CYSTEINE ARYLATION

by

Chi Zhang
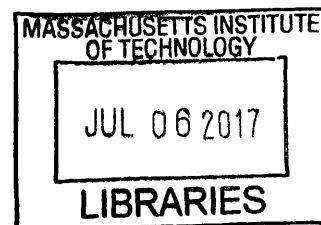
Bachelor of Science in Chemistry

Peking University, China, 2012

Submitted to the Department of Chemistry

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2017

© 2017 Massachusetts Institute of Technology

Signature redacted

Signature of Author......................................

Department of Chemistry

16 May 2017

Signature redacted

Certified by..

Bradley L. Pentelute

Pfizer-Laubach Career Development Professor of Chemistry

Thesis Supervisor

Signature redacted

Accepted by..

Robert W. Field

Haslam and Dewey Professor of Chemistry

Chairman, Departmental Committee for Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Signature redacted

JoAnne Stubbe
Novartis Professor of Chemistry
Thesis Committee Chair

Signature redacted

Bradley L. Pentelute
Pfizer-Laubach Career Development Professor of Chemistry
Thesis Supervisor

Signature redacted

Stephen L. Buchwald
Camille Dreyfus Professor of Chemistry

CYSTEINE ARYLATION

by

Chi Zhang

Submitted to the Department of Chemistry
on 16 May 2017 in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

Abstract

Proteins are the chemical and biological foundation of life. The longstanding goals of chemical biology are to understand the structure and function of proteins and to use these biomolecules for applications in chemistry, biology, medicine, and material science. To achieve such goals, highly efficient, selective, and robust chemical reactions are desired to modify proteins. For decades, cysteine-based reactions with maleimides and alkyl halides are the primary methods for selectively tagging proteins with fluorescent dyes, affinity and radio labels, drug molecules, and polymers and nanocomposites. These traditional reactions generate sulfur-sp^3 carbon bonds between the cysteine thiol and the labeling reagents. The goal of this thesis is to develop new cysteine arylation reactions to generate sulfur–sp^2 carbon bonds on proteins. These reactions are used to make novel peptide and protein therapeutics.

Two mechanistically complementary approaches are developed to arylate cysteine thiol. First, through a nucleophilic aromatic substitution (S_NAr) mechanism, fluorinated aromatic reagents are used for the regioselective arylation of a single cysteine in the presence of many. An enzyme-tag pair (glutathione S-transferase (GST) and glutathione (GSH)) and a self-labeling short peptide (Phe-Cys-Pro-Phe, the π-clamp) are respectively developed to recognize the fluorinated aromatic reagents and to promote the arylation reaction in aqueous solution. The second approach utilizes organometallic palladium reagents to chemoselectively install electron-neutral and electron-rich aryls on cysteine thiols.

These cysteine arylation reactions are applied to the synthesis of macrocyclic peptides and antibody-drug conjugates (ADCs). Long unprotected macrocyclic peptides up to forty residues are efficiently synthesized using the GST enzyme. Using bispalladium reagents, macrocyclic peptides bearing aryl linkers are synthesized via crosslinking of two cysteine thiols. π-Clamp antibodies enable a one-step synthesis of site-specific antibody-drug conjugates to selectively kill cancer cells. Organometallic palladium reagents are used to synthesize linker-free ADCs where the drug molecules are directly linked to cysteine thiols in antibodies.

Thesis Supervisor: Bradley L. Pentelute
Pfizer-Laubach Career Development Professor of Chemistry

Acknowledgements

Nothing in this thesis would be possible without the support from my advisors, colleagues, friends, and family. All of these wonderful people that I work together and live with enabled me to take this big step towards becoming a scientist.

I am most grateful to my Ph. D. advisor Professor Brad Pentelute. Coming to MIT to work with Brad is the best decision that I have ever made in my research career. I have benefited tremendously from the outstanding research environment built and led by Brad. He has spent countless hours and efforts to help me design my research projects, to guide me through intellectual challenges, and to help me improve my communication skills. Most importantly, Brad has been encouraging me to think big and to tackle tough problems with a fearless and optimistic mind.

Professor JoAnne Stubbe and Professor Steve Buchwald have been excellent mentors of mine. I am more than grateful for their advice and support. Their inspiring perspectives on science truly enriched my understanding of chemistry.

It has been a great journey and a lot of fun to work with everyone in the Pentelute lab. I am indebted to the early members of the lab, Dr. Amy Rabideu, Dr. Mark Simon, Dr. Xiaoli Liao, Dr. Alex Spokoyny, Dr. Yekui Zou, Ms. Jingjing Ling, and Mr. Rocco Policapo for their help that was instrumental for my smooth start of graduate research. I have shared the excitement for science and worked side-by-side with Mr. Peng Dai, Dr. Ekaterina Vinogradova, Dr. Daniel Cohen, Dr. Michael Santos, and Dr. Tim Senter. Their efforts were instrumental in the success of our collaborative projects. I cherish every discussion, argument, and brainstorming with Dr. Alex Vinogradov, Dr. Zak Gates, Dr. Anthony Quatoraro, and Mr. Suan Tuang.

MIT is a wonderful place for collaborations. My thesis was made possible by outstanding collaborators from the labs of Professor Steve Buchwald, Professor Troy Van Voorhis, and Professor Dane Wittrup. I am indebted to Dr. Matt Welborn, Mr. Tianyu Zhu, and Dr. Nicole Yang for helping and teaching me so much outside of my field.

My family has been the rock of my life. Science is hard, and I could not make it so far without the support of my family. Thank you, Mom and Dad, for providing me with the freedom to chase my dream. Especially, I would like to thank my wife Li, she has been a big support of my graduate school, and she will be my lasting support for the years to come.

Table of Contents

11

List of Figures

15

17

19

List of Tables

Scope of Thesis

This thesis aims to: (1) develop cysteine arylation reactions for the site-selective, efficient, and robust modification of proteins with diverse synthetic molecules; and, (2) apply the developed reactions to the synthesis of peptide and protein therapeutics.

Chapter 1 is a review of arylation chemistry for bioconjugation. We outline and provide perspectives on sulfur, nitrogen, selenium, oxygen, and carbon arylative bioconjugation strategies and their applications to modify peptides, proteins, sugars, and nucleic acids.

