Title
Development of Site-Selective Protein and Peptide Modification Strategies

Author
Dolan, Nicholas

Publication Date
2023

Peer reviewed|Thesis/dissertation

Development of Site-Selective Protein and Peptide Modification Strategies

by

Nicholas Dolan

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 Richmond Sarpong
Professor Phillip Messersmith

Summer 2023

Development of Site-Selective Protein and Peptide Modification Strategies

Abstract

Development of Site-Selective Protein and Peptide Modification Strategies

by

Nicholas Dolan

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Matthew B. Francis, Chair

The goal of protein modification is to take advantage of the structural complexity and binding specificity of a protein and pair it with the additional properties that a synthetic molecule can provide. Many strategies have been developed that target native amino acids present on the protein surface, such as lysine, but these modification methods tend to produce heterogeneous mixtures of products since there are often multiples copies of the residue displayed. To better control the level of modification, the N-terminus has become a popular target since each protein has only one. Our lab has recently developed an N-terminal modification strategy that uses 2-pyridinecarboxaldehyde (2PCA) to form an imidazolidine product. The reaction proceeds through an imine intermediate, which is then cyclized with the second amino acid residue to form the product. The reaction can be performed with all 20 native amino acids at the N-terminus, and high levels of conversion can be obtained. Despite this generality, we have found that over time the imidazolidinone conjugate with 2PCA can reverse to liberate the free, unmodified protein, and that the rate of reversal is dependent on the identity of the N-terminal residue. To better understand the factors that influence product stability, we undertook a mechanistic study of the reaction, using NMR kinetics and DFT calculations, to explore the most likely pathway leading to product formation. Through these studies, we found that N-terminal proline residues create the most stable product, and that glycine at position two can also have a secondary stabilizing effect. DFT calculations supported these findings and allowed for structural analysis of the reactive species and transition states that contributed to this stabilization. With a better understanding of how the reaction proceeds, future goals will be toward the development of 2PCA derivatives that can capitalize on these effects to form irreversible conjugates.

In addition to N-terminal modification, our lab has also developed several methods for site-selective tyrosine modification using a tyrosinase enzyme. Tyrosinase is able to oxidize the phenol of tyrosine to a reactive quinone intermediate, which can then be captured by a variety of nucleophiles. Our goal was to use this chemistry and apply it toward the synthesis

of cyclic peptides. We have demonstrated that linear peptides bearing a tyrosine and cysteine residue are able to be oxidized with tyrosinase, and the subsequently formed quinone can be coupled with the cysteine thiol to form a cyclic product. In addition to peptide substrates, we have also demonstrated that this chemistry works on peptide sequences displayed at the N- or C-terminus of a protein. Initial work has shown that the level of tyrosinase activity and selectivity is heavily influenced by the identity of the amino acids next to the tyrosine residue, and future work will be focused on engineering new tyrosinase variants that are less sequence dependent.

Dedicated to my parents, Sheryl and Steve, who have always supported me throughout my entire academic journey.

Table of Contents

Acknowledgments

I firstly would like to thank Matt Francis for accepting me into his lab and for providing the guidance and lab environment that made all this work possible. With respect to research, he constantly encouraged me to explore my own interests and ideas and has always created time to discuss and analyze results. Outside of research, I have greatly appreciated that he strives to create a welcoming atmosphere within the lab and that a focus has always been on diversity and inclusion. Even through tough times, I have always felt that I could confide in him when I needed support.

Secondly, I would like to thank all the members of the Francis group, both present and former alumni. Our lab continues to be a collaborative environment, and all members at one point or another have helped me throughout my Ph.D.

Johnathan joined the group the same year that I did, and he not only became a great labmate but also an incredible friend. We worked together with Merck and helped each other troubleshoot our reactions, but we also shared many other experiences like learning how to ski, taking nature hikes, and going to countless drag shows that really created a memorable time while here at Cal.

Sarah was also someone I spent a lot of time with, and she was always someone I looked up to. It seemed like there was nothing in lab she couldn't do, and her endless excitement and hard work constantly motivated me.

Adel and I also worked together with Merck, and he and I shared our love for synthetic chemistry. He was always available to brainstorm new synthetic routes or talk about the latest paper, and I appreciated having someone else in the lab who was as enthusiastic as I was about chemistry.

Kristin helped me integrate into the lab when I first joined and contributed a lot of work to the 2PCA research. She was a great resource while learning how to work with proteins and prepare for presentations.

Vanessa and I worked in the same lab room, and it was great having her around. From late nights in lab, to sharing vegan food, to rock climbing at Ironworks, I am thankful we got to spend so much time together.

Rachel was always so positive and friendly, and I am thankful for all our talks over tea and chocolates.

Finally, I would like to thank Paul for helping to keep our lab run. He is always willing to give up time to help fix and maintain our instruments and without him progress would be much slower.

I also want to thank my friends Paul, Alex, Timmy, and Stephen for everything they have done for me. We have shared many amazing experiences together and I'm very grateful to have them in my life. They have been an invaluable support outside of lab and are largely the reason I have come to love living in the Bay Area.

Finally, I want to thank my parents who have been my biggest support since day one. It is because of them that I was able to pursue a Ph.D., and they have stood by every decision I had to make to get here.

Chapter 1

Protein Bioconjugation Strategies using Native Amino Acid Residues

Abstract

Proteins are complex molecules that have virtually unlimited structural combinations and can achieve a vast variety of functions. Because of this, bioconjugation is a vital tool in exploiting these molecules and has been applied to many fields, such as the synthesis of materials and therapeutics. Given the importance of these constructs, many strategies have been developed to modify a protein using the reactive amino acids present on the surface. This chapter will highlight some of the most common methods that target lysine, cysteine, and tyrosine residues, as well as the N-terminus, and provide examples of a few more recently developed strategies.[1]

[1]Portions of this chapter will appear in Dolan, N. S.; Maza, J. C.; Ramsey, A. V.; Chapter 14: The Chemical and Enzymatic Modification of Proteins. *In Advanced Chemical Biology: Chemical Dissection and Reprogramming of Biological Systems*; Hang, H.C.; Pratt, M. R.; Prescher, J. P. Eds.; John Wiley & Sons, 2023.

1.1 Introduction

Proteins play an essential role in cellular functions for all living systems. The complex architectures of these biomolecules can accelerate chemical reactions dramatically, transduce signals across large distances, and generate long-range structures through self-assembly. Because of this, a growing area of research has been in the development of chemical tools to modify proteins through the attachment of additional functional groups, such as fluorophores, radiolabels, NMR probes, drug molecules, polymers, nanoparticles, and surfaces. The resulting bioconjugates have proven to be essential tools for the elucidation of biological pathways, and they have been utilized as new platforms for both therapeutics and materials.

Although many protein modification strategies already exist, new methods are constantly being developed to expand the repertoire of techniques available to attach synthetic groups to proteins. The need for such a diverse selection is critical because each protein has a unique reactivity that depends on the almost infinite combination of amino acid side chains presented on its surface. This is further complicated by the dynamic behavior that proteins display in solution and by the changes they often undergo under different reaction conditions, such as pH and ionic strength. The synthetic groups to be attached may also have functional groups that would interfere with some reaction strategies.

1.2 General Considerations

When designing a chemical strategy for synthesizing protein conjugates, reaction conditions, such as pH and temperature, need to be considered. Small perturbations can drastically alter the stability of a protein, and correct folding is essential for protein function. For most proteins, this requires that reactions take place between pH 6–8.5 and from 0–37 °C. In addition to pH and temperature, organic cosolvents, like dimethylsulfoxide, *N,N'*-dimethylformamide, and acetonitrile, should be avoided, since these can also lead to protein denaturation by disrupting the balance of hydrophobic and hydrophilic interactions that are important for proper folding. As such, most protein bioconjugation reactions are best performed in buffered, aqueous solutions with no organic solvents, or with at maximum 10% of the solution volume if organic solvents are needed to dissolve the small molecule coupling reagent. Another considerable challenge for protein modification is the high degree of protein dilution (usually in the low micromolar range) relative to most organic reactions. This, coupled with the fact that most modification reagents are used in low concentrations as well due to their (often) high cost and low solubility, requires that developed chemistries exhibit highly favorable reaction kinetics to function at very low substrate concentrations.

Most protein modification reactions display chemoselectivity that is specific to a single residue type or to a bioorthogonal reactive group that has been previously installed. As a result, the outcome of the chemistry is strongly influenced by the number of copies of this group that are accessible on a given protein. When multiple residues are present, as is usually the case with lysine, tyrosine, and carboxylic acid residues, heterogeneous product

mixtures result that vary both in the numbers and positions of the chemical modifications. While this may be suitable for some applications, this can lead to differing activities, binding properties, and solubilities in many cases. Site-selective chemistries that modify a single amino acid residue within a protein can produce well-defined and homogenous products, but these approaches usually require manipulation of the protein sequence during expression. This is most often achieved in the context of cysteine residues, artificial amino acids, or specific N-terminal residue combinations. Thus, when choosing a chemistry for protein modification, it is important to consider the intended application of the protein conjugate and whether a heterogeneous or homogeneous product is required.

In addition to protein modification, many of the reactions developed can also be applied to peptide modification. The shorter chains of peptides generally lack higher order structures, allowing them to accommodate a wider scope of reaction conditions. Many peptide modification strategies occur in high yield at temperatures >37 °C, in pure organic solvent, and using microwave irradiation. The advancement of solid phase synthesis techniques has also allowed for a wider range of bioorthogonal handles to be introduced into the peptide sequence for modification.

This chapter will cover some of the traditional modification strategies for commonly targeted residues, such as lysine, cysteine, tyrosine, and the N-terminus, as well as highlight a few more recently developed methods.

1.3 Modification Chemistries

1.3.1 Lysine Modification

Lysine is the 7th most abundant amino acid and occurs in over 90% of proteins.[1] Because of this commonality, lysine residues are frequently targeted for protein modification, especially in cases where the protein sequence is unknown or cannot be altered easily. Most lysine residues are surface-exposed and have a pK_a of roughly 10.55,[2] so at physiological pH the amines of lysine residue side chains remain largely protonated and exhibit low reactivity. In practice, the optimum pH for lysine bioconjugation is roughly 8.5-9.6,[3] which increases the number of free amines while maintaining protein stability. Since hydrolysis of amine-reactive reagents can still occur readily under these conditions, it is common to use a large stoichiometric excess—particularly when the targeted lysine residue is less reactive due to either limited solution exposure or an increased pK_a value.

While lysine modification chemistry is widely applicable across a vast range of proteins, important limitations do exist. The high occurrence of this residue almost always results in heterogeneous modification products. This highlights the key trade-off in lysine bioconjugation chemistry: convenience versus heterogeneity. For example, lysine chemistry is often used as a facile and rapid method for protein immobilization on surfaces, but the inherent heterogeneity can lead to randomly oriented proteins. An additional limitation of lysine bioconjugation is that other nucleophilic residues, such as exposed cysteines, can cross react with amine-reactive

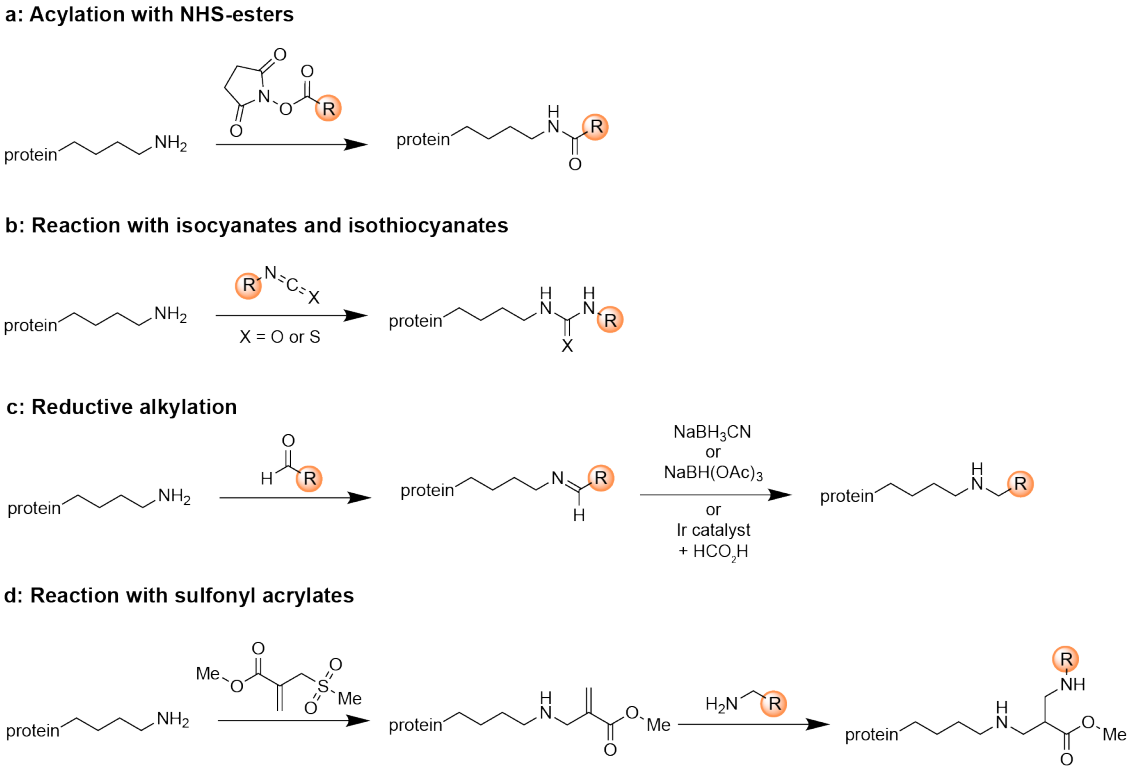


Figure 1: Chemical modification of lysine residues. **(a)** Lysine residues can be acylated with NHS-esters. **(b)** Isocyanates and isothiocyanates react readily with lysine residues; however, isothiocyanates are more easily available. **(c)** Reductive alkylation of lysine residues can be carried out with borane- and borohydride-based reducing reagents or an iridium-based catalyst and formic acid. **(d)** A more recent lysine strategy involves reaction with sulfonyl acrylates.

reagents. Despite these limitations, lysine chemistry remains the most common approach for protein modification with tracking labels, polymers, and many other chemical functional groups.

Perhaps the easiest route for lysine modification is the use of *N*-hydroxysuccinimide (NHS) activated esters (Figure 1a).[4] The electrophilic carbonyl of these esters reacts readily with side chain amines to form highly stable amide bonds. While the extent of modification can be controlled by tuning the equivalents of the NHS ester component, heterogenous modification is typically obtained; however, NHS chemistry is very convenient as reactions are typically completed in 1 h at room temperature under mild aqueous reaction conditions. Additionally, there exist many commercially available NHS esters, including fluorophores, biotin, and a variety of cross-linkers. NHS esters can also be synthesized from carboxylic acids by initial activation with dicyclohexylcarbodiimide (DCC), diisopropylcarbodiimide (DIC), or *N*-3-dimethylaminopropyl-*N'*-ethylcarbodiimide (EDC), to form an anhydride intermediate, followed by reaction with NHS-OH. For protein immobilization, there are many commercially available resins that are pre-activated with NHS esters. The wide accessibility of these derivatives and ease of use make NHS modification a common and widely used strategy.

Lysine residues can also be acylated using isocyanates and isothiocyanates (Figure 1b). Many isothiocyanates, especially fluorophores, are commercially available, while there are fewer isocyanates due to the difficulty in storing these reagents. Reaction conditions for isothiocyanates are comparable to those for NHS esters, but the thiourea linkage that forms is often less hydrolytically stable than an amide bond and can degrade over time; thus NHS chemistry is typically preferred.

Reductive alkylation is another method used for lysine modification (Figure 1c). This strategy involves the condensation of an aldehyde with the side chain amine to form a transient imine, which is then reduced with a mild reducing agent, such as sodium cyanoborohydride, to form a more hydrolytically stable alkylamine linkage. Despite its widespread use, sodium cyanoborohydride is water sensitive and requires a desiccant for storage. In addition, this reductant has been found to reduce disulfides as well as imines, making it potentially problematic for the modification of antibodies and other proteins with critical disulfide bonds. As an alternative, a variety of borane and borohydride reductants have been developed that display varying water sensitivity, solubility, and reactivity.[5] Additionally, an iridium catalyst that is less water sensitive than traditional reagents has been reported for reductive alkylation under mild reaction conditions with formic acid as a hydride source.[6]

An important advantage of reductive alkylation strategies is their ability to maintain the overall charge of a protein in contrast to acylation strategies that reduce the degree of positive charge with each modification. However, even with effective reducing reagents, reductive alkylation can suffer from long reaction times (often up to 24 h) and moderate-to-low conversion, making it a low efficiency reaction. Despite these challenges, reductive alkylation has found great utility in coupling amines to oligosaccharides.[7,8]

Although several long-standing chemistries remain in use, new methods for lysine modification are still being developed. Recently, a chemo- and regioselective two-step modification strategy was shown to target a single lysine residue on five different proteins, including the anti-cancer monoclonal antibody (mAb) trastuzumab.[9] The first step is the addition of a sulfonyl acrylate reagent, which is attacked by the most reactive lysine on the protein (Figure 1d). This reaction proceeds rapidly at mild reaction conditions (1-2 h, pH 8, 37 °C), requires only 1 equivalent of the sulfonyl acrylate, and results in quantitative and irreversible conversion. Selectivity is driven by a hydrogen bond between the side chain amino group and the sulfonyl group. The second step is modification through an aza-Michael addition that can be carried out with a variety of amine-based compounds, producing conjugates with fluorophores, drugs, and PEG chains attached. New reactions like these pave the way for the site-selective targeting of this ubiquitous and highly accessible amino acid, as do efforts to improve our understanding of how the local environment can influence the ability of lysine residues to participate in reactions.

1.3.2 Cysteine Modification

The most common methods for site-specific protein modification target cysteine residues. The relatively low abundance of reduced cysteines on protein surfaces compared to other

residues, such as lysine, generally allows for selective modification at a single site. Moreover, single cysteine residues can often be introduced readily through genetic engineering to provide new reactive handles. With an average pK_a of 8, an appreciable fraction of sulfhydryl groups is deprotonated even at neutral pH, providing strongly nucleophilic thiolate anions that can react with a variety of electrophiles. In addition, the relatively weak S-H bond compared to the O-H bond of serine and threonine (87 kcal/mol vs ~105 kcal/mol, respectively) allows for selective, single-electron processes. Taking advantage of the wide range of chemistries in which cysteine residues can participate, many modification strategies have been developed based on redox reactions, substitution and addition reactions, and radical reactions.

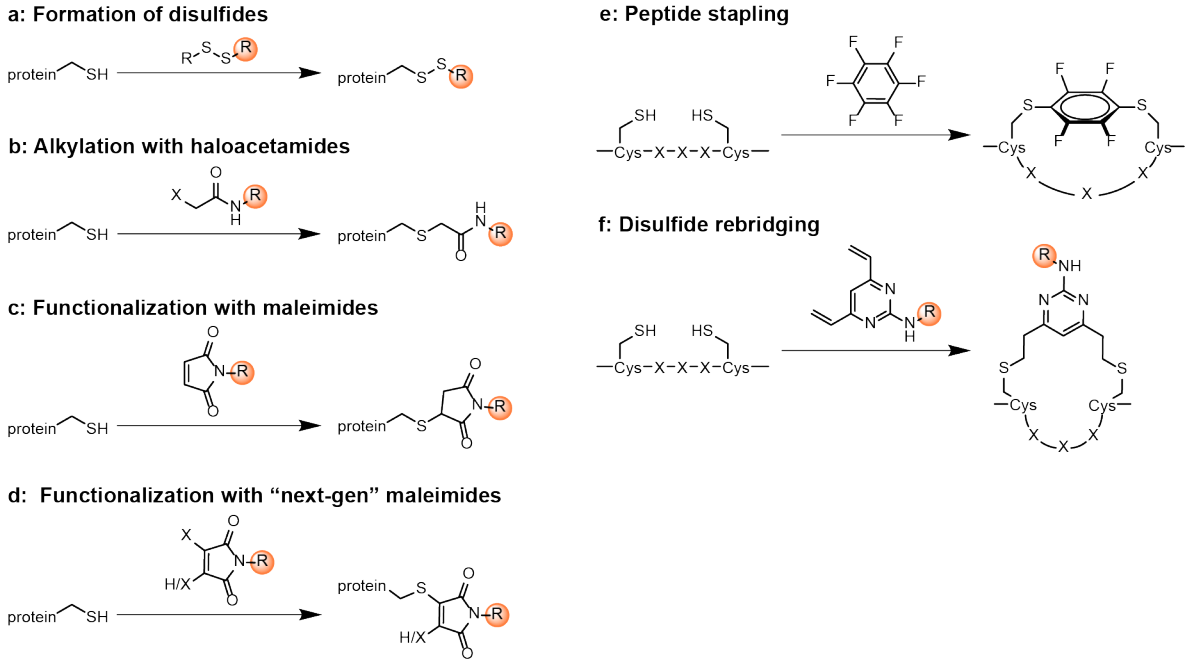


Figure 2: Chemical modification of cysteine residues. **(a)** The sulfhydryl group of cysteine can undergo disulfide exchange, as well as react with a variety of electrophiles, such as **(b)** haloacetamides, **(c)** maleimides, and **(d)** next-gen halogen-substituted maleimides. **(e)** Perfluorobenzene can undergo nucleophilic aromatic substitution with a reduced disulfide to provide a *para*-substituted linker for peptide stapling. **(f)** Disulfide rebridging of partially reduced antibody disulfide bonds can be carried out with divinylpyrimidine linkers.

A common method for cysteine modification involves the formation of disulfide bonds (Figure 2a). The most used reagents for this purpose are 5,5'-dithiobis(2-nitrobenzoic acid) (Ellman's reagent) and 2,2'-dithiodipyridine, which are added in excess to form activated, mixed disulfide intermediates. These species can be isolated, if desired, and are highly reactive towards the formation of new disulfides with additional thiols, since 2-nitro-5-thiobenzoate and 2-thiopyridine both serve as excellent leaving groups. This overall strategy can be used to form disulfides between cysteine residues, or it can be used to attach probe molecules and drug cargo. In the latter case, these molecules can be released upon entering a reducing environment, such as the cytoplasm of mammalian cells.[10] This method can also be used

to protect the sulfhydryl group and prevent cross reactivity during subsequent modification of other amino acids.[11] When desired, the disulfides can be reduced in the presence of mild reductants like tris(2-carboxyethyl)phosphine (TCEP) or dithiothreitol (DTT) to liberate the free thiols.

Another modification method for cysteine residues takes advantage of the strongly nucleophilic thiolate anion that is formed when the sulfhydryl group is deprotonated. This anion can undergo subsequent reactions with a variety of electrophiles, including S_N2 substitutions with haloacetamides (Figure 2b) and Michael additions with maleimides (Figure 2c). Haloacetamides are routinely used in protein digests to cap cysteine residues and occasionally for alkylating intact proteins with exposed cysteines. Reactivity with other nucleophilic side chains, such as lysine and histidine, can occur, so care must be taken to maintain a pH between 7.5–8 to favor thiol deprotonation. Maleimide conjugation has been routinely used to attach payload molecules to IgG1 antibodies, which is typically done through reduction of the four antibody disulfide chains.[12,13] This produces eight unique thiols for potential modification, and by controlling the stoichiometry of reductant and maleimide, researchers can generate antibodies conjugated with either four or eight drug molecules. An example of this approach is Adcetris, an antibody therapy for Hodgkin's lymphoma synthesized by maleimide conjugation.[14]

Maleimide-containing reagents are readily formed through condensation with a desired amine and maleic anhydride, followed by an additional cyclization step. The reliability and scalability of this transformation makes it heavily used; however, one drawback is that the thiosuccinimide product can undergo further reactions, such as a retro-Michael addition or ring-opening hydrolysis.[15] Studies with maleimide-conjugated antibodies have shown that the maleimide linker can exchange with other thiols present on proteins, such as albumin, reducing antibody efficacy and potentially leading to new toxicity.[16] This instability has spurred the development of next-generation maleimides (NGMs) that are substituted at one or both of the maleimide carbons (Figure 2d).[17] These NGM conjugates display a higher level of stability over a wider range of conditions, but can also be released in the presence of TCEP or by thiol exchange.[18]

While engineered surface thiols have commonly been used for single-site modification, cysteine residues are also canonically present on proteins in disulfide bridges. Disulfide functionalization at both thiols in the bridge is an emerging area of cysteine chemistry and has been used for peptide stapling and disulfide rebridging.[19] These methods begin with disulfide bond reduction to generate free thiols, which are subsequently modified with a single electrophile containing two reactive handles. In the case of peptide stapling, the linkage usually imparts rigidity to alpha helical secondary structures but can also be used as a method for synthesizing peptide macrocycles. One class of molecules used in peptide stapling is perfluorinated aryl and biaryl compounds (Figure 2e).[20] When perfluorobenzene is used, the free thiols add into the aryl ring by nucleophilic aromatic substitution (S_NAr) to provide *para*-substituted products exclusively. While these reactions are chemoselective and proceed in high yield, the major drawback is the high percentage of organic solvent (DMF) needed to solubilize the hydrophobic perfluoro compounds, making this strategy largely limited to

peptides.

