DDI-based protein immobilization have used gold or coated plastic substrates.[13] For these studies we selected glass slides because of their advantages for spectroscopic and microscopic analysis. We used silanization with 3-(4-azidophenyl)-*N*-(3-trimethoxysilylpropyl) propanamide followed by TCEP reduction in order to derivatize glass with aniline functional groups.[25] We then synthesized the aminophenol modified strand A (AP-A), a 20 base oligomer (Figure 4.1a), and coupled it to the aniline surface in the presence of potassium ferricyanide as an oxidizing agent (Figure 4.1b). Once the glass slides were modified with AP-A, each surface was incubated with complementary strand, A', which had a fluorophore conjugated to its 5' end (A'*). Non-complementary AP-B was also synthesized to be used as a negative control for the surface modification. Following hybridization and rinsing, fluorescent images were collected to confirm that the DNA mediated hybridization between A and A'* was specific (Figure 4.1c). Fluorescence intensity on these slides was 50-fold greater than that of the mismatched control between B and A'* (Figure 4.1d), indicating that there was only nominal non-specific binding of A'* onto the glass surface. DNA sequences used in these studies are highlighted in Table 4.1. It is worth noting that the 20 base DNA sequences that were selected did not exhibit any hairpin or dimer formation at room temperature. It was important to verify that

Name	Sequence	T_m (°C)
A	5′ - CCC TAG AGT GAG TCG TAT GA - 3′	52.6
Ax	5′ - CCC TAG AGT GAG TCG TAT GAA AAA A - 3′	54.4
A′	5′ - TCA TAC GAC TCA CTC TAG GG - 3′	52.6
Ax′	5′ - TTT TTT CAT ACG ACT CAC TCT AGG G - 3′	54.4
A′*	5′ - AlexaFluor488 TCA TAC GAC TCA CTC TAG GG - 3′	52.6
B	5′ - AGT GAC AGC TGG ATC GTT AC - 3′	54.4

Table 4.1. DNA sequence information

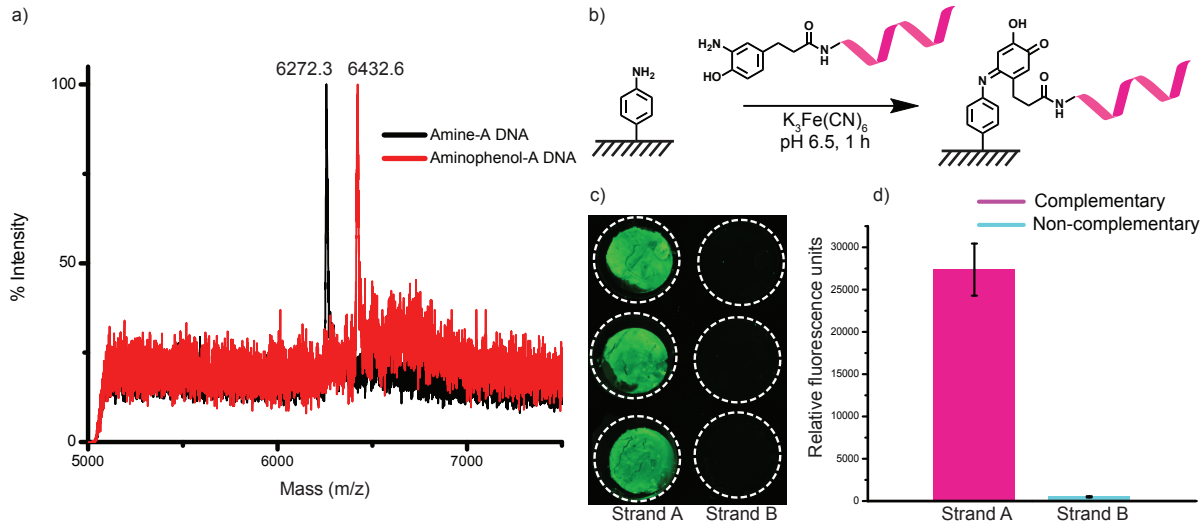


Figure 4.1. Modification of aniline coated glass slides with single stranded aminophenol-DNA. (a) MALDI-TOF analysis of DNA before (black) and after (red) aminophenol attachment. (b) Schematic of an oxidative coupling reaction for single stranded DNA attachment to aniline modified surfaces. (c) Fluorescence studies to verify DNA strand hybridization. After attachment of strand A or strand B, glass slides were incubated with a fluorescently tagged DNA (A'*). Fluorescence signal was evaluated when there was sequence complementarity (left column, strands A and A'*) and when there was not (right column, strands B and A'*). (d) Plot of fluorescence intensities, conducted in triplicate.

we selected sequences that did not undergo hairpin formation or self-dimerization because both would interfere with complementary strand hyridizations, leading to little, if any surface attachment.

4.2.2. Modifying Aldolase with A' DNA and Evaluating its Activity.

A fructose-bisphosphate aldolase (ALD) from rabbit muscle was chosen as a protein of particular interest for these studies. Aldolase is a protein involved in a series of enzymes within the glycolytic pathway, as it catalyzes the reversible breakdown of fructose-1, 6-bisphosphate (FBP) to glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP). We first carried out in solution activity assays of aldolase to determine Michaelis-Menten parameters of the enzyme (Figure 4.2). Because it is involved in a C-C bond breaking and (in the microscopic reverse) bond forming reaction, it is of particular importance in industrial processes that can engineer the enzyme to be promiscuous and catalyze other C-C bond processes.[30] It is a homotetramer with D_2 symmetry (PDB: 6ALD). All four of the N-termini are solvent exposed, with two N-termini in proximity to one another and the other two N-termini on the opposite face. As a result of this configuration, there are two possible ways for the protein to be immobilized via its N-terminal positions. Fortunately, these would be expected to display the protein with highly similar orientations. Additionally, it retains a proline residue at its N-terminus, which has a favorable reactivity towards the oxidative coupling reaction.[29]

Aldolase was first capped with *N*-ethyl maleimide (NEM) to prevent any reactivity with free cysteine residues (Figure 4.3). Then, AP-A' was synthesized and coupled to the aldolase N-terminus using potassium ferricyanide mediated oxidative coupling (Figure 4.4a). Free DNA was removed through spin concentration, and a BCA assay was used to quantify the total protein remaining. Additionally, gel densitometry was used to determine the level of modification of the DNA-aldolase bioconjugate. Concentrations were quantified for a single monomer, and since aldolase is a tetramer, results of ~25% modification indicated that the mixture predominantly contained a single DNA strand per tetramer (Figure 4.4b).

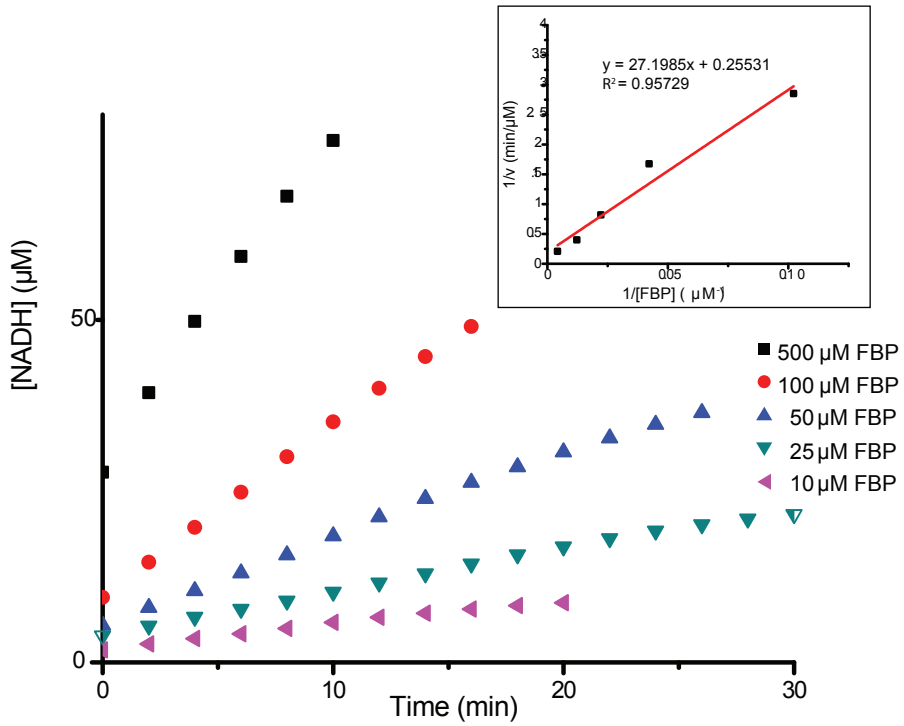


Figure 4.2. Michaelis-Menten analysis of aldolase. Activity of 20 nM aldolase was measured with 10-500 μM fructose 1,6-bis-phosphate (FBP) and the initial linear region of the activities were plotted. Inset: Lineweaver-Burk double reciprocal plot to determine the Michaelis-Menten constants. Michaelis-Menten constants found to be kcat = 3.26 s-1, Km = 1.07 x 10-4 M, and kcat/Km = 3.06 x 104 M-1s-1. All activity assays were run at 37° C, and conducted in triplicate.

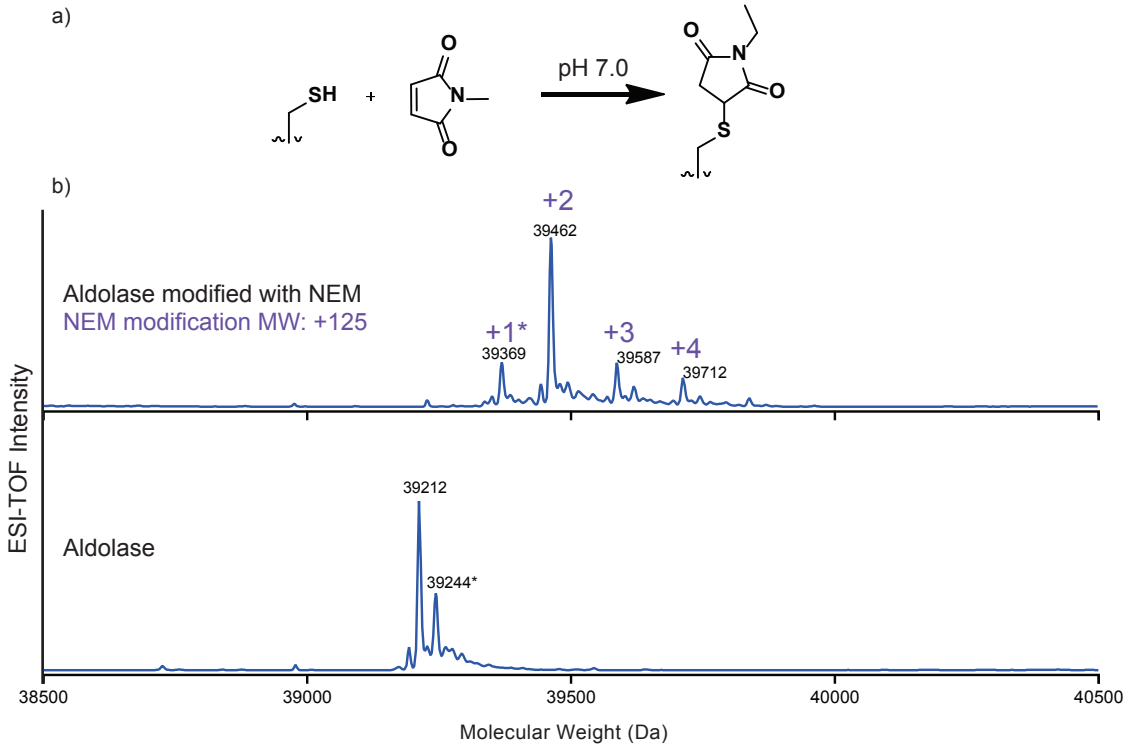


Figure 4.3. Modifying aldolase with N-ethyl maleimide (NEM) to cap free cyteine side chains before modifying at the N-terminus with DNA. (a) Chemical reaction between thiol groups on cysteine side chains and N-ethyl maleimide. (b) Deconvoluted ESI-TOF mass spectra of aldolase with and without N-ethyl maleimide capping. Unmodified aldolase has a mass of 39212 Da. MS analysis of pure aldolase presented two peaks; the parent peak at 39212 Da and a shoulderpeak at 39244 Da. This shoulder peak was also observed to get modified by NEM, as indicated by the * labels.