Chapter 2 describes enzyme-catalyzed cysteine arylation. Glutathione *S*-tranferases catalyze the arylation of glutathione in aqueous buffer using perfluoroaromatic reagents. The glutathione tag placed at the *N*-terminus of the target peptide or protein can be selectively arylated under GST-catalysis without affecting other thiol species on the same molecule of interest. This glutathione S-transferase catalyzed arylation reaction is applied to the macrocyclization of long unprotected peptides.

Chapter 3 presents the π-clamp, a four-residue peptide sequence (Phe-Cys-Pro-Phe), that promotes a self-arylating reaction at its own cysteine residue with perfluoroaromatic reagents. The π-clamp is small and fully genetically encodable, enabling the synthesis of site-specific π-clamp antibody-drug conjugates that selectively kill breast cancer cells.

Chapter 4 documents a palladium-mediated approach to cysteine bioconjugation. Palladium(II) oxidative addition complexes are developed to arylate cysteine residues in peptides and proteins in aqueous solution. This approach allows the synthesis of aryl-linked macrocyclic peptides and the direct attachment of drug molecules to antibodies without using linkers.

Chapter 1. Arylation Chemistry for Bioconjugation

The work presented in this chapter is a part of the following manuscript and is reproduced here with permission from the authors.

Zhang, C.; Vinogradova, E. V.; Buchwald, S. L.; Spokoyny, A. M.; Pentelute, B. L.

Arylation Chemistry for Bioconjugation

To be submitted.

1.1. Introduction

Bioconjugation processes involve chemical reactions to form either covalent or well-defined non-covalent attachments between two molecular entities, at least one of which is a biomolecule.[1,2] Specifically, the biomolecule involved in a bioconjugation process is usually obtained from natural sources or via genetic means. Development of novel bioconjugation approaches provides new opportunities for modulation of biomolecule structure and function through the use of synthetic chemistry both *in vitro* and in living systems. The marriage of biomolecules with synthetic modalities makes possible the discovery of new medicines,[3–5] the exploration of complex biological processes,[6] the generation of new biomaterials,[7,8] and the invention of novel diagnostic and medical tools.[9]

Bioconjugation reactions that form covalent bonds normally require conditions that are more stringent and milder compared to those used in small molecule synthesis. The use of low temperatures (< 37 °C), close to neutral pH, and aqueous buffer preserves the structure and function of biomolecules. In addition, bioconjugation reactions need to be fast, and reach full conversion in hours with micromolar concentrations of substrates. Reaction rates of most bioconjugation reactions are in the range of 1-1000 $M^{-1}s^{-1}$, with a few fast reactions having rates of more than 1000 $M^{-1}s^{-1}$.[10,11] In some instances less stringent conditions (i.e., the use of organic solvents, prolonged reaction time, and heating) can be used for structurally simple and robust biomolecules such as peptides, oligosaccharides, and oligonucleotides.

Besides the restrictions on reaction conditions, another major challenge of developing efficient bioconjugation processes comes from the intrinsic complexity of biomolecules containing multiple reactive functional groups. Indeed, many bioconjugation reactions utilize nucleophiles in biomolecules (e.g., amines, hydroxyl groups, carboxylates, and thiols). Yet, there are often tens or hundreds of nucleophilic sites in a single biomolecule, making the control of chemoselectivity and regioselectivity difficult. As a result, many traditional bioconjugation reactions are nonselective, and form heterogeneous products. This, clearly, leads to further challenges for the characterization and purification of products.

The development of bioconjugation protocols has been largely inspired by early work in protein chemistry.[12–14] Procedures to chemically modify proteins were used for practical purposes prior to the scientific conception of bioconjugation and the understanding of the underlying

chemical transformations. For example, formaldehyde was used for tanning in the leather industry long before knowing its ability to crosslink proteins.[15] Pioneering protein chemists developed reagents to modify amino acids to determine the compositions of proteins.[16–19] These reagents were used to covalently bind to the active sites of proteins[20] and to develop protein sequencing techniques such as Edman degradation.[21] Early work in the area of protein crystallography also relied heavily on the use of various heavy-atom-containing reagents (e.g., U, Pb, Hg, Au, Ag, and I) that bind amino acid residues thereby aiding the structure refinement process.[22]

Selectivity is currently recognized as the most important and challenging requirement in devising new bioconjugation processes. The ability to perform bioconjugation in a predictable manner significantly simplifies the purification and characterization of the resulting products. For many bioconjugation reactions, chemoselectivity stems from the differences in the intrinsic reactivity of different functional groups in biomolecules.[23,24] Regioselective modification of a single site (e.g., a particular cysteine thiol among many) is challenging and often achieved through enzyme-mediated reactions.[25–30] More recently, The genetic encoding of unnatural handles makes possible to judiciously choose two completely abiotic coupling partners for bioconjugation.[31–33] For example, following the landmark introduction of the term "click chemistry" by Sharpless[34] in 2001 to describe reactions that are high yielding, easy to perform, wide in substrate scope, and with easy-to-remove byproducts, synthetic chemists quickly started to develop highly efficient and selective reactions for the modification of biomolecules using this strategy. An even higher selectivity bar was set for bioconjugation reactions used to study biomolecules in a complex biological milieu. Introduced by Bertozzi in 2003, bio-orthogonal chemistry provides highly selective bioconjugation reactions that can be applied in living cells and animals.[35]
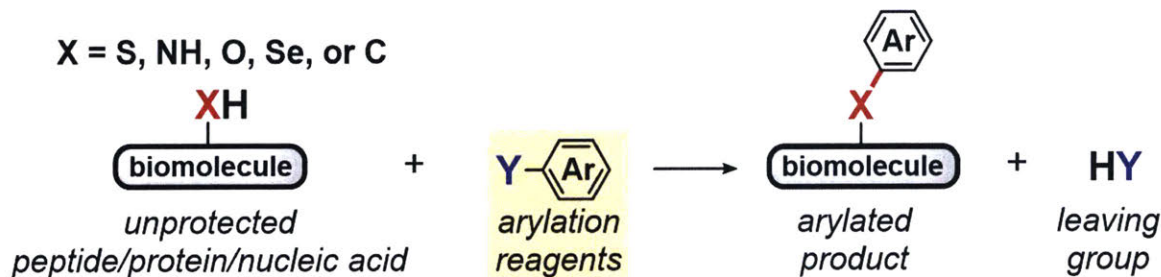
Figure 1.1. The generic arylation process for bioconjugation.