Disulfide rebridging has been used to synthesize antibody-drug conjugates (ADCs), where the cargo is installed between the reduced, interchain disulfide bonds of an immunoglobulin.[21] Since there are four or more well-defined disulfide bridges in most antibody types, this strategy provides a reliable way to synthesize ADCs with definitive modification sites. The number of interchain disulfide bonds that are reduced can be controlled by the choice of reductant, stoichiometry, and reaction time, leading to drug-to-antibody ratios (DAR) of two for partial reduction and four for complete reduction. One example of this approach has been the modification of trastuzumab, an anti-HER2 antibody, using a substituted divinylpyrimidine (DVP) linker (Figure 2f).[22] After partial reduction using TCEP, treatment with a slight excess of DVP provided the rebridged antibody with good conversion, and the product displayed excellent stability in the presence of free thiols.

1.3.3 Tyrosine Modification

Historically undervalued for its potential in protein conjugation, tyrosine has recently emerged as an exciting chemical handle for the construction of protein bioconjugates. Tyrosine is a relatively rare amino acid and is one of three aromatic amino acids available in nature. Due to its aromaticity, modification typically proceeds through electrophilic aromatic substitution (EAS), a route distinct from other amino acids. The hydroxyl group of tyrosine has a pK_a of $\sim$10.5, making tyrosine neutrally charged over the common physiological pH range. Due to its ability to participate in pi-stacking interactions, H-bond, as well as its neutral charge, tyrosine residues are often buried within the interior of the protein, which helps stabilize the protein structure. This often causes them to be inaccessible to conjugation chemistries; however, genetic engineering, often at the C- or N-terminus, can easily introduce solvent accessible tyrosine residues for modification.

One of the earliest tyrosine modification strategies involves coupling to an aryldiazonium salt through an EAS reaction to form a new azo linkage (Figure 3a). The chemistry requires activation of the phenol to the phenolate anion, so reactions are best performed at pH 8-10 to reach high levels of conversion. Also, electron-withdrawing groups can be added at the *para* position of the diazonium salt to greatly accelerate the reaction. Although some cross reactivity can be observed with histidine residues, diazonium coupling has been successfully employed to modify tyrosine residues on complex protein assemblies, like the viral capsids MS2 and TMV, which lack solvent exposed histidines.[23,24]

Another common strategy for tyrosine modification is tyrosine crosslinking (Figure 3b). This involves generation of a tyrosyl radical through single electron abstraction, which can then undergo dimerization with a second tyrosine residue to form a cross-linked product. Generation of the tyrosyl radical can either occur enzymatically, such as with horseradish peroxidase,[25] or photochemically, such as with $Ru(bpy)^{32+}$.[26] Crosslinking tyrosine residues has found wide-ranging commercial uses, including the synthesis of antibody-protein conjugates, the photocatalyzed crosslinking of transcription factors, and protein stabilization in the food processing industry.

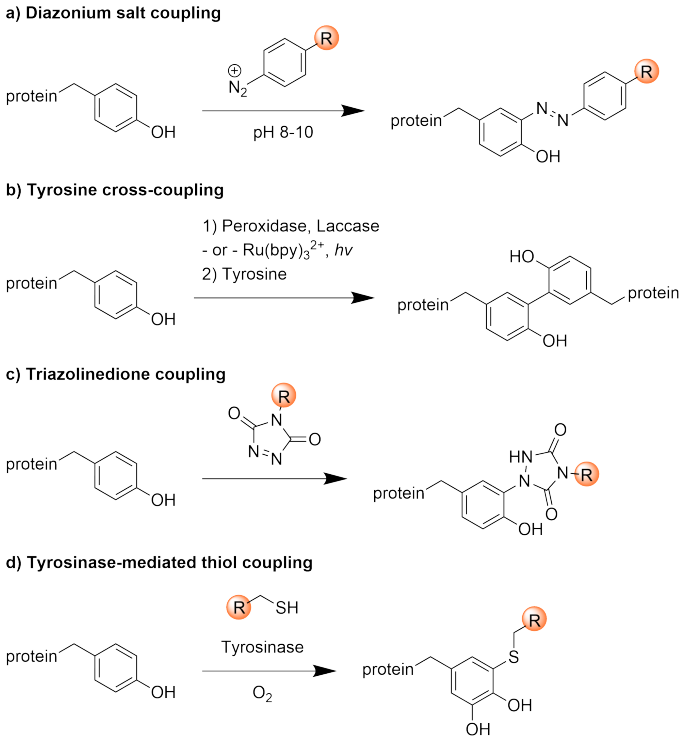


Figure 3: Chemical and enzymatic modification of tyrosine residues. **(a)** Diazonium salts form a new C-N bond with tyrosine, but are often generated immediately before use. **(b)** Nearby tyrosine rings can be crosslinked using either chemical or enzymatic means. This process is facilitated by the ability of tyrosine to accommodate a single radical species. **(c)** Triazolinediones couple to tyrosine phenols, generating a stable C-N bond. **(d)** The phenol of tyrosine can be oxidized by tyrosinase to a quinone intermediate, which can then be coupled with a thiol.

Triazolinediones (TADs) have recently emerged as a useful reagent for the site-selective modification of tyrosine residues (Figure 3c). The highly reactive TAD is prepared by treatment of cyclic diazodicarboxamides with an oxidizer, such as NBS or dibromo-dimethyl-hydantoin,[27] and reacts with a tyrosine residue through an EAS-type pathway.[28] Since its initial development, phenyltriazolinedione (PTAD)-based coupling reagents have flourished due to the high degree of stability exhibited by the C-N bond formed using this reagent. Studies have shown that the C-N bond between PTAD and *p*-cresol is stable in both strongly acidic (10% HCl in MeOH) and basic conditions (10% NaOH in MeOH) for 24 h at room temperature or for 1 h at 120 °C. These results suggest that PTAD-tyrosine conjugates are more robust than the more traditional thiol-maleimide conjugates, which is supported by retention of the PTAD-tyrosine linkage in human plasma over one week. This chemistry has been successfully used in the modification of peptide and antibody substrates, and has even been used in the pharmaceutical industry for the conjugation of carbohydrate antigens to carrier proteins for vaccine development.[29] Another recently developed tyrosine modification is the tyrosinase-mediated coupling of tyrosine residues to thiols.[30] The tyrosinase enzyme is able to oxidize phenols to activated quinone intermediates, which can then be coupled to

thiols on small molecules, peptides, and proteins. Further discussion of this strategy can be found in chapter 3.

1.3.4 N-Terminal Modification

N-terminal modification has become an attractive method to label protein substrates due to its unique properties relative to other potentially reactive functional groups. Unlike lysine residues, which can be found in high abundance on the protein surface, single-chain proteins possess only one N-terminus. The N-terminus is generally solvent exposed and, if necessary, can often be genetically extended while retaining protein function and overall structure.[31] The N-terminal amine is also less basic than the amines of lysine side chains, with a pK_a of 6-8 versus 10.5, respectively.[32] Selective N-terminal acylations can thus be performed if few lysine residues are present, but it is usually difficult to achieve high selectivity for proteins containing several lysines residues.

One common strategy takes advantage of specific N-terminal amino acids by involving the side chain groups as additional reactive handles. For example, N-terminal serine and threonine residues can undergo oxidative cleavage with sodium periodate to generate α-ketoaldehydes, which can further react with alkoxyamines and hydrazine to form oxime and hydrazone conjugates, respectively (Figure 4a).[33] A variety of alkoxyamine and hydrazine-containing cargo are readily available, including fluorophores, cytotoxic molecules, and polymers. One drawback of this method is the need for sodium periodate, which can cause unwanted oxidation of cysteine residues, methionine residues, and glycans.

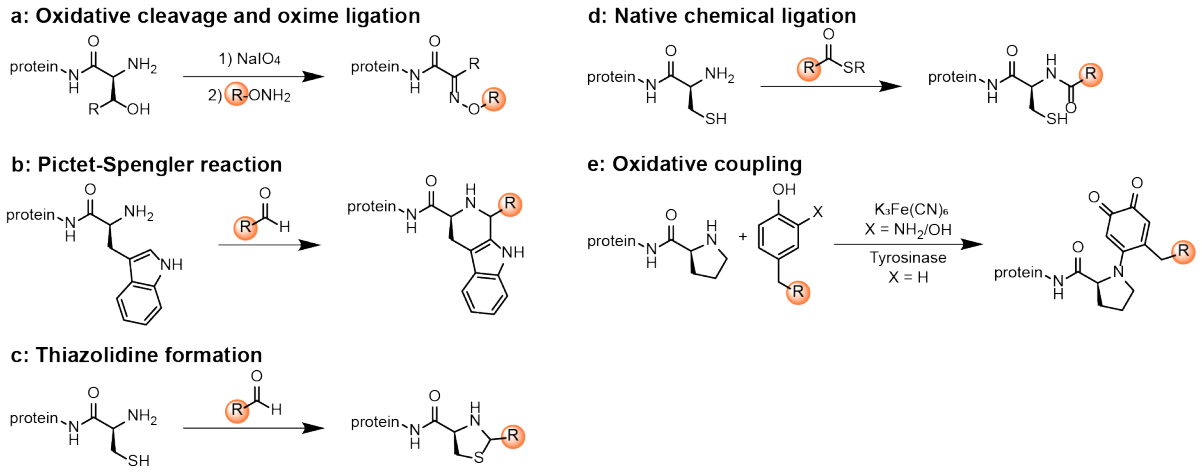


Figure 4: Modification of specific amino acids at the N-terminus. **(a)** N-terminal serine and threonine can be oxidatively cleaved with periodate, followed by oxime ligation with alkoxyamines. **(b)** N-terminal tryptophan can undergo a Pictet-Spengler reaction with aldehydes. **(c)** N-terminal cysteine can be modified through condensation with aldehydes, followed by cyclization to form thiazolidine products, as well as **(d)** through native chemical ligation with thioesters. **(e)** N-terminal proline can be coupled with aminophenols and catechols with a ferrocyanide oxidant, and phenols can be catalytically oxidized and coupled using a tyrosinase enzyme and molecular oxygen.

Some N-terminal amino acids can perform unique chemistries with aldehyde-bearing cargo molecules. When tryptophan is the N-terminal amino acid, a Pictet-Spengler reaction can be used (Figure 4b).[34] The N-terminal amine undergoes a condensation reaction with an aldehyde, and the transient imine is trapped by the indole side chain. N-terminal tryptophan can be obtained by proteolysis or with synthetic polypeptides; however, it cannot be genetically engineered at the N-terminus, since wild-type methionine aminopeptidase (MetAP) is unable to cleave methionine before bulky residues like tryptophan.[35] In the case of cysteine, the N-terminal amine and an aldehyde condense to form an imine, which can be trapped by the cysteine sulfhydryl group to produce a thiazolidine (Figure 4c).[36] N-terminal cysteine can also condense with 2-cyanobenzothiazole containing reagents, producing a new thiazole-thiazole bond at the protein N-terminus that has been exploited for many applications, including cell imaging.[37] Although N-terminal cysteine residues are rarely found in naturally expressed proteins, they can be genetically incorporated at the second position of the sequence by cleavage of the starting methionine residue with MetAP. For both modification methods with tryptophan and cysteine, cyclization with the side chain group provides the selectivity of the reactions as well as the stability of the products.

N-terminal cysteine can also react with a thioester in native chemical ligation (NCL) reactions to synthesize large peptides and proteins (Figure 4d).[38] NCL begins with transthioesterification, which occurs between the sulfhydryl group of the cysteine and an introduced thioester, followed by S→N acyl transfer to the N-terminal amine. This powerful transformation has found a wide range of applications, including incorporation of post-translational modifications, such as ubiquitin[39] and polysaccharides,[40] access to biomolecules like peptide-oligonucleotide conjugates,[41] and synthesis of materials, such as biopolymers,[42] hydrogels,[43] and peptide-based dendrimers.[44]

N-terminal proline residues can couple with *ortho*-quinones that are oxidatively generated *in situ* (Figure 4e). Aminophenols and catechols can be oxidized to quinones by potassium ferrocyanide, which is a mild enough oxidant to prevent unwanted oxidative byproducts.[45] The reaction is extremely fast (generally less than 10 min) and proceeds in high yield. As mentioned previously, phenols can also be oxidized directly to *ortho*-quinones by the enzyme tyrosinase.[46] Unlike aminophenol and catechol reagents, which can be oxidized by air and polymerize in solution, phenols exhibit much higher stability and are easier to implement. An additional advantage is that the tyrosinase enzyme uses molecular oxygen as the sole oxidant in the reaction. Whether enzymatic or produced by small molecule oxidation, the *ortho*-quinones generated have been successfully used to attach a variety of cargo, including fluorophores, metal chelators, and oligonucleotides[47] to proline N-termini for applications in cell imaging, drug delivery, and protein immobilization.

More general methods for N-terminal modification have also been developed that do not rely on the participation of a particular amino acid side chain. Transamination of the N-terminal amine with pyridoxal-5'-phosphate (PLP) produces an α-ketoaldehyde, which can react further (as described above) with alkoxyamines to form oxime-ligated products (Figure 5a).[48] This approach avoids the use of a strong oxidant like periodate and is largely unaffected by the identity of the amino acid; however, long reaction times are generally required and

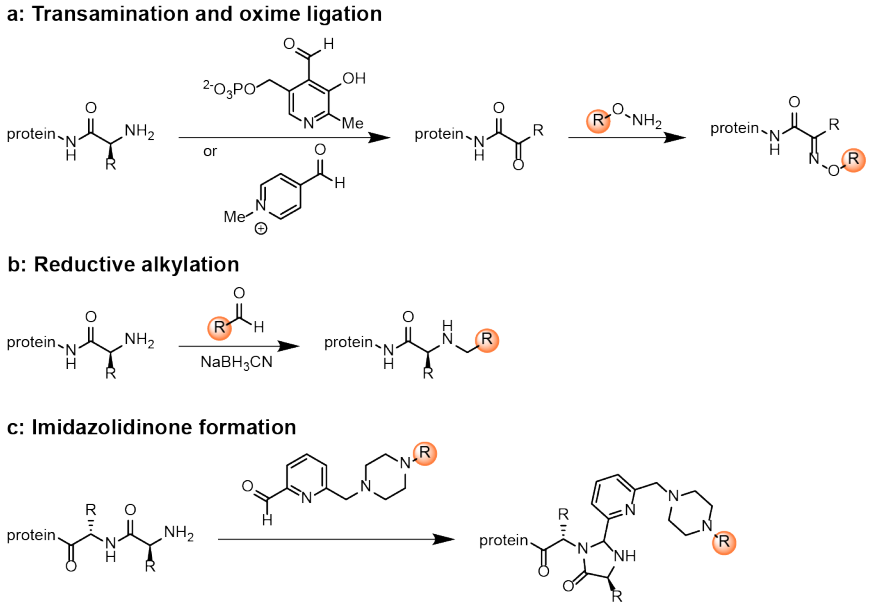


Figure 5: General modification of amino acids at the N-terminus. **(a)** Transamination with pyridoxal-5'-phosphate (PLP) or Rapoport's salt (RS) produces an α-ketoaldehyde that can subsequently react with alkoxyamines. **(b)** N-terminal amines can undergo reductive alkylation with aldehydes in the presence of a borohydride reducing agent. **(c)** 2-pyridinecarboxaldehyde (2PCA) derivatives react to form an imine intermediate with the N-terminus, which can then cyclize to produce an imidazolidinone product.

side reactions, such as aldol additions, can be observed. Another general method is the reductive alkylation of an aldehyde with the N-terminal amine in the presence of sodium cyanoborohydride (Figure 5b). However, complete site selectivity is hard to control since lysine side chains can also participate in the reaction. Lower pH (down to 4.5) can help improve selectivity, but often at the cost of the reaction yield. Finally, 2-pyridinecarboxaldehyde (2PCA) can be used to modify the N-terminus to form an imidazolidinone conjugate (Figure 5c).[49] Even though lysine side chains can also form imines with 2PCA, these products are transient since cyclization to the imidazolidinone cannot occur. This strategy is also compatible with unprotected cysteine residues elsewhere in the protein, creating a facile approach to the incorporation of dual modifications when used in tandem. Further discussion of this reaction can be found in chapter 2.

1.4 Conclusion

Protein modification continues to be a heavily researched area of interest, and the construction of protein bioconjugates has been widely employed in many fields. The desired modification strategy largely depends on the application, and varying levels of selectivity can be achieved through different routes. The overall sequence of the protein will determine which methods can be used to target native amino acid residues exposed on the surface.

Common residues like lysine can be easily modified and are present on almost all proteins; however, heterogeneous product mixtures are often produced which may not be desired for some applications. Since multiple copies of a specific amino acid may be present, genetic engineering can also be used to install residues with unique, bioorthogonal reactivity. Other modification strategies that target residues like tyrosine or the N-terminus often provide a much higher degree of selectivity. Even with all the current strategies that exist, new methods are always needed since the vast diversity of protein structures constantly provides endless challenges for site-selective modification.

1.5 References

(1) Nelson, D.; Cox, M. *Principles of Biochemistry, 7th ed.* **2017**.

(2) Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. *Protein Science* **2009**, *18*, 247–251.

(3) Koniev, O.; Wagner, A. *Chem. Soc. Rev.* **2015**, *44*, 5495–5551.

(4) Anderson, G. W.; Zimmerman, J. E.; Callahan, F. M. *J. Am. Chem. Soc.* **1964**, *86*, 1839–1842.

(5) Ambrogelly, A.; Cutler, C.; Paporello, B. *Protein J.* **2013**, *32*, 337–342.

(6) McFarland, J. M.; Francis, M. B. *J. Am. Chem. Soc.* **2005**, *127*, 13490–13491.

(7) Stefanetti, G.; Rondini, S.; Lanzilao, L.; Saul, A.; MacLennan, C.; Micoli, F. *Vaccine* **2014**, *32*, 6122–6129.

(8) Rondini, S.; Micoli, F.; Lanzilao, L.; Gavini, M.; Alfini, R.; Brandt, C.; Clare, S.; Mastroeni, P.; Saul, A.; MacLennan, C. *Infect. Immun.* **2015**, *83*, 996–1007.

(9) Matos, M. J.; Oliveira, B. L.; Martınez-Sáez, N.; Guerreiro, A.; Cal, P. M.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J., et al. *J. Am. Chem. Soc.* **2018**, *140*, 4004–4017.

(10) Baldwin, A. D.; Kiick, K. L. *Bioconjugate Chem.* **2011**, *22*, 1946–1953.

(11) Rosen, C. B.; Kwant, R. L.; MacDonald, J. I.; Rao, M.; Francis, M. B. *Angew. Chem* **2016**, *128*, 8727–8731.

(12) Alley, S. C.; Okeley, N. M.; Senter, P. D. *Curr. Opin. Chem. Biol.* **2010**, *14*, 529–537.

(13) Lyon, R. P.; Meyer, D. L.; Setter, J. R.; Senter, P. D. In *Meth. Enzymol.* Elsevier: 2012; Vol. 502, pp 123–138.

(14) Feuillâtre, O.; Gély, C.; Huvelle, S.; Baltus, C. B.; Juen, L.; Joubert, N.; Desgranges, A.; Viaud-Massuard, M.; Martin, C. *ACS omega* **2020**, *5*, 1557–1565.

(15) Renault, K.; Fredy, J. W.; Renard, P.-Y.; Sabot, C. *Bioconjugate Chem.* **2018**, *29*, 2497–2513.

(16) Dadova, J.; Galan, S. R.; Davis, B. G. *Curr. Opin. Chem. Biol.* **2018**, *46*, 71–81.

(17) Smith, M. E.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R. *J. Am. Chem. Soc.* **2010**, *132*, 1960–1965.

(18) Nathani, R. I.; Chudasama, V.; Ryan, C. P.; Moody, P. R.; Morgan, R. E.; Fitzmaurice, R. J.; Smith, M. E.; Baker, J. R.; Caddick, S. *Org. Biomol. Chem.* **2013**, *11*, 2408–2411.

(19) Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.

(20) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. *Nature* **2015**, *526*, 687–691.

(21) Walsh, S. J.; Omarjee, S.; Galloway, W. R.; Kwan, T. T.-L.; Sore, H. F.; Parker, J. S.; Hyvönen, M.; Carroll, J. S.; Spring, D. R. *Chem. Sci.* **2019**, *10*, 694–700.

(22) Maruani, A.; Smith, M. E.; Miranda, E.; Chester, K. A.; Chudasama, V.; Caddick, S. *Nat. Commun* **2015**, *6*, 6645.

(23) Hooker, J. M.; Kovacs, E. W.; Francis, M. B. *J. Am. Chem. Soc.* **2004**, *126*, 3718–3719.

(24) Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B. *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.

(25) Minamihata, K.; Goto, M.; Kamiya, N. *Bioconjugate Chem.* **2011**, *22*, 74–81.

(26) Fancy, D. A.; Kodadek, T. *PNAS* **1999**, *96*, 6020–6024.

(27) De Bruycker, K.; Billiet, S.; Houck, H. A.; Chattopadhyay, S.; Winne, J. M.; Du Prez, F. E. *Chem. Rev.* **2016**, *116*, 3919–3974.

(28) Ban, H.; Gavrilyuk, J.; Barbas III, C. F. *J. Am. Chem. Soc.* **2010**, *132*, 1523–1525.

(29) Hu, Q.-Y.; Allan, M.; Adamo, R.; Quinn, D.; Zhai, H.; Wu, G.; Clark, K.; Zhou, J.; Ortiz, S.; Wang, B., et al. *Chem. Sci.* **2013**, *4*, 3827–3832.

(30) Lobba, M. J.; Fellmann, C.; Marmelstein, A. M.; Maza, J. C.; Kissman, E. N.; Robinson, S. A.; Staahl, B. T.; Urnes, C.; Lew, R. J.; Mogilevsky, C. S.; Doudna, J. A.; Francis, M. B. *ACS Cent. Sci.* **2020**, *6*, 1564–1571.

(31) Sereda, T. J.; Mant, C. T.; Quinn, A. M.; Hodges, R. S. *J. Chromatogr. A* **1993**, *646*, 17–30.

(32) Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chem.* **1992**, *3*, 138–146.

(33) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett.* **2000**, *41*, 4069–4073.

(34) Varland, S.; Osberg, C.; Arnesen, T. *Proteomics* **2015**, *15*, 2385–2401.

(35) Zhang, L.; Tam, J. P. *Anal. Biochem.* **1996**, *233*, 87–93.

(36) Liang, G.; Ren, H.; Rao, J. *Nat. Chem* **2010**, *2*, 54–60.

(37) Tang, S.; Liang, L.-J.; Si, Y.-Y.; Gao, S.; Wang, J.-X.; Liang, J.; Mei, Z.; Zheng, J.-S.; Liu, L. *Angew. Chem.* **2017**, *129*, 13518–13522.

(38) Bloom, S.; Liu, C.; Kölmel, D. K.; Qiao, J. X.; Zhang, Y.; Poss, M. A.; Ewing, W. R.; MacMillan, D. W. *Nat. Chem* **2018**, *10*, 205–211.

(39) Reif, A.; Siebenhaar, S.; Tröster, A.; Schmälzlein, M.; Lechner, C.; Velisetty, P.; Gottwald, K.; Pöhner, C.; Boos, I.; Schubert, V., et al. *Angew. Chem., Int. Ed.* **2014**, *53*, 12125–12131.

(40) Diezmann, F.; Eberhard, H.; Seitz, O. *Pept. Sci.* **2010**, *94*, 397–404.

(41) Paramonov, S. E.; Gauba, V.; Hartgerink, J. D. *Macromolecules* **2005**, *38*, 7555–7561.

(42) Hu, B.-H.; Su, J.; Messersmith, P. B. *Biomacromolecules* **2009**, *10*, 2194–2200.

(43) Dirksen, A.; Meijer, E.; Adriaens, W.; Hackeng, T. M. *ChemComm* **2006**, 1667–1669.

(44) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, *136*, 9572–9579.

(45) Maza, J. C.; Bader, D. L.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Krska, S. W.; Parish, C. A.; Francis, M. B. *J. Am. Chem. Soc.* **2019**, *141*, 3885–3892.

(46) Palla, K. S.; Hurlburt, T. J.; Buyanin, A. M.; Somorjai, G. A.; Francis, M. B. *J. Am. Chem. Soc.* **2017**, *139*, 1967–1974.

(47) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 5307–5311.

(48) Hung-ChieháChou, D. et al. *Chem. Sci.* **2017**, *8*, 2717–2722.

(49) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol* **2015**, *11*, 326–331.

Chapter 2

Mechanistic Study of N-Terminal Modification with 2PCA

Abstract

Protein modification has become an invaluable tool used in a myriad of diverse applications due to its ability to combine the unique structure and function of a protein with the properties of a synthetic molecule. To synthesize these protein conjugates, many strategies have been developed that target native amino acids on the protein surface, such as lysine and cysteine; however, it is difficult to control the number of modifications since multiple copies are generally present. One approach to achieve site-specificity is to target the N-terminus, due to its unique reactivity and position on a protein. Our lab has previously demonstrated one such method that uses 2-pyridinecarboxaldehyde (2PCA) to site-selectively modify the N-terminus. Despite the generality of this reaction, we have observed that over time the imidazolidinone conjugate can reverse to produce unmodified protein and that the rate of reversal depends on the identity of the N-terminal amino acid. In this work, we sought to characterize the conjugation products and determine the reaction mechanism with NMR kinetic experiments and DTF calculations to gain better insight into the factors that contribute to product formation and stability.