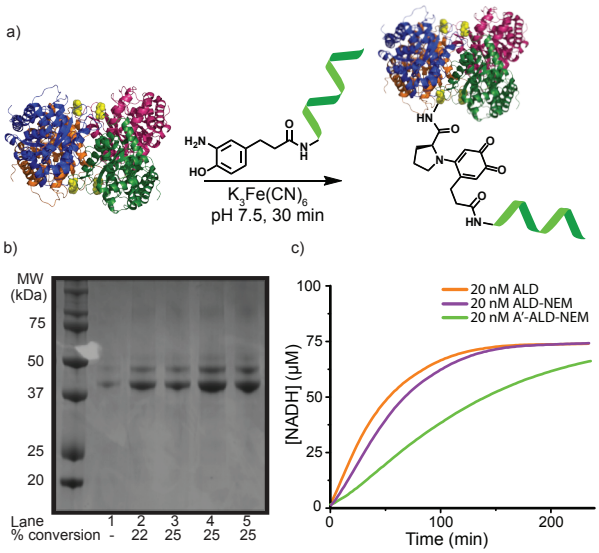


Figure 4.4. Aldolase modification with DNA. (a) Schematic of protein modiification at the N-termini (yellow) with aminophenol modified DNA. (b) SDS-PAGE gel showing the attachment of a 20 base single stranded DNA to NEM capped aldolase. (c) Quantification of solution activity of unmodified aldolase (orange), aldolase with reactive cysteines capped with NEM (purple), and NEM capped aldolase after modification with A' DNA (green). Samples were analyzed in triplicate and data points were collected every two minutes. Initial rates were 1.08 ± 0.01 μM/min for unmodified aldolase, 0.780 ± 0.004 μM/min for NEM capped aldolase, and 0.376 ± 0.002 μM/min for NEM capped aldolase modified with DNA. All assays were run at 37 °C and conducted in triplicate.

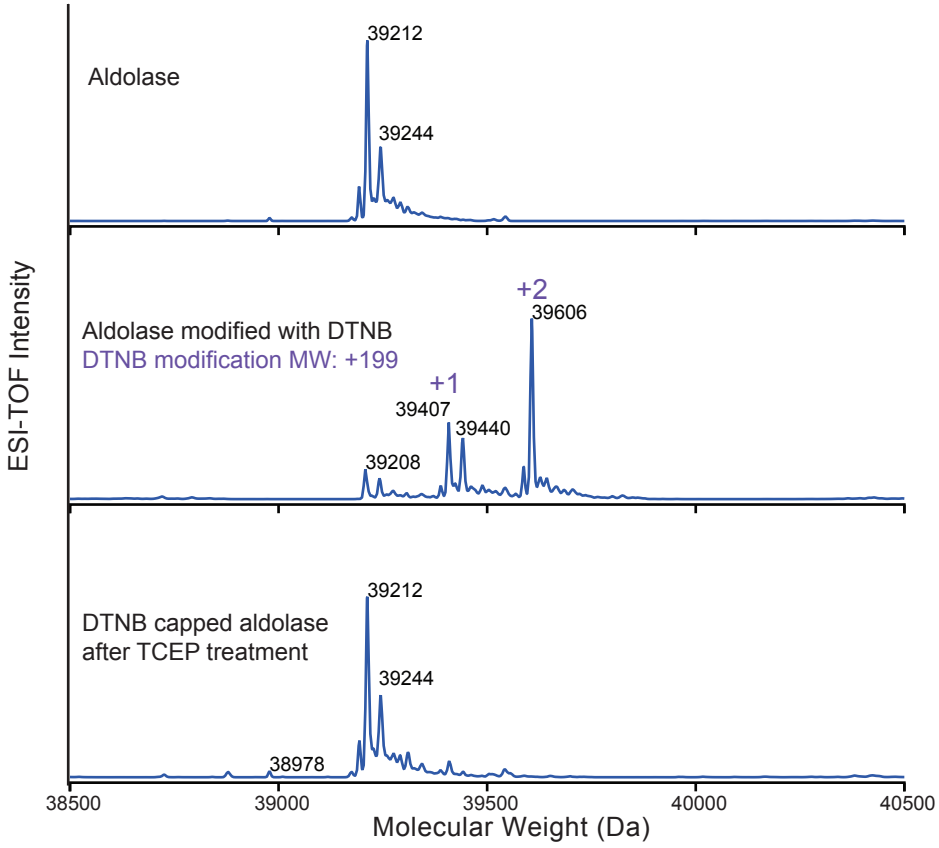


Figure 4.5. Modifying aldolase with 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) to block free cyteine side chains before modifying the N-terminus with DNA. Deconvoluted LC-MS spectra of aldolase, aldolase after DTNB capping, and DTNB capped aldolase after TCEP treatment to remove DTNB. Unmodified aldolase has a mass of 39212 Da.

Solution activity assays were carried out on aldolase, aldolase capped with NEM, and NEM capped aldolase that was modified with AP-A' DNA to determine how the modification itself as well as reaction conditions impacted the enzymatic activity. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was paired with aldolase in the activity assays. GAPDH catalyzes the conversion of G3P to 1,3-bisphosphoglycerate in the presence of NAD^+ as a cofactor. The conversion of NAD^+ to NADH can be monitored by the increase in absorbance at 340 nm and can be used to quantify aldolase activity. As illustrated in Figure 4.4c, NEM capped aldolase after modification with DNA retained about 48% of its enzymatic activity. This difference could be attributed to negative-charge

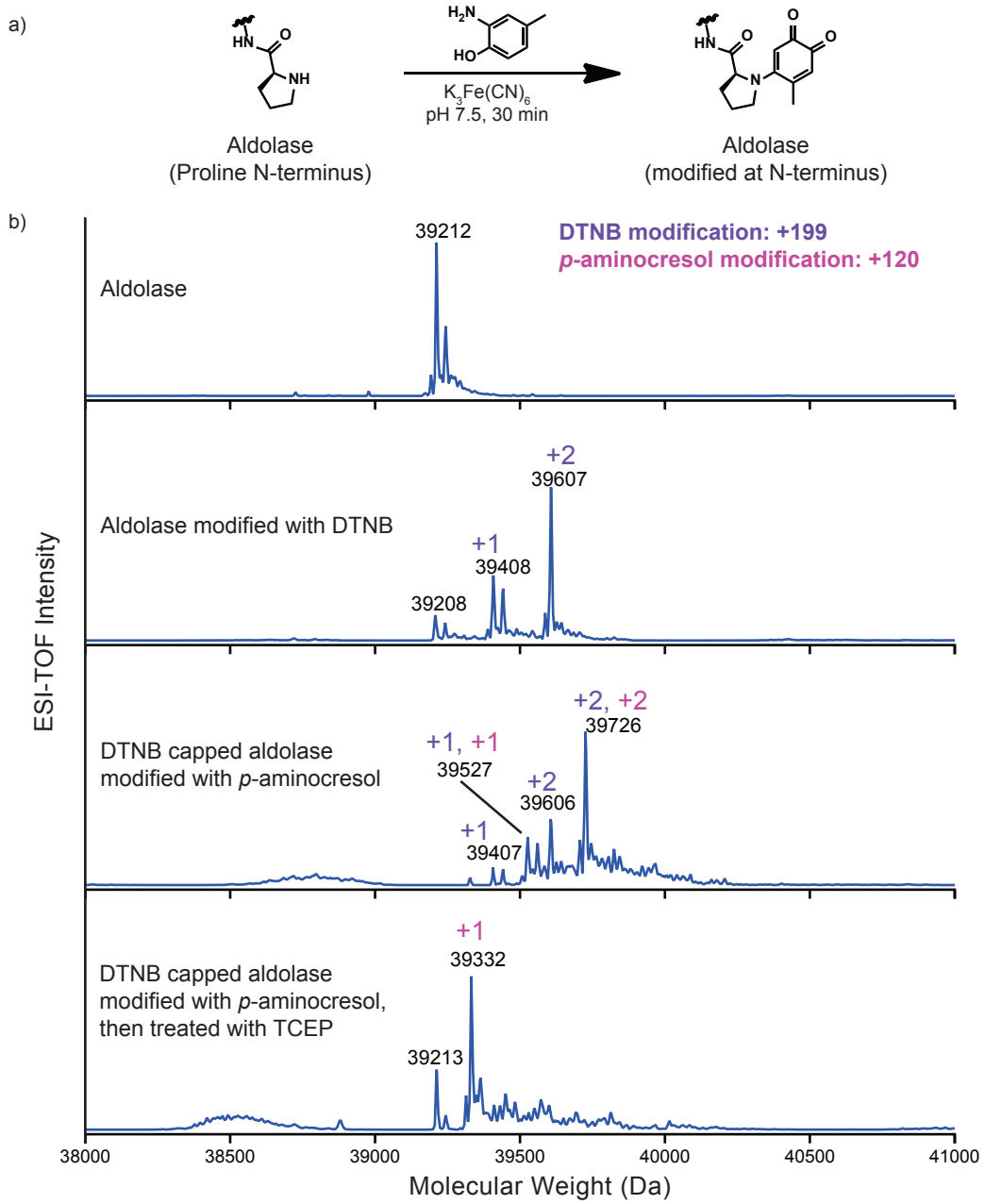


Figure 4.6. Modifying DTNB-aldolase at the N-terminus. (a) Oxidative couping reaction with the N-terminal proline residue of aldoase and *p*-aminocresol in the presence of potassium ferricyanide. (b) Deconvoluted LC-MS spectra of aldolase, DTNB-aldolase, DTNB-aldolase after oxidative couping, and DTNB-aldolase after oxidative coupling and subsequent TCEP treatment to remove DTNB. Only a single modification with *p*-aminocresol is observed. Unmodified aldolase has a mass of 39212 Da.

repulsion between the aldolase substrate, FBP, and the DNA strand, resulting in destabilization of the enzyme.[31]

Aldolase capped with 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) was modified at the N-terminus with a small molecule aminophenol (2-amino-*p*-cresol). After removal of the cysteine cap, the bioconjugate was analyzed by ESI-TOF mass spectrometry and showed high yields of a single *p*-aminocresol addition. These iterative reactions on the protein are illustrated in Figures 4.5 and 4.6. A trypsin digestion of this sample confirmed that the modification was occurring at the N-terminal peptide fragment (Figure 4.7 and 4.8) and subsequent MS/MS analysis confirmed that the oxidative coupling mediated modification was occurring site-selectively at the N-terminal proline residue of aldolase (Figure 4.9).

a) PHSHPALTPEQKKELSDIAHRIVAPGKGILAADESTGSIAKRLQSIGTENTEENRRFYRQLLTAD
DRVNPCIGGVILFHETLYQKADDGRPFPQVIKSKGGVVGIKVDKGVVPLAGTNGETTTQGLDGL
SERCAQYKKDGADFAKWRCVLKIGEHTPSALAIMENANVLARYASICQQNGIVPIVEPEILPDGD
HDLKRCQYVTEKVLAAVYKALSDHHIYLEGTLLKPNMVTPGHACTQKYSHEEIAMATVTALRRT
VPPAVTGVTFLSGGQSEEEASINLNAINKCPLLKPWALTFSYGRALQASALKAWGGKKENLKAA
QEEYVKRALANSLACQGKYTPSGQAGAAASESLFISNHAY

b)