Among the existing strategies, chemistry used for the arylation of nucleophiles in biomolecues (Figure 1.1) has been relatively less developed for bioconjugation until recently. This

is somewhat surprising, given how routinely the formation of nucleophile–sp^2 carbon bonds has been used in small molecule chemistry over the past few decades.[36,37] Arylation reactions generate an X–C(sp^2) bond (X is a carbon or a heteroatom) which has significantly different chemical properties compared to other X–C(sp^2, sp^3, or sp) bonds produced with traditional bioconjugation reactions.[23] Recent work has shown that diverse arylation reactions developed for small molecule chemistry can be utilized in bioconjugation processes (Figure 1.2). For example, electron-deficient aryl halides react with some nucleophilic groups (e.g., thiols, hydroxyl groups, amines, and selenols) in biomolecules through nucleophilic aromatic substitution (S$_N$Ar) reactions (Figure 1.2A).[38] In addition, nucleophiles in biomolecules react with transition metal-based arylation reagents (Figure 1.2B).[39,40] Nucleophilic selenols are oxidized and turned into electrophiles, and the resulting dichalogenides are arylated through umpolung approaches using aryl boronic acid-derived nucleophiles. Finally, widely used palladium-catalyzed cross-coupling reactions[39,40] and C–H activation strategies[41] are used to forge carbon-carbon bonds onto biomolecules (Figure 1.2D).

Here we provide an account of processes that result in the formation of X–C(sp^2) bond (X is a carbon or a heteroatom) which we termed arylative bioconjugation reactions. Because of the application-oriented nature of bioconjugation, we categorize these reactions according to their target biomolecules in following sections. Note that the arylation of small biomolecules including amino acids, monosaccharides, natural products, and nucleotides is not covered in this review. The nucleophilic aromatic substitution (S$_N$Ar) and metal-mediated arylation of cysteine thiol in peptides and proteins are first discussed. We then describe the N-arylation of lysine ε-amine. Arylation of other natural and unnatural amino acid side chains is also discussed with a focus on the umpolung arylation of selenocysteine and the palladium-catalyzed cross-coupling reactions for protein modification. Finally, we discuss metal-mediated arylation chemistry for the post-synthetic modification of nucleic acids. We conclude with a discussion on limitations of current methods and future avenues to broaden the scope and application of arylative bioconjugation.

Throughout this review, we try to emphasize the fundamental chemistry features of arylative bioconjugation reactions. The complex nature of bioconjugates requires well-designed reagents and conditions as well as the accurate product characterization prior to translation of the resulting conjugates for biological applications. We believe that there is no perfect bioconjugation reaction that fulfills all application needs. We hope that this review will serve to complement other

recent reviews on bioconjugation[42–52] and to provide the reader with an expanded understanding of bioconjugation techniques available today.
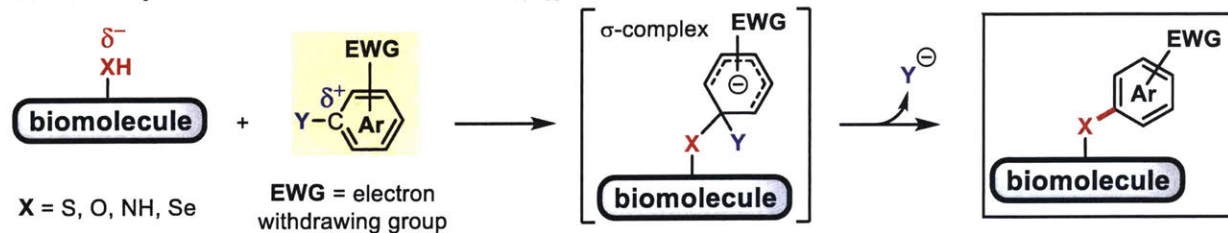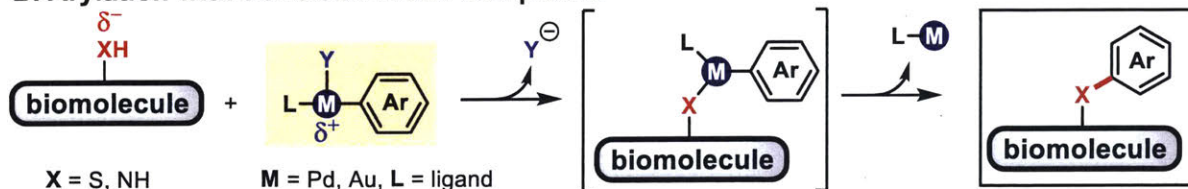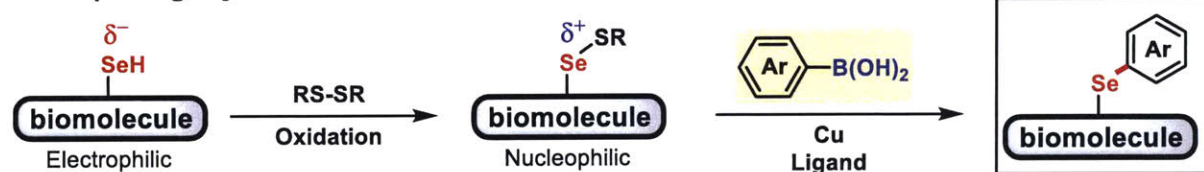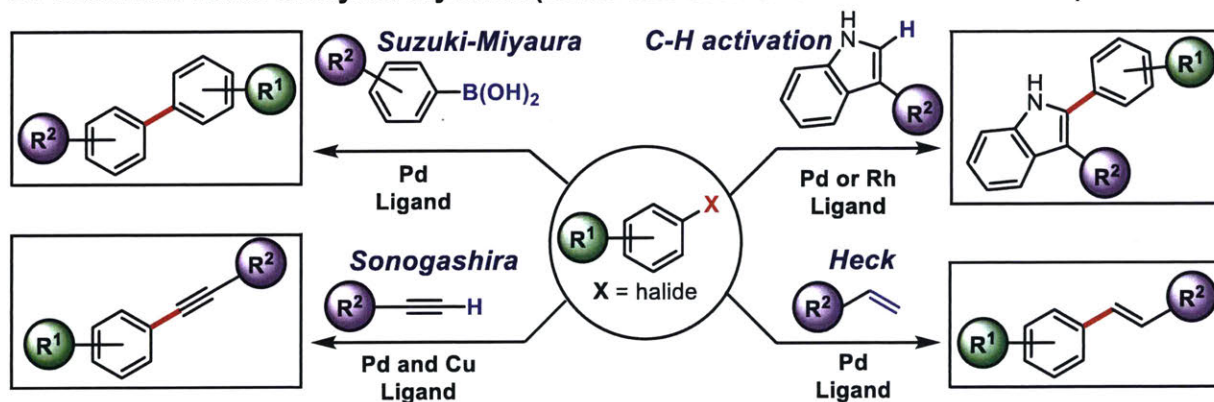
Figure 1.2. Mechanistically distinct routes for arylation of biomolecules.