2.1 Introduction

The site-specific modification of biomolecules depends on a rapidly expanding set of techniques that can be used for a variety of outputs, from generating modern biotherapeutics[1,2] to studying cellular functions.[3] The attachment of synthetic molecules to proteins, while maintaining their desired structure and function, allows for the formation of hybrid materials that can capitalize on the properties of both components.[4] Synthesis of these bioconjugate constructs requires the use of chemoselective reactions that can proceed in aqueous solutions under mild pH and temperature conditions. Certain native amino acid side chains provide unique reactive handles for such reactions; however, the high frequency of most amino acid side chains on the surfaces of proteins can limit the site-specificity of modification. A growing set of chemical reactions have been developed to control the site of attachment and the number of modifications, such as the alkylation of genetically introduced cysteine residues,[5] the targeting of artificial amino acids with distinct reactivity,[6] native chemical ligations,[7] and enzymatic labeling techniques.[8] While the degree to which these amino acids are found on a given protein surface is either known or can be controlled, modification still requires varying levels of protein engineering.

As a result, novel protein modification techniques target distinct reactive sites, such as the N-terminal amine, due to its unique environment and pK_a.[9] N-terminal protein modification strategies offer significant advantages for bioconjugate preparation as they can be used for a wide range of protein targets produced by virtually any expression system. Some of these strategies target specific amino acid residues at the N-terminus; tryptophan residues can be modified by Pictet-Spengler reactions,[10] serine and threonine residues can yield reactive ketones or aldehydes after periodate oxidation,[11] and cysteine residues can condense with aldehydes to form thiazolidines.[12] Our group has reported a site-specific transamination reaction that introduces reactive ketones or aldehydes at the N-terminus,[13,14] as well as an oxidative coupling reaction between aminophenols and N-terminal proline residues.[15] Powerful as they are, however, all these methods have varying degrees of product stability, place constraints on the specific N-terminal amino acids that can be present, and require multiple chemical steps in several cases.

In previous reports, our group has demonstrated that 2-pyridinecarboxaldehyde (2PCA) can be utilized in a simple, one-step method to selectively modify the N-terminus of a broad scope of structurally and chemically varied proteins.[16] This modification reaction features mild reaction conditions, excellent site specificity, and no inherent requirement for genetic engineering. We have reported that the imidazolidinone formation reaction is reversible over extended time periods after the 2PCA reagent has been removed. Thus, we recommend that conjugates be used promptly after preparation, or stored in the presence of the 2PCA reagent until needed. Upon reversal the original unmodified protein is obtained, thus making this a "traceless" modification technique that could be useful for delivery applications. In other contexts, however, such as fluorophore labeling or enzyme immobilization, non-reversible attachment chemistry would be preferable.

To control this aspect of the chemistry more thoroughly, we embarked on a combined

experimental and computational study of the reaction to understand (1) the overall energetics of imidazolidinone formation, (2) the effects of differing amino acid side chains on the reaction performance, (3) the likely mechanisms through which the forward and reverse reactions proceed, and (4) likely explanations for why 2-pyridyl-substituted aldehydes are more effective for this chemistry than other analogs. As a result of these studies, we have emerged with a significantly clearer picture of how this reaction proceeds and design criteria for making conjugates more stable for applications that require it.

2.2 Results and Discussion

2.2.1 Measurement of Equilibrium Constants for Protein Substrates

In our experiments, we have consistently noted that 2PCA concentrations between 1 and 20 mM are required to obtain high levels of product formation. It was hypothesized that these concentrations were required to favor product equilibrium, which could differ slightly from substrate to substrate. To measure these values, protein modification reactions were run at 37 °C using varying concentrations of 2PCA over extended periods of time to ensure that maximum levels of conversion were obtained. Three different protein substrates were explored at 20 µM concentrations: RNase A with a KET N-terminal sequence and two modified ubiquitin samples with AAM and PMQ N-terminal sequences. For each substrate and 2PCA concentration, the reaction conversion plateaued at different levels, which are listed in Figure 1.

One consideration for these experiments is the tendency of 2PCA to form hydrates in aqueous solution (Figure 2), which has the effect of lowering the actual concentration of aldehyde reagent that is available for the reaction. The 2PCA aldehyde-to-hydrate ratio can be determined using NMR and was used to correct the initial 2PCA concentrations for equilibrium constant calculations. Using these values, the thermodynamic energies of the reactions can be determined for the different proteins.

RNase A and AAM-ubiquitin afforded similar values of $\Delta G^\circ = -7.9$ and –7.4 kcal/mol, respectively. The addition of a proline residue at the N-terminus led to a less-favorable equilibrium, affording $\Delta G^\circ =$ –6.0 kcal/mol. This observation agreed with previous studies that indicated lower levels of conversion for proline-terminated peptides using lower concentrations of the 2PCA reagent. Fits of the experimental data to the energy values are shown in Figure 1. An equilibrium constant was also measured for an ala-ala-ala (AAA) tripeptide as a model substrate. In this case, $\Delta G^\circ =$ –6.8 kcal/mol. This value is quite consistent with the protein data for RNase A and AAM-ubiquitin, suggesting that similar thermodynamic considerations apply to both cases.

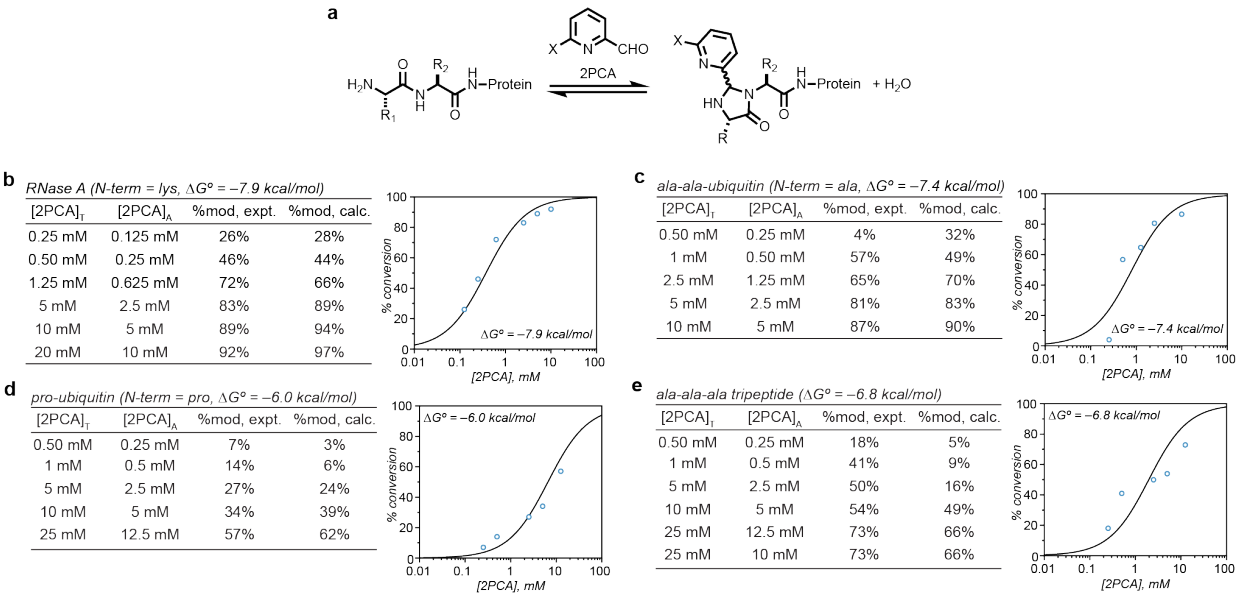


Figure 1: Comparison of K_{eq} values for imidazolidinone formation reactions. RNAse A, two N-terminally extended ubiquitin proteins, and an ala-ala-ala tripeptide were reacted with 2PCA to equilibrium. Conversions for each set of conditions were determined using ESI-TOF MS. The actual 2PCA concentrations ($[2PCA]_A$ values) differed from the total amount added ($[2PCA]_T$) by the amount of hydrate that was present, as shown in Figure 2a. These corrected values were used in the data analysis. The resulting data were matched to curves corresponding to different K_{eq} values. Reaction conditions: 50 uM protein or peptide, X mM 2PCA, 50 mM phosphate buffer pH 7.5, 37 °C, 72 h

2.2.2 Identification of Product Diastereomers

In our initial report, imidazolinone formation was postulated to introduce a new stereogenic center into the product. Assuming there are energetic consequences for the two product isomers (both in the forward and reverse direction), we sought to characterize the initial product isomer ratio and to identify if one product offered better stability. To determine the inherent diastereoselectivity of the reaction, the reaction was followed for the model AAA peptide using ^{1}H NMR. Fortuitously, many of the peptide and conjugate peaks are well resolved, with the key 2-pyridylmethine proton signals being particularly distinct between 5.5 and 5.75 ppm (signals a and a' in Figure 2b). For the AAA substrate, a pair of diastereomers was observed in a 3:1 ratio. To determine the identity of the major diastereomer, a product sample was evaporated to dryness and redissolved in deuterated acetonitrile for further structural analysis (see SI Figure 1). A 2D NOESY experiment was particularly diagnostic, indicating that the major 2-pyridylmethine proton peak correlated with a methyl group substituent of the imidazolidinone (see SI Figure 2). This transannular interaction was elucidated to be the *trans* isomer. The minor product peak showed correlation of the 2-pyridylmethine proton to both an alpha proton of the tripeptide and a methyl group. From this, we determined this was the *cis* isomer. Additional correlations arise from the *syn* interaction of the two imidazolidinone ring protons, as well as from an interaction with the

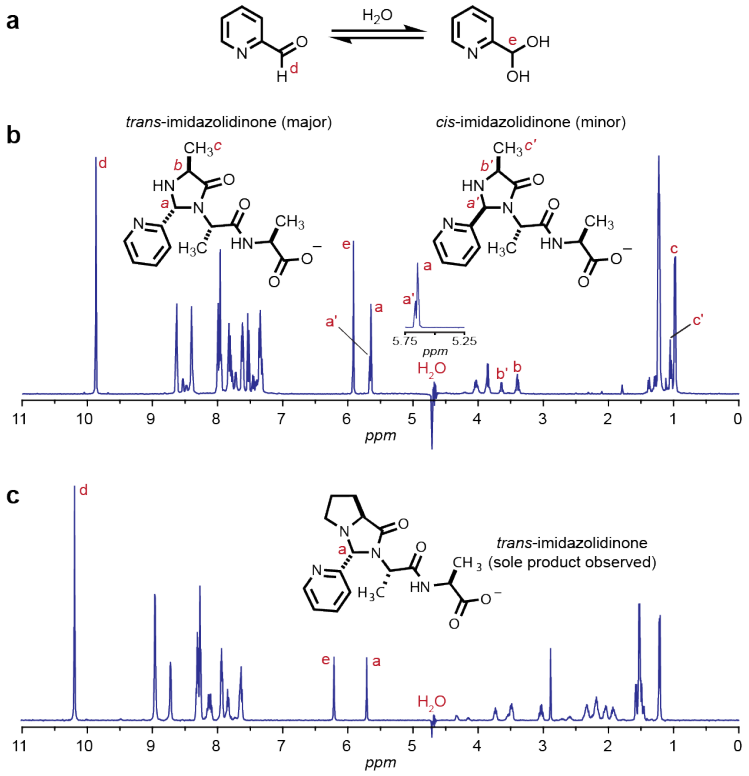


Figure 2: ^{1}H NMR spectra of imidazolidinone formation reactions. **(a)** An appreciable amount of aldehyde hydrate was observed under the reaction conditions. **(b)** NMR analysis of the imidazolidinone products resulting from an ala-ala-ala tripeptide showed a 2.2:1 mixture of *trans*-to-*cis* isomers at 40 °C. **(c)** In contrast, an N-terminal proline peptide yield only the *trans* imidazolidinone isomer. The spectrum in (a) was acquired at 40 °C and the spectrum in (b) was acquired at 60 °C.

methyl side chain of the second amino acid.

A similar NMR experiment was conducted for the PAA tripeptide, which produced a single product diastereomer (Figure 2b). Analysis using a 2D NOESY experiment showed correlations between the 2-pyridylmethine signal of the imidazolidinone and two of the protons in the proline ring side chain, as well as to one of the methyl groups (see SI Figure 3). This suggests that the observed product is again the *trans* isomer, where the imidazolidinone peak and the alpha proton of the ring are in an *anti* configuration.

2.2.3 Identification of Major Products for N-Terminal Serine and Cysteine Residues

Formation of the imidazolidinone product occurs by an intramolecular cyclization of the penultimate amino acid into the imine intermediate; however, some amino acids bearing beta-functional groups offer additional reaction pathways that could compete with imidazolidinone formation. In the case of N-terminal serine, the hydroxymethyl group could be envisioned

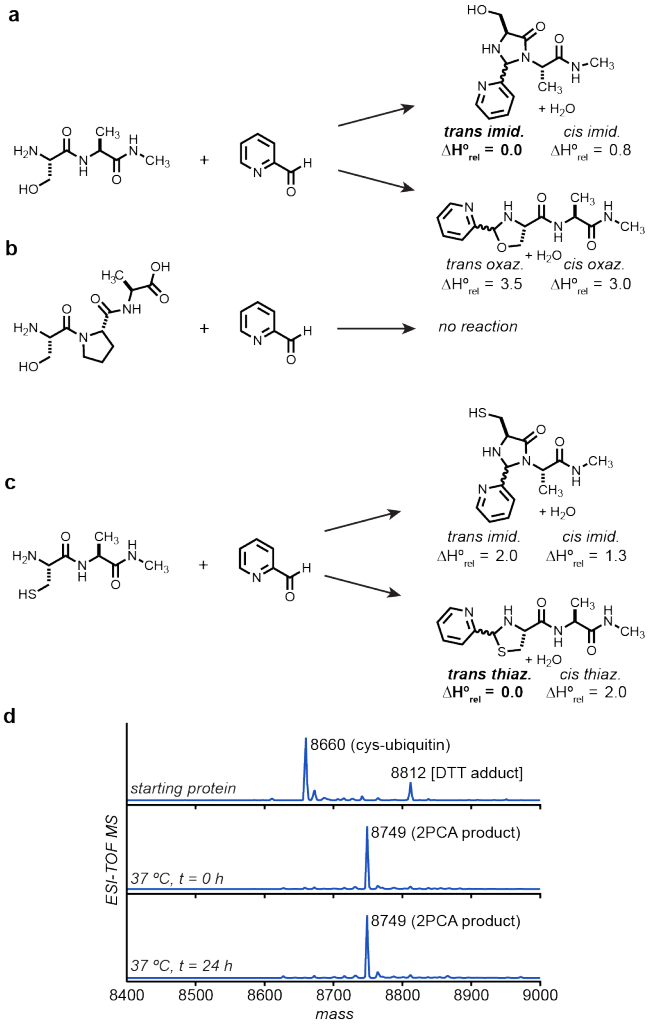


Figure 3: Product formation for peptides bearing N-terminal serine and cysteine residues. **(a)** For serine, computational analysis indicated preferential formation of the imidazolidinone products over the corresponding oxazolidines. **(b)** Exposure of a ser-pro-ala peptide to 2PCA yielded no reaction products, which is consistent with preferential imidazolidinone formation with this residue. **(c)** Peptides with N-terminal cysteine residues are predicted to form thiazolidine products more favorably. **(d)** The increased thermodynamic stability of the thiazolidine products was evident for ubiquitin with an N-terminal cysteine residue. No product degradation was observed even upon prolonged heating. Computations were conducted using ωB97M-V/ 6-311^{++}G(3df,3pd) and a CPCM solvent model. All energies are reported in kcal/mol.

to compete for addition into the imine to form an oxazolidinone product instead of the imidazolidinone. To determine whether this is the case, we compared the reactions of tripeptides SGG and SPA with 2PCA (Figure 3, see also SI Figure 4). When proline is the penultimate amino acid (as with SPA), cyclization to form the imidazolidinone product is not possible, so any product formed must instead involve participation of the hydroxymethyl group. By ^{1}H-NMR, the SGG-PCA conjugate displayed a single set of diastereomers with 2-pyridylmethine protons at 5.56 and 5.63 ppm. This chemical shift is very similar to that

obtained for the imidazolidinone with AAA. In contrast, no product peaks were observed by ^{1}H-NMR when SPA was reacted with 2PCA. This suggests that imidazolidinone formation is the operating pathway when serine is the N-terminal amino acid.

In addition to these experimental results, we also sought to support this finding computationally using density functional theory (DFT) as the method for determining the reaction energetics. One significant challenge for these studies was the high degree of conformational flexibility in the tripeptide substrates; therefore, as model compounds, we chose to use dipeptides bearing C-terminal amides, since these should reveal the interplay between the side chain identities and the imidazolidine ring while maintaining computational practicality. Using this approach for the serine peptide analogs, ΔH°_{rel} values were calculated for the *cis* and *trans* isomers of the oxazolidine and imidazolidinone products (Figure 3a). Both oxazolidine products were higher in energy than the imidazolidinone products, supporting the experimental results that formation of the imidazolidinone is the preferred pathway.

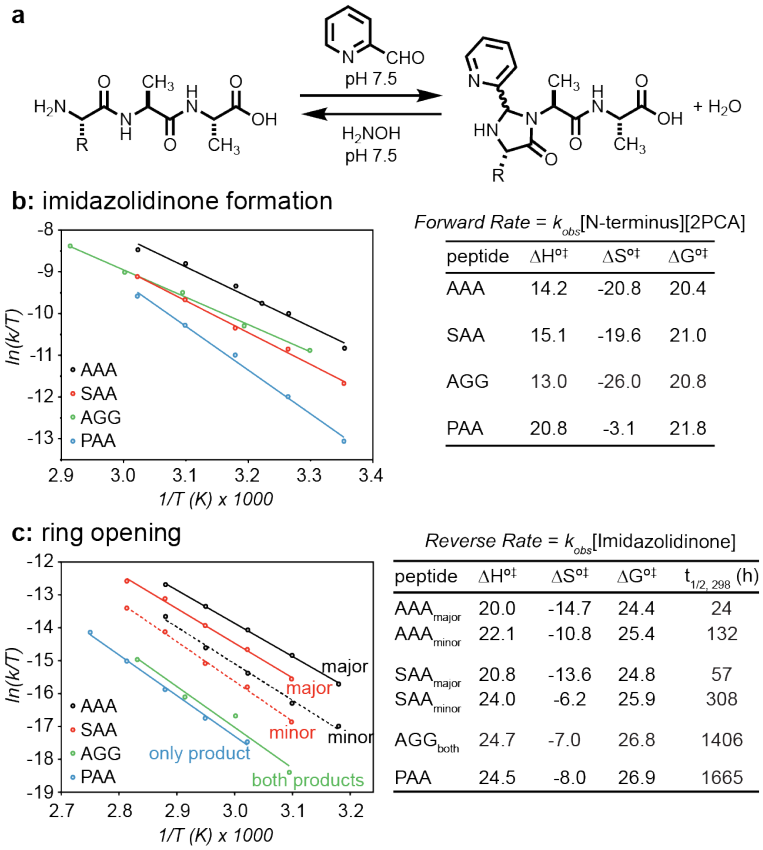
When cysteine is the N-terminal residue, the sulfhydryl group is known to participate in modification reactions such as condensation with 2-cyanobenzothiazole (CBT) to form thiazolines[17] or with aldehydes to form thiazolidines.[18] For the cysteine peptide analog, it was found computationally that formation of the *trans* thiazolidine product with 2PCA was more favorable than imidazolidinone formation (2.0 kcal/mol and 1.3 kcal/mol higher for the *trans* and *cis* isomers, respectively) (Figure 3b). To explore this increased stability of the thiazolidine product further, a ubiquitin protein was expressed with an N-terminal cysteine residue (Figure 3d). After initial conjugation with 2PCA and removal of excess reagent, no reversibility was observed after 24 h at 37 °C.

2.2.4 Determination of Reaction Kinetics

The distinct 2-pyridylmethine proton signals in the ^{1}H NMR spectrum provide a convenient and accurate way to monitor reaction conversion, using a known amount of DMSO as an internal standard. Using this approach with varying substrate concentrations, the rate law for the AAA imidazolidinone formation reaction was determined to be

$$r = k_{obs}[peptide][2PCA]$$

Next, kinetic experiments were performed for a series of tripeptides at varying temperatures in the NMR probe. Second-order rate constants were measured at five different temperatures for AAA, SAA, AGG, and PAA, and these data were used to derive kinetic activation parameters from an Eyring plot (Figure 4a and 4b; see also SI Table 1). In all cases, excellent linear fits were obtained for plots of ln(k/T) vs. 1/T ($R^2 \geq 0.98$). The $\Delta G^{\circ\ddagger}$ value for the reaction with AAA was found to be 20.4 kcal/mol, with $\Delta H^{\circ\ddagger}$ = 14.2 kcal/mol and $\Delta S^{\circ\ddagger}$ = –20.8 e.u. Similar activation parameters were obtained for SAA and AGG, with a small, expected increase in $\Delta S^{\circ\ddagger}$ for the AGG case. Both SAA and AAA afforded major and minor product isomers (∼3:1), but the AGG peptide provided an equal mixture of the product diastereomers. Distinct activation parameter values were observed for PAA, with a significantly higher $\Delta H^{\circ\ddagger}$ value (20.8 kcal/mol) and a more modest $\Delta S^{\circ\ddagger}$ penalty (–3.1 e.u.).



Forward Rate = k_{obs}[N-terminus][2PCA]

peptide	$\Delta H^{o\ddagger}$	$\Delta S^{o\ddagger}$	$\Delta G^{o\ddagger}$
AAA	14.2	-20.8	20.4
SAA	15.1	-19.6	21.0
AGG	13.0	-26.0	20.8
PAA	20.8	-3.1	21.8

Reverse Rate = k_{obs}[Imidazolidinone]

peptide	$\Delta H^{o\ddagger}$	$\Delta S^{o\ddagger}$	$\Delta G^{o\ddagger}$	$t_{1/2,\ 298}$ (h)
AAA_{major}	20.0	-14.7	24.4	24
AAA_{minor}	22.1	-10.8	25.4	132
SAA_{major}	20.8	-13.6	24.8	57
SAA_{minor}	24.0	-6.2	25.9	308
AGG_{both}	24.7	-7.0	26.8	1406
PAA	24.5	-8.0	26.9	1665

Figure 4: Determination of the activation parameters for the forward and reverse reactions using Eyring plots. **(a)** Forward reactions were run at a variety of temperatures for a series of tripeptides (2 mM) using 10 mM 2PCA at pH 7.5. To follow the reverse reactions, the remaining 2PCA was reacted with hydroxylamine. Upon doing so, the reactions slowly reversed to yield the starting peptides. Conversion to the major product was monitored using ^{1}H NMR. **(b)** Eyring plots for the forward reaction yielded the activation parameters. Notably different values were obtained for the N-terminal proline case. **(c)** The reverse reaction rates were monitored for the two product diastereomers, except for the proline case (which yielded a single product). The proline-terminated peptide also exhibited the slowest reverse reaction rate.

To study the ring-opening kinetics of the imidazolidinone product, a method was needed to remove the participation of 2PCA without significant perturbation of the reaction solution. We found that hydroxylamine rapidly and irreversibly forms an oxime with the excess 2PCA when added to the reaction; however, this reagent was not observed to react with the imidazolidinone itself. It is recognized that hydroxylamine could also react with transient imine species formed immediately after ring opening, but this species is presumably after the rate determining step and would not alter the observed kinetic behavior.

Using this approach, the rate law for the reverse reaction presumably becomes

$$r = k_{obs}[\text{imidazolidinone}]$$

and the disappearance of product can again be followed using ^{1}H NMR compared to a

DMSO internal standard. Using this approach, Eyring plots were again used to determine the activation parameters for the ring opening of the AAA, SAA, AGG, and PAA imidazolidinone products (Figure 4c, see also SI Table 2). Excellent linear fits were again obtained ($R^2 \geq 0.97$) in all cases. For AAA and SAA, the major and minor products were tracked independently. In both cases, the minor isomers were in fact more kinetically stable. In the case of AGG, the two product isomers opened at equal rates. Interestingly, the ring opening rate of the AGG peptide was found to be substantially slower than the AAA and SAA substrates, suggesting that a stabilizing effect can be obtained by omitting the side chain of the second amino acid. Product stabilization can also be achieved by introducing a proline residue in the terminal position. The $\Delta G^{\circ\ddagger}$ for PAA was found to be 26.9 kcal/mol, which was similar in magnitude to that of AGG. Thus, these studies revealed two different ways to use the amino acid sequence to achieve lower reversibility rates.