Sequence	Modifications observed	Observed $[M+H]^+$	Expected $[M+H]^+$
PHSHPALTPEQK		1341.691	1341.691
KELSDIAHR		1068.581	1068.58
IVAPGK		-	*584.377*
GILAADESTGSIAK		1332.701	1332.701
RLQSIGTENTEENRR		1802.912	1802.91
FYR		-	*485.251*
QLLLTADDRVNPCIGGVILFHETLYQK	C13(Carbamidomethyl)	3113.649	3170.66914
ADDGRPFPQVIKSKGGVVGIK		2168.216	2168.219
VDKGVVPLAGTNGETTTQGLDGLSER		2614.332	2614.332
CAQYKK	C1(Carbamidomethyl)	797.398	854.41769
DGADFAK		723.331	723.331
WR		-	*361.198*
CVLK		-	*462.275*
IGEHTPSALAIMENANVLAR		2107.099	2107.096
YASICQQNGIVPIVEPEILPDGDHDLKR	C5(Carbamidomethyl)	3176.609	3233.62923
CQYVTEK	C1(Carbamidomethyl)	927.424	984.444303
VLAAVYK		763.472	763.471
ALSDHHIYLEGTLLKPNMVTPGHACTQK	C25(Carbamidomethyl)	3131.575	3188.59534
YSHEEIAMATVTALRR		1847.943	1847.943
TVPPAVTGVTFLSGGQSEEEASINLNAINK		3043.562	3043.558
CPLLKPWALTFSYGR	C1(Carbamidomethyl)	1808.952	1865.97163
ALQASALK		801.482	801.483
AWGGK		-	*518.272*
KENLK		-	*631.377*
AAQEEYVK		1093.564	937.46233
RALANSLACQGK	C9(Carbamidomethyl)	1288.679	1345.69914
YTPSGQAGAAASESLFISNHAY		2242.045	2242.041

Figure 4.7. Tryptic digest results of unmodified aldolase. (a) Amino acid sequence of aldolase from rabbit muscle. (b) Peptide fragments accounting for the full aldolase sequence are tabulated. Aldolase used in the tryptic digest was first evaluated by intact MS, as seen in Figure 4.6b. In italics are the fragments that were not observed in the analysis. Only the longest, unique fragments observed are shown. A 93.1% sequence coverage was observed. During the digestion protocol, cysteines were capped with iodoacetamide, as observed by the carbamidomethyl modifications.

Sequence	Modifications observed	Observed $[M+H]^+$	Expected $[M+H]^+$
PHSHPALTPEQK	P1 (+120 Da)	1461.713	1461.711
PHSHPALTPEQK		1341.691	1341.691
KELSDIAHR		1068.581	1068.580
IVAPGKGILAADESTGSIAK		1898.059	1898.059
RLQSIGTENTEENRR		1802.912	1802.910
FYR		-	*485.251*
QLLLTADDR		1044.569	1044.568
VNPCIGGVILFHETLYQK	C4(Carbamidomethyl)	2088.096	2088.093
ADDGRPFPQVIK		1342.712	1342.711
SKGGVVGIK		844.526	844.525
VDKGVVPLAGTNGETTTQGLDGLSER		2614.335	2614.332
CAQYKK	C1(Carbamidomethyl)	797.398	797.396
DGADFAK		723.331	723.331
WR		-	*361.198*
CVLK		-	*462.275*
IGEHTPSALAIMENANVLAR		2107.096	2107.096
YASICQQNGIVPIVEPEILPDGDHDLKR	C5(Carbamidomethyl)	3176.605	3176.603
CQYVTEK	C1(Carbamidomethyl)	927.425	927.423
VLAAVYK		763.472	763.471
ALSDHHIYLEGTLLKPNMVTPGHACTQK	C25(Carbamidomethyl)	3131.579	3131.575
YSHEEIAMATVTALRR		1847.946	1847.943
TVPPAVTGVTFLSGGQSEEEASINLNAINK		3043.563	3043.558
CPLLKPWALTFSYGR	C1(Carbamidomethyl)	1808.952	1808.950
ALQASALK		801.483	801.483
AWGGK		-	*518.272*
KENLK		-	*631.377*
AAQEEYVK		937.463	937.463
RALANSLACQGK	C9(Carbamidomethyl)	1288.679	1288.679
YTPSGQAGAAASESLFISNHAY		2242.044	2242.041

Figure 4.8. Tryptic digest results of aldolase modified with 2-amino-*p*-aminocresol (+120). Peptide fragments accounting for the full aldolase sequence are tabulated. N-terminally modified aldolase used in the tryptic digest was first evaluated by intact MS, as seen in Figure 4.6b. In red is the N-terminal tryptic peptide, showing an expected mass addition of 120 Da. In italics are the fragments that were not observed in the analysis. Only the longest, unique fragments observed are shown. A 94.8% sequence coverage was observed. During the digestion protocol, cysteines were capped with iodoacetamide, as observed by the carbamidomethyl modifications.

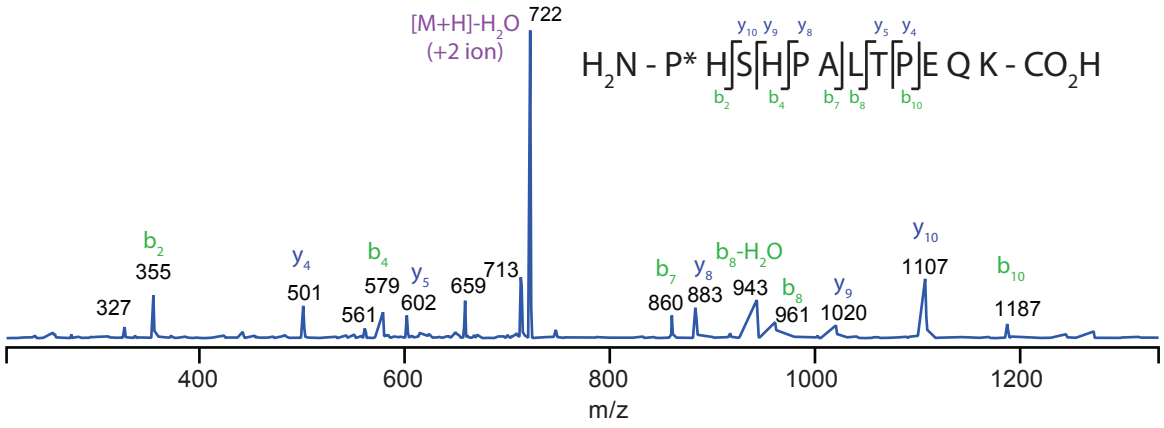


Figure 4.9. MS/MS analysis of the N-terminal tryptic peptide of aldolase. The y ions are shown in blue, the b ions are shown in green (with neutral losses of water). The analysis is consistent with the expected modification of +120 Da at the N-terminal proline residue, as represented by the * in the peptide sequence.

4.2.3. Evaluating the Activity of Surface Immobilized Aldolase.

The A'-aldolase bioconjugate was incubated on A-modified (complementary) surfaces to allow for DNA directed immobilization. Enzymatic activity was subsequently evaluated and the results of these assays are shown in Figure 4.10. Because hybridization to the surface is an internal purification tool, unmodified aldolase did not need to be purified away from DNA modified aldolase. In order to evaluate any non-specific binding that was occurring, a control was included where the DNA-protein conjugate was incubated on B-modified (non-complementary) surfaces. Incubation concentrations were decided upon after studies evaluating non-specific adsorption were carried out (Figure 4.11). As compared to the control, we observed that when A'-aldolase was incubated on the surface displaying its complementary strand, aldolase activity was significantly elevated, and background activity was minimal. This confirmed that aldolase was successfully immobilized with very low levels of non-specific adsorption and that the surface immobilization itself did not destroy its quaternary structure. An additional control was included with 20 nM free A'-aldolase in solution to ensure that the assay was functioning as expected. Fluorescence studies were also conducted to visualize each step qualitatively, and are depicted in Figures 4.12 and 4.13.

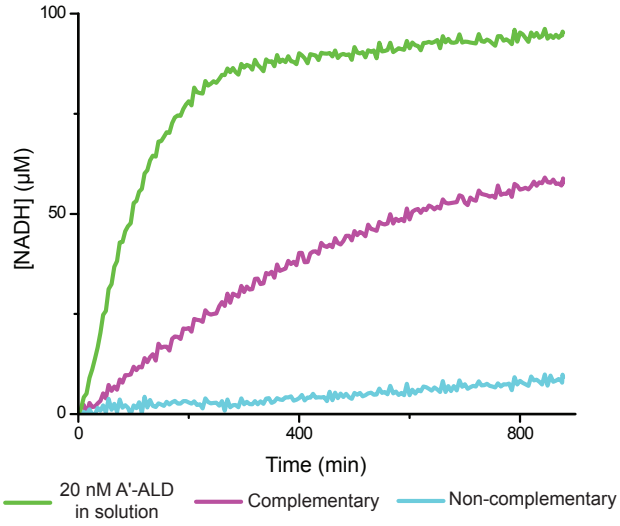


Figure 4.10. Testing the activity of DNA-aldolase conjugates immobilized after hybridization to the glass surface. Activity assay of A'-aldolase exposed to a glass surface displaying the complementary DNA strand (A, pink), the non-complementary DNA strand (B, blue), and free in solution at a concentration of 20 nM (green). All assays were run at 37 °C and conducted in triplicate.

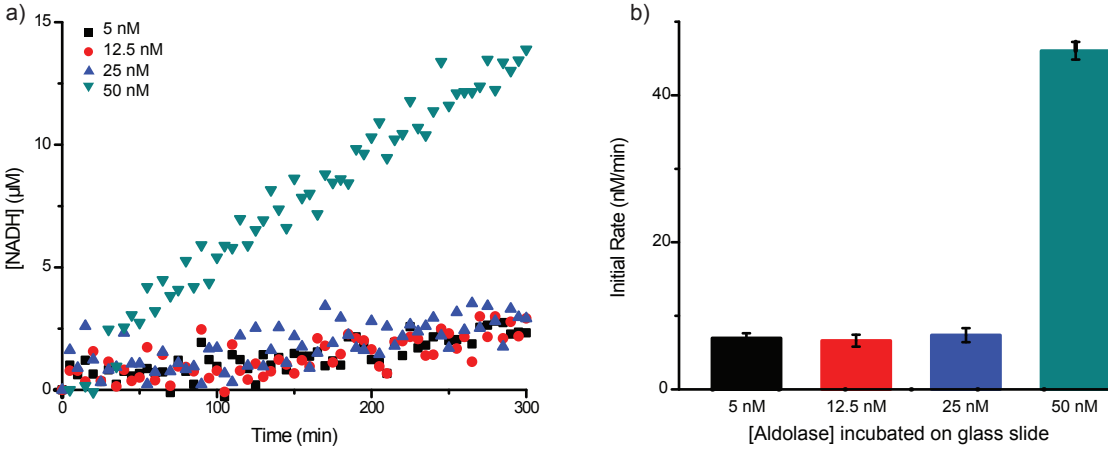
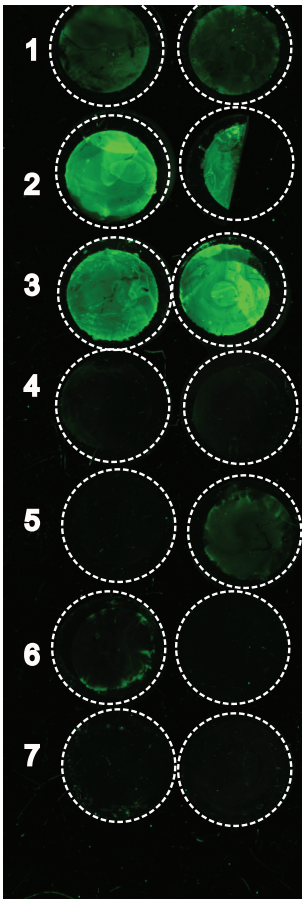