1.2. Cysteine arylation

As the only thiol-containing canonical amino acid, cysteine is often used for chemical protein modification.[53,54] In aqueous buffer, cysteine's thiol group ($pK_a \approx 8$) is more acidic than other common protein nucleophiles such as serine's hydroxyl group ($pK_a \approx 13$), tyrosine's phenol group ($pK_a \approx 10$), or lysine's ε-amine ($pK_a \approx 10$ for RNH_3^+, $pK_a \approx 36$ for RNH_2). Moreover, the sulfur is more polarizable (soft) than oxygen or nitrogen. Together, these chemical features make the cysteine thiol/thiolate more reactive than other protein nucleophiles towards electrophiles such as maleimides and alkyl halides, enabling the chemoselective modification of cysteine thiol in the presence of other amino acid side chains. Cysteine is also a relatively rare amino acid within the human proteome, making it possible to generate homogeneous conjugates from a protein that contain only one chemically accessible cysteine thiol. Commercially available maleimide and alkyl halide reagents are routinely used to conjugate fluorescent dyes, affinity labels, and drugs to peptides and proteins. Classic and some more modern methods for bioconjugation via cysteine have been summarized in several recent review articles.[53,54]

A key feature of cysteine arylation is that it generates a $S-C(sp^2)$ bond that is different from the $S-C(sp^3)$ bonds produced using maleimides or alkyl halides. Electron-deficient aryl halides such as chloronitrobenzenes and perfluoroarenes react with cysteine via a nucleophilic aromatic substitution (S_NAr) mechanism. Alternatively, palladium-mediated S-arylation approaches are developed to introduce electron-neutral and electron-rich aryls. Compared to S_NAr processes, palladium-mediated reactions have a broader substrate scope because they are relatively indifferent to the electronic nature of the aryl group. Through these two complementary approaches, a wide range of aryl groups can be conjugated to cysteine thiols in peptides and proteins, enabling the development of covalent protein modulators, macrocyclic peptides, antibody-drug conjugates, and other hybrid protein-based constructs.

1.2.1. Cysteine arylation via nucleophilic aromatic substitution (S_NAr)

Nucleophilic aromatic substitution (S_NAr) reactions generally require strong electron-withdrawing groups on the aryl ring to stabilize the negatively-charged σ-complex (Figure 1.2A). As a result, aryl and heteroaryl electrophiles with strong electron-withdrawing groups such as nitro, fluoro, and sulfonyl are used to arylate cysteine thiols in peptides and proteins. The arrangement

of these electron-withdrawing substitutions on the aryl ring greatly affect the S$_N$Ar reactivity of the aryl electrophiles. For example, o- and p-nitroaryl halides are more reactive than m-nitro variants because of the strong resonance effect of o- and p-nitro substitutions that are lacking in the m-nitro substituted compounds.

(a) Chemoselective cysteine S$_N$Ar arylation

The chemoselectivity of the S$_N$Ar arylation of cysteine is usually high because of the higher nucleophilicity of thiol/thiolate compared to other protein nucleophiles. However, the selectivity varies with reaction conditions because arylation reaction with amines and hydroxyls can occur competitively in the presence of strong bases. The selectivity of this reaction can also be altered as a function of solvent polarity. In addition, the microenvironment around an amino acid side chain in a protein effects its nucleophilicity, resulting in different reactivity profiles for the same amino acid side chains at different protein sites.

Aryl chlorides as cysteine-selective covalent protein modulators

The thiol group of cysteine is utilized by nature to perform a wide range of biochemical functions.[55–57] Many proteins have essential cysteine residues that are either directly involved in catalysis or critical for maintaining the protein structure. This provides the possibility for designing and discovering aryl halides as cysteine-selective reagents to covalently modulate the structure and function of these proteins.

Thiol-selective aryl halides were initially developed to identify cysteines in the active sites of enzymes. Price and Cohn used 4-chloro-7-nitrobenzofurazan (NBD-chloride) to investigate the different environments of two cysteines (Cys-25 and Cys-187) in porcine muscle adenylate kinase (Figure 1.3).[58] Both cysteine residues were arylated with excess NBD-chloride (Figure 1.3, top). Surprisingly, the more solvent-exposed Cys-187 reacted with NBD-chloride ($k = 205$ M^{-1}s^{-1}) about 40-fold more slowly than the buried Cys-25 ($k = 8580$ M^{-1}s^{-1}). Regioselective arylation of Cys-25 was achieved using a stoichiometric amount of NBD-chloride. In contrast, both cysteine thiols were labeled with similar reaction rates with other thiol-reactive reagents such as Ellman's reagent (5,5'-dithiolbis-(2-nitrobenzoic acid)). The Cys-25-labeled enzyme was completely inactive, indicating that the Cys-25 was in close proximity to the active site of the enzyme. Together, these results highlight the utility of arylation reagents for studies of protein structure-function.

Importantly, this example also illustrates how the local environment of an amino acid side chain strongly affects its reactivity.
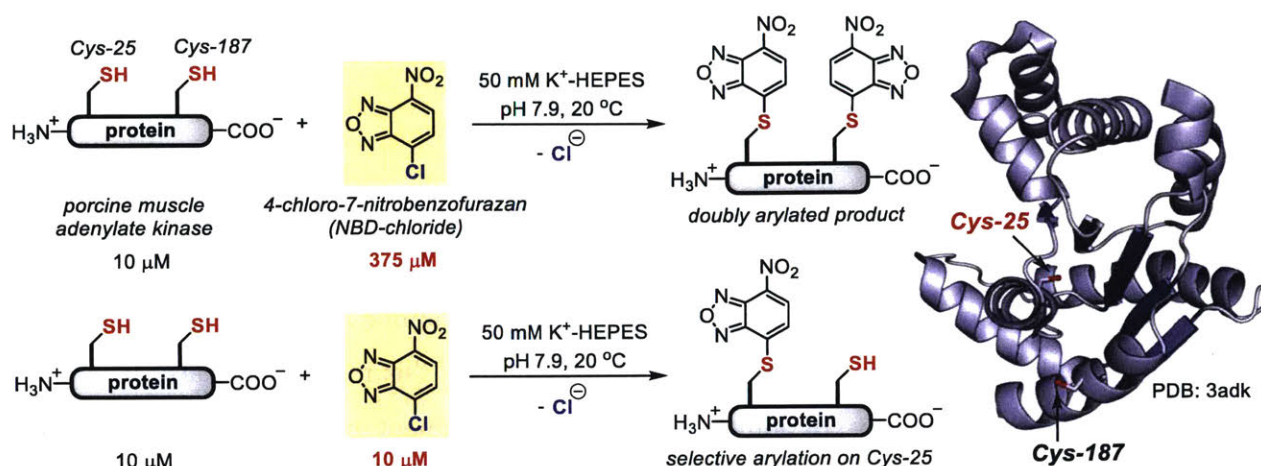
Figure 1.3. Chemo- and regio-selective arylation of cysteine thiols in porcine muscle adenylate kinase using NBD-chloride.