2.2.5 Determination of Rate-Limiting Steps

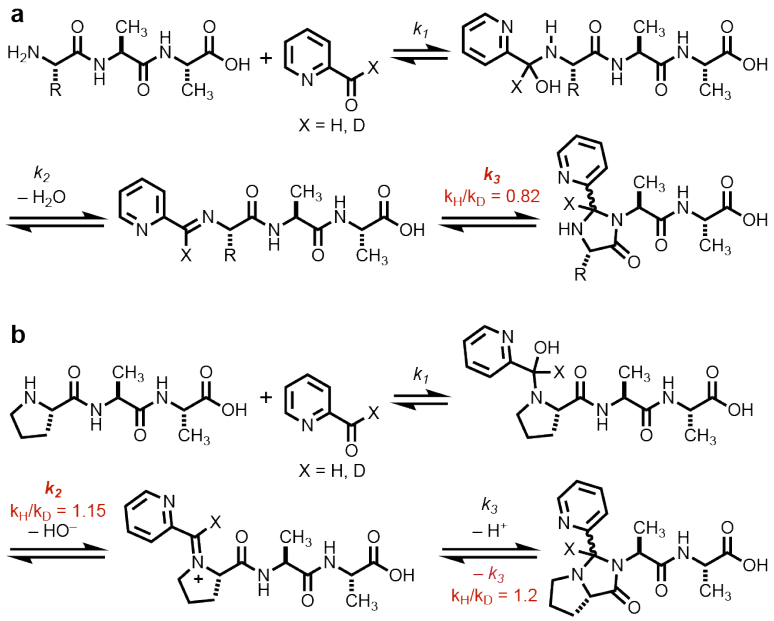


Figure 5: Individual reaction steps for imidazolidinone formation, along with observed kinetic isotope effects for 2PCA bearing a deuterium substitution on the aldehyde group. **(a)** An inverse secondary isotope effect was observed for an ala-ala-ala tripeptide, consistent with an sp^2-to-sp^3 hybridization change in the rate deterining step. This is assigned to imidazolidinone ring formation. **(b)** A normal secondary isotope effect was observed for a pro-ala-ala peptide, which indicates an sp^3-to-sp^2 hybridization change in the rate determining step. This presumably corresponds to iminium ion formation, followed by fast cyclization. A normal secondary isotope effect is also observed for the reverse ring opening reaction, similarly corresponding to iminium ion formaion (albeit in a different reaction step). Kinetic isotope ratios were determined using NMR.

The imidazolidinone product is formed through a transient imine intermediate and loss of

water, which then undergoes the cyclization reaction. The cyclization of this intermediate can be inferred through the kinetic analysis of an isotopically labeled 2PCA substrate. As shown in Figure 5, the aldehyde proton of 2PCA was replaced with a deuterium atom, and the reaction rates were compared using ^{1}H NMR (compared to a DMSO internal standard) (see also SI Figure 5). For AAA, the ratio of k_H/k_D was 0.82, which is consistent with an inverse secondary isotope effect. This observation is most consistent with an sp^2 to sp^3 hybridization change in the rate-determining step, as shown for the cyclization event itself in Figure 5. In contrast, imidazolidinone formation for PAA afforded a k_H/k_D ratio of 1.15, which is consistent with a hybridization change from sp^3 to sp^2 in the rate determining step. This suggests that the collapse of the 2PCA hemiaminal intermediate is rate limiting, presumably because this leads directly to a higher-energy iminium intermediate. When run in the presence of hydroxylamine, the ring-opening reaction for the PAA substrate afforded a k_H/k_D ratio of 1.2, suggesting that the formation of the iminium intermediate is again rate limiting.

2.2.6 Computational analysis of imidazolidinone products

From the NMR experiments, we have found that in the case of AAA and SAA, two imidazolidinone products are formed after 2PCA conjugation in a 2.2:1 ratio of *trans* to *cis* isomers; however, we have also noted that the 2PCA-AGG conjugate is formed in an almost equal 1:1 mixture of isomers. To understand the origins of this difference, we modeled the imidazolidinone products and the acyclic imine precursors using DFT to calculate the energetics of these species. With the AA peptide analog, imidazolidinone formation of both the *trans* and *cis* isomers was found to be exothermic (ΔH° = -7.5 kcal/mol and -6.4 kcal/mol, respectively, relative to the ΔH° of the imine precursor) (Figure 6a). These values also indicate that the *trans* isomer is more stable by 1.1 kcal/mol and should be preferentially formed, which corresponds well to the experimentally observed product ratio. We then modeled the AG peptide analog and found that the *trans* and *cis* imidazolidinones are almost equal in energy (-7.8 kcal/mol and -8.1 kcal/mol, respectively) (Figure 6b), which would explain the ∼1:1 ratio of observed product isomers. These products are also more favorable than those with the AA analog, and to understand this difference we examined the structures of the imidazolidinones. The AA analog has a methyl group on the α carbon, which forces an unfavorable steric interaction between the pyridine ring and the second alanine residue. This residue preferentially sits with the amine backbone *syn* to the pyridine over the methyl group because there is a smaller energetic penalty. Since the AG analog doesn't have an α carbon substituent, these unfavorable interactions can be avoided by facing the amine backbone away, leaving the α proton *syn* to the pyridine ring. Finally, we also examined the reaction of 4PCA with the AA peptide analog and found that both the *trans* and *cis* imidazolidinones are less favorable than with 2PCA (Figure 6b). This is due to unavoidable steric interactions between the C-H bonds of the pyridine ring and the second residue, which aren't present with 2PCA.

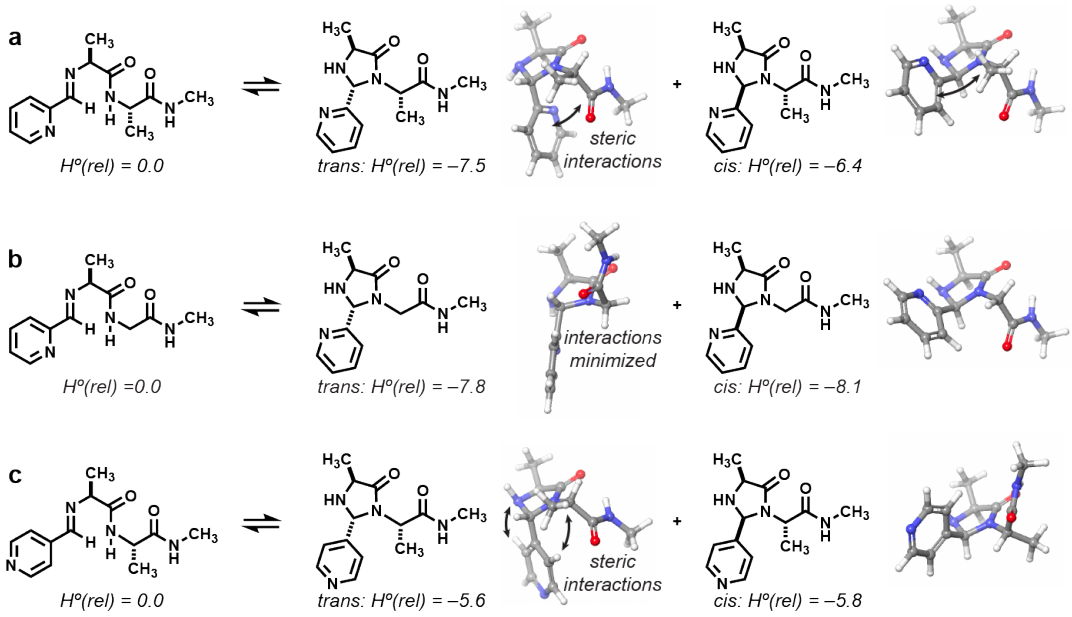


Figure 6: Computational analysis of imidazolidinone products. The enthalpies of cyclization were calculated relative to the acyclic imine precursors. **(a)** Both imidazolidinone formation pathways with an ala-ala dipeptide derivative are exothermic, with a 1.1 kcal/mol preference for the *trans* product. This corresponds well to the product ratio determined in Figure 2. The pyridine ring undergoes steric interactions with the second alanine residue in both cases, as indicated by the arrows, with a smaller energetic penalty for the amine backbone than the methyl side chain group. **(b)** Imidazolidinones formed for an ala-gly dipeptide analog avoid these interactions by orienting the peptide backbone away from the pyridine ring in both isomers. This results in greater product favorability overall and the observed ∼1:1 ratio of product isomers. **(c)** An analogous analysis for 4PCA shows greater steric interactions in all products, largely due to the replacement of the 2-pyridyl nitrogen with a more encumbering C-H bond. Computations were conducted using ωB97M-V/ 6-311^{++}G(3df,3pd) and a CPCM solvent model. All energies are reported in kcal/mol.

2.2.7 Computational analysis of imidazolidinone ring-opening reactions

The combined kinetic and isotope information provides a platform for understanding how imidazolidinone formation occurs, as well as comparative data for model validation. As noted above, imidazolidinone cyclization should proceed through the formation of a 2PCA-imine intermediate, which is an isomer of the reaction product. It is further likely that the nitrogen atom of the imine species must be protonated for the reaction to occur, both to activate the resulting iminium carbon for nucleophilic attack and to prevent the formation of a nitrogen anion as the addition occurs. What is less clear is whether the cyclization proceeds through the direct attack of the amide nitrogen lone pair on the iminium carbon, or through the attack of an imidate tautomer. Although this species would be expected to be higher in energy than the amide, it would likely be lower in energy than the transition state for cyclization, and thus could still be a viable intermediate. One can use similar considerations to determine how amide protons exchange with the bulk solvent either through direct protonation of the amide

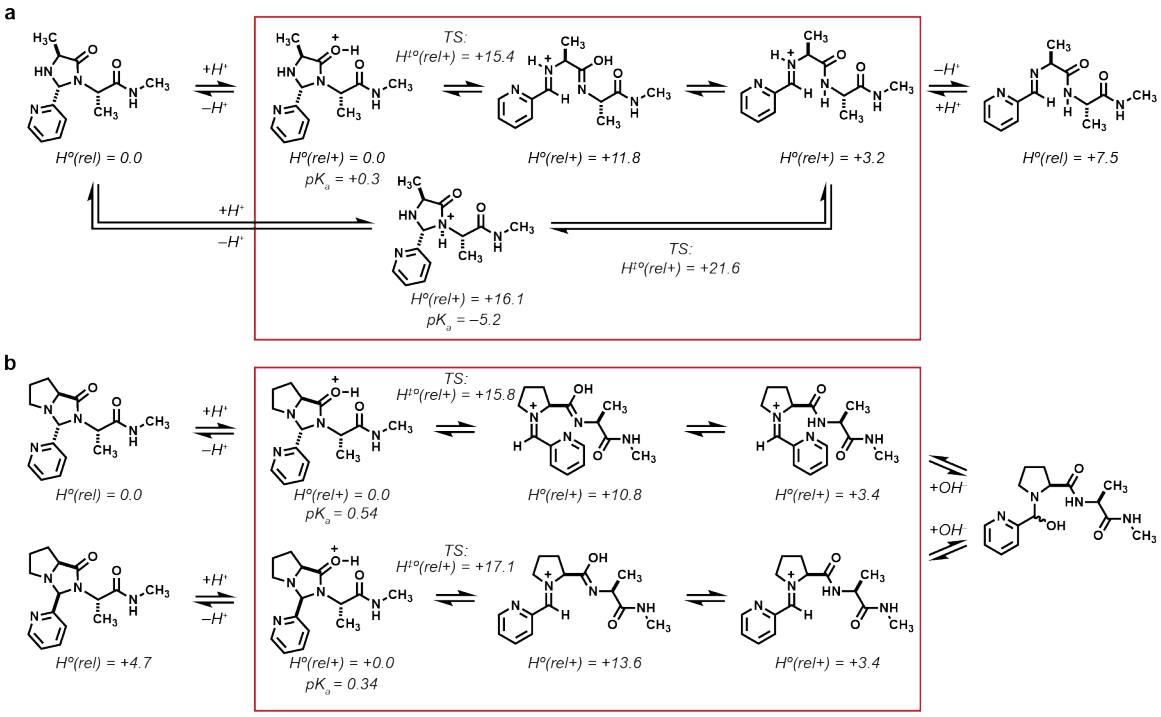


Figure 7: Computational analysis of imidazolidinone ring-opening reactions for ala-ala and pro-ala dipeptide analogs. The reactions can be assumed to proceed through an initial protonation step, leading to the sets of cationic isomers contained in the red boxes. These intermediates can be compared using the H°(rel+) values listed. **(a)** For the ala-ala peptide analog, the species with a protonated carbonyl oxygen atom can ring-open to form a transient imidate intermediate. This species can then tautomerize to yield an iminium ion that is deprotonated to yield the neutral imine product. The *N*-protonated amide species can ring open to form the same iminium species directly. The *O*-protonated pathway is more likely because it is much more favorable to protonate the amide carbonyl oxygen atom than the nitrogen atom, as reflected by both thermodynamic energies and pK_a calculations. Similarly, the ring-opening transition state (TS) is considerably lower in energy for the *O*-protonated isomer. The structure of the identified *O*-protonated transition state appears in Figure 8a. **(b)** A similar analysis is presented for a pro-ala dipeptide analog, considering only the mechanistic pathways for the *trans* and *cis* *O*-protonated species. As shown on the left, the *trans* isomer is more stable thermodynamically than the *cis* isomer by 4.7 kcal/mol, explaining the high product preference. This energetic favorability is also reflected in the imidate intermediates (which are themselves *cis* and *trans* isomers) and the transition states leading to them. Energetic computations were conducted using ωB97M-V/ 6-311^{++}G(3df,3pd) and a CPCM solvent model. All energies are reported in kcal/mol.

nitrogen, or through a tautomerization mechanism. Evidence for both pathways has been reported in the literature, with the former pathway predominating at lower pH conditions and the tautomerization occurring preferentially at higher pH.[19] With these considerations, we initiated a computational study of the reaction, with the overall goal of determining the most likely reaction pathway.

The ring-opening reaction begins with an initial protonation of the imidazolidinone product. From here, all tautomerizations and ring-opening steps form cationic isomers, which

can be directly compared using H° values (Figure 7 red box). For the ala-ala peptide analog (Figure 7a), there are two potential sites for this initial protonation. If the carbonyl of the amide is protonated, then ring opening can occur to produce an iminium cation and a transient imidate intermediate, which can further undergo tautomerization to form an amide. Alternatively, the imidazolidinone could be protonated on the nitrogen of the amide, which leads to a direct ring opening to form the iminium cation. Comparing the H° values, protonation of the amide nitrogen is 16.1 kcal/mol higher in energy than protonation of the amide carbonyl, so the latter pathway is more likely. This is also supported by the calculated pK_a values (+0.3 vs -5.2 for the amide carbonyl and nitrogen, respectively). In addition, the transition state energy for ring opening is significantly lower when going through the pathway with the amide carbonyl protonated.

For the pro-ala peptide analog, a comparison of the *cis* and *trans* imidazolidinone isomers leading to the ring opened product was made, only considering the pathway from amide carbonyl protonation. Looking at the neutral species, the *cis* isomer is 4.7 kcal/mol higher in energy than the *trans* isomer, suggesting that the latter is more thermodynamically favored.

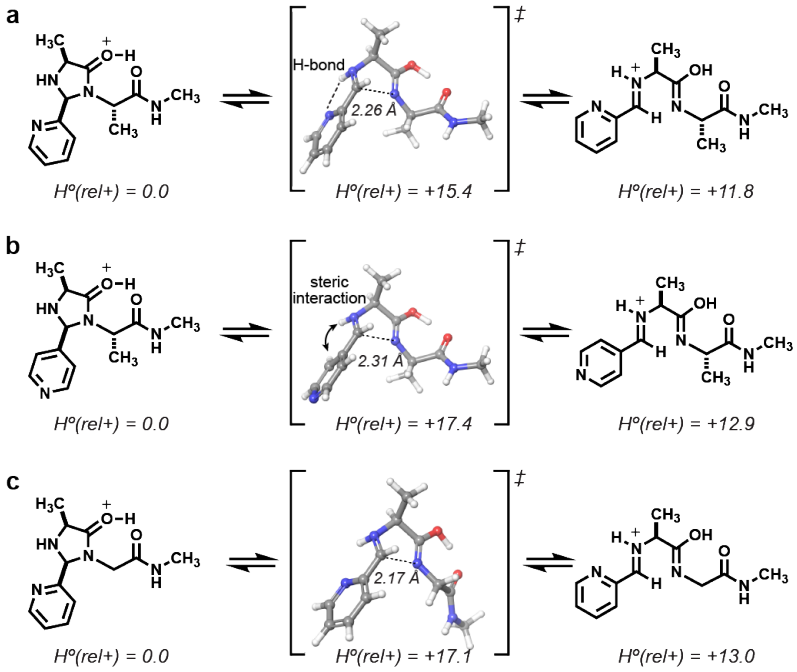


Figure 8: Comparisons of calculated transition states for imidazolidinone ring-opening reactions. Transition state structures were identified between the *O*-protonated carbonyl species and the imidate intermediates using DFT methods. Only pathways for the *trans* isomers are shown. **(a)** For the product corresponding to an ala-ala dipeptide analog, a favorable hydrogen bond is observed during the bond formation step. **(b)** This hydrogen bond is replaced by unfavorable steric interactions with the two *ortho* protons for a 4PCA-derived product, leading to a higher energy transition state. **(c)** For an ala-gly dipeptide analog, the transition state structure is higher in energy relative to the protonated imidazolidinone, presumably due to ground state effects described for the products in Figure 6b. Energetic computations were conducted using ωB97M-V/ 6-311^{++}G(3df,3pd) and a CPCM solvent model. All energies are reported in kcal/mol.

This large difference most likely explains why only one product isomer is observed for the reaction. The preference for *trans* over *cis* can also be seen in the energies of the imidate intermediates after ring opening (10.8 kcal/mol vs 13.6 kcal/mol, respectively), as well as in the transition state energies (15.8 kcal/mol vs 17.1 kcal/mol, respectively).

With clearer insight into the mechanism through which ring opening occurs, we next sought to understand what leads to the different product stabilities by examining the transition states between the protonated amide carbonyl species and the imidate intermediates using DFT. For the AA peptide analog with 2PCA, a favorable hydrogen bond is observed between the nitrogen lone pair of the pyridine ring and the protonated iminium cation in the bond forming step (Figure 8a). With 4PCA, this hydrogen bond can no longer occur and is instead replaced by unfavorable steric interactions with the two C-H protons, leading to a higher energy transition state (Figure 8b). For the AG peptide analog with 2PCA, the relative energy difference between the transition state and iminium cation is greater than the AA peptide due to a more stable ground state (Figure 8c), and we believe this difference leads to a higher product stability.

2.2.8 2PCA Modification of anti-HER2 Nanobodies with Various N-Terminal Extensions

Having used peptide substrates initially to gain a better understanding of the reaction, we then wanted to explore if these trends in amino acid identity at the N-terminus would also apply to protein substrates. To this end, we chose the anti-HER2 nanobody (HER2-Nb) as our protein of interest for 2PCA modification. Nanobodies are monomeric antibody fragments consisting of the antigen binding region of heavy chain antibodies. Due to their small size, high stability, excellent tissue penetration, and ease of production, nanobodies have gained great interest as an alternative to monoclonal antibodies for drug delivery.[20]

We began by genetically engineering the HER2-Nb to include N-terminal extensions with our sequences of interest. To mimic the tripeptide substrates, an ala-ala-ala (AAA) extension was added, as well as a pro-ala-ala (PAA) extension to determine if the N-terminal proline would display a higher conjugate stability on a protein with 2PCA. We also included ala-gly-gly (AGG) and pro-gly-gly (PGG) extensions to explore if the additive stabilization effects of glycine at position two would be observed. Following each tripeptide extension, a His6 tag was also added for purification.

With the desired proteins in hand, we next modified each with 2PCA, and good to excellent levels of modification were achieved (see SI Figure 6). The levels of conversion were all normalized to 100%, and then each sample was monitored over 3 days at 37 °C after removal of excess 2PCA (Figure 9). After 24 h, the modification level of the 2PCA-AAA Nb conjugate decreased by 47%, and after 72 h it decreased by 61% of the initial amount. In comparison, the 2PCA-PAA Nb conjugate only decreased by 8% after 24 h and 15% after 72 h, providing evidence that N-terminal proline residues on proteins create much more stable conjugates with 2PCA. When looking at the glycine effect at the second position, the

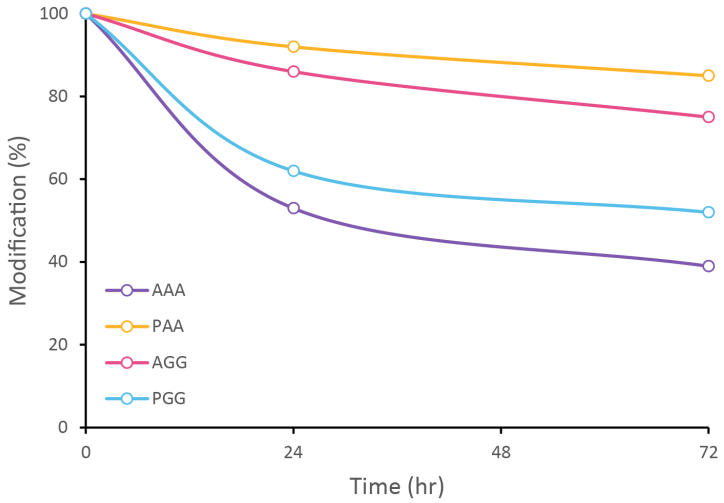


Figure 9: Reversibility of various anti-HER2 nanobodies with 2PCA. Four nanobodies were expressed with either an AAA, PAA, AGG, or PGG N-terminal extension. Each was modified with 2PCA, then excess 2PCA was removed and the modification levels were monitored over time at 37 °C. The 2PCA-PAA conjugate displayed the highest stability over three days. The AGG and PGG conjugates with 2PCA were also more stable than the AAA peptide.

2PCA-AGG Nb conjugate decreased by 14% after 24 h and 25% after 72 h. This stability falls in between the AAA and PAA Nb conjugates, suggesting that the glycine effect does help create a more stable conjugate when compared to the AAA Nb, but the effect is not as great as when proline is the N-terminal residue. This observed trend agrees with the thermodynamic values calculated for tripeptide modification. Finally, the modification level of the 2PCA-PGG Nb conjugate decreased by 38% after 24 h and 48% after 72 h. Although the PGG sequence did increase the conjugate stability, the additive effect of an N-terminal proline followed by a glycine residue was not observed. This could be because the bicyclic product structure with a proline residue interacts differently with the second amino acid than the monocyclic product with alanine, and the different steric interactions might lead to different stabilities. From these studies with the HER2-Nb mutants, it is clear that an N-terminal proline residue greatly increases the stability of the conjugate. Further investigation of the glycine effect is needed to determine how the adjacent amino acids affect the structure and stability of the conjugate.

2.3 Conclusion

N-terminal modification with 2PCA is a highly useful reaction that facilitates the single, site-specific chemical modification of a biomolecule of interest. As a result of these ongoing studies, we have emerged with a clearer picture of how this reaction proceeds and designed criteria for tuning the modification of the synthesized conjugate material. Furthermore, as

site-specific protein modification techniques are of utmost interest, especially those that are easily controlled, applications of this bioconjugation reaction have shown great potential, from drug delivery and enzyme remediation of pollutants to novel heterobifunctional linkers. Because chemical modification of biomolecules is an important set of techniques, especially for therapeutic applications, second-generation 2PCA molecules will help enhance the ever-expanding bioconjugation field.

2.4 Supporting Information

2.4.1 Reagents and Materials

Unless otherwise noted, all reagents were obtained from commercial sources and used without any further purification. 2-pyridinecarboxaldehyde (2PCA) was purchased from Sigma-Aldrich and stored as a 100 mM solution in ddH_2O at 4 °C. RNAse A was purchased from Sigma-Aldrich and dissolved in ddH_2O to an approximate concentration of 1 mM based on tryptophan absorbance at 280 nm. The dissolved peptides were stored at 4 °C to prevent degradation. Both reaction components were diluted with the appropriate buffer immediately before the reaction. Analytical thin layer chromatography (TLC) was performed on EM Reagent 0.25 mm silica gel 60-F254 plates and visualized by ultraviolet (UV) irradiation at 254 nm and/or staining with potassium permanganate. Purifications by flash silica gel chromatography were performed using EM silica gel 60 (230–400 mesh). All organic solvents were removed under reduced pressure using a rotary evaporator. Water (dd-H_2O) used in all procedures was deionized using a NANOpureTM purification system (Barnstead, USA). Centrifugations were performed with an Eppendorf 5424 R at 4 °C (Eppendorf, Hauppauge, NY). Peptides were procured from GenScript (Piscataway, NJ).

2.4.2 Instrumentation Methods

Nuclear Magnetic Resonance (NMR) Spectroscopy. ^{1}H spectra were measured with a Bruker AV-500 (500 MHz, 150 MHz) spectrometer. ^{1}H NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to residual CH_3CN (δ 2.10, singlet) or H_2O (δ 4.72, singlet).

High Performance Liquid Chromatography (HPLC). HPLC was performed on Agilent 1200 series HPLC systems (Agilent Technologies, USA) equipped with an in-line diode array detector (DAD), fluorescence detector (FLD), and automatic fraction collector. Semi-preparative reverse-phase chromatography was achieved using a C8 stationary phase (5 micron, 250x10 mm Syncronis column, Thermo Scientific) and a H_2O/CH_3CN with 0.1% TFA gradient mobile phase at a flow rate of 3.0 mL/min.

Liquid chromatography mass spectrometry (LC-MS) analysis. Proteins and protein conjugates were analyzed on an Agilent 6224 Time-of-Flight (TOF) mass spectrometer with a dual electrospray source (ESI) 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/CH_3CN gradient mobile phase containing 0.1% formic acid. Mass spectra of proteins and protein conjugates were deconvoluted with MassHunter Qualitative Analysis Suite B.05 (Agilent Technologies, USA). Small molecule LC-MS analysis was performed on a C18 (1.7 micron, 150x2.5 mm, Gemini column, Phenomenex) with a gradient mobile phase containing 0.1% formic acid.