Figure 4.11. Evaluation of non-specific protein adsorption on glass slides. Glass slides were modified with single stranded DNA (sequence A), and then incubated with free aldolase for 1 h at room temperature. Concentrations of aldolase incubated on the slides were 5 nM, 12.5 nM, 25 nM, and 50 nM (based on the monomer). Slides were rinsed and an activity assay was carried out the determine residual activity levels on non-specifically adsorbed aldolase. (a) Activity assay of non-specifically adsorbed aldolase. (b) Initial rates for non-specifically adsorbed aldolase. All activity assays were run at 37 °C. For subsequent aldolase incubations, 25 nM aldolase (based on the monomer) was used.

a)



b)

Name	Sequence	T_m (°C)
A	5′ - CCC TAG AGT GAG TCG TAT GA - 3′	52.6
A′	5′ - TCA TAC GAC TCA CTC TAG GG - 3′	52.6
A′*	5′ - AlexaFluor488 TCA TAC GAC TCA CTC TAG GG - 3′	52.6
B	5′ - AGT GAC AGC TGG ATC GTT AC - 3′	54.4
C′*	5′ - AlexaFluor488 ACT GAT GGT AAT CTG CAC CT - 3′	53.5

c)

#	Surface DNA	Incubation 1: Protein	Incubation 2: Fluorescently tagged DNA
1	A	A′-ALD	A′*
2	A	-	A′*
3	A	ALD	A′*
4	A	A′-ALD	C′*
5	A	-	C′*
6	A	ALD	C′*
7	B	A′-ALD	A′*

d)

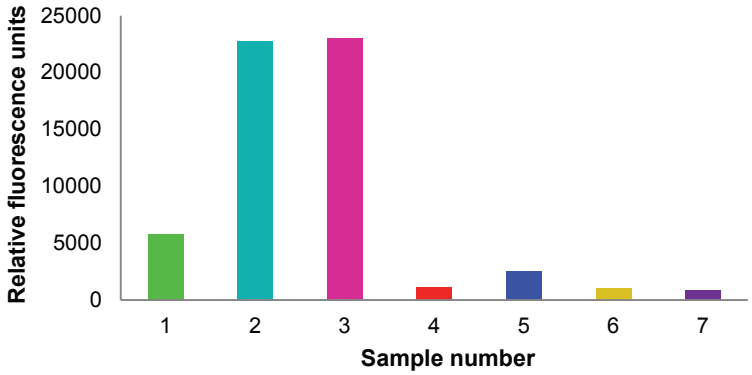
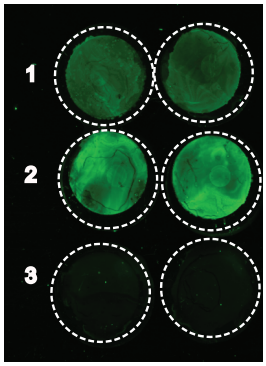


Figure 4.12. Using fluorescent DNA to visualize protein immobilization. To verify the attachment of aldolase through DNA directed immobilization, glass slides were modified with either strand A or B. Then, surfaces were incubated with either A′-aldolase (A′-ALD) or just aldolase (ALD). Following the first incubation, a second incubation was carried out with either complementary (A′*) or non-complementary DNA that was fluorescently labelled with AlexaFluor488 (C′*). Two replicates are shown for each set of experimental conditions. Incubation 2 allowed for backfilling of sites not occupied by A′-ALD, as seen in sample 1. Samples 2 and 3 were positive controls and 4-7 were negative controls. Because the fluorescence in sample 1 is lower than sample 2, it indicates that hybridization of A′-ALD occupied sites on the A surface and thus led to fewer sites being accessible for hybridization during incubation step 2. Additionally, sample 3 confirms that there is little, if any, non-specific binding to the surface, or at the very least it does not prevent DNA hybridization from occurring.

a)



b)

#	Surface DNA	Incubation sample
1	A	A'-ALD*
2	A	A'*
3	B	A'-ALD*

c)

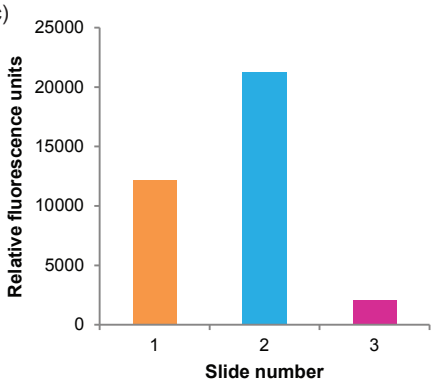


Figure 4.13. Using a fluorescent aldolase DNA conjugate to visualize protein immobilization. Slides modified with single stranded DNA were modified with fluorophore (Oregon Green 488) labeled aldolase (A'-ALD*). Slides modified with A showed a fluorescence increase due to complementary strand hybridization (1) and slides modified with B did not show a significant increase due to non-complementarity between the seqences (2). As a positive control, complementary DNA with AlexaFluor 488 conjugated onto it (A'*) was hybridized to slides with strand A. The fluorescence increase observed in slide 1 over slide 3 indicated DNA directed attachment of aldolase onto the glass surface. Fluorescence data were quantified using ImageJ software, and the data plotted are the average of two replicates.

4.2.4. Reusing the Protein Immobilized Surfaces.

Given the successful immobilization of aldolase onto the glass slides, we were interested in investigating the reusability of the surfaces. We ran each cycle for 15 h, rinsed reagents from the wells and repeated the assay using the same surface. These data are shown in Figure 4.14. It can be seen that, while we do see a drop in activity in each subsequent cycle, about half of the activity is maintained from one run to the next. We hypothesized that because the assay was conducted at 37 °C, the temperature could be contributing to inactivation of the protein over time. To test this, we incubated unmodified aldolase in solution at 37 °C for lengths of time equivalent to each iterative cycle, and we observed that the drop in activity was in fact a result of the protein being exposed to the elevated temperatures for extended periods of time (Figure 4.15).

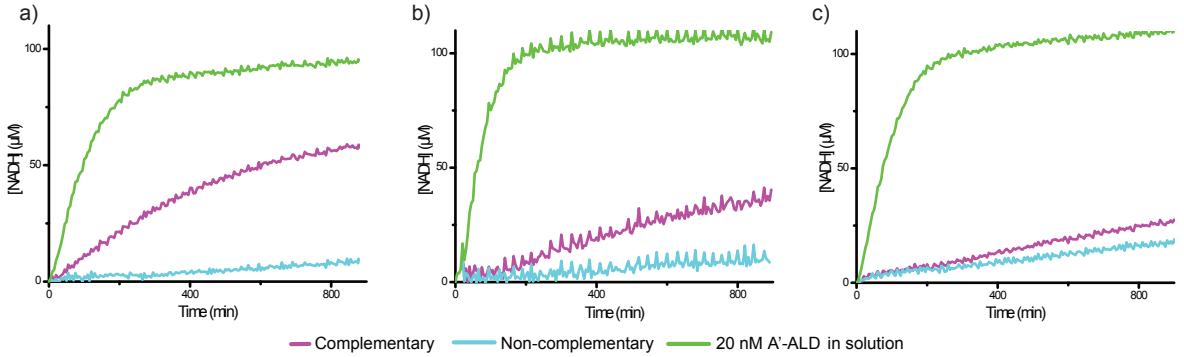


Figure 4.14. Testing for reusability of aldolase modifed surfaces. (a) Cycle 1: activity assay of A′-aldolase immobilized on a glass surface with the complementary DNA strand, A (pink), with the non-complementary DNA strand, B (blue), and free in solution at a concentration of 20 nM (green). (b) Cycle 2: activity assay after 24 h on surfaces used in a, after rinsing of glass sides and addition of fresh substrates. (c) Cycle 3: activity assay after 24 h on surfaces used in b, after rinsing of glass slides and addition of fresh substrates. All samples were run in triplicate, and data shown are the averages observed. All trials were run at 37 °C.

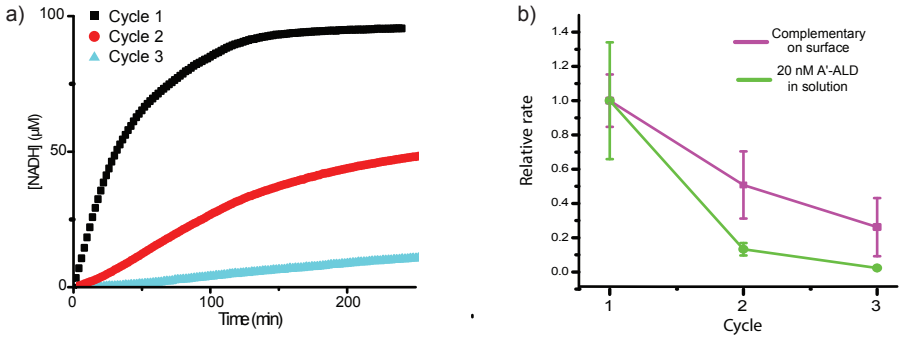


Figure 4.15. Testing the effect of elevated temperature for extended time on aldolase activity. (a) A 96-well plate was loaded with three sets of an aldolase activity assay solution. The first set was initiated by the addition of FBP and the activity was measured over 15 h at 37 °C (black). The 96-well plate was stored at room temperature for 9 h and then the second set of aldolase assays were initiated by the addition of FBP. The activity of this batch was measured over a subsequent 15 h at 37 °C (red). This was repeated for a third cycle (blue). (b) Compiled data for the initial rates of activity for multiple cycles of aldolase in solution and surface immobiized aldolase. All samples were run in triplicate at 37 °C, and data shown are the averages observed.

4.2.5. Surface Characterization with Atomic Force Microscopy.

Atomic Force Microscopy (AFM) studies were carried out at the various stages of the surface modification process on mica surfaces. Mica was chosen due to its atomically flat nature.[32] This allowed for verification that any changes to the surface morphology were due solely to the chemistries we applied and not the underlying morphology of the substrate. Additionally, free surface silanol groups on mica allowed for identical surface chemistry to that used on the glass

slides. AFM images were taken of surfaces functionalized with a) aniline, b) single stranded DNA (sequence A), c) double stranded DNA (sequence A' hybridized to sequence A), and d) immobilized aldolase via DNA hybridization (A'-aldolase hybridized to A) (Figure 4.16). It was seen that the aniline functionalized surface was uniformly flat, showing minimal variation in height over the observed region. Upon the attachment of single stranded DNA, the surface became rougher, showing a high density of small features of increased height. These features became larger in area upon addition of the complementary strand of DNA. The addition of the A'-aldolase conjugate to a surface displaying single stranded A DNA continued the trend of increasing morphological heterogeneity. These images verified that the surface was becoming more complex at each stage of the modification process, and thus the morphologies are changing in a manner consistent with what we expected to see based on the fluorescence imaging studies.