Blanchard[59] and Li[60] independently reported two structurally similar covalent antagonists for the nuclear hormone receptor peroxisome proliferator activated receptor gamma (PPARγ) (Figure 1.4, entries 1 and 2). PPARγ regulates several genes involved in adipogenesis and is the molecular target of the thiazolidinedione family of antidiabetic drugs.[61,62] Both antagonists (GW9662 and T0070907) block ligand binding to PPARγ through the selective arylation of Cys-285 that is located in the orthosteric ligand binding pocket of PPARγ. These two compounds showed antagonist activity both *in vitro* and in cell culture. However, these two antagonists were ineffective for inhibiting the binding of PPARγ ligands to an allosteric site other than the orthosteric pocket on the ligand binding domain.[63] Through expansion of the size of the 2-chloro-5-nitrobenzamidyl scaffold, Kojetin developed a dual-site PPARγ antagonist (Figure 1.4, entry 3) that simultaneously blocked the ligand binding to both the orthosteric pocket and the allosteric site.[64] Together, this family of 2-chloro-5-nitrobenzamidyl-based compounds are useful chemical tools to modulate PPARγ function in biological processes.

Entry	Inhibitor	Name	Target protein and residue	Ref.
1		GW9662		59
2		T0070907	peroxisome proliferator activated receptor-γ (PPARγ), Cys-285	60
3		SR16832		64
4		4-chloro-2-hydroxyl-methylpyridine	dimethylarginine dimethylaminohydrolase (DDAH), Cys-249	66
5		RB-2-cb	β-tublin (TUBB), Cys-239	67
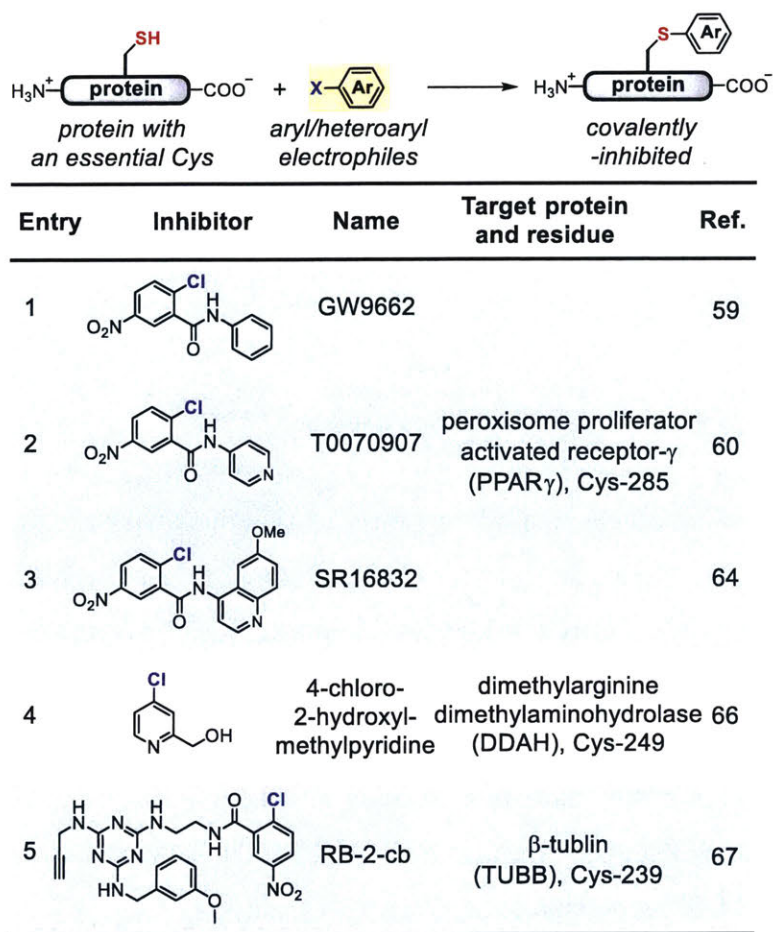
Figure 1.4. Representative aryl and heteroaryl halides as cysteine-selective covalent protein modulators.

Besides chloronitrobenzenes, 4-halopyridines have also been explored as cysteine-selective covalent protein inhibitors. Through screening a library of 4000 compounds, Fast found a class of 4-halopyridines that could serve as covalent inhibitors for dimethylarginine dimethylaminohydrolase (DDAH) through the arylation of its active site Cys-249.[65] Inhibition assays and mutation studies indicated that Asp-66 of DDAH was required for the inactivation of the protein by 4-chloro-2-hydroxylmethylpyridine (Figure 1.4, entry 4). The carboxylate residue of Asp-66 was proposed to stabilize the pyridinium form (computationally predicted $pK_a = 3.6$) of the inhibitor thus promoting the arylation reaction at pH 7.5. This ensures the high selectivity of the compound for DDAH, because the free compound in solution mainly resides in its pyridine form and is thus unreactive for other thiol species. Although computer modeling supported this proposed mechanism of inactivation, an X-ray structure of inactivated DDAH showed no close

contact between the carboxylate of Asp-66 and the pyridine ring of the inhibitor, suggesting possible conformational changes of DDAH after the covalent inhibition.[66]
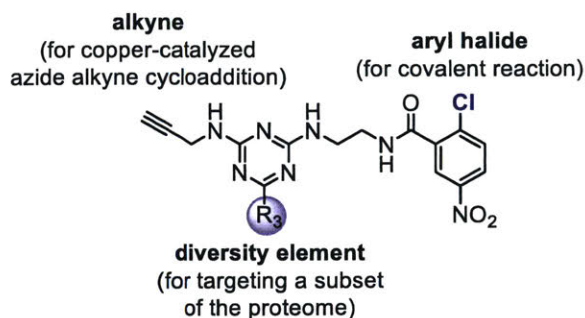
Figure 1.5. Triazine-derived electrophilic probes for reactions with the proteome.