2.4.3 Experimental Procedures

General method for modification of proteins with 2PCA. The reaction was prepared in a 0.6 mL microcentrifuge tube. A 25 µL aliquot was taken from a 100 µM solution of protein (2.5 nmol, final concentration 50 µM) and added to 15–24.75 µL of various buffers at various pH values. To the resulting solution was added a 0.25-10 µL aliquot from a 100 mM solution of 2PCA in ddH_2O (25–1,000 nmol, final concentration 0.5–20 mM). The reaction was briefly agitated to ensure proper mixing and incubated at room temperature or 37 °C without further agitation. After various time points, the reaction was purified using repeated (five times) centrifugal filtration against a 0.5 mL Amicon Ultra-4 Centrifugal Filter spin concentrator with an appropriate molecular weight cutoff (EMD Millipore, USA). Modification was monitored by ESI-TOF LC-MS.

General method for modification of commercial and synthetic peptides with 2PCA. A 2 µl aliquot was taken from a 1 mM solution of peptide (2 nmol, final concentration 100 µM) and added to 16 µl of 10 mM phosphate buffer at pH 7.5. To the resulting solution was added a 2 µL aliquot from a 100 mM solution of 2PCA (200 nmol, final concentration 10 mM). The reaction was briefly agitated to ensure proper mixing and incubated at room temperature or 37 °C without further agitation. After various time points, excess 2PCA was quenched by the addition of hydroxylamine and purified using Waters Sep-Pak C18 cartridge (1 cc 50 mg) following manufacturer's instructions. Modification was monitored by ESI-TOF LC-MS or RP-HPLC.

General method for the NMR analysis of 2PCA-tripeptide modification. Tripeptide (2 mM) was mixed with 2PCA (10 mM) in phosphate buffer (50 mM) at pH 7.5. A solution of DMSO (1 µL, 0.5 mM) as internal standard and D_2O (10% v/v) were added, and the mixture was then transferred to an NMR tube.

Data acquisition notes. Spectra were acquired in a 500 or 600 MHz Advance series Bruker NMR spectrometer set to maintain a constant temperature. The noted reaction temperature

refers to that calibrated with an ethylene glycol sample within one week of the experiment (either neat or 80% in DMSO-d_6). NMR spectra were acquired with a 90° pulse and 8 scans per timepoint at set intervals of 5-30 min (varied depending on the signal-to-noise ratio and half-life).

Data analysis notes. The concentrations of starting materials and products were determined by integration against the internal standard. First-order rate constants were calculated by fitting the plotted data to the formula

$$\ln[\text{imidazolidinone}] = -kt + \ln[\text{imidazolidinone}]_o$$

Second-order rate constants were calculated from

$$\ln\frac{[\text{2PCA}][\text{peptide}]_o}{[\text{peptide}][\text{2PCA}]_o} = k([\text{2PCA}]_o - [\text{peptide}]_o)t$$

where

$$[\text{peptide}] = [\text{peptide}]_o - [\text{imidazolidinone}]$$

and

$$[\text{2PCA}]_{\text{actual}} = [\text{2PCA}]_{\text{measured}} + [\text{hydrate}]_{\text{measured}}$$

The data from at least one half-life were used in the fit. The free energy of activation was calculated from the equation

$$\ln\frac{k}{T} = \frac{-\Delta H^{\ddagger}}{R}\frac{1}{T} + \ln\frac{k_B}{h} + \frac{\Delta S^{\ddagger}}{R}$$

and

$$\Delta G^{\circ\ddagger} = \Delta H^{\circ\ddagger} - T\Delta S^{\circ\ddagger}$$

General method for the computational analysis of 2PCA-peptide products. Molecular mechanics methods (Macromodel, OPLS4 force field) were used to generate and minimize large populations of conformers to identify the lowest energy candidates. For each compound under study, the geometries of all conformers within 3 kcal/mol of the global minimum were optimized using at the B3LYP-D3/6-31G** level.[21] At this stage, the resulting global minima were subjected to a second geometry optimization and vibrational spectrum calculation (B3LYP-D3/6-31G**) to determine the zero-point energies, enthalpy corrections, and the internal entropy values at 298.15 K. Finally, refined electronic energy calculations were performed on each optimized geometry using an expanded basis set ωB97M-V/ 6-311^{++}G(3df,3pd). For calculations including solvation, a CPCM model was used with Bondi van der Waals radii

and a SAS solvent probe radius of 1.4 Å.[22,23] In these cases, solvation was introduced only into the final electronic energy calculations.

Ubiquitin expression and purification. Ubiquitin, and all ubiquitin mutants, were expressed and purified as previously described.[24] Briefly, Rosetta II (DE3) pLysS *E. coli* cells were transformed with a pET28a vector containing the ubiquitin gene from *S. cerevisiae* under control of a T7 promoter. Cells were grown in Terrific Broth supplemented with 1% glycerol at 37 °C until OD600 = 1.5–2.0 and were induced with 0.5 mM IPTG overnight at 16 °C. Cells were harvested by centrifugation, and pellets were frozen at -80 °C. Purification was carried out as described previously,34 but scaled up with minor modifications. The lysis buffer contained 50 mM Tris-HCl, pH 7.6, 0.02% NP-40, 2 mg/mL lysozyme, benzonase (Novagen), and protease inhibitors (aprotinin, pepstatin, leupeptin and PMSF). Cells were lysed by sonication (on ice) twice, followed by centrifugation to remove cellular debris. The supernatant was subjected to salting out; 60% perchloric acid was slowly added to a final concentration of 0.5%, and the solution was stirred on ice for a total of 20 min. A 5-mL HiTrap SP FF column (GE Life Sciences) was used for cation-exchange chromatography, and ubiquitin-containing fractions were pooled and exchanged into storage buffer (20 mM Tris-HCl, pH 7.6, and 150 mM NaCl) by repeated dilution and concentration in Amicon Ultra 3000 MWCO spin concentrators (Millipore).

Ubiquitin mutagenesis. QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, USA) was used to extend the N-terminus of ubiquitin. Mutated plasmids were transformed into XL1Blu *E. coli* cells, cultured overnight, and subsequently miniprepped to verify incorporation of the amino acid(s) by sequencing. The following primers were used for the respective extensions:

Ala-Ala extension (M-A-A-M-Q-I-F):
Forward: 5'-GAAGGAGATATACCCATATGGCTGCTATGCAGATTTTCG-3'
Reverse: 5'-GACGAAAATCTGAGCAGCCATCATATGGGTATATCTCC-3'

Pro-Ala extension (M-P-A-M-Q-I-F):
Forward: 5'-GGAGATATACCCATATGCCGGCTATGCAGATTTTCGTCAG-3'
Reverse: 5'-GACGAAAATCTGCATAGCCGGCATATGGGTATATCTCC-3'

Cys extension (M-C-M-Q-I-F):
Forward: 5'-GATATACCCATATGTGTATGCAGATTTTCGTCAAGAC-3'
Reverse: 5'-GACGAAAATCTGCATACACATATGGGTATATCTCC-3'

Nanobody expression and purification. Sequenced plasmids were transformed into BL21(DE3) Star competent cells via a 45 s, 42 °C heat shock and plated on LB/kanamycin. Resulting colonies were used to inoculate 5 mL LB/kanamycin cultures (50 µg/mL) and grown overnight at 37 °C, 225 RPM. The culture was then added to 500 mL of LB media and grown at 37 °C

to an OD600 of 0.6- 0.8. Expression was induced with a final concentration of 0.1 mM IPTG, switching the temperature to 18 °C. After 18-24 h the cells were collected by centrifugation at 8000 rpm for 15 min at 4 °C, then frozen at -80 °C if not purified immediately. The cells were resuspended in 15 mL of Ni-NTA equilibration buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4) with 0.1 mM - 1 mM PMSF. They were then sonicated for 6.67 min on ice, with 2 s bursts followed by 4 s rest time (20 min total time per sample) at 85% amplitude. Lysed cells were centrifuged at 12,000 x g at 4 °C for 8 min. Next, the supernatant was loaded onto a spin column with 3 mL Ni-NTA resin and rotated for 30 min at 4 °C. The resin was washed 10x with 6 mL of Ni-NTA wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4), and the protein was eluted 3x with 3 mL Ni-NTA elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4). The fractions were collected, buffer-exchanged into PBS, and concentrated using Amicon Ultra 10 kD MWCO centrifugal concentrators (MilliporeSigma).

Nanobody mutagenesis. Protein gene blocks were ordered from Integrated DNA Technologies, codon-optimized for *E. coli* expression. BsaI cut sites were present at either end of the gene sequence for cloning into the pET28b golden gate entry. This vector enabled green / white screening for colonies successfully transformed with plasmids bearing the inserted gene and provided resistance to kanamycin. The gene block sequences for the nanobodies contained a three amino acid extension at the N-terminus, followed by a His6 tag for purification.

Gene block for AAA-His6 Extension:

1	GGTCTCCATG	GCTGCTGCGC	ACCATCACCA	TCACCATGAG	GTACAGCTGG	TTGAATCTGG
61	AGGGGGGCTC	GTACAGGCGG	GCGGTAGTTT	ACGTCTTTCA	TGCGCCGCAT	CCGGTATTAC
121	GTTTAGCATT	AATACCATGG	GCTGGTATCG	TCAGGCGCCG	GGAAAACAAC	GCGAACTGGT
181	TGCACTGATC	TCATCGATTG	GTGACACGTA	CTATGCCGAT	TCCGTCAAAG	GCCGCTTCAC
241	TATCAGTCGT	GACAACGCCA	AAAATACCGT	GTACCTGCAA	ATGAACTCGT	TGAAGCCAGA
301	AGATACGGCG	GTGTATTATT	GTAAACGCTT	TCGGACCGCA	GCGCAGGGCA	CTGATTACTG
361	GGGTCAGGGC	ACCCAAGTCA	CAGTGAGCAG	CTAAAGAGAC	C	

Gene block for PAA-His6 Extension:

1	GGTCTCCATG	CCGGCGGCTC	ACCATCACCA	CCATCATGAA	GTACAACTGG	TTGAATCAGG
61	CGGTGGTTTA	GTGCAAGCGG	GCGGTTCCCT	GCGTCTCAGT	TGCGCTGCCT	CCGGAATTAC
121	CTTCTCAATC	AATACCATGG	GGTGGTATCG	GCAGGCGCCG	GGCAAACAGC	GTGAGCTGGT
181	AGCGCTTATT	TCTAGTATCG	GGGATACTTA	CTATGCCGAT	TCGGTTAAAG	GACGTTTTAC
241	GATTAGCCGC	GATAATGCCA	AAAACACGGT	CTACTTGCAG	ATGAACAGCC	TGAAGCCAGA
301	AGACACCGCA	GTGTACTATT	GTAAACGCTT	TCGCACCGCA	GCACAGGGCA	CAGACTATTG
361	GGGTCAAGGC	ACTCAGGTCA	CGGTGTCGAG	CTAAAGAGAC	C	

Gene block for AGG-His6 Extension:

1	GGTCTCCATG	GCTGGCGGTC	ATCATCACCA	CCATCACGAG	GTGCAACTGG	TTGAAAGCGG
61	AGGCGGCCTT	GTACAGGCTG	GCGGTTCCCT	GCGTTTATCT	TGTGCGGCGA	GCGGTATTAC
121	CTTCAGTATT	AATACGATGG	GGTGGTATCG	TCAGGCGCCG	GGTAAACAAC	GCGAACTGGT
181	CGCACTCATT	TCGTCCATCG	GAGATACTTA	CTATGCCGAT	AGTGTTAAAG	GTCGCTTTAC
241	AATCTCGCGT	GATAACGCCA	AAAATACGGT	GTACTTGCAG	ATGAACAGCC	TGAAACCAGA
301	AGACACGGCC	GTGTACTATT	GCAAGCGCTT	TCGGACCGCG	GCACAGGGCA	CCGACTATTG
361	GGGGCAGGGC	ACTCAAGTCA	CCGTATCATC	ATAAAGAGAC	C	

Gene block for PGG-His6 Extension:

1	GGTCTCCATG	CCGGGGGGCC	ATCACCACCA	CCATCATGAA	GTGCAACTGG	TAGAATCGGG
61	AGGTGGCTTA	GTCCAGGCGG	GCGGTTCGCT	GCGTCTCAGC	TGCGCTGCAT	CCGGAATTAC
121	GTTCAGTATT	AACACCATGG	GGTGGTATCG	CCAGGCCCCG	GGTAAACAGC	GGGAATTGGT
181	CGCACTGATC	TCCAGCATTG	GTGACACCTA	TTATGCCGAT	AGCGTTAAAG	GTCGTTTTAC
241	TATCTCTCGC	GATAATGCCA	AAAATACCGT	TTACCTTCAA	ATGAACAGTC	TGAAGCCAGA
301	GGACACGGCT	GTGTATTACT	GTAAACGCTT	TCGTACCGCG	GCGCAGGGCA	CAGATTACTG
361	GGGCCAAGGC	ACGCAGGTGA	CTGTATCATC	ATAAAGAGAC	C	

2.5 References

(1) Gunnoo, S. B.; Madder, A. *Org. Biomol. Chem.* **2016**, *14*, 8002–8013.

(2) Li, F.; Mahato, R. I. *Mol. Pharmaceutics* **2017**, *14*, 1321–1324.

(3) Baalmann, M.; Ziegler, M. J.; Werther, P.; Wilhelm, J.; Wombacher, R. *Bioconjugate Chem.* **2019**, *30*, 1405–1414.

(4) Witus, L. S.; Francis, M. B. *Acc. Chem. Res.* **2011**, *44*, 774–783.

(5) Chalker, J. M.; Bernardes, G. J.; Lin, Y. A.; Davis, B. G. *Chem. Asian J.* **2009**, *4*, 630–640.

(6) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782–11783.

(7) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.

(8) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. *PNAS* **2009**, *106*, 3000–3005.

(9) Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew. Chem., Int. Ed.* **2006**, *45*, 5307–5311.

(10) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. *Tetrahedron Lett.* **2000**, *41*, 4069–4073.

(11) Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chem.* **1992**, *3*, 138–146.

(12) Tam, J. P.; Yu, Q.; Miao, Z. *Pept. Sci.* **1999**, *51*, 311–332.

(13) Witus, L. S.; Netirojjanakul, C.; Palla, K. S.; Muehl, E. M.; Weng, C.-H.; Iavarone, A. T.; Francis, M. B. *J. Am. Chem. Soc.* **2013**, *135*, 17223–17229.

(14) Dixon, H. B.; Fields, R. In *Meth. Enzymol.* Elsevier: 1972; Vol. 25, pp 409–419.

(15) Obermeyer, A. C.; Jarman, J. B.; Francis, M. B. *J. Am. Chem. Soc.* **2014**, *136*, 9572–9579.

(16) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. *Nat. Chem. Biol* **2015**, *11*, 326–331.

(17) Cui, L.; Rao, J. *Site-Specific Protein Labeling: Methods and Protocols* **2015**, 81–92.

(18) Liang, G.; Ren, H.; Rao, J. *Nat. Chem* **2010**, *2*, 54–60.

(19) Perrin, C. L. *Acc. Chem. Res.* **1989**, *22*, 268–275.

(20) Panikar, S. S.; Banu, N.; Haramati, J.; del Toro-Arreola, S.; Leal, A. R.; Salas, P. *J Control Release* **2021**, *334*, 389–412.

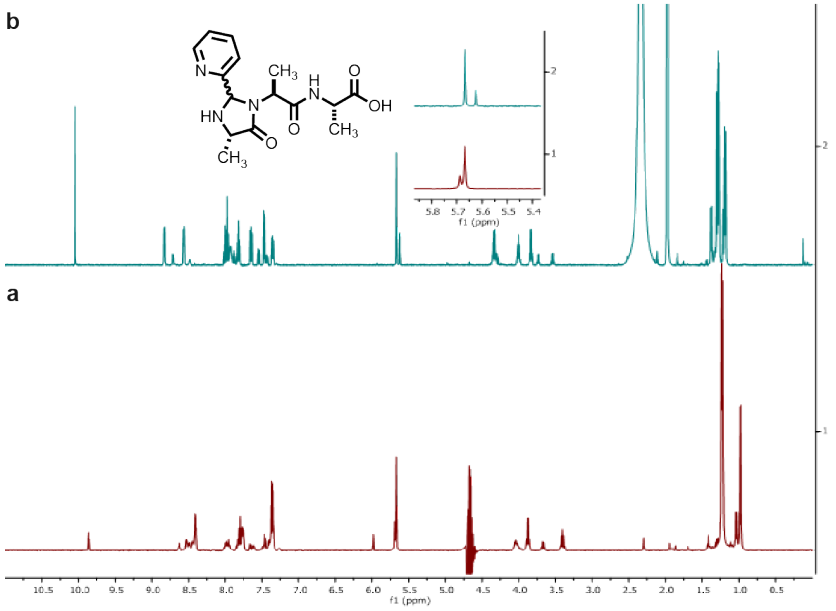
(21) Mardirossian, N.; Head-Gordon, M. *Mol. Phys.* **2017**, *115*, 2315–2372.

(22) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.

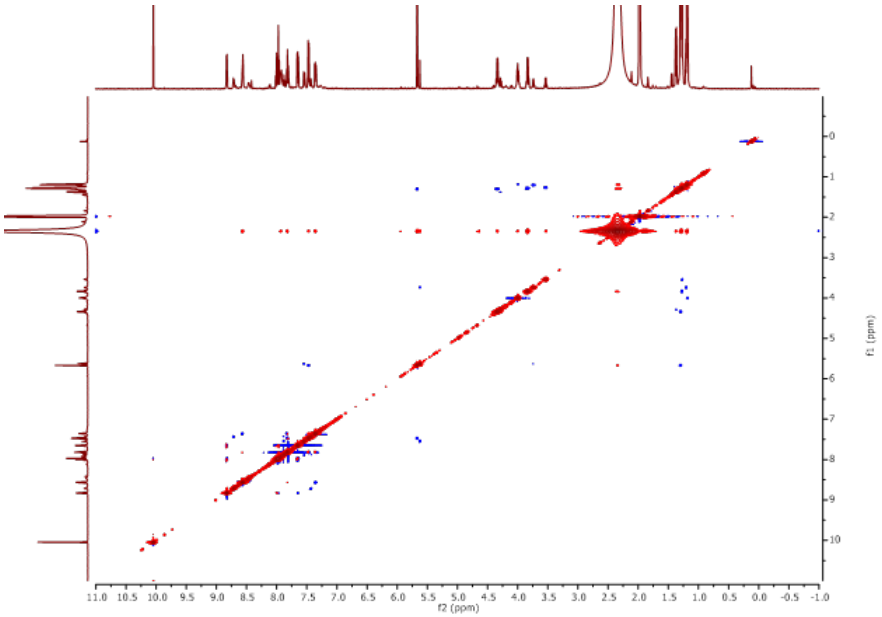
(23) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669–681.

(24) Worden, E. J.; Padovani, C.; Martin, A. *Nat. Struct. Mol* **2014**, *21*, 220–227.

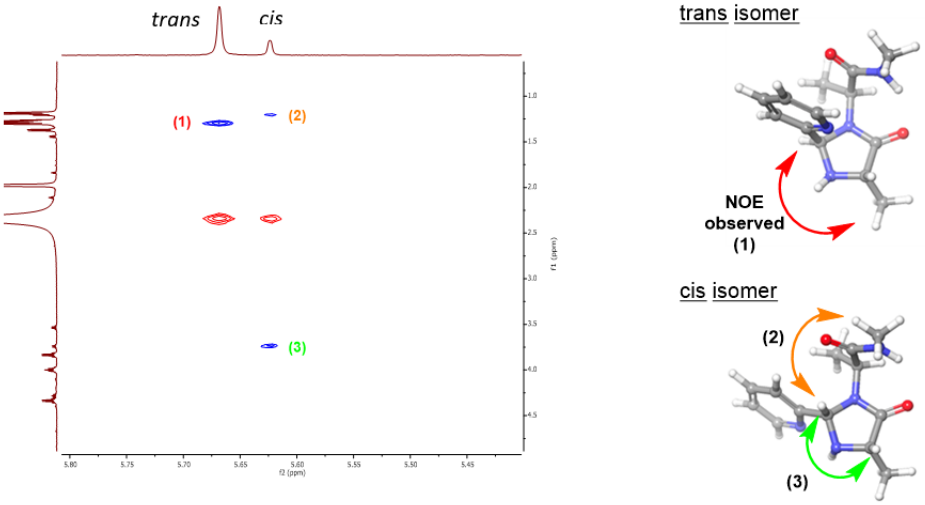
2.6 Supplementary Figures



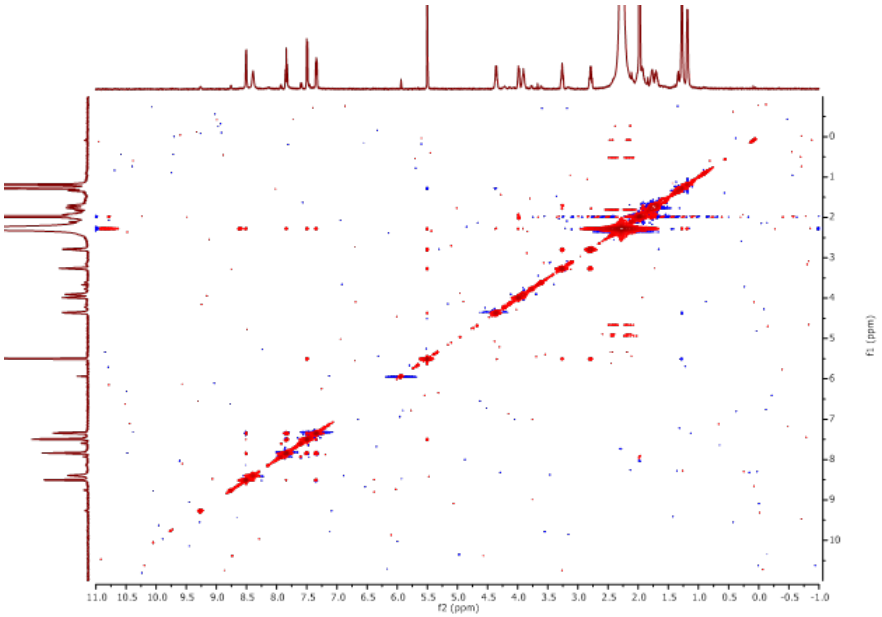
SI Figure 1: ^{1}H-NMR spectra of AAA-PCA conjugate in **(a)** D_2O and **(b)** CD_3CN. AAA (2 mM) and 2PCA (10 mM) in pH 7.5 PB (20 mM) were reacted overnight at 37 °C, then dried under vacuum. The lyophilized sample was redissolved in deuterated solvent and the spectra were acquired on a 500 mHz spectrometer. Spectra in CD_3CN afforded more resolved peaks of the major and minor isomers.



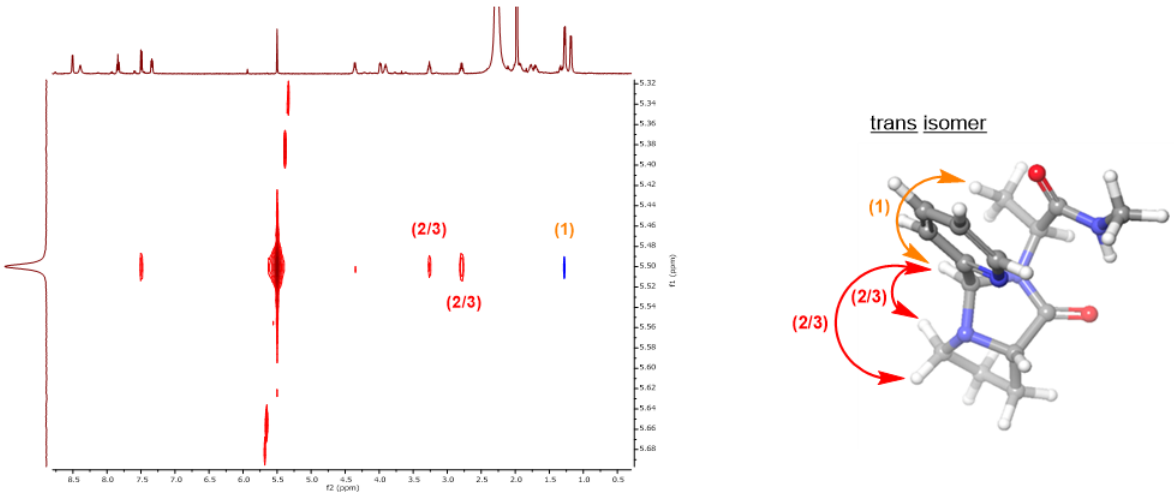
SI Figure 2.1: 2D NOESY spectrum of AAA-PCA conjugate. AAA (2 mM) and 2PCA (10 mM) in pH 7.5 PB (20 mM) were reacted overnight at 37 °C, then dried under vacuum. The lyophilized sample was redissolved in CD_3CN and the spectrum was acquired on a 500 mHz spectrometer.