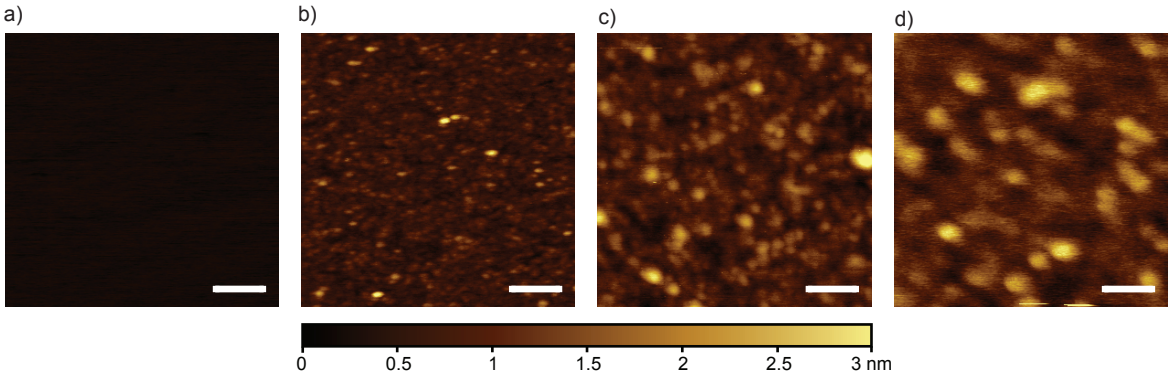


Figure 4.16. Height AFM images taken in non-contact mode of (a) aniline modified mica; (b) mica functionalized with single-stranded DNA; (c) complementary DNA hybridized to mica functionalized with single-stranded DNA; and (d) complementary DNA-aldolase conjugate hybridized to mica functionalized with single-stranded DNA. Scale bars are 50 nm.

4.2.6. Hybridization Temperature Modulates Immobilization Levels.

Previous reports have suggested that levels of modification on the surface play a significant role in activity levels, and that higher surface coverage does not always correlate to higher activity due to the effects of over-crowding and blockage of enzyme active sites.[33] Given this information, having a method to tune the level of modification in either direction could prove useful for tailoring these surfaces for different proteins. For all of the experiments described thus far, hybridization between DNA-aldolase and DNA modified glass surfaces was carried out at room temperature. Interestingly, when hybridization temperatures were varied (4, 23 and 37 °C), it was observed that the increasing temperatures resulted in increasing levels of protein immobilization. This was first determined through backfilling of open ssDNA sites after aldolase had been immobilized, where an expected trend of decreasing fluorescence with increasing annealing temperature was observed when the average fluorescence was quantified for the total surface area of the glass slides (Figure 4.17a, b). We hypothesize that the closer the hybridization temperature is to the melting temperature of the DNA strands (52.6 °C), the more efficient the thermal annealing becomes because the strands can melt and rehybridize to reach optimal coverage. At a lower temperature the strands are more restricted to the first location of hybridization, leading to lower levels of surface coverage. We also observed an increase in aldolase activity at the higher levels of modification, and could use this approach to refine our surfaces further (Figure 4.17c). Additionally, in the future, we envision using temperature variations to determine the levels at which we can saturate surfaces with enzymes without impeding their catalytic activity, in an effort to obtain surfaces with maximum efficiency.[34]

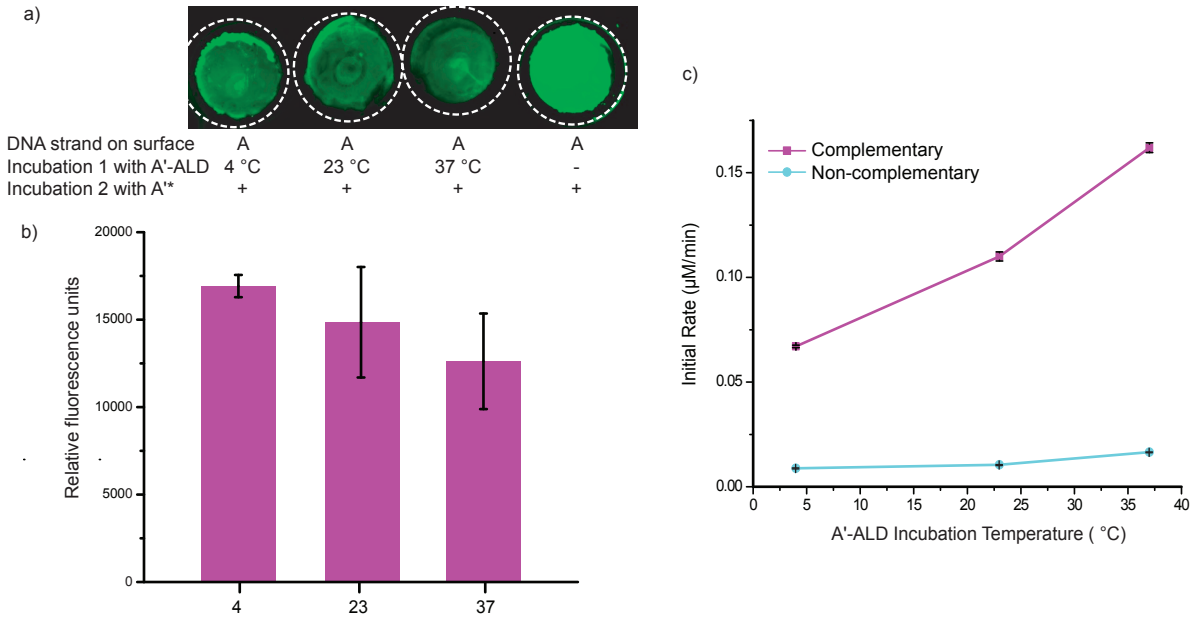


Figure 4.17. Impact of annealing temperature on surface hybridization levels. (a) Fluorescence images of glass slides modified with A, then incubated with A'-aldolase at varying temperatures. Subsequent incubation with fluorescently tagged DNA, A'*, allowed for backfilling of open surface DNA sites. (b) Quantification of fluorescent slides. Lower fluorescence corresponds to greater hybridization of the A'-aldolase conjugate. (c) Initial rates of A'-aldolase activity when immobilized on surfaces with strand A (complementary) or strand B (non-complementary) at varying incubation temperatures, conducted in triplicate.

4.2.7. Regenerating and Recycling the Single Stranded DNA Modified Surfaces.

A particularly noteworthy advantage of using DDI to orient proteins onto solid surfaces is that the DNA strands can be separated in order to remove the DNA-aldolase conjugate and regenerate the surface bearing single stranded DNA. This allows for storage of the slides for an

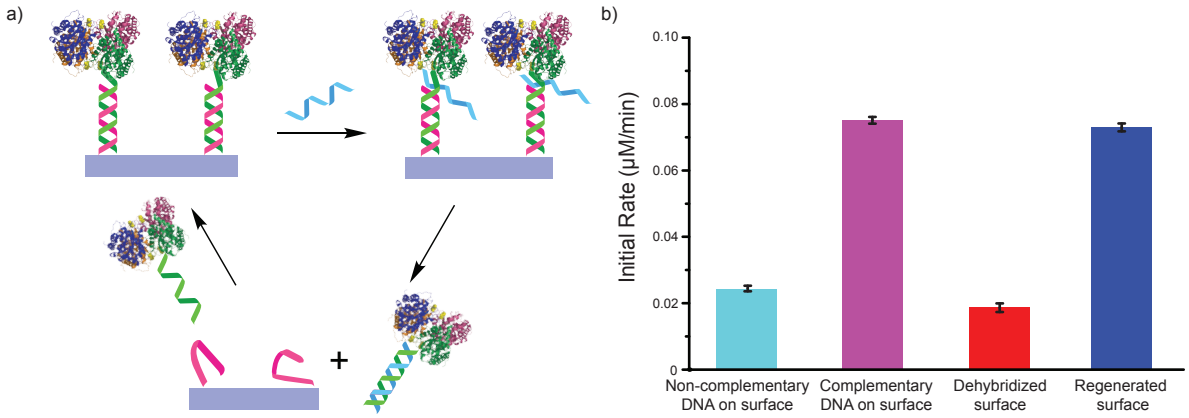
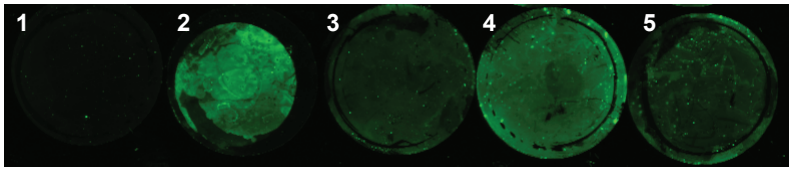


Figure 4.18. Testing the activity of regenerated DNA-aldolase conjugates immobilized on glass surfaces. (a) Schematic of DNA strand displacement mediated surface regeneration. (b) Initial rates of the activity of Ax'-aldolase exposed to a glass surface displaying the non-complementary DNA strand (strand B), and the complementary DNA strand (strand A) were obtained. Then, DNA strand displacement was carried out to remove the Ax'-aldolase conjugate from the glass surface. The regenerated surfaces were then incubated with a new batch of Ax'-aldolase. Initial rates of activity for both of these surfaces were obtained. All assays were run at 37 °C and conducted in triplicate.

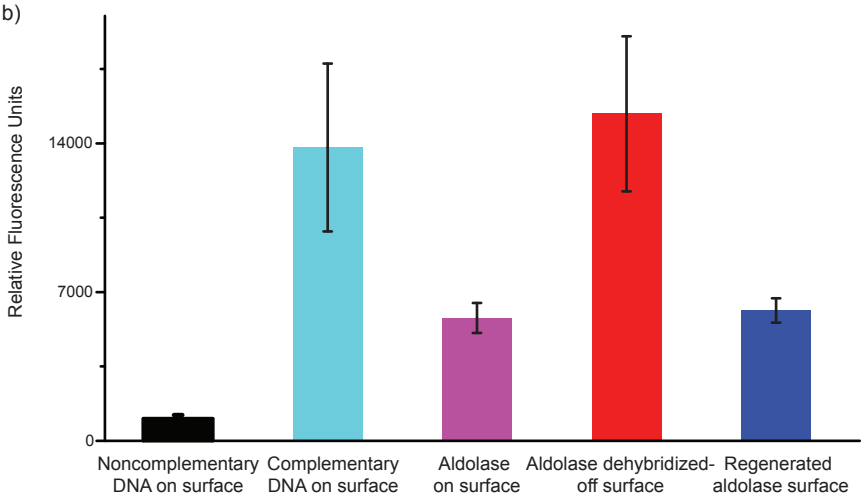
extended period of time due to the stability of DNA, and ultimately the ability to reuse them in future assays. We used DNA strand displacement to remove the DNA-aldolase conjugate and then rehybridized a fresh batch of DNA-aldolase. In this assay, we conjugated aldolase onto 25 base strand Ax', where 20 bases were complementary to A, but the remaining 5 served as an overhang. Ax'-aldolase was immobilized to glass slides displaying A. Then, Ax, a 25 base strand with complete complementarity to Ax', was used to displace Ax'-aldolase from the surface. This surface was then reused, and a fresh batch of Ax'-aldolase was immobilized. Activity assays of the surfaces at each stage were carried out, and as seen in Figure 4.18, the activity of aldolase immobilized on fresh glass slides remained consistent with the activity observed on aldolase that was immobilized on a regenerated glass slide. Fluorescence studies also corroborated these trends (Figure 4.19).

a)



Surface DNA	B	A	A	A	A
First incubation	-	-	Ax′-ALD	Ax′-ALD	Ax′-ALD
Second incubation	-	-	-	Ax	Ax
Third incubation	-	-	-	-	Ax′-ALD
Fourth incubation	Ax′*	Ax′*	Ax′*	Ax′*	Ax′*

b)



c)

Name	Sequence	T_m (°C)
A	5′ - CCC TAG AGT GAG TCG TAT GA - 3′	52.6
Ax	5′ - CCC TAG AGT GAG TCG TAT GAA AAA A - 3′	54.4
Ax′	5′ - TTT TTT CAT ACG ACT CAC TCT AGG G - 3′	54.4
Ax′*	5′ - AF488 TTT TTT CAT ACG ACT CAC TCT AGG G - 3′	54.4
B	5′ - AGT GAC AGC TGG ATC GTT AC - 3′	54.4

Figure 4.19. Using fluorescent DNA to visualize protein immobilization and regeneration. (a) Fluorescence images of various modified glass slides backfilled with DNA labelled with AlexaFluor488 (Ax′*). Slides were modified with: the non-complementary DNA strand (B, 1); the complementary DNA strand (A, 2); the complementary DNA strand (A) and then incubated with Ax′-aldolase (Ax′-ALD, 3); strand A, then incubated with Ax′-ALD, and then incubated with the complement to Ax′ (Ax) to dehybridize the DNA-enzyme conjugate off the surface (4); strand A, then incubated with Ax′-ALD, then incubated with Ax, and reincubated with Ax′-ALD to test for the regenerability of the ssDNA modified glass slides (5). (b) Quantification of the fluorescence using ImageJ software. A greater level of fluorescence corresponds to more open DNA sites for the fluorescently labelled DNA to hybridize to. Each experimental condition was run in triplicate. (c) DNA sequences used in this experiment.