Weerapana investigated the reactions of *p*-chloronitrobenzene-based probes with the proteome through quantitative activity-based protein profiling.[67–69] Using triazine as the scaffold, different electrophilic probes were prepared to target subsets of the proteome (Figure 1.5).[67] The reactions between the probes and the proteome were qualitatively assessed using in-gel fluorescence and quantitatively measured through enrichment and mass spectrometry. Initial studies with these probes resulted in the discovery of RB-2-cb that covalently arylated Cys-239 of β-tublin and inhibited tublin polymerization (Figure 1.4, entry 5).[67]
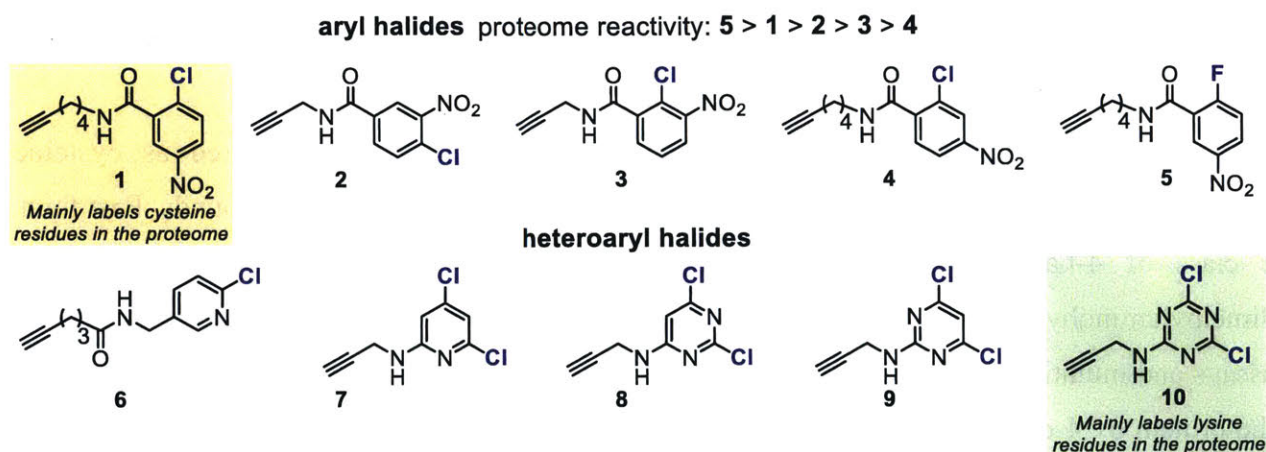
Figure 1.6. Reactivity and selectivity of aryl halides towards the proteome.

Using a similar activity-based protein profiling approach, Weerapana and co-workers expanded their studies to other aryl halides (Figure 1.6).[68] Because of the previously mentioned resonance effects, the *p*-nitro- and *o*-nitro-chlorobenzene electrophiles (**1**, **2**, and **3**) are more reactive towards the proteome than the *m*-nitro variant (**4**). This result corroborated the dominant

role of *p*-nitrochlorobenzene in previously reported cysteine-selective covalent protein inhibitors (Figure 1.4). Additionally, *p*-nitrofluorobenzene (**5**), chloropyridine (**6** and **7**), chloropyrimidines (**8** and **9**), and chlorotriazine (**10**) were investigated for their reactions with the proteome. In-gel fluorescence assay showed that electrophiles **1**, **5**, and **10** reacted with the proteome at micromolar concentrations, indicating that they might be used for developing more complex covalent protein modifiers. Of importance, mass spectrometry analysis of the labeled proteomes showed that *p*-nitrochlorobenzene (**1**) was cysteine-selective while dichlorotriazine (**10**) favored arylation on lysines. However, electrophile **10** reacted with N- and C-terminal protected cysteine much faster than electrophile **1**. This disparity between reactions with the proteome and with simple amino acids indicates that a subset of the proteome might possess lysine residues that are highly reactive towards probe **10**. Further structural and mechanistic characterization of the labeled proteins identified from mass spectrometry will likely provide more understanding of the reactivity and selectivity of these aryl electrophiles toward the proteome.

Dichlorotetrazine for S$_N$Ar stapling and photo-triggered unstapling of peptides and proteins

Hochstrasser and Smith identified 1,4-*S*,*S*-tetrazine (i.e., a tetrazine with sulfur substituents at the 1 and 4 positions) as a fast photochemical triggerfor the generation of thiocyanate, enabling the study of dynamics of peptide and protein folding.[70,71] This *S*,*S*-tetrazine unit was incorporated into peptides through either on-resin or in-solution S$_N$Ar reactions of 1,4-dichlorotetrazine with cysteine thiols.[72,73]

Based on these early studies, Smith and co-workers recently reported a method for the rapid stapling and photo-triggered unstapling of peptides and proteins.[74] They used 1,4-dichlorotetrazine to chemoselectively crosslink two cysteine thiols in an unprotected peptide (**P1**), generating the *S*,*S*-tetrazine-stapled peptide macrocycles within one minute (Figure 1.7A). The resulting stapled peptide (**P2**) was unstapled under UV irradiation. The cyano groups were removed from the peptide bisthiocyanate (**P3**) through reactions with free cysteine to restore the starting unprotected peptide. Of note is that the *S*,*S*-tetrazine also served as a bioconjugation handle for the inverse electron demand Diels-Alder reaction (Figure 1.7B). Although dichlorotetrazine chemoselectively reacts with the thiols of cysteines in the presence of other nucleophilic groups contained in the side chains of amino acid, it is not compatible with the disulfide reducing reagent *tris*(2-carboxyethyl)