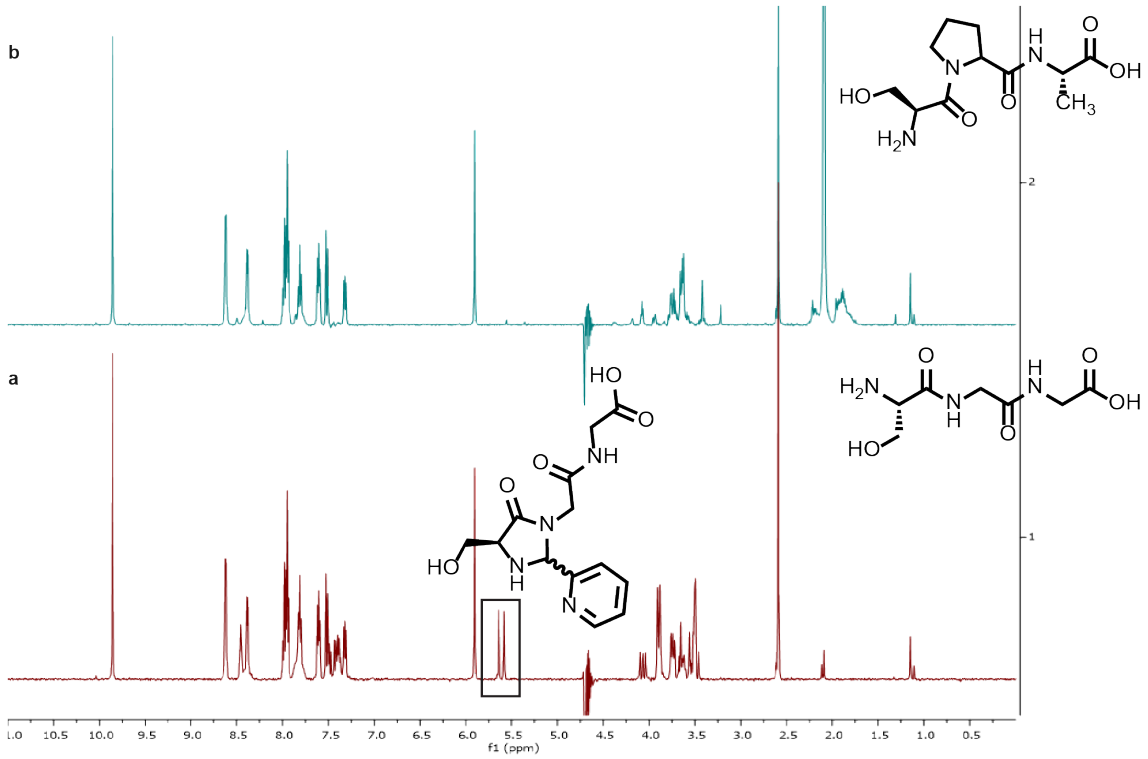
SI Figure 2.2: Enlargement of imidazolidinone peak correlations. NOE correlations suggest that the *trans* isomer is the major product formed.



SI Figure 3.1: 2D NOESY spectrum of PAA-PCA conjugate. PAA (2 mM) and 2PCA (10 mM) in pH 7.5 PB (20 mM) were reacted overnight at 37 °C, then dried under vacuum. The lyophilized sample was redissolved in CD_3CN and the spectrum was acquired on a 500 mHz spectrometer.



SI Figure 3.2: Enlargement of imidazolidinone peak correlations. NOE correlations suggest that the *trans* isomer is the major product formed.



SI Figure 4: ^{1}H-NMR spectra of Ser-X-G\A conjugates with 2PCA. SGG or SPA (2 mM) and 2PCA (10 mM) in pH 7.5 PB (20 mM) were reacted overnight at 37 °C. Hydroxylamine (11 mM) was added and the spectra were acquired on a 500 mHz spectrometer with excitation sculpting. **(a)** With the SGG tripeptide, peaks at 5.56 and 5.63 ppm were observed, which is consistent with the formation of the imidazolidinone products. **(b)** Proline at position two of the SPA tripeptide prevents cyclization to form the imidazolidinone products, and the absence of other peaks in that region suggests that the hydroxymethyl side chain of serine does not react to form the oxazolidinone.

SI Table 1: Rate of product formation at various temperatures. All temperatures are based on a calibration curve using 80% ethylene glycol in DMSO-d_6 as a standard.

AAA with PCA			
T (K)	1/T (K-1)	k (M-1)	ln(k/T)
298.1	0.00335	5.91E-03	-10.8
306.3	0.00326	1.39E-02	-10.0
310.4	0.00322	1.80E-02	-9.8
314.5	0.00318	2.76E-02	-9.3
322.6	0.00310	4.84E-02	-8.8
330.8	0.00302	6.93E-02	-8.5
SAA with PCA			
T (K)	1/T (K-1)	k (M-1)	ln(k/T)
298.2	0.00335	2.54E-03	-11.7
306.4	0.00326	5.95E-03	-10.8
314.6	0.00318	1.01E-02	-10.3
322.7	0.00310	2.04E-02	-9.7
330.9	0.00302	3.64E-02	-9.1
PAA with PCA			
T (K)	1/T (K-1)	k (M-1)	ln(k/T)
298.2	0.00335	6.37E-04	-13.1
306.4	0.00326	1.90E-03	-12.0
314.6	0.00318	5.29E-03	-11.0
322.7	0.00310	1.11E-02	-10.3
330.9	0.00302	2.28E-02	-9.6
AGG with PCA			
T (K)	1/T (K-1)	k (M-1)	ln(k/T)
303.2	0.00330	5.69E-03	-10.9
313.2	0.00319	1.06E-02	-10.3
323.2	0.00310	2.43E-02	-9.5
333.2	0.00300	4.07E-02	-9.0
343.2	0.00291	7.87E-02	-8.4

SI Table 2: Rate of ring opening of the two isomers at various temperatures. All temperatures are based on a calibration curve using 80% ethylene glycol in DMSO-d_6 as a standard. The AGG-2PCA conjugates formed in almost equal amounts and had similar stabilities, so only one is shown. The PAA-2PCA conjugate formed only one observable isomer under experimental conditions.

AAA with 2PCA		Major product		Minor product	
T (K)	1/T (K-1)	k (M-1)	ln(k/T)	k (M-1)	ln(k/T)
306.3	0.00326	1.83E-05	-16.6	6.43E-06	-17.7
314.5	0.00318	4.71E-05	-15.7	1.31E-05	-17.0
322.6	0.00310	1.16E-04	-14.8	2.70E-05	-16.3
330.8	0.00302	2.57E-04	-14.1	6.89E-05	-15.4
339.0	0.00295	5.39E-04	-13.4	1.53E-04	-14.6
347.2	0.00288	1.07E-03	-12.7	4.06E-04	-13.7
AGG with 2PCA		Major product			
T (K)	1/T (K-1)	k (M-1)	ln(k/T)		
323.2	0.00310	3.31E-06	-18.4		
333.2	0.00300	1.90E-05	-16.7		
343.2	0.00291	3.47E-05	-16.1		
353.2	0.00283	1.11E-04	-15.0		
SAA with 2PCA		Major product		Minor product	
T (K)	1/T (K-1)	k (M-1)	ln(k/T)	k (M-1)	ln(k/T)
298.2	0.00335	3.08E-06	-18.4	1.17E-06	-19.4
306.4	0.00326	8.20E-06	-17.4	3.55E-06	-18.3
314.6	0.00318	2.30E-05	-16.4	8.95E-06	-17.4
322.7	0.00310	5.65E-05	-15.6	1.53E-05	-16.9
330.9	0.00302	1.42E-04	-14.7	4.53E-05	-15.8
339.1	0.00295	3.02E-04	-13.9	9.51E-05	-15.1
PAA with 2PCA		Major product			
T (K)	1/T (K-1)	k (M-1)	ln(k/T)		
306.4	0.00326	5.80E-07	-20.1		
314.6	0.00318	1.56E-06	-19.1		
322.7	0.00310	3.35E-06	-18.4		
330.9	0.00302	8.55E-06	-17.5		
339.1	0.00295	1.80E-05	-16.8		
347.3	0.00288	4.41E-05	-15.9		
355.4	0.00281	1.07E-04	-15.0		
363.6	0.00275	2.63E-04	-14.1		

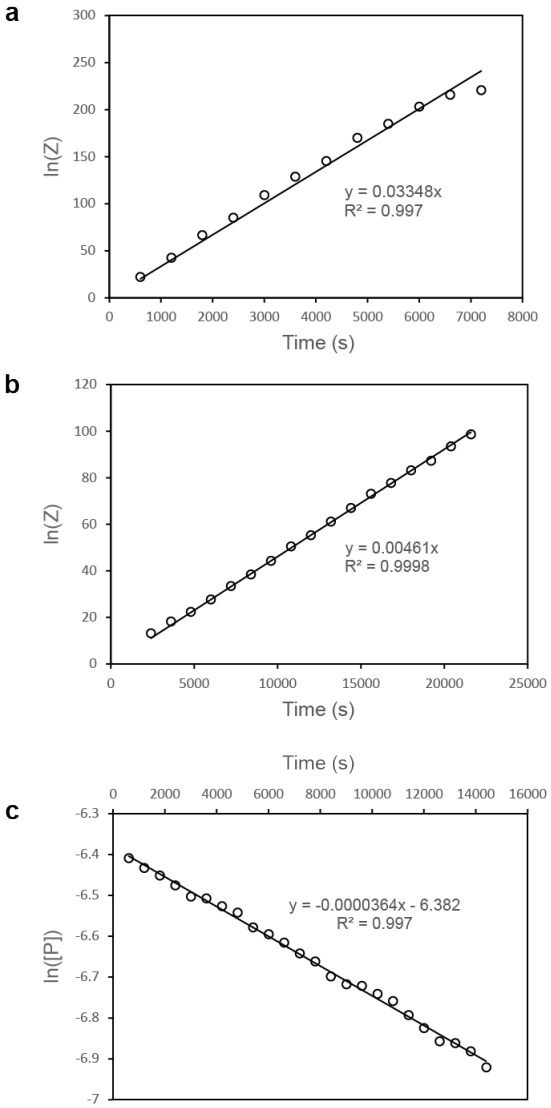
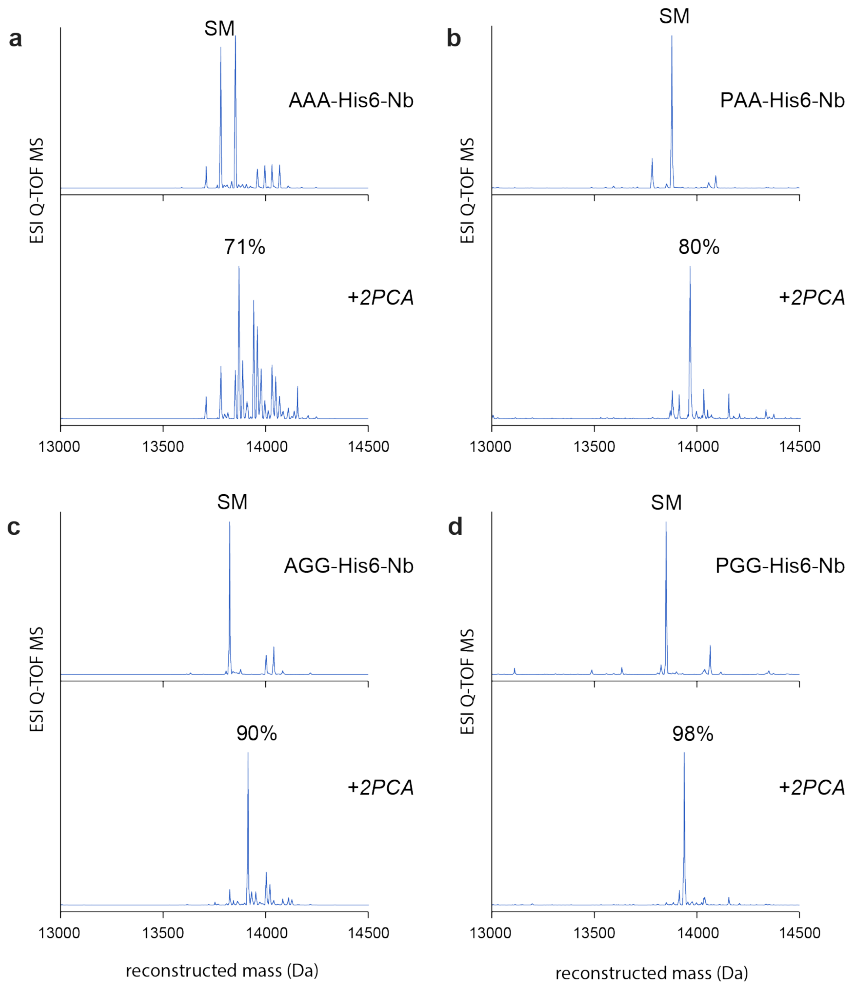


Figure 5: Kinetic isotope experiments with PCA-d. **(a)*** The rate constant for imidazolidinone formation at 41 °C with AAA and PCA-d was found to be 0.0335 $M^{-1}s^{-1}$, which leads to an inverse secondary isotope effect. **(b)*** A rate constant of 0.00461 $M^{-1}s^{-1}$ was found for conjugation of PAA with 2PCA-d at 41 °C, so a normal secondary isotope effect was observed. **(c)** The rate constant for ring opening of the deuterated 2PCA-PAA conjugate at 66 °C was found to be $3.64x10^{-5}$ s^{-1}, which leads to a normal secondary isotope effect. *For these graphs, the y-axis label units are ln(Z) where $Z = \frac{[2PCA][peptide]_o}{[peptide][2PCA]_o}$.



SI Figure 6: Modification of N-terminal anti-HER2 nanobody mutants with 2PCA. The top trace is the deconvoluted mass spectrum of the starting HER2-Nb and the bottom spectrum is after modification with 2PCA (20 mM) overnight. **(a)** AAA-His6-Nb. After expression and purification, a mixture was obtained with the desired protein as well as a higher molecular weight impurity. **(b)** PAA-His6-Nb **(c)** AGG-His6-Nb **(d)** PGG-His6-Nb

Chapter 3

Synthesis of Cyclic Peptides by a Tyrosinase-Mediated Oxidative Cyclization

Abstract

Peptide cyclization improves the stability of the peptide as well as the pharmacokinetic properties for therapeutic development, and many peptide cyclization methods have been developed over the years, such as chemoselective and enzymatic ligations. Chemoselective ligations generally rely on the incorporation of a reactive bioorthogonal functional group for cyclization, which can complicate the synthesis of the linear precursors. Enzymatic ligations are highly selective due to the need for a recognition sequence; however, this can greatly limit the sequence diversity of the peptide. Our approach to the synthesis of cyclic peptides uses the enzyme tyrosinase to oxidatively couple a tyrosine and cysteine residue through an *ortho*-quinone intermediate. In addition to peptide substrates, we also examined this reaction on peptide sequences expressed at the N- and C-terminus of a protein. A variety of peptide lengths and amino acid sequences appended to the termini of GFP were able to be oxidized by tyrosinase and successfully cyclized.

3.1 Introduction

Cyclic peptides have become an attractive class of compounds due to their unique properties compared to linear peptides and have found wide use in a range of applications, including nanomaterials,[1] imaging,[2,3] and most notably as therapeutics.[4,5] One of the advantages to cyclic peptides is a more rigid conformation, which preorganizes and restricts the conformation of the peptide. This both lowers the entropic cost of receptor binding and increases the binding affinity and specificity for its target.[6] With a more defined structure, cyclic peptides have been used to disrupt extended binding sites, such as protein-protein interactions, which are generally characterized as "undruggable" for conventional small molecule therapeutics.[7] Another advantage to cyclic peptides is their increased resistance toward protease degradation, which makes them more metabolically stable.[8]

Given their advantageous properties, developing methods for peptide cyclization has become a growing field of interest. Some of the most common ways include macrolactamization of fully and semi-protected peptides, chemoselective ligations with engineered functional groups,[9] and enzymatic ligations.[10] Macrolactamization of linear peptides with coupling agents generally occurs in a head-to-tail fashion, in which cyclization involves the amine and carboxylate of the N and C-termini, respectively. Other reactive functional groups present in the peptide sequence, such as the amine side chain of lysine, the thiol side chain of cysteine, and the carboxylate side chain of aspartate and glutamate, can also participate in the reaction and are therefore left protected until after cyclization. Although straightforward, this approach generally has low synthetic efficiency due to the high dilution conditions needed to prevent unwanted oligomerization, low solubility of the fully protected peptide, and potential epimerization of the activated C-terminal carboxylate group.

To overcome some of the challenges with direct macrolactamization, many chemoselective ligations have been developed for the synthesis of cyclic peptides. These methods utilize bioorthogonal functional groups incorporated into the peptide sequence for cyclization, alleviating the need to protect most side chain groups that would interfere with macrolactamization. One example is the copper-catalyzed cyclization of a peptide containing an alkyne and azide functional group.[11] While this method generates a nonpeptide backbone linkage, the resulting triazole was found to effectively mimic the native *trans*-amide bond.[12] Additionally, the reaction can take place in water under mild conditions. Other cyclization methods that form nonpeptide backbone linkages include thioether formation between a cysteine thiol and a bromo- or chloroacetyl functional group[13] and hydrazone formation between a hydrazide and an aldehyde.[14] To synthesize cyclic peptides with natural peptide linkages, native chemical ligation can be used,[15] as well as enzymatic approaches that incorporate recognition sequences for proteases such as sortase,[16] peptiligase,[17] and butelase.[18] Although these approaches are highly selective, the need for bioorthogonal functional groups requires incorporation of amino acids with unnatural side chains, which can complicate the synthesis of the linear peptide precursors.

We envisioned a new peptide cyclization method that would utilize a tyrosinase enzyme and only require the incorporation of a tyrosine and cysteine residue within the peptide

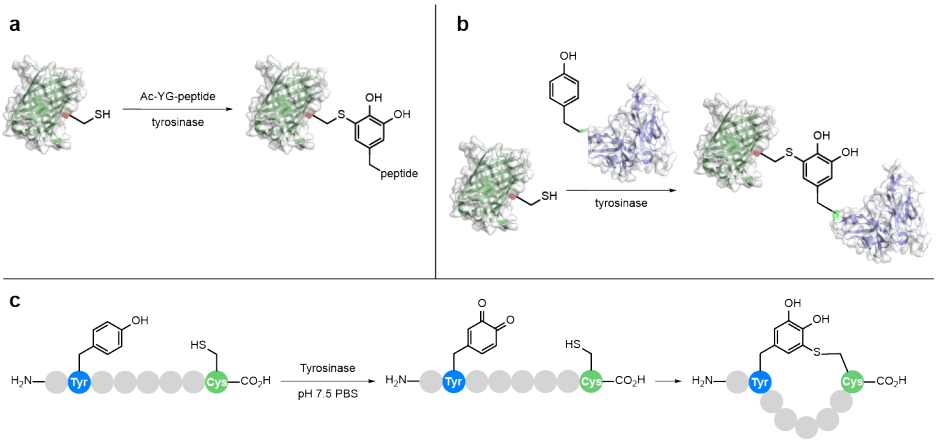


Figure 1: Coupling methods utilizing tyrosinase. **(a)** Tyrosine-tagged peptides can by oxidized by tyrosinase and coupled to proteins containing a solvent-exposed cysteine residue. **(b)** Protein dimers can be synthesized by coupling a cysteine-bearing protein and a protein with an N-terminal tyrosine tag after exposure to tyrosinase. **(c)** Our proposed approach to the synthesis of peptide macrocycles: A peptide containing both a tyrosine and cysteine residue can be exposed to tyrosinase, which would oxidize the tyrosine residue followed by an intracmolular cyclization with the cysteine thiol.

sequence. Tyrosinase is an enzyme involved in the synthesis of melanin and uses molecular oxygen to catalyze the oxidation of tyrosine to a dopaquinone intermediate, which then cyclizes intramolecularly to form dopachrome.[19] Taking advantage of this reactivity, our lab has previously utilized tyrosinase to form peptide-protein conjugates (Figure 1a).[20] In this strategy, a peptide with an acylated, N-terminal tyrosine residue was oxidized with tyrosinase to the activated quinone. This intermediate could then be captured by a cysteine thiol, which was engineered into a solvent-accessible loop of GFP. In addition to peptide-protein conjugates, our lab has also applied this method to the synthesis of protein dimer and trimer complexes (Figure 1b).[21] For this approach, nanoluciferase was engineered to display an N-terminal MYGGS tag, which was oxidized by tyrosinase to the quinone. GFP with an engineered cysteine residue could then be directly coupled to nanoluciferase to form a protein dimer. Following these examples, we wanted to develop a peptide cyclization method in which a tyrosine residue could be oxidized by tyrosinase and then subsequently captured by a cysteine thiol located on the same peptide (Figure 1c). This strategy would allow for greater peptide sequence diversity by alleviating the need for unnatural amino acid incorporation or enzyme recognition sites, and the only oxidant needed for the chemistry is molecular oxygen found in solution, which reduces the risk of side reactions.

3.2 Results and Discussion

3.2.1 Peptide Cyclization

The initial peptide sequence was based on the human angiotensin II peptide and included a tyrosine residue following the N-terminal methionine and a cysteine residue at the C-terminus (MYHIVRDC) (Figure 2a). The reverse analog was also synthesized with an N-terminal cysteine and a tyrosine residue before the C-terminal methionine (CDRVIHYM) (Figure 2b). Exposure to tyrosinase isolated from the common button mushroom *Agaricus bisporus* (abTYR) showed no reactivity with either peptide substrate (Figure 2, middle spectra). Although abTYR has been used by our lab in the past to couple tyrosine residues on peptides and proteins with proteins containing surface exposed cysteine residues, we have also found that small molecule thiols can inhibit the reactivity of abTYR when present in excess. We believe that the smaller size of the peptide substrate, as well as the high local concentration of the thiol after tyrosine binding to tyrosinase, causes the cysteine in the peptide sequence to inhibit this variant of tyrosinase. We then attempted to oxidize the tyrosine residue with tyrosinase isolated from *Bacillus megaterium* (megaTYR), and complete conversion of the peptide to a single oxidized product was observed by mass spectrometry after 1 h at room temperature (Figure 2, bottom spectra). When the cysteine residue is positioned at the C-terminus, the cyclized product is formed with the expected mass (m/z = 1050); however,

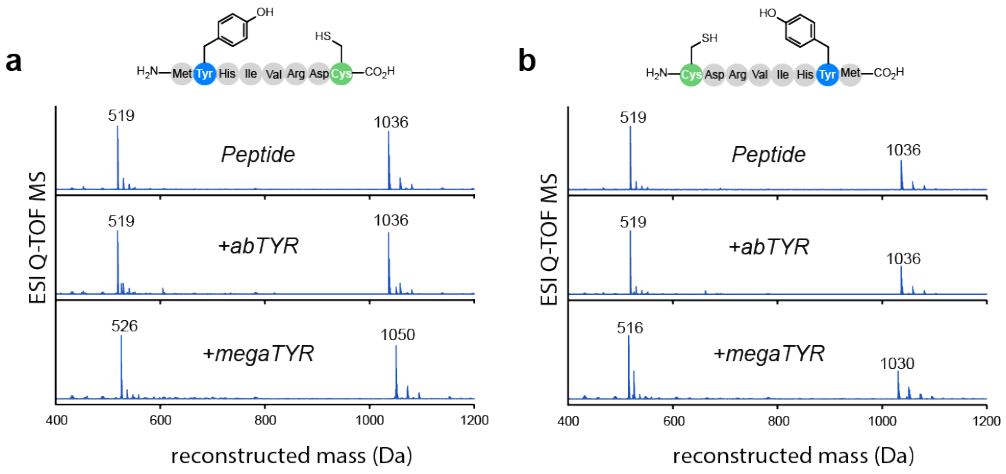


Figure 2: Oxidation of peptide substrates with tyrosinase enzymes. **(a)** Peptide sequence with the tyrosine residue following the N-terminal methionine and the cysteine residue at the C-terminus. **(b)** Peptide sequence with the cysteine residue at the N-terminus and the tyrosine residue before the C-terminal methionine. *(top)* Mass spectra of the unmodified peptides. *(middle)* No oxidation occurred when the peptides were exposed to abTYR. *(bottom)* Complete conversion was observed when the peptides were treated with megaTYR. Reactions were run for 1 h at room temperature.

when the cysteine residue is at the N-terminus, a mass of 1030 Da is observed. This occurs because after phenol oxidation and thiol addition, the free amine of the N-terminus can also be added into the quinone in a 1,2-addition. The enhanced reactivity of the megaTYR

enzyme is most likely due to its smaller active site pocket compared to the abTYR variant, which prevents unwanted thiol binding and inhibition of the enzyme. Finally, although mass spectrometry can provide evidence for tyrosine oxidation by tyrosinase to the activated quinone, there is no differentiation between this and the product after thiol addition since the masses are identical. To discern which product was formed, we then confirmed the cyclized structure of the peptide by ^{1}H-NMR. The starting peptide has two pairs of aromatic protons from the tyrosine residue, and after cyclization only two proton peaks with a small coupling are observed (see SI Figure 1). This showed that the thiol did indeed add into the quinone in a 1,6-addition, and that the product then reduced to the catechol in solution.

3.2.2 C-Terminal Oxidative Cyclization

After showing the utility of this method to synthesis cyclic peptides, we then wanted to turn our attention toward cyclizing peptide sequences expressed at the N- or C-terminus of a protein, since protein-peptide conjugates have been synthesized in the past to improve protein targeting and delivery.[22] Cell-penetrating peptides (CPPs) attached to proteins allow for the complex to translocate across the cell membrane, delivering the protein inside the cell; however, linear CPPs can suffer from poor stability and overall suboptimal cell penetration. To overcome these challenges, cyclic CPPs have been developed that provide superior intracellular delivery and improved physicochemical characteristics.[23] We envisioned that using a tyrosinase-mediated oxidative cyclization on peptides expressed directly on proteins would provide a new method to access peptide-tagged proteins for targeting and delivery. This approach introduces an additional challenge of targeting only the desired tyrosine residue within the peptide sequence over any other tyrosine residues on the protein; however, previous work in our group has shown that megaTYR is selective for tyrosine residues engineered at the N- or C-terminus over other surface-exposed tyrosine residues.