4.3. Conclusions

In this work, we have prepared a single class of aminophenol substituted DNA strands and used them to modify both glass surfaces and protein N-termini. By piecing together each component, we have developed a step-by-step platform for producing oriented displays of proteins on glass surfaces. We have qualitatively verified our chemistry at each stage through fluorescence studies and AFM, and have also demonstrated its utility by testing the enzymatic activity of surface immobilized aldolase. This allows for the oriented immobilization of proteins with an adjustable spacer, where the enzyme can be reused in multiple cycles. Additionally, through DNA strand diplacement, we have successfully regenerated the single stranded DNA bearing glass surfaces, and we have shown them to be reusable in subsequent hybridization assays. The chemistry involved in attaching the first strand of DNA to the surface is convenient, quick, and stable. The conjugation of the complementary DNA strands to proteins is biocompatible, quick, and only requires low concentrations of the coupling partners. In addition, because the native N-terminal amine is being targeted as the attachment site, it can be applied to a large scope of proteins with minimal genetic engineering being required.

Because we have used DNA hybridization as our mode of protein attachment, the easily accessible diversity of DNA strands offer a wide range of attachment handles, both in terms of linker length and rigidity. Additionally, the generalizable nature of this DDI method should facilitate the immobilization of a variety of proteins, with the only stipulation being that the protein have a proline at the N-terminus. Given these advantages, we seek to enhance this platform further in the future.

Taking advantage of the transparent nature of the glass surfaces used in these studies, we are also seeking to characterize these surfaces using alternative spectroscopic techniques, such as Sum Frequency Generation and X-Ray Photoelectron Spectroscopy, to gain information about the orientation and coverage of the protein.[35–37]

4.4. Materials and Methods

4.4.1. General Procedures and Materials

Unless otherwise noted, the chemicals and solvents used were of analytical grade and were used as received from commercial sources. Water (dd-H_2O) used as a reaction solvent and in biological procedures was deionized using a Barnstead NANOpure purification system (ThermoFisher, Waltham, MA). Aldolase from rabbit muscle, glyceraldehyde-3-phosphate dehydrogenase from rabbit muscle, NAD^+, NADH and Fructose-1,6-bisphosphate was obtained from Aldrich (St. Louis, MO). Single stranded 5' aminated or fluorophore labeled DNA molecules were purchased from Integrated DNA Technologies. Absorbance measurements of samples in 24 and 96–well plates were obtained on a Tecan Infinite 200 Pro plate reader. Fluorecence images of glass slides were taken on a Typhoon 9410 variable mode imager (Amersham Biosciences).

4.4.2. Instrumentation and Sample Analysis

NMR. 1H spectra were measured with a Bruker AVQ-400 (400 MHz, 100 MHz) spectrometer.

Mass Spectrometry of DNA strands. Matrix assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed on a Voyager-DE system (PerSeptive

Biosystems, USA) and data were analyzed using Data Explorer software. Oligonucleotide samples were co-crystallized using 3-hydroxypicolinic acid: ammonium citrate solution (9:1) in 1:1 acetonitrile (MeCN) to H_2O.

Atomic Force Microscopy. The Agilent 5500 system (Keysight Technologies Inc., Santa Rose, CA, USA) was used for AFM high resolution imaging of the surface modification on mica. Non-contact AFM images were obtained under dry nitrogen conditions (<1%RH) using Tap150Al-G probes (Innovative Solutions Bulgaria Ltd., Sofia, Bulgaria) with a nominal resonant frequency value of 150 kHz. Freshly cleaved muscovite mica, V1 quality (Electron Microscopy Sciences, Hatfield, PA, USA) was used as AFM substrates for surface modification. The AFM images were analyzed using Gwyddion SPM data analysis software.

Full Length Protein Mass Spectrometry. Proteins and protein conjugates were analyzed on an Agilent 6224 Time-of-Flight (TOF) mass spectrometer with a dual electrospray source connected in-line with an Agilent 1200 series HPLC (Agilent Technologies, USA). Chromatography was performed using a Proswift RP-4H (Thermo Scientific, USA) column with a H_2O/MeCN gradient mobile phase containing 0.1% formic acid. Mass spectra of proteins and protein conjugates were deconvoluted with the MassHunter Qualitative Analysis Suite B.05 (Agilent Technologies, USA).

Liquid Chromatography/Tandem Mass Spectrometry. High-resolution electrospray ionization (ESI) and liquid chromatography with tandem mass spectrometry detection (LC-MS/MS) mass spectra were obtained at the UC Berkeley QB3/Chemistry Mass Spectrometry Facility.

Trypsin-digested protein samples were analyzed using a Thermo Dionex UltiMate3000 RSLCnano liquid chromatograph that was connected in-line with an LTQ-Orbitrap-XL mass spectrometer equipped with a nanoelectrospray ionization (nanoESI) source (Thermo Fisher Scientific, Waltham, MA). The LC was equipped with a C18 analytical column (Acclaim® PepMap RSLC, 150 mm length × 0.075 mm inner diameter, 2 μm particles, 100 Å pores, Thermo) and a 1 μL sample loop. Acetonitrile, formic acid (Fisher Optima grade, 99.9%), and water purified to a resistivity of 18.2 MΩ·cm (at 25 °C) using a Milli-Q Gradient ultrapure water purification system (Millipore, Billerica, MA) were used to prepare mobile phase solvents. Solvent A was 99.9% water/0.1% formic acid and solvent B was 99.9% acetonitrile/0.1% formic acid (v/v). The elution program consisted of isocratic flow at 2% B for 4 min, a linear gradient to 30% B over 38 min, isocratic flow at 95% B for 6 min, and isocratic flow at 2% B for 12 min, at a flow rate of 300 nL/min. Full-scan mass spectra were acquired in the positive ion mode over the range m/z = 350 to 1600 using the Orbitrap mass analyzer, in profile format, with a mass resolution setting of 60,000 (at m/z = 400, measured at full width at half-maximum peak height, FWHM). In the data-dependent mode, the eight most intense ions exceeding an intensity threshold of 20,000 counts were selected from each full-scan mass spectrum for tandem mass spectrometry (MS/MS) analysis using collision-induced dissociation (CID). Data acquisition and analysis were performed using Xcalibur (version 2.0.7) and Proteome Discoverer software (version 1.3, Thermo), respectively. Peptide identifications were validated by manual inspection of MS/MS spectra, i.e., to check for the presence of y-type and b-type fragment ions that identify the peptide sequences.[38]

4.4.3. Preparation of Aniline Functionalized Glass Slides

Synthesis of 3-(4-azidophenyl)-*N*-(3-trimethoxysilylpropyl) Propanamide (Phenylazide Silane). Phenylazide silane was synthesized following a previously pubished protocol.[2] Briefly, to a

solution of 3-(4-azidophenyl) propionic acid (574 mg, 3.0 mmol) in 90 mL of anhydrous THF under positive nitrogen pressure were added DIPEA (0.7 mL, 7.5 mmol, excess) and pentafluorphenyl trifluoroacetate (0.64 mL, 3.75 mmol), which was added slowly over 20 min, resulting in the formation of dense, white fumes. The reaction was stirred for 2 h at RT. To this solution was added 3-aminopropyl trimethoxysilane (0.57 mL, 3.3 mmol), and the resulting solution was stirred under nitrogen at RT for 18 h. The solvent was then removed using a rotary evaporator. After purification by flash chromatography (100% EtOAc), a light yellow oil was obtained, and confirmed by NMR.

Modification of Glass Slides with 3-(4-azidophenyl)-*N*-(3-trimethoxysilylpropyl) Propanamide. Circular glass slides (Fisher Scientific, 15 mm diameter, 0.17-0.25 mm thick) were sonicated (Cole-Parmer Ultrasonic Cleaner 8890-R-MTH) for 2 min in acetone and isopropanol consecutively, then immersed in Nanopure water. The slides were dried with N_2 and then cleaned with oxygen plasma (Plasma Equipment Technical Services, Inc, PETS reactive ion etcher, RIE-1) at 100 W (~0.2 Torr) for 5 min. Following plasma cleaning, the slides were immersed in a solution of 25 mM 3-(4-azidophenyl)-*N*-(3-trimethoxysilylpropyl) propanamide in methanol containing 1% v/v water and 148 mM acetic acid. The slides were incubated in this solution at RT for 2 h with stirring, after which they were rinsed in a solution of 148 mM acetic acid in methanol for 2 min. The slides were then rinsed with pure methanol, and finally dried with a stream of N_2. After drying, the slides were cured in an oven at 110 °C for a minimum of 12 h to promote covalent modification of the glass surface. These phenylazide coated glass slides were stored in the dark at RT in a dessicator until the azide groups were reduced to the anilines, as described below. The silanization solution could be reused a minimum of six times with no change in performance if stored at -20 °C between uses.

Reduction of Phenylazide Functionalized Glass Slides to Aniline Functionalized Glass Slides. Phenylazide modified slides were submerged in a solution of 50 mM TCEP in 200 mM sodium phosphate buffer at pH 7.5 and incubated at RT with stirring for 30 min. After reduction, the slides were rinsed in Nanopure water and dried with a stream of N_2. The aniline displaying glass slides were stored at RT in a desiccator until use.

4.4.4. Synthesis of Aminophenol-DNA.

Amine-modified DNA was dissolved to 2.5 mM. The reaction conditions describe the generalized procedure for the modification of amine functionalized DNA. [Representative DNA strand A: 5'/5AmMC6/ CCC TAG AGT GAG TCG TAT GA-3' (5AmMC6 = 6-aminohexyl phosphate)]. To a solution of 125 µL of 100 mM pH 8 phosphate buffer was added 125 µL of a 2.5 mM amine DNA solution (0.31 µmol). To this were added 150 µL of DMF and 100 µL of a ~500 mM solution of nitrophenol-NHS (~50 µmol) in DMSO. The 3(4-hydroxy-3-nitrophenyl) propionic acid NHS ester (nitrophenol-NHS) was prepared following a previously published protocol.[3] The solution was shaken for at least 4 h. This solution was added to a NAP-5 column (GE Healthcare), equilibrated and eluted with nanopure water. To the resulting solution was added 53 µL of 200 mM $Na_2S_2O_4$ to reduce the nitrophenol to the desired aminophenol. The $Na_2S_2O_4$ was stored in a dessicator at RT, and a 200 mM stock solution was prepared fresh in 0.2 M phosphate buffer, pH 6.5 before addition. The solution was shaken for at least 20 minutes before being directly loaded onto a NAP-10 column, and the elution process was repeated. The eluent was lyophilized, yielding ~1.0 mg of a white solid (50%). The DNA was prepared via C-18 ziptip for MADLI-TOF analysis and modification was confirmed. Aminophenol-DNA stock solutions were prepared at 1 mM in water

and stored at -20 °C for future use.[25]

4.4.5. Patterning Single Stranded DNA on Aniline Gunctionalized Slides Using Potassium Ferricyanide Mediated Oxidative Coupling.