phosphine (TCEP). Therefore, purification of the reduced dithiol species is required prior to the stapling reaction.
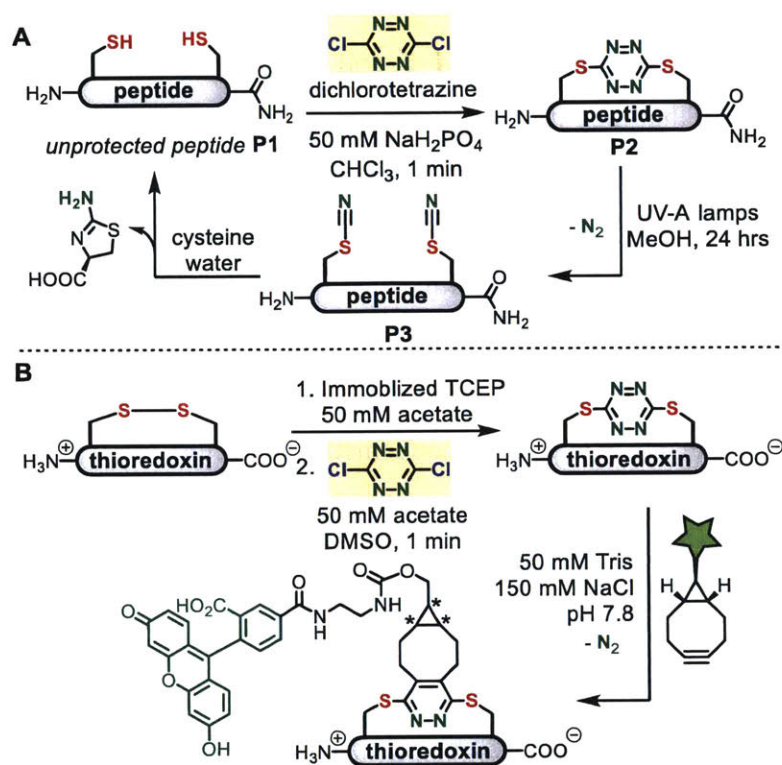
Figure 1.7. Peptide and protein stapling with dichlorotetrazine.

(**A**) Unprotected peptides with two cysteines were stapled via S$_N$Ar reaction with dichlorotetrazine. The resulting tetrazine staple could be removed under photo-triggered conditions. (**B**) Thioredoxin was first stapled with dichlorotetrazine; the resulting dithiol-tetrazine motif was further utilized as a bioconjugation handle for Diels-Alder reaction with a cyclooctyne. *Note that a mixture of diastereomers was formed through cycloaddition from either face of the tetrazine ring.

Heteroaryl methylsulfones for cysteine arylation

Heteroaromatic methylsulfones were developed to generate heteroaryl-thiol linkages in proteins. Xian first reported the use of methylsulfonyl benzothiazole (MSBT) as a thiol-selective reagent for blocking of cysteines in proteins (Figure 1.8A).[75] MSBT showed more product formation when reacted with N- and C-terminal protected cysteine when compared to other substituted benzothiazoles including those with halides and diazo groups at the C-2 position. Because halides and diazo group are better leaving groups than sulfonyl group in S$_N$Ar reactions,[76] it is possible that these more reactive electrophiles are hydrolyzed in aqueous solution before react

with the cysteine derivative. MSBT showed excellent chemoselectivity for reaction with the cysteine thiol; no reaction was observed for other protein nucleophiles including those in the side chains of Ser, Tyr, Trp, Met, His, and Lys. The reaction rate depended on the pH of the solution, consistent with the idea that the thiolate might be more reactive with MSBT than thiol. The labeling reaction was protein-compatible, as demonstrated by blocking the essential thiol of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) under mild conditions.

Inspired by the high chemoselectivity and mild conditions manifested by reactions of MSBT, Barbas developed a series of heteroaryl methylsulfone reagents and investigated their reactions with cysteine residues in proteins (Figure 1.8B).[77] These heteroaryl methylsulfonyl reagents reacted with proteins in a thiol-specific manner, enabling site-selective conjugation of fluorescent dyes and polyethylene glycols (PEGs) to cysteine thiols in proteins. The labeling reaction rates were comparable to cysteine-maleimide conjugation. Importantly, these heteroaryl-cysteine conjugates are chemically stable toward acids, bases, and external thiols, resulting in exceptionally more stable conjugates compared to the corresponding maleimide adducts.
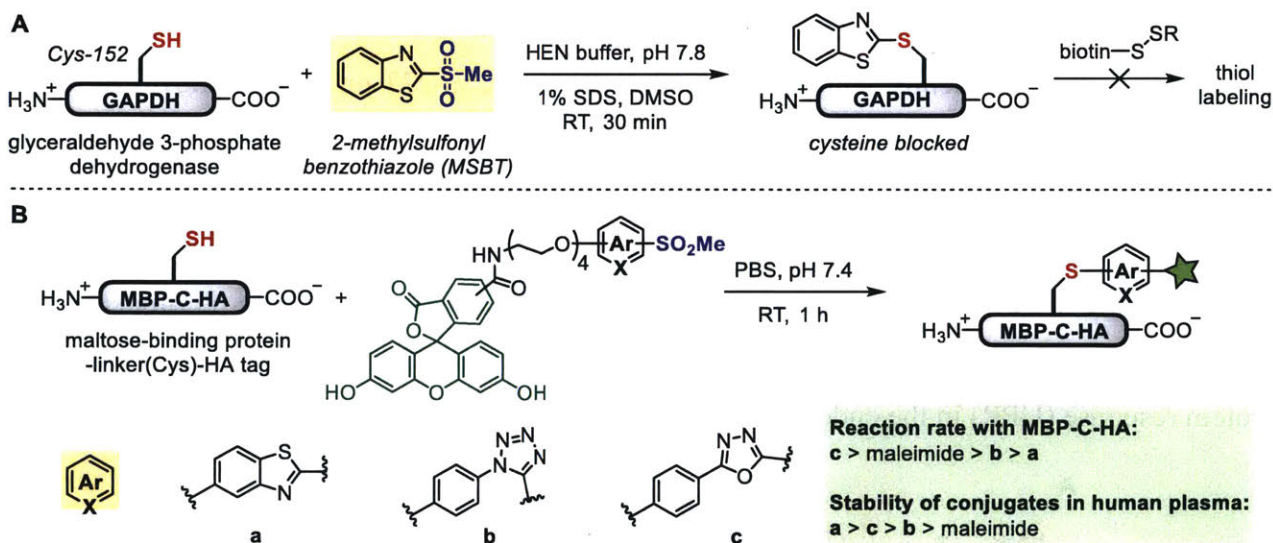
Figure 1.8. Heteroaryl methylsulfonyl reagents for protein cysteine arylation.
(A) Methylsulfonyl benzothiazole (MSBT) as a protein thiol blocking reagent. (B) Stable protein bioconjugates generated from thiol modification with heteroaryl methylsulfonyl reagents.