To explore this approach, we began by engineering the C-terminus of superfolder GFP to include the peptide sequence from above (CDRVIHY). The construct consisted of GFP, followed by a His6 tag for purification, a TEV recognition sequence, and the linear peptide sequence. The TEV recognition sequence was included to allow for the potential isolation of the cyclic peptide after cyclization was carried out on the protein substrate. Using the initial oxidation conditions developed for the peptide substrates, a mass difference of 28 was observed, indicating that two oxidation events took place (see SI Figure 2). Taking a time course of the reaction, it was found that even after just 5 min all the starting material was oxidized, and a mixture of products with one and two oxidations was observed (58% and 42%, respectively). After 30 min, the product with two oxidations was predominantly formed (83%). The reaction took place much faster than expected, so for greater selectivity, we cooled the reaction to 4 °C. After 30 min at 4 °C, 41% of the singly oxidized product was formed (Figure 3a). After 90 min, almost all the starting protein was consumed, and the desired product was obtained in 82% yield, with only 16% of the doubly oxidized product observed. To verify that the cyclization occurred, 5,5'-dithiobis(2-nitrobenzoic acid) (Ellman's reagent) was used to confirm the presence or absence of a free thiol (see SI Figure 3a). When the starting

protein was exposed to Ellman's reagent, a change in mass was observed corresponding to the formation of a disulfide with the cysteine thiol side chain. When the protein was first exposed to megaTYR and then treated with Ellman's reagent, no disulfide formation was observed, indicating that the thiol must have been added into the quinone intermediate.

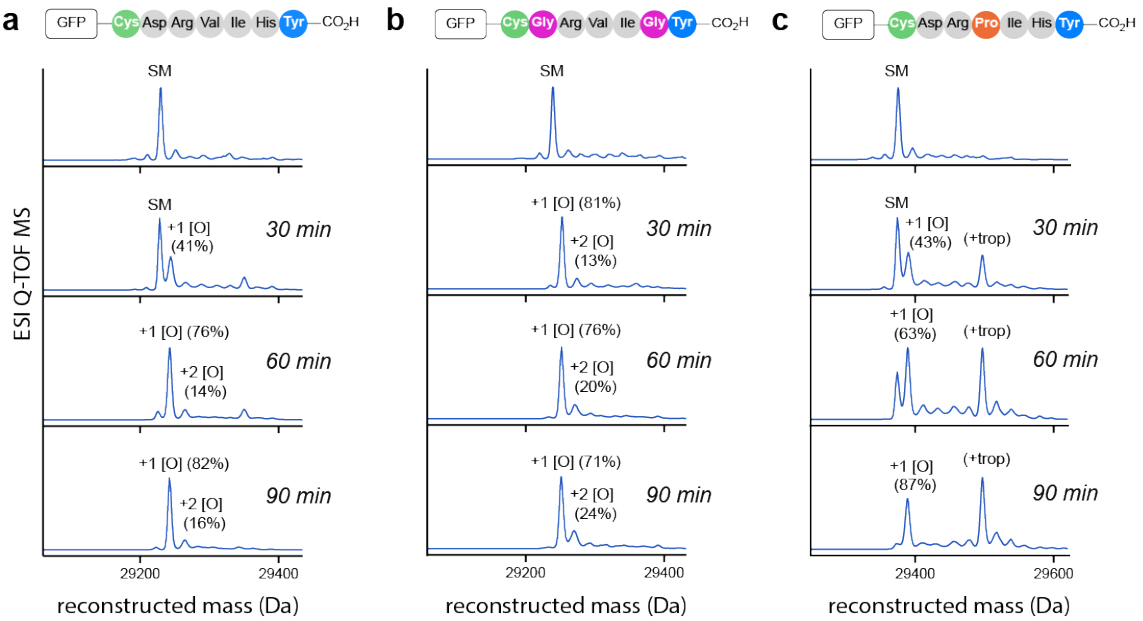


Figure 3: Oxidative cyclization of peptide sequences at the C-terminus of GFP. GFP was expressed with a C-terminal His6 tag and a TEV recognition sequence, followed by the peptide sequence of interest. **(a)** The model sequence showed complete conversion after 60 min. **(b)** Two glycine substitutions were made at the sites adjacent to the tyrosine and cysteine residues, which reduced the reaction time to 30 min. **(c)** A proline substitution was made in the middle of the peptide sequence, which diplayed a slower reactivity. The addition of tropolone, used the quench the reaction, can be seen here. Experiments have shown that this only occurs after the tyosine residue is oxidized.

With greater control of the reaction, we then wanted to explore the effects on cyclization with different amino acid substitutions within the peptide sequence. One mutant was designed and expressed with glycine substitutions adjacent to the cysteine and tyrosine residues. The purpose of these substitutions was to enable greater flexibility within the peptide backbone as well as decrease the steric hindrance around the reacting amino acid side chains. We also included a mutant with a proline residue in the middle of the sequence to induce a turn within the peptide. After exposure to tyrosinase, the mutant with the glycine substitutions was almost fully converted in 30 min, with 81% and 13% of the singly and doubly oxidized products, respectively, observed (Figure 3b). Under the same conditions, only 43% of the product was obtained with the proline substitution, with 63% and 87% obtained after 60 and 90 min, respectively (Figure 3c). Interestingly, we thought that introducing a turn in the peptide sequence with the proline substitution would bring the two reactive centers closer

together and increase the rate of cyclization; however, formation of the cyclized peptide with a proline residue was the slowest. This is mostly likely because although the thiol and phenol were closer together for cyclization, the initial oxidation of the tyrosine by tyrosinase was slower due to increased steric bulk.

3.2.3 N-Terminal Oxidative Cyclization

Next, we wanted to see if this oxidative cyclization method could be applied to peptide sequences expressed at the N-terminus of a protein, since this construct more closely resembles the peptide display complexes used in library screening, such as mRNA display. We began by mutating GFP to include an N-terminal peptide sequence (MYHIVRDC), again followed by a TEV recognition sequence and a His6 tag. We also wanted to screen different peptide lengths, so three additional mutants were expressed with one, two, and three serine-alanine extensions between the arginine and aspartate residues. This in turn would provide cyclic peptides with a ring size of 7, 9, 11, and 13 amino acids, respectively. After cloning and expression, the four mutants were isolated in high purity; however, under the expression conditions, between 17-24% of each construct had the methionine residue cleaved, leaving an N-terminal tyrosine (see SI Figure 4). Although these mutants could also be oxidized by tyrosinase, it has previously been shown that when tyrosine is the N-terminal amino acid, the free amine of the N-terminus can cyclize with the quinone instead. Since the samples consisted primarily of the desired protein constructs, we carried forward with the mixture for initial screening.

When the parent construct was exposed to tyrosinase under the reaction conditions optimized for the C-terminally tagged proteins (30 min at 4 °C), no reaction was observed; however, complete conversion could be obtained when the temperature was increased to room temperature for 2 h (Figure 4a). This slower rate of reaction is most likely due to the increased steric bulk of the methionine and histidine residues flanking the tyrosine residue, compared to the more standard N-terminal GYGG sequence our lab has used in the past for other applications. Although the GYGG tag provides a faster reaction, we chose a more difficult substrate in the hopes of developing a method that would be less dependent on the identity of the adjacent amino acids. This in turn could lead to more variety in the sequences of the peptides used in a library. To confirm that the cyclized product was formed after oxidation, Ellman's reagent was added, and similar to the C-terminally tagged protein, no shift was observed indicating the absence of a free thiol (see SI Figure 3b). When the mutants containing one and two serine-alanine extensions were exposed to tyrosinase at room temperature, excellent conversions were obtained after just 1 h (Figure 4b and c). The extensions increase the distance of the tyrosine residue from the body of the protein and make it more solvent accessible, and this most likely leads to a decrease in reaction time compared to the parent construct. Exposure of the mutant with three serine-alanine extensions to megaTYR under the same reaction conditions yielded 78% of the desired product (Figure 4d).

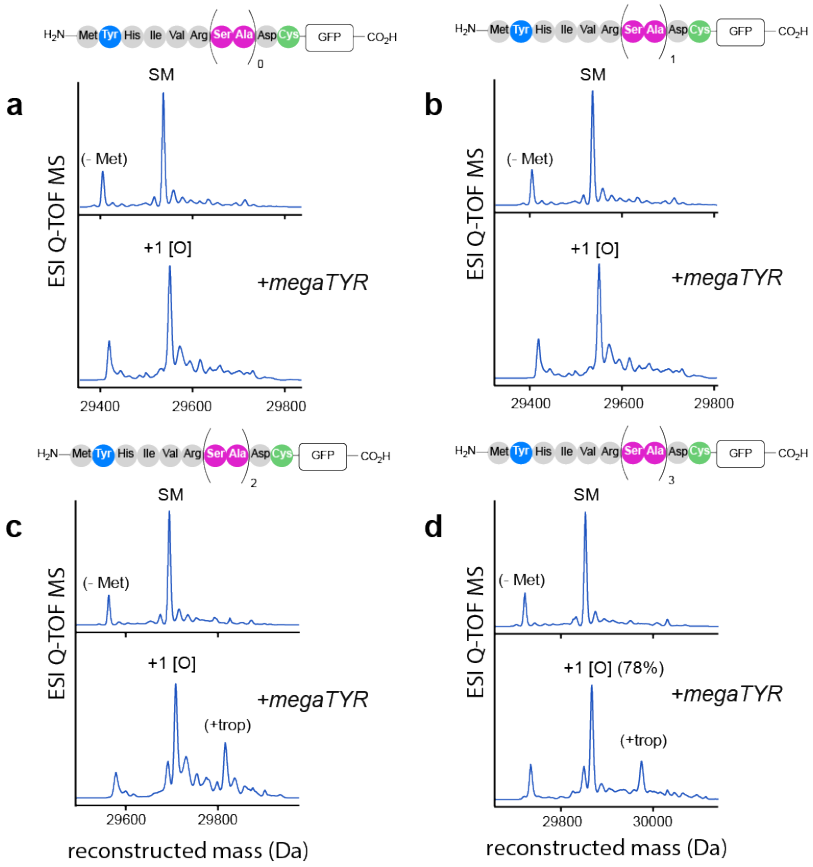


Figure 4: Oxidative cyclization of peptide sequences at the N-terminus of GFP. GFP was expressed with an N-terminal peptide sequence, followed by a TEV recognition sequence and a His6 tag. After expression, each sample was isolated as a mixture of the desired protein and a protein cleaved after the methionine residue, leaving an N-terminal tyrosine. Complete conversion to the oxidized product was observed after exposure to megaTYR with **(a)*** the model sequence, **(b)**$^{+}$ one serine-alanine insertion, and **(c)**$^{+}$ two serine-alanine insertions. **(d)**$^{+}$ After the mutant with three serine-alanine insertions was treated with megaTYR, the oxidized product was observed in 78% yield. The cleaved protein was also fully oxidized for all samples. *The reaction was conducted at room temperature for 2 h. $^{+}$The reaction was conducted at room temperature for 1 h.

With the optimized conditions for oxidative cyclization of N-terminal peptide sequences in hand, we then wanted to examine the substrate scope of the reaction with various other amino acids next to the tyrosine residue. From this we could gain insight into how the sequence and local environment affect tyrosinase activity. To begin, the methionine residue in the original sequence (MYHIVRDC) was changed to a glycine residue, which provided a single product after expression and purification (Figure 5b). We then designed a series of point mutations at the third amino acid position (GY-X-IVRDC) with different charges and polarities (Figure 5a). For charged amino acids, we included an arginine residue in addition to the parent histidine sequence, as well as a glutamate residue for negative charge. For polar residues we included sequences with serine and glutamine, and for hydrophobic residues we included

leucine and phenylalanine. Cloning of the constructs was successful for all but the glutamine mutant, with matching DNA sequences to the designed vector sequences. Of the mutants that had matching sequences, all but the arginine mutant were then successfully expressed and purified (Figure 5b). The serine, glutamate, and leucine mutants were isolated in high purity, but the phenylalanine mutant contained a higher molecular weight impurity that was consistent across multiple expression batches. Various attempts were made to optimize the expression of the arginine mutant with different media, temperatures, and expression volumes, but none provided the desired protein after purification. Preliminary attempts to oxidize the Gly-Tyr-X sequences are shown in SI Figure 4. All mutants showed reactivity

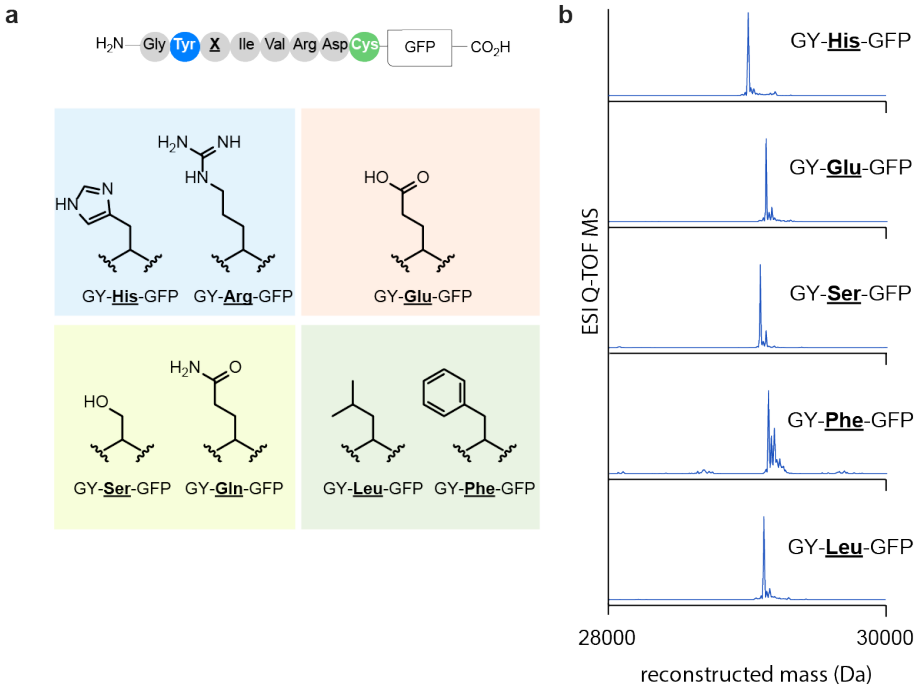


Figure 5: Panel of GFP mutants with amino acid substitutions after the tyrosine residue. **(a)** Various amino acids were substituted next to the tyrosine residue to examine the tolerance of tyrosinase activity to different local environments, including positively and negatively charged side chains, polar side chains, and hydrophobic side chains **(b)** Of the panel of proteins, these five were successfully cloned and expressed. The GY-F mutant was isolated as a mixture of the desired protein with a higher molecular weight protein impurity.

with tyrosinase to the desired oxidized product. After 30 min at room temperature, the serine and glutamate mutants were completely converted, with 35% and 0% of the secondary oxidation observed, respectively, while the histidine mutant required 60 min and had 34% of the +2 oxidation. The Leu mutant displayed a much slower reactivity. After 90 min, 81% of the oxidized product was observed, but no secondary oxidation occurred. With the phenylalanine mutant, we had thought that the large steric bulk might inhibit tyrosinase activity, but surprisingly, complete oxidization was observed after 60 min; however, most of the product mixture was the doubly oxidized product.

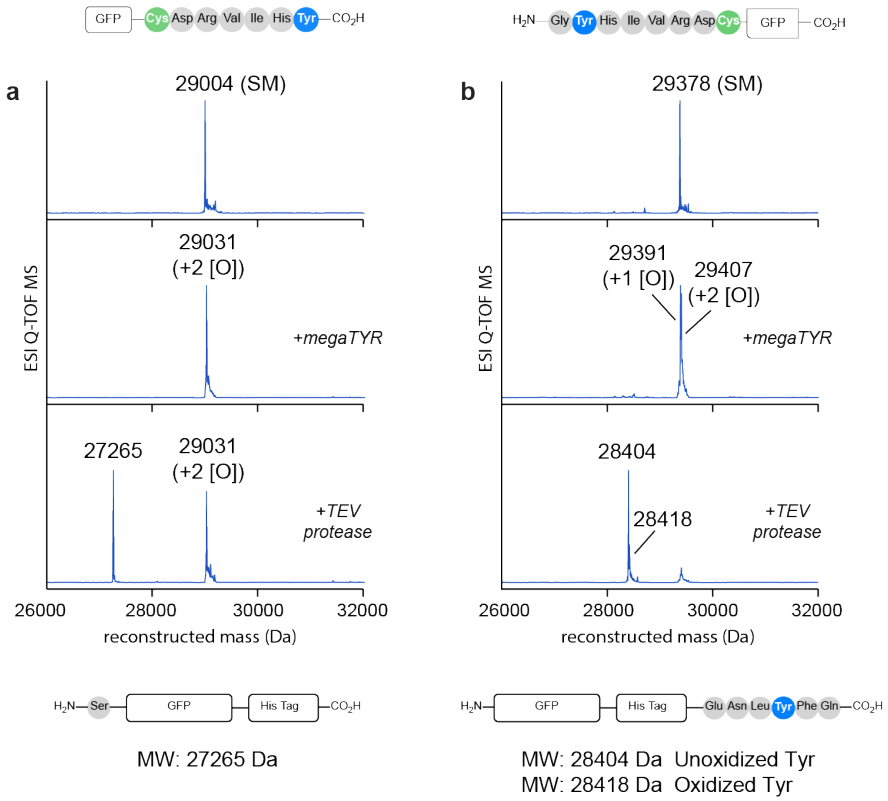


Figure 6: Determination of the secondary oxidation site. **(a)** GFP with C-terminally tagged tyrosine residue. **(b)** GFP with N-terminally tagged Gly-Tyr residues. *(top)* Mass spectra of the unmodified peptides. *(middle)* Long exposure to megaTYR with the C-terminally tagged GFP shows complete conversion to a product with two oxidations. Under similar condictions, the N-terminally tagged GFP provides an almost even mixture of two products with one and two oxidations. *(bottom)* After exposure to TEV protease, the cleaved protein fragment from the C-terminally tagged GFP displays a mass corresponding to unmodified protein, suggesting the two oxidations occured on the peptide fragment. The N-terminally tagged GFP provided a mixture of protein fragments after treatment with TEV protease, corresponding to an unmodified fragment (28404 Da) and a fragment with one oxidation (28418 Da).

Finally, although excellent selectivity is observed when the reaction time is monitored, we wanted to determine where the second oxidation occurs within the sequence when the reaction time is lengthened. One possible site is the tyrosine residue that is introduced within the TEV recognition sequence. To test this, we first exposed the N-terminally tagged protein to tyrosinase and let the reaction run until all the starting material was converted to a single product with two oxidations (Figure 6a, middle spectrum). The peptide was then cleaved with TEV protease, and the resulting protein mass was 27265 Da (Figure 6a, bottom spectrum), which matches the expected mass of unoxidized GFP. This provides evidence that the two oxidation sites are located on the cleaved peptide sequence and most likely occurred at the N-terminal tyrosine and the tyrosine within the TEV recognition sequence. We also conducted a similar experiment with the C-terminally tagged protein. Extended

reaction time with tyrosinase provided a mixture of products with one and two oxidations (Figure 6b, middle spectrum). After cleavage with TEV protease, a mixture of products was also obtained: one with no oxidation (28404 Da) and another with one oxidation (28418 Da) (Figure 6b, bottom spectrum). This is because the TEV protease cleaves between the glutamine and serine residues located after the tyrosine residue, and so the tyrosine within the recognition sequence remains appended to GFP. When this tyrosine was oxidized, the oxidized mass was also observed with GFP after cleavage.

3.3 Conclusion

In conclusion, we have developed a reaction to synthesize cyclic peptides using a tyrosinase-mediated oxidative cyclization of sequences containing a tyrosine and cysteine residue. The reaction can be performed on peptide substrates, as well as peptide sequences appended to the N- and C-termini of proteins. The reaction is performed under mild conditions, and great selectivity can be achieved; however, certain challenges remain, which will be addressed in the future. The first is eliminating the secondary oxidation that can occur over time. Since we have evidence that the site of oxidation is the tyrosine residue within the TEV recognition sequence, we will express new constructs with this sequence taken out. Although this no longer allows for isolation of the cyclized peptide after oxidation, the benefit is that the product would only contain a single site of modification. The second challenge to address is the overall tolerability of the reaction with different peptide sequences. One approach to this is to mutate the active site of tyrosinase to explore how the local charge and active site size affects the binding of tyrosine residues within various sequences. Overall, this oxidative cyclization strategy streamlines the synthesis of cyclic peptides, requiring no exogenous coupling reagent or the use of protecting groups for the chemistry to take place. After further optimization, we hope to apply this to the generation of cyclic peptide libraries and the development of library display technologies.

3.4 Supporting Information

3.4.1 Reagents and Materials

All reagents were obtained from commercial sources without further purification. Tyrosinase isolated from *Agaricus bisporus* (abTYR, both 25 kU [SKU = T3824- 25KU] and 50 kU [SKU = T3824-50KU]) were purchased from Sigma-Aldrich (St. Louis, MO). TCEP and Dulbecco's Phosphate Buffered Saline (no CaCl or NaCl) (DPBS) were purchased from Sigma. PBS 10X was purchased from Corning (Corning, NY). Milli-Q H_2O was obtained from a Millipore purification system. Spin concentrators with 10 and 100 kDa molecular weight cutoffs (MWCO) and sterile spin filters with 0.22 µm pores were purchased from Millipore (Billerica, MA).

3.4.2 Instrumentation Methods

Liquid chromatography mass spectrometry (LC-MS) analysis. Acetonitrile (Optima grade, 99.9%, Fisher,Waltham, MA), formic acid (1 mL ampules, 99+%, Pierce, Rockford, IL), 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 for LCMS. Electrospray ionization mass spectrometry (ESI-MS) of protein bioconjugates was performed using an Agilent 1260 Infinity II liquid chromatograph outfitted with an Agilent 6530 quadrupole time-of-flight (Q-TOF) LC-MS system (Santa Clara, CA). The LC was equipped with a Proswift RP-4H (monolithic phenyl, 1.0 mm × 50 mm, Dionex) 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 15 to 30 picomoles of analyte were injected onto the column. Following sample injection, a 5-100% B elution gradient was run at a flow rate of 0.40 mL/min over 2 min at 55 °C. Data were collected and analyzed by deconvolution of the entire elution profile to provide reconsctructed mass spectra that are representative of the entire sample using Agilent MassHunter BioConfirm Software 10.0. Percent modification was determined through integration of MS peaks using opensource Chartograph software (www.chartograph.com). The integration of the completely unmodified protein peak served as an internal standard in determining the percent modification.

UV-Vis measurements. A NanoDrop 1000 (Thermo Fisher Scientific) was used to quantify the samples in this work based on absorbance values at 280 nm.

HPLC purification. Semiprep HPLC purification of synthesized peptides was performed on Agilent 1260 Infinity series equipped with UV-Vis detector and fraction collector. Solvent A was Milli-Q H_2O with 0.1% trifluoroacetic acid (v/v) and solvent B was acetonitrile with 0.1% trifluoroacetic acid (v/v).

Nuclear Magnetic Resonance (NMR) Spectroscopy. ^{1}H spectra were measured with a Burker AV-500 (500 MHz, 150 MHz) spectrometer. ^{1}H NMR chemical shifts are reported as δ in units of parts per million (ppm) relative to residual H_2O (δ 4.72, singlet).

3.4.3 Experimental Procedures

General tyrosinase oxidation. To a 20 µM solution of protein or 100µM solution of peptide in PBS (pH = 7.2) was added a 5 mM solution of TCEP to a final concentration of 500 µM. The mixture was allowed to sit for 5 min, then a 2 µM solution of tyrosinase (abTYR or megaTYR) in 50 mM phosphate buffer (pH = 7.2) was added to reach a final enzyme concentration of 400 nM. After the reaction proceeded, it was quenched with the addition of 1 mM tropolone. To test for the presence of a free cysteine side chain thiol after oxidation, a 5 mM solution of DTNB was added to a final concentration of 500 µM and was allowed to sit for 30 min. The resulting samples were analyzed using ESI-QTOF MS.

TEV protease cleavage. To a solution of modified GFP in PBS (pH = 7.2) was added a solution of TEV protease (2 mg/mL) to a final concentration of 1:100 w/w. The mixture was stirred overnight on a rotator at 4 °C and then analyzed using ESI-QTOF MS.

GFP expression and purification. Sequenced plasmids were transformed into BL21(DE3)Star competent cells via a 45 s, 42 °C heat shock and plated on LB/kanamycin. Resulting colonies were used to inoculate 5 mL LB/kanamycin cultures (50 µg/mL) and grown overnight at 37 °C, 225 RPM. The culture was then added to 500 mL of LB media and grown at 37 °C to an OD600 of 0.6- 0.8. Expression was induced with a final concentration of 0.1 mM IPTG, switching the temperature to 18 °C. After 18-24 h the cells were collected by centrifugation at 8000 rpm for 15 min at 4 °C, then frozen at -80 °C if not purified immediately. The cells were resuspended in 15 mL of Ni-NTA equilibration buffer (20 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4) with 0.1 mM - 1 mM PMSF. They were then sonicated for 6.67 min on ice, with 2 s bursts followed by 4 s rest time (20 min total time per sample) at 85% amplitude. Lysed cells were centrifuged at 12,000 x g at 4 °C for 8 min. Next, the supernatant was loaded onto a spin column with 3 mL Ni-NTA resin and rotated for 30 min at 4 °C. The resin was washed 10x with 6 mL of Ni-NTA wash buffer (20 mM sodium phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4), and the protein was eluted 3x with 3 mL Ni-NTA elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4). The fractions were collected, buffer-exchanged into PBS, and concentrated using Amicon Ultra 10 kD MWCO centrifugal concentrators (MilliporeSigma).