Aniline functionalized 15mm circular glass slides were modified with ssDNA. A 4.5 μL drop of 50 μM aminophenol-modified DNA, 1 mM $K_3Fe(CN)_6$, and 250 mM NaCl in 10 mM pH 6.5 phosphate buffer was placed on the center of the aniline glass slide. Another unmodified glass slide was placed on top of the drop causing the DNA solution to spread over the slide in a sandwich. Slides were incubated in the dark at RT for 1 h. Then, slides were dipped in water and the unmodified glass slides were removed from the DNA-modified glass. DNA-modified glass slides were rinsed in 0.4% SDS and then 10 mM pH 6.5 phosphate buffered saline (PBS) each at RT for 5 min with stirring. Slides were rinsed in Nanopure water and dried with a stream of N_2. The single stranded DNA displaying glass slides were stored at RT in a desiccator until use.

4.4.6. Annealing of Complementary DNA Strands on Single Stranded DNA Modified Glass Slides.

Glass slides (15mm in diameter) coated with ssDNA were incubated with the complementary strand for hybridization. PDMS wells were used to form a well on top of the glass slides, and 200 μL of 0.05 μM DNA in 5x SSC + 0.1% Tween 20 was added to the well, just enough to cover the top of the slide. Slides were incubated for 1h in humidifying conditions on an orbital shaker at RT. Following incubation, the PDMS wells were removed and the glass slides were rinsed in 5x SSC + 0.1% Tween 20 three times for 1 min and then in 1x SSC + 0.1% Tween 20 two times for 10 min. Following the rinses, glass slides were rinsed in water, dried with nitrogen and stored for subsequent use and analysis. Slides with fluorescent DNA strands were analyzed using a Typhoon 9410 variable mode imager. Controls involved DNA strand mismatches chosen such that hybridization should not occur. (20X SSC buffer: 0.3 M sodium citrate dihydrate, 3 M NaCl, pH 7.0)

4.4.7. Capping Free Cysteines on Aldolase with *N*-ethyl maleimide.

Free cysteines were capped from potential modification during subsequent oxidative coupling steps by reaction with *N*-ethyl maleimide (NEM). To a solution of aldolase from rabbit muscle (100 μL of a 118 μM solution in 100 mM pH 7.0 phosphate buffer) was added NEM (5 μL of a 100 mM solution in DMSO) and 95 uL of 100 mM pH 7.0 phosphate buffer. The reaction mixture was incubated at RT for 3 h and then the excess NEM was removed by repeated (6 times) centrifugal filtration against a 30 kDa MWCO membrane.

4.4.8. Synthesis of DNA-Aldolase Bioconjugate.

To the aldolase (20 μM) in 10 mM phosphate buffer, pH 7.5 was added 5 equiv. of the *o*-aminophenol modified DNA (100 μM). The solution was briefly vortexed and then 10 equiv. (relative to the *o*-aminophenol) of $K_3Fe(CN)_6$ (as a 10 mM stock solution in water) was added. After 30 min, the reaction was purified by repeated (>12 times) centrifugal filtration against a 30 kDa MWCO membrane (Millipore), allowing for purification of aldolase and aldolase-DNA conjugate from free DNA. Modification was monitored by a combination of SDS-PAGE and a BCA assay (kit purchased from Thermo Scientific).

4.4.9. Synthesis of Fluorescent DNA-Aldolase Conjugate.

To aldolase-DNA conjugate (15 μM based on DNA concentration) in 100 mM carbonate buffer, pH 10.0 was added 25 equiv. of commercially purchased Oregon Green NHS ester. This solution was vortexed for 1 h. The reaction was purified by repeated (>6 times) centrifugal filtration against a 30 kDa MWCO membrane (Millipore) into 25 mM phosphate buffer, pH 8.5, allowing for purification of fluorescent aldolase-DNA conjugate from free dye.

4.4.10. Characterization of DNA-Aldolase Conjugate.

For protein analysis, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a Mini Gel Tank apparatus from Life Technologies following the manufacturer's protocols. MOPS buffer purchased from Life Technologies was used as the electrode buffer. All protein electrophoresis samples were heated for at least 15 min at 100 °C in the presence of 1,4-dithiothreitol (DTT) to ensure reduction of any disulfide bonds. NuPAGE Bis-Tris Mini Gels (12%) were run for 50 min at 200 V to allow good resolution of bands. Commercially available markers (Bio Rad) were applied to at least one lane of each gel for assignment of apparent molecular masses. Visualization of protein bands was accomplished by staining with Coomassie Brilliant Blue R-250. Gel imaging was performed on a Gel Doc EZ™ Imager (Bio Rad). Image Lab software was used to determine the level of modification by optical densitometry.

Once the level of modification was determined, a BCA assay was run following the manufacturer's protocols to determined the total protein concentration of the stock solution.

4.4.11. Activity Assay of Aldolase in Solution.

Activities of aldolase, unmodified and modified, were meaured in solution in a 96-well plate. Each well had aldolase, NAD^+, GAPDH, and FBP added to it such that the final concentrations were 20 nM aldolase, 1 mM NAD^+, 1.35 μM GAPDH, and 100 μM FBP in 250 μL total reaction volume of the activity assay buffer (20 mM potassium phosphate, 10 mM potassium pyrophosphate, 3 μM dithiothreitol (DTT) pH 8.5). Prior to adding the FBP, the plate was equilibrated to 37 °C. After the addition of FBP, the absorbance at 340 nm was measured every 2 min for 5 h while holding the temperature at 37 °C (Tecan Infinite 200 Pro plate reader). Samples were run in triplicate.

4.4.12. Immobilization of DNA-Aldolase Onto Glass Surfaces and Analysis of Activity.

Glass slides (15mm in diameter) coated with ssDNA (strands A and B) were loaded in a 24 well plate. These slides were incubated with DNA-aldolase conjugate (strand A') for hybridization. PDMS wells were used to form a well on top of the glass slides while in the 24-well plate, and 200 μL of 0.025 μM conjugate (based on DNA concentration) in 5x SSC + 0.1% Tween 20 was added to the well, just enough to cover the top of the slide. Slides were incubated for 1 h in humidifying conditions on an orbital shaker at RT (unless otherwise noted). Following incubation, 2 mL of 5x SSC + 0.1% Tween 20 were added to the wells and the PDMS wells were removed. The glass slides were rinsed in 5x SSC + 0.1% Tween 20 via repeated pipetting and then the plate was shaken twice for ten minutes with the wells filled with 2 mL of 1x SSC + 0.1% Tween 20. At this point, the PDMS wells were removed from the glass slides, but the slides remained immersed in solution within the 24-well plate.

After rinsing with SSC, the buffer was exchanged into the activity assay buffer (20 mM

potassium phosphate, 10 mM potassium pyrophosphate, 3 μM dithiothreitol (DTT) pH 8.5). Each well had NAD^+, GAPDH, and FBP added to it such that the final concentrations were 1 mM NAD^+, 1.35 μM GAPDH, and 100 μM FBP in 500 μL total volume. Prior to adding the FBP, the plate was equilibrated to 37 °C. A positive control was also run with the addition of 20 nM DNA-aldolase conjugate (strand A') in solution with wells containing a glass slide that had the non-complementary DNA strand attached to it. Immediately after the addition of FBP, the absorbance at 340 nm was measured every 5 min for 16 h while holding the temperature at 37 °C (Tecan Infinite 200 Pro plate reader). A standard curve with NADH was also prepared on the same plate, in concentrations ranging from 0-100 μM. All samples were run in triplicate.

4.4.13. Reusing Surfaces with Immobilized Aldolase.

After running the activity assay, the 24-well plate was removed from the plate reader and stored at 4 °C until further use (~6 h). The used wells had 2 mL of fresh activity assay buffer added to them and then 2 mL were drawn out, using a pipet. This was repeated at least two more times, so that the wells were rinsed without allowing the slides to dry out. After rinsing, each well had NAD^+, GAPDH, and FBP added to it such that the final concentrations were 1 mM NAD^+, 1.35 μM GAPDH, and 100 μM FBP in 500 μL total volume. Prior to adding the FBP, the plate was heated to 37 °C. After the addition of FBP, the absorbance at 340 nm was measured every 5 min for 16 h while holding the plate at 37 °C (Tecan Infinite 200 Pro plate reader). All samples were run in triplicate.

4.4.14. Modification of Glass Slides Regenerating Surfaces with Immobilized Aldolase.

Glass slides displaying strand A were modified with DNA-aldolase (strand Ax') and were tested for activity in a 24-well plate assay, as described above. Each well had 1 mL of 25 μM complementary DNA (strand Ax) in 5x SSC + 0.1% Tween 20 added to it to dehybridize the DNA-aldolase from the surface. Slides were incubated for 1 h in humidifying conditions on an orbital shaker at 37 °C. Following incubation, 2 mL of 5x SSC + 0.1% Tween 20 were added to the wells. The glass slides were rinsed in 5x SSC + 0.1% Tween 20 via repeated pipetting and then the plate was shaken twice for ten minutes with the wells filled with 2 mL of 1x SSC + 0.1% Tween 20. All liquid was drawn out of each well. PDMS wells were used to form a well on top of the glass slides while in the 24-well plate, and 200 μL of 0.025 μM conjugate (based on DNA concentration) in 5x SSC + 0.1% Tween 20 was added to the well, just enough to cover the top of the slide. Slides were incubated and rinsed as before.

After rinsing with SSC, the buffer was exchanged into the activity assay buffer (20 mM potassium phosphate, 10 mM potassium pyrophosphate, 3 μM dithiothreitol (DTT) pH 8.5). Each well had NAD^+, GAPDH, and FBP added to it such that the final concentrations were 1 mM NAD^+, 1.35 μM GAPDH, and 100 μM FBP in 500 μL total volume. Prior to adding the FBP, the plate was equilibrated to 37 °C. A positive control was also run with the addition of 20 nM DNA-aldolase conjugate (strand Ax') in solution. After the addition of FBP, the absorbance at 340 nm was measured every 5 min for 16 h while holding the temperature at 37 °C (Tecan Infinite 200 Pro plate reader). A standard curve with NADH was also prepared on the same plate, in concentrations ranging from 0-100 μM.

4.4.15. Atomic Force Microscopy studies.

Mica surfaces were used in the AFM studies. Surface modification protocols on mica were identical to those outlined above for glass slides. In lieu of sonication, a fresh mica layer was exposed just before immersing it into the 3-(4-azidophenyl)-*N*-(3-trimethoxysilylpropyl) propanamide solution. All subsequent steps remained unchanged when preparing surfaces functionalized with a) aniline, b) single stranded DNA (sequence A), c) double stranded DNA (sequence A' hybridized to sequence A), and d) immobilized aldolase via DNA hybridization (A'-aldolase hybridized to A).

4.4.16. Capping of Free Cysteines With 5,5'-dithio-bis-(2-nitrobenzoic Acid) (DTNB).

Free cysteines were protected from potential modification by disulfide formation with 5,5'-dithiobis-(2-nitrobenzoic acid), (DTNB). To a solution of aldolase (780 μL of a 100 μM solution) was added DTNB (20 μL of a 20 mM solution in 100 mM phosphate buffer, pH 7.2 with 1 mM EDTA). The reaction mixture was incubated at RT for 15-30 min and then the excess DTNB was removed by repeated (6 times) centrifugal filtration against a 10 kDa MWCO membrane. To reduce the disulfide, approximately 25 equiv. of TCEP (as a 0.5 M solution, pH 7.0) was added to the protein sample. The reaction mixture was incubated at RT for 15 min and then purified using a 0.5 mL centrifugal filter with a 10 kDa MWCO.