GFP mutagenesis. Protein gene blocks were ordered from Integrated DNA Technologies, codon-optimized for *E. coli* expression. BsaI cut sites were present at either end of the gene sequence for cloning into the pET28b golden gate entry. This vector enabled green / white

screening for colonies successfully transformed with plasmids bearing the inserted gene and provided resistance to kanamycin. The gene block sequences for GFP contained either the peptide sequence extension at the N terminus, followed by a His6 tag for purification, a TEV cleavage sequence, and the GFP gene, or the GFP gene, followed by a His6 tag, TEV cleavage sequence, and peptide sequence extension. Mutated plasmids were transformed into XL1-Blu competent *E. coli* cells, cultured overnight, and subsequently miniprepped to verify incorporation of the amino acids by sequencing.

Gene block for superfolder GFP:

1	ATGCGTAAAG	GCGAAGAGCT	GTTCACTGGT	GTCGTCCCTA	TTCTGGTGGA	ACTGGATGGT
61	GATGTCAACG	GTCATAAGTT	TTCCGTGCGT	GGCGAGGGTG	AAGGTGACGC	AACTAATGGT
121	AAACTGACGC	TGAAGTTCAT	CTGTACTACT	GGTAAACTGC	CGGTACCTTG	GCCGACTCTG
181	GTAACGACGC	TGACTTATGG	TGTTCAGTGC	TTTGCTCGTT	ATCCGGACCA	TATGAAGCAG
241	CATGACTTCT	TCAAGTCCGC	CATGCCGGAA	GGCTATGTGC	AGGAACGCAC	GATTTCCTTT
301	AAGGATGACG	GCACGTACAA	AACGCGTGCG	GAAGTGAAAT	TTGAAGGCGA	TACCCTGGTA
361	AACCGCATTG	AGCTGAAAGG	CATTGACTTT	AAAGAAGACG	GCAATATCCTG	GGCCATAAGC
421	TGGAATACAA	TTTTAACAGC	CACAATGTTT	ACATCACCGC	CGATAAACAA	AAAAATGGCA
481	TTAAAGCGAA	TTTTAAAATT	CGCCACAACG	TGGAGGATGG	ATCTGTGCAG	CTGGCTGATC
541	ACTACCAGCA	AAACACTCCA	ATCGGTGATG	GTCCTGTTCT	GCTGCCAGAC	AATCACTATC
601	TGAGCACGCA	AAGCGTTCTG	TCTAAAGATC	CGAACGAGAA	ACGCGATCAT	ATGGTTCTGC
661	TGGAGTTCGT	AACCGCAGCG	GGCATCACGC	ATGGTATGGA	TGAACTGTAC	AAA

Corresponding to amino acid sequence:

1	MRKGEELFTG	VVPILVELDG	DVNGHKFSVR	GEGEGDATNG	KLTLKFICTT	GKLPVPWPTL
61	VTTLTYGVQC	FARYPDHMKQ	HDFFKSAMPE	GYVQERTISF	KDDGTYKTRA	EVKFEGDTLV
121	NRIELKGIDF	KEDGNILGHK	LEYNFNSHNV	YITADKQKNG	IKANFKIRHN	VEDGSVQLAD
181	HYQQNTPIGD	GPVLLPDNHY	LSTQSVLSKD	PNEKRDHMVL	LEFVTAAGIT	HGMDELYK

N-terminal extensions. Below are listed the amino acid sequences of N-terminal extensions, followed by the corresponding DNA sequences added to the GFP gene block. Each DNA sequence encodes for either (1) the peptide sequence of interest, a TEV recognition sequence, and a His6 tag or (2) the peptide sequence and a TEV recognition sequence with the His6 tag at the C-terminus.

MY-$(SA)_0$ extension (MYHIVRDC-)[1]

ATGTATCATA TTGTGCGCGA TTGCGAAAAC CTATATTTTC AGAGTCATCA CCACCATCAT CAC

MY-$(SA)_1$ extension (MYHIVRSADC-)[1]

ATGTATCATA TTGTGCGCTC TGCCGATTGC GAAAACCTAT ATTTTCAGAG TCATCACCAC CATCATCAC

MY-$(SA)_2$ extension (MYHIVRSASADC-)[1]

ATGTATCATA TTGTGCGCTC TGCGTCCGCT GATTGCGAAA ACCTATATTT TCAGAGTCAT CACCACCATC ATCAC

MY-$(SA)_3$ extension (MYHIVRSASASADC-)[1]

ATGTATCATA TTGTGCGCTC TGCGTCCGCT TCAGCTGATT GCGAAAACCT ATATTTTCAG AGTCATCACC ACCATCATCA C

GY-H extension (GYHIVRDC-)[2]

ATGGGTTATC ATATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

GY-R extension (GYRIVRDC-)[2]

ATGGGTTATC GTATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

GY-E extension (GYEIVRDC-)[2]

ATGGGTTATG AAATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

GY-S extension (GYSIVRDC-)[2]

ATGGGTTATT CTATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

GY-Q extension (GYQIVRDC-)[2]

ATGGGTTATC AAATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

GY-F extension (GYFIVRDC-)[2]

ATGGGTTATT TTATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

GY-L extension (GYLIVRDC-)[2]

ATGGGTTATT TAATTGTGCG CGATTGCGAA AACCTATATT TTCAGAGT

C-terminal extensions. Below are the amino acid sequences of C-terminal extensions, followed by the corresponding DNA sequences added to the GFP gene block. Each DNA sequence encodes for a His6 tag, a TEV recognition sequence, and the peptide sequence of interest.

Extension 1 (-CDRVIHY)

CATCACCACC ATCATCACGA AAACCTGTAC TTTCAGAGCT GCGATCGCGT GATTCATTAT

Extension 2 (-CDRVIGY)

CATCACCACC ATCATCACGA AAACCTGTAC TTTCAGAGCT GCGATCGCGT GATTGGCTAT

Extension 3 (-CGRVIGH)

CATCACCACC ATCATCACGA AAACCTGTAC TTTCAGAGCT GCGGCCGCGT GATTCATTAT

Extension 4 (-CGRVIGY)

CATCACCACC ATCATCACGA AAACCTGTAC TTTCAGAGCT GCGGCCGCGT GATTGGCTAT

Extension 5 (-CGGGGGY)

CATCACCACC ATCATCACGA AAACCTGTAC TTTCAGAGCT GCGGCGGTGG CGGTGGCTAT

Extension 6 (-CDRPIHY)

CATCACCACC ATCATCACGA AAACCTGTAC TTTCAGAGCT GCGATCGCCC GATTCATTAT

3.5 References

(1) Brea, R. J.; Reiriz, C.; Granja, J. R. *Chem. Soc. Rev.* **2010**, *39*, 1448–1456.

(2) Staderini, M.; Megia-Fernandez, A.; Dhaliwal, K.; Bradley, M. *Bioorg. Med. Chem.* **2018**, *26*, 2816–2826.

(3) Roxin, A.; Zheng, G. *Future Med. Chem.* **2012**, *4*, 1601–1618.

(4) Morrison, C. *Nat. Rev. Drug Discovery* **2018**, *11*, 531–533.

(5) Apeinou, A.; Matsoukas, M. T.; Simal, C.; Tselios, T. *Biopolymers* **2015**, *104*, 453–461.

(6) Horton, D. A.; Bourne, G. T.; Smythe, M. L. *J. Comput. Aided Mol. Des.* **2002**, *16*, 415–431.

(7) Bruzzoni-Giovanelli, H.; Alezra, V.; Wolff, N.; Dong, C. Z.; Tuffery, P.; Rebollo, A. *Drug Discov. Today* **2018**, *23*, 272–285.

(8) Tyndall, J. D.; Nall, T.; Fairlie, D. P. *Chem. Rev.* **2005**, *105*, 973–999.

(9) Chow, H. Y.; Y., Z.; E., M.; X., L. *Chem. Rev.* **2019**, *119*, 9971–10001.

(10) Nuijens, T.; Toplak, A.; Schmidt, M.; Ricci, A.; Cabri, W. *Front. Chem.* **2019**, *7*, 829.

(11) Angell, Y.; Burgess, K. *J. Org. Chem.* **2005**, *70*, 9595–9598.

(12) Bock, V. D.; Speijer, D.; Hiemstra, H.; van Maarseveen, J. H. *J. Org. Chem.* **2007**, *5*, 971–975.

(13) Goto, Y.; Ohta, A.; Sako, Y.; Yamagishi, Y.; Murakami, H.; Suga, H. *ACS Chem. Biol.* **2008**, *3*, 120–129.

(14) Avrutina, O.; Schmoldt, H. U.; Gabrijelcic-Geiger, D.; Wentzel, A.; Frauendorf, H.; Sommerhoff, C. P.; Diederichsen, U.; Kolmar, H. *ChemBioChem* **2008**, *9*, 33–37.

(15) Zhang, L.; Tam, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 2363–2370.

(16) Wu, Z.; Guo, X.; Guo, Z. *Chem. Commun.* **2011**, *47*, 9218–9220.

(17) Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. *Biochem.* **1991**, *30*, 4151–4159.

(18) Nguyen, G. K.; Kam, A.; Loo, S.; Jansson, A. E.; Pan, L. X.; Tam, J. P. *J. Am. Chem. Soc.* **2015**, *137*, 15398–15401.

(19) Land, E. J.; Ramsden, C. A.; Riley, P. *Acc.Chem. Res.* **2003**, *36*, 300–308.

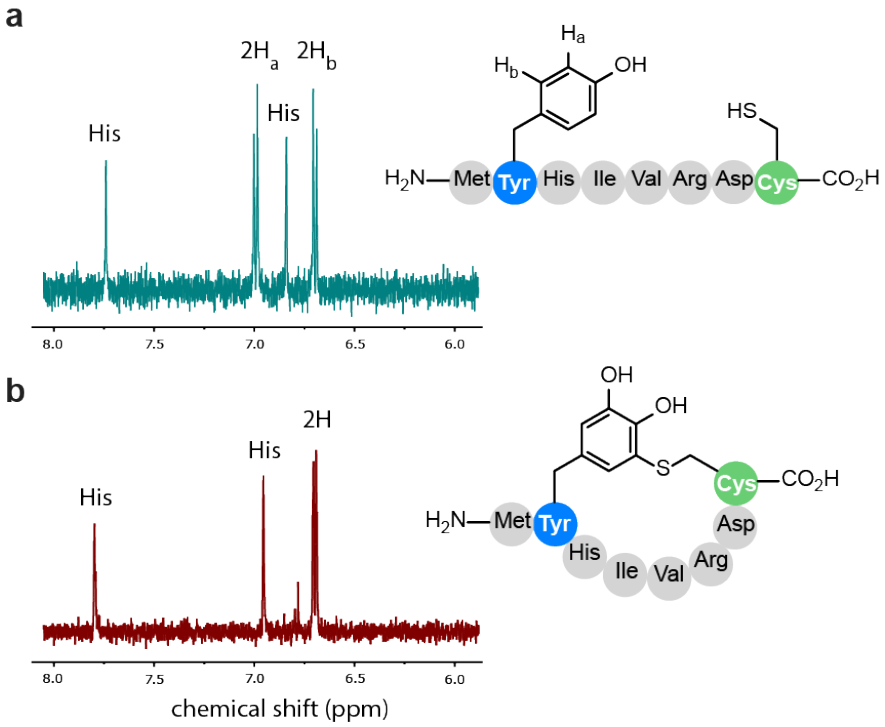
(20) Maza, J. C.; Bader, D. L. V.; Xiao, L.; Marmelstein, A. M.; Brauer, D. D.; ElSohly, A. M.; Smith, M. J.; Krska, S. W.; Parish, C. A.; Francis, M. B. *J. Am. Chem. Soc.* **2019**, *141*, 3885–3892.

(21) Lobba, M. J.; Fellmann, C.; Marmelstein, A. M.; Maza, J. C.; Kissman, E. N.; Robinson, S. A.; Staahl, B. T.; Urnes, C.; Lew, R. J.; Mogilevsky, C. S.; Doudna, J. A.; Francis, M. B. *ACS Cent. Sci.* **2020**, *6*, 1564–1571.

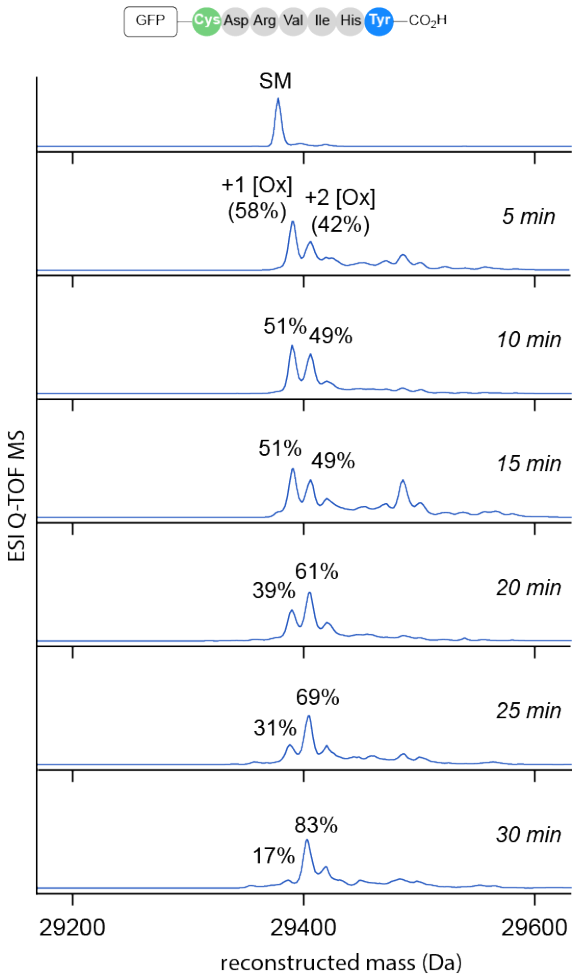
(22) Dinca, A.; Chien, W.-M.; Chin, M. T. *Int. J. Mol. Sci.* **2016**, *17*, 263.

(23) Park, S. E.; Sajid, M. I.; Parang, K.; Tiwari, R. K. *Mol. Pharmaceutics* **2019**, *16*, 3727–3743.

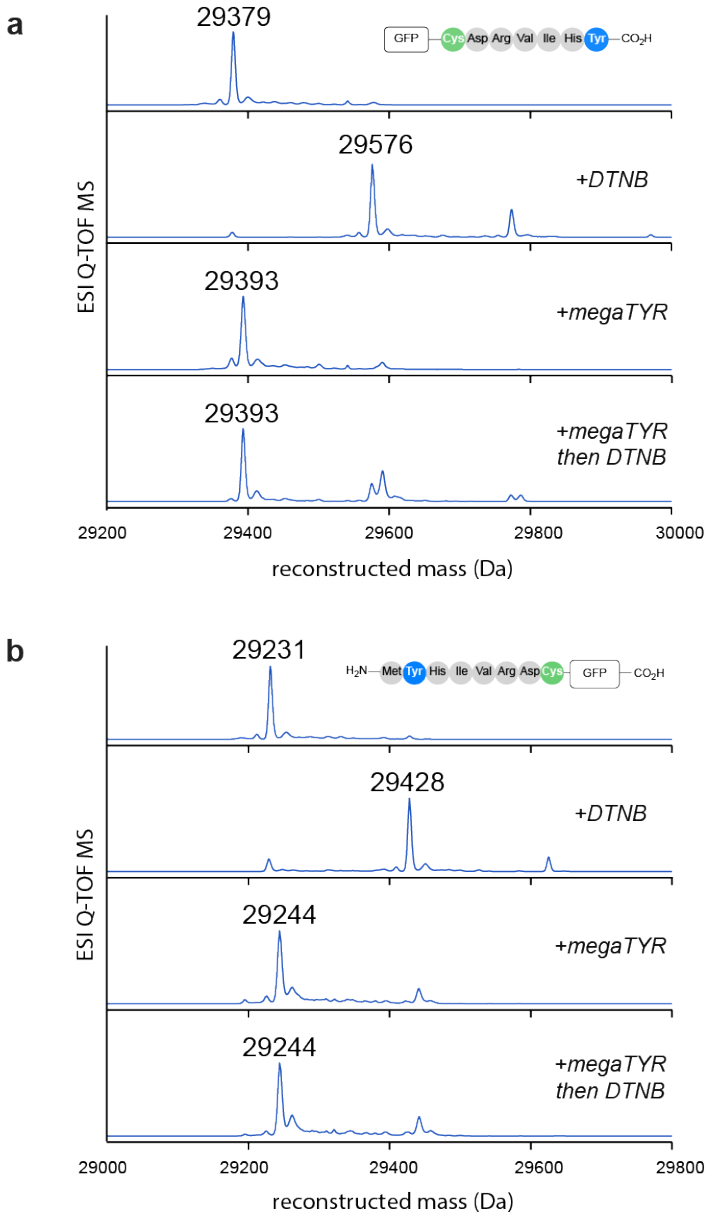
3.6 Supplementary Figures



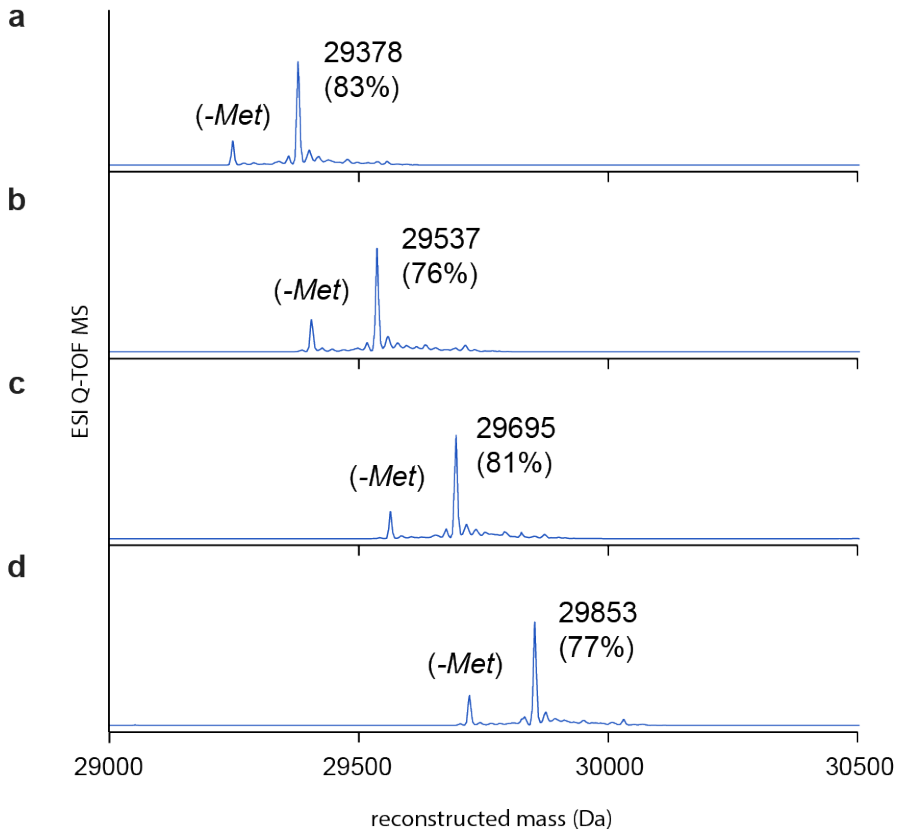
SI Figure 1: ^{1}H NMR aromatic region of the linear and cyclized peptide using excitation sculpting water suppression. **(a)** The aromatic region contains the two phenol doublets of the starting peptide, as well as the histidine imidazole peaks. **(b)** After exposure to tyrosinase, the four protons of the phenol ring become two with very similar chemical shifts, suggesting that the product sits as the catechol. This provides evidence that the thiol does add into the quinone to form the cyclized product.



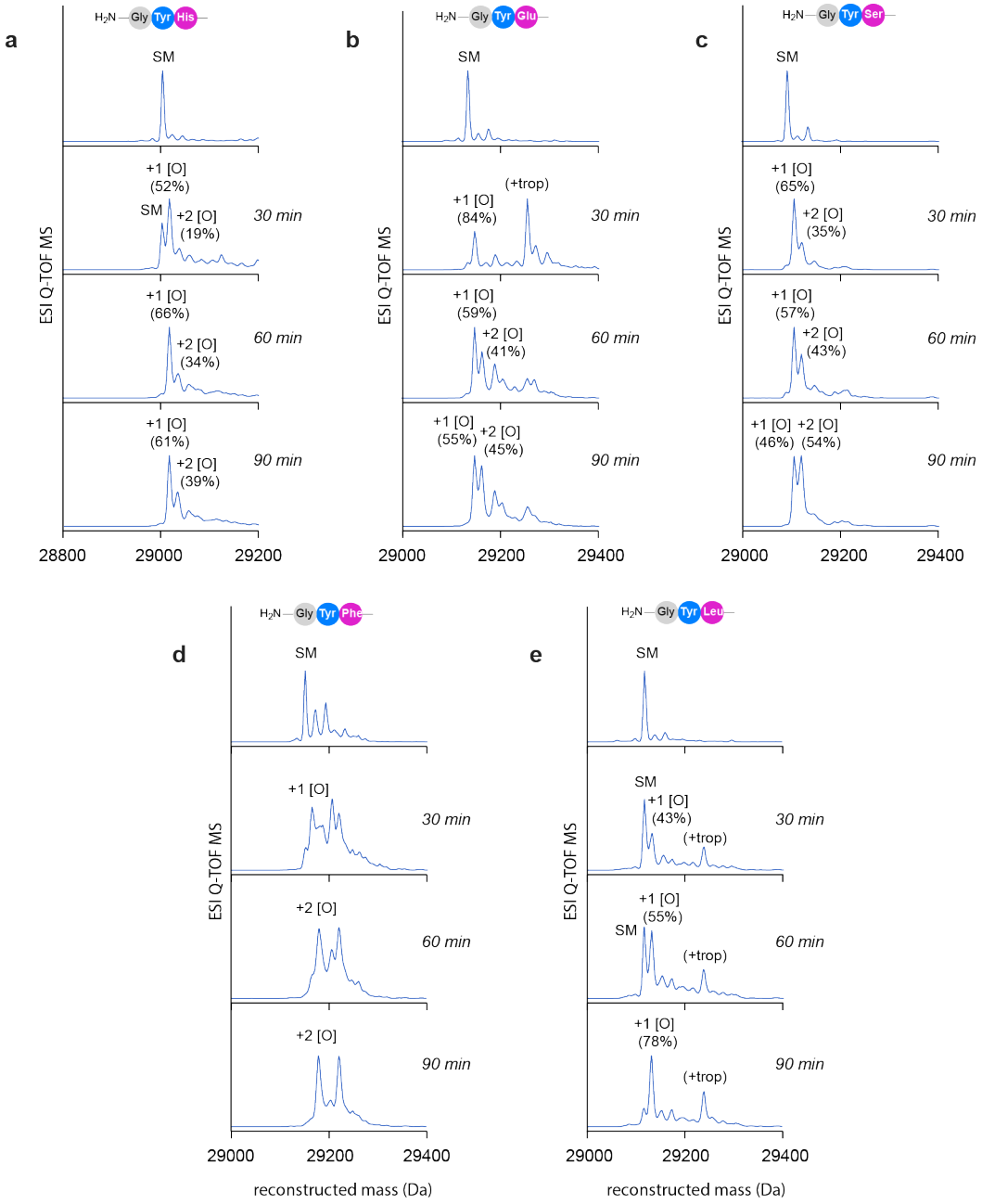
SI Figure 2: Reaction time course for the oxidation of C-terminally tagged protein with megaTYR. The reaction took place over 30 min at room temperature. For each time point, an aliquot was taken and immediately quenched with tropolone.



SI Figure 3: Confirmation of thiol cyclization after megaTYR oxidation. **(a)** The C-terminally tagged peptide sequence. *(second spectrum)* After treatment with Ellman's reagent, the cysteine thiol is capped and a shift is observed. *(third spectrum)* The tyrosine residue is fully oxidized after exposure to megaTYR. *(fourth spectrum)* After treatment of the product with Ellman's reagent, no shift is observed, indicating that the thiol has cyclized into the quinone intermediate. **(b)** The N-terminally tagged peptide sequence. *(second spectrum)* When treated with Ellman's reagent, a shift is observed with the unmodified protein; *(fourth spectrum)* however, no shift if observed after exposure to megaTYR.



SI Figure 4: Expression of N-terminal peptide sequences with various (Ser-Ala) extensions. During expression, a small percentage of protein was cleaved after the methionine residue, resulting in an N-terminal tyrosine. **(a)** MYHIVRDC- **(b)** MYHIVRSADC- **(c)** MYHIVRSASADC- **(d)** MYHIVRSASASADC-



SI Figure 5: Oxidative cyclization of GY-X- mutants with megaTYR. Each mutant was exposed to megaTYR at room temperature for 90 min, with timepoints being taken every 30 min. **(a)** GYH **(b)** GYE **(c)** GYS **(d)** GYF. No integrations were measured since the peak resolution was poor. **(e)** GYL