4.4.17. Modification of Aldolase With a Small Molecule *o*-aminophenol Reagent at the N-terminus for Mass Spectrometry Analysis.

To a solution of aldolase (20 μM) in 10 mM phosphate buffer, pH 7.5 was added 5 equiv. of 2-amino-*p*-cresol (100 μM). The solution was briefly vortexed and then 10 equiv. of $K_3Fe(CN)_6$ (1 mM as a solution in 10 mM phosphate buffer, pH 7.5) was added. After 30 min, the reaction was purified using a 0.5 mL centrifugal filter with a 10 kDa MWCO.

4.4.18. Trypsin Digestion of a Small Molecule Modified Aldolase for MS/MS Analysis.

For the tryptic digest, a procedure from UC Berkeley QB3 Mass Spectrometry Facility was followed.[39-41] To 5 μL of a 765 μM protein stock was added 20 μL of 8M urea and 0.5 μL of 500 mM DTT, all in 50 Tris buffer, pH 7.0. After incubation at 55 °C for 20 minues, 6 μL of a freshly made 100 mM iodoacetamide solution was added. The resulting solution was incubated at RT for 30 minutes in the dark. After incubation, 2 uL of 500 mM DTT were added. After incubation at RT for 20 minutes, 6 μL of the alkylyated protein from this sample was used for digestion. To 6 uL was added 2.5 uL of 1 M Tris, pH 7.0, 0.5 μL of 100 mM $CaCl_2$ 0.5 μL of 1 μg/μl trypsin (Promega), and 41.5 μL water. This solution was allowed to incubate overnight at RT for protein digestion to occur prior to sample analysis.

4.5. References

1. Phizicky, E.; Bastiaens, P. I. H.; Zhu, H.; Snyder, M.; Fields, S., Protein analysis on a proteomic scale. *Nature* **2003**, *422*, 208–215.

2. Brady, D.; Jordaan, J., Advances in enzyme immobilization. *Biotechnol. Lett.* **2009**, *31*, 1639–1650.

3. Jonkheijm, P.; Weinrich, D.; SchrÖder, H.; Niemeyer, C. M.; Waldmann, H., Chemical strategies

for generating protein biochips. *Angew. Chem. Int. Ed.* **2008**, *47*, 9618–9647.

4. Cha, T.; Guo, A.; Zhu, X., Enzymatic activity on a chip: The critical rold of protein orientation. *Proteomics* **2005**, *5*, 416–419.

5. Vallières, K.; Chevallier, P.; Sarra-Bournet, C.; Turgeon, S. Laroche, G., AFM imaging of immobilized fibronectin: Does the surface conjugation scheme affect the protein orientation/ conformation? *Langmuir* **2007**, *23*, 9745–9751.

6. Camarero, J. A.; Kwon, Y.; Coleman, M. A., Chemoselective attachment of biologically active proteins to surfaces by expressed protein ligation and its application for "protein chip" fabrication. *J. Am. Chem. Soc.* **2004**, *126*, 14730–14731.

7. Buhl, M.; VonhÖren, B.; Ravoo, B. J., Immobilization of enzymes via microcontact printing and thio-ene chemistry. *Bioconjugate Chem.* **2015**, *26*, 1017–1020.

8. Kindermann, M.; George, N.; Johnsson, N.; Johnsson, K., Covalent and site-specific immobilization of fusion proteins. *J. Am. Chem. Soc.* **2003**, *125*, 7810–7811.

9. Wong, L. S.; Thirlway, J.; Micklefield, J., Direct site-selective covalent immobilization catalyzed by a phosphopantetheinyl transferase. *J. Am. Chem. Soc.* **2008**, *130*, 12456–12464.

10. Abad, J. M.; Mertens, S. F. L.; Pita, M.; Fernandez, V. M.; Schiffrin, D. J., Functionalization of thioctic acid-capped gold nanoparticles for specific immobilization of histidine tagged proteins. *J. Am. Chem. Soc.* **2005**, *127*, 5689–5694.

11. Reynolds, N. P.; Tucker, J. D.; Davison, P. A.; Timney, J. A.; Hunter, C. N.; Leggett, G. J., Site-specific immobilization and micrometer and nanometer scale photopatterning of yellow fluorescent protein on glass Ssurfaces. *J. Am. Chem. Soc.* **2009**, *131*, 896–897.

12. Hutsell, S. Q.; Kimple, R. J.; Siderovski, D. P.; Willard, F. S.; Kimple, A. J., High affinity immobilization of proteins using biotin and GST-based coupling strategies. *Methods Mol. Biol.* **2010**, *627*, 75–90.

13. Nimse, S. B.; Song, K.; Sonawane, M. D.; Sayyed, D. R.; Kim, T., Immobilization techniques for microarray: challenges and applications. *Sensors* **2014**, *14*, 22208–22229.

14. Niemeyer, C. M., Semisynthetic DNA-protein conjugates for biosensing and nanofabrication. *Angew. Chem. Int. Ed.* **2010**, *49*, 1200–1216.

15. Beaucage, S. L., Strategies in the preparation of DNA oligonucleotide arrays for siagnostic applications. *Curr. Med. Chem.* **2001**, 8, 1213–1244.

16. Niemeyer, C. M.; Boldt, L; Ceyhan, B; Blohm, D., DNA-directed immobilization: Efficient, reversible, and site-selective surface binding of proteins by means of covalent DNA- streptavidin conjugates. *Anal. Biochem.* **1999**, *268*, 54–63.

17. Fruk, L.; MÜller, J.; Niemeyer, C. M., Kinetic analysis of semisynthetic peroxidase enzymes containing a covalent DNA-heme adduct as the cofactor. *Chem. Eur. J.* **2006**, *12*, 7448–7457.

18. Wold, E. D.; McBride, R.; Axup, J. Y.; Kazane, S. A.; Smider, V. V., Antibody microarrays utilizing site-specific antibody-oligonucleotide conjugates. *Bioconjugate Chem.* **2015**, *26*, 807–811.

19. Behrens, C. R.; Hooker, J. M., Obermeyer, A. C.; Romanini, D. W.; Katz, E. M.; Francis, M. B., Rapid chemoselective bioconjugation through the oxidative coupling of anilines and aminophenols. *J. Am. Chem. Soc.* **2011**, *133*, 16398–16401.

20. Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M,; Weng, C. H.; Iavarone, A. T.; Francis, M. B., Site-specific protein transamination using N-methylpyridinium-4-carboxaldehyde. *J. Am. Chem. Soc.* **2013**, *135*, 17223–17229.

21. 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.

22. Palla, K. S.; Witus, L. S.; Mackenzie, K. J.; Netirojjanakul, C.; Francis, M. B., Optimization and expansion of a site selective N-methylpyridinium-4-carboxaldehyde mediated transamination for bacterially expressed proteins. *J. Am. Chem. Soc.* **2015**, *137*, 1123–1129.

23. Tong, G. J.; Hsiao, S. C.; Carrico, Z. M.; Francis, M. B., Viral capsid DNA aptamer conjugates as multivalent cell-targeting vehicles. *J. Am. Chem. Soc.* **2009**, *131*, 11174–11178.

24. Netirojjanakul, C.; Witus, L. S.; Behrens, C. R.; Weng, C.; Iavarone, A. T.; Francis, M. B., Synthetically modified Fc domains as building blocks for immunotherapy applications. *Chem. Sci.* **2013**, *4*, 266–272.

25. El Muslemany, L. M.; Twite, A. A.; ElSohly, A. M.; Obermeyer, A. C.; Mathies, R. A.; Francis, M. B., Photoactivated bioconjugation between ortho-azidophenols and anilines: A facile approach to biomolecular photopatterning. *J. Am. Chem. Soc.* **2014**, *136*, 12600–12606.

26. Capehart, S. L.; Coyle, M. P.; Glasgow, J. E.; Francis, M.B., Controlled integration of gold nanoparticles and organic fluorophores using synthetically modified MS2 viral capsids. *J. Am. Chem. Soc.* **2013**, *135*, 3011–3016.

27. Capehart, S. L.; ElSohly, A. M.; Obermeyer, A. C.; Francis, M. B., Bioconjugation of Gold nanoparticles through the oxidative coupling of ortho-aminophenols and anilines. *Bioconjugate Chem.* **2014**, *25*, 1888–1892.

28. Obermeyer, A. C.; Jarman, J. B.; Netirojjanakul, C.; Francis, M. B., Mild Bioconjugation Through the Oxidative coupling of ortho-aminophenols and anilines with ferricyanide. *Angew. Chem. Int. Ed.* **2014**, *53*, 1057–1061.

29. Obermeyer, A.C.; Jarman, J.B.; Francis, M.B., N-Terminal modification of proteins with *o*-aminophenols. *J. Am. Chem. Soc.* **2014**, *136*, 9572–9579.

30. Windle, C. L.; MÜller, M.; Nelson, A.; Berry, A. Engineering aldolases as biocatalysts. *Curr. Opin. Chem. Biol.* **2014**, *19*, 25–33.

31. Cheriyan, M.; Toone, E. J.; Fierke, C. A., Mutagenesis of the phosphate binding pocket of KDPG aldolase enhances selectivity for hydrophobic substrates. *Protein Sci.* **2007**, *16*, 2368–2377.

32. Sasou, M.; Sugiyama, S.; Yoshino, T.; Ohtani, T., Molecular flat mica surface silanized with methyltrimethoxysilane for fixing and straightening DNA. *Langmuir* **2003**, *19*, 9845–9849.

33. Saha, B.; Saikia, J.; Das, G., Correlating enzyme density, conformation and activity on nanoparticle surfaces in highly functional bio-nanocomposites. *Analyst* **2015**, *140*, 532–542.

34. Seeman, N. C., Nanomaterials based on DNA. *Annu. Rev. Biochem.* **2010**, *79*, 65–87.

35. Phillips, D. C.; York, R. L.; Mermut, O.; McCrea, K. R.; Ward, R. S.; Somorjai, G. A., Side chain, chain Length, and sequence effects on amphiphilic peptide adsorption at hydrophobic and hydrophilic surfaces studied by sum-frequency generation vibrational spectroscopy and quartz crystal microbalance. *J. Phys. Chem. C* **2007**, *111*, 255–261.

36. York, R. L.; Browne, W. K.; Geissler, P. L.; Somorjai, G. A., Peptides adsorbed on hydrophobic surfaces—A sum frequency generation vibrational spectroscopy and modeling study. *Isr. J. Chem.* **2007**, *47*, 51–58.

37. Baio, J. E.; Weidner, T.; Baugh, L.; Gamble, L. J.; Stayton, P. S.; Castner, D. G., Probing the orientation of electrostatically immobilized protein G B1 by time of flight aecondary ion spectrometry, sum frequency generation and near-edge x-ray adsorption fine structure spectroscopy. *Langmuir* **2012**, *28*, 2107–2112.

38. Roepstorff, P.; Fohlman, J., Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biol. Mass Spectrom.* **1984**, *11*, 601.

39. Rebecchi, K. R.; Go, E. P.; Xu, L.; Woodin, C. L.; Mure, M.; Desaire, H., A general protease digestion procedure for optimal protein sequence coverage and post-translational modifications analysis of recombinant glycoproteins: application to the characterization of human lysyl oxidase-like 2 glycosylation. *Anal. Chem.* **2011**, *83*, 8484–8491.

40. Hervey, W. J., IV; Strader, M. B.; Hurst, G. B., Comparison of digestion protocols for microgram quantities of enriched protein samples. *J. Proteome Res.* **2007**, *6*, 3054–3061.

41. Strader, M. B.; Tabb, D. L.; Hervey, W. J.; Pan, C.; Hurst, G. B., Efficient and specific trypsin digestion of microgram to nanogram quantities of proteins in organic-aqueous solvent systems. *Anal. Chem.* **2006**, *78*, 125–134.