The high speed, selectivity, and stability of these heteroaryl methylsulfone-cysteine reactions enabled their wide application in bioconjugation reactions. Using the most reactive phenyloxadiazole sulfone linker, Barbas prepared site-specific antibody conjugates with improved

37

stability in human plasma compared to the corresponding maleimide conjugates.[78] Mindt prepared ^{18}F-containing phenyloxadiazole sulfone reagents and showed their rapid conjugation to targeting peptides and proteins for positron emission tomography (PET).[79] Alternatively, ^{18}F-labeled integrin binding peptides (RGD peptides) were synthesized by combining the phenyloxadiazole sulfone-cysteine reaction with a fluorinase-catalyzed transhalogenation process.[80] The high thiol-specificity of these sulfone reagents was further highlighted in selective modifications of complex molecules such as polyacrylamide hydrogels[81] and protein enzymes[82].

Methylsulfonyl-based arylation reagents were also developed for thiol conjugation in living cells. Fang reported 4-methylsulfonyl-N-n-1,8-naphthalimide (MSBN) as a highly fluorogenic reagent for the specific detection and labeling of thiol species and proteins (Figure 1.9A).[83] More than 100-fold increase in fluorescence was observed upon reacting MSBN with thiols, enabling imaging of thiol species and detection of reversible protein thiol modification in living cells.

Fersht screened a fragment library of 2-sulfonylpyrimidines and identified a compound named PK11000 as a cysteine-selective arylation reagent for the oncogenic target p53 (Figure 1.9B).[84] PK11000 covalently conjugated to Cys-182 and Cys-277 on p53 as shown by mass spectrometry analysis of the labeling reactions of p53 and its cysteine-to-serine mutants. The labeled p53 showed increased stability and retained DNA binding *in vitro*, which potentially explained the high potency of PK11000 against p53-deactivated or p53-mutated cancer cells. Interestingly, PK11000 was also found to exert its anti-tumor function through p53-independent pathways such as increasing the level of reactive oxygen species (ROS) or inducing the unfolded protein response (UPR) in the endoplasmic reticulum (ER).
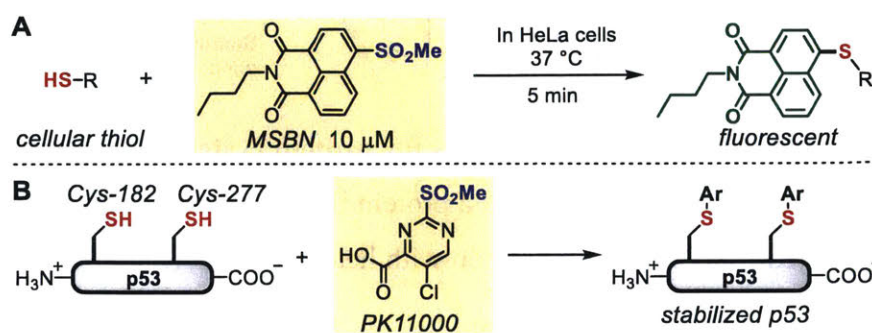
Figure 1.9. Arylation of thiols in living cells using aryl and heteroaryl methylsulfonyl reagents.

Perfluoroarenes for cysteine arylation

In the 1960s, several research groups whose primary focus was synthetic fluorine chemistry carried out fundamental studies that provided the ground rules for the selectivity, reactivity, and substitution patterns of nucleophilic aromatic substitution (SNAr) reactions between perfluoroarenes and thiolates.[85,86] Tatlow found that the 1,4-disubstituted product predominantly formed when reacting an excess amount of potassium phenylsulfide with hexafluorobenzene in refluxing pyridine (Figure 1.10A).[85] Later, systematic and mechanistic studies by Pitts established the activating effect of *p*-substitutions on the perfluorobenzene ring.[86] The rate of SNAr reaction with potassium phenylsulfide was more than three orders of magnitude higher for phenylthiopentafluorobenzene compared to that with pentafluorobenzene (Figure 1.10B). The thioether substituent was activated toward attack at the *o*- and *p*- positions due to the resonance stabilization of the negative charge in the σ-complex intermediate (Figure 1.2A). Regioselective *p*-substitution was favored because the *o*-positions were hindered by the relatively large thiophenyl group.
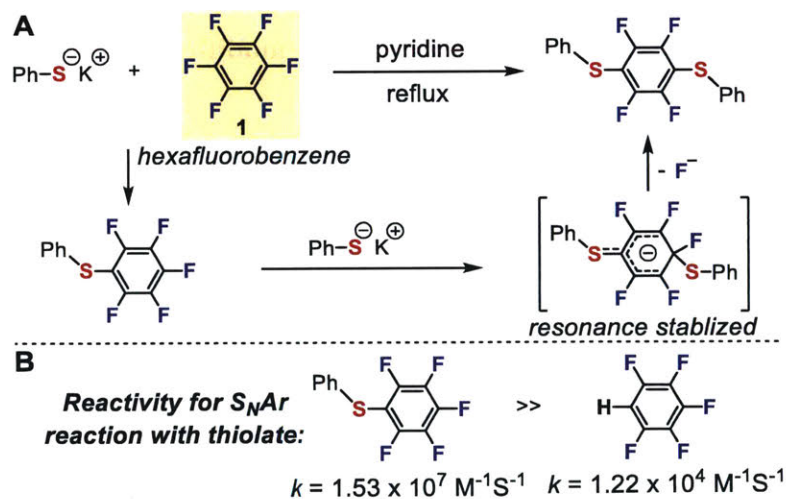
Figure 1.10. Nucleophilic aromatic substitution reactions between thiol species and hexafluorobenzene.

In the following decades, applications of the perfluoroaryl-thiol reaction mainly focused on synthetic molecules, especially modifications of porphyrinoids[87] and synthetic polymers[88–90] (Figure 1.11). In 1991, Mansuy reported SNAr reactions of tetra-(pentafluorophenyl)porphyrin with various nucleophiles.[91] Using triethylamine as the base, in refluxing DMF, *p*-fluorines on the tetra-(pentafluorophenyl)porphyrin were selectively replaced with thiol nucleophiles (Figure

1.11A). This efficient reaction provided a facile and divergent way to modify the peripheries of different porphyrinoid cores, enabling the synthesis of various modified porphyrinoids for applications in catalysis[92], materials science[93], and as drug carriers[94–96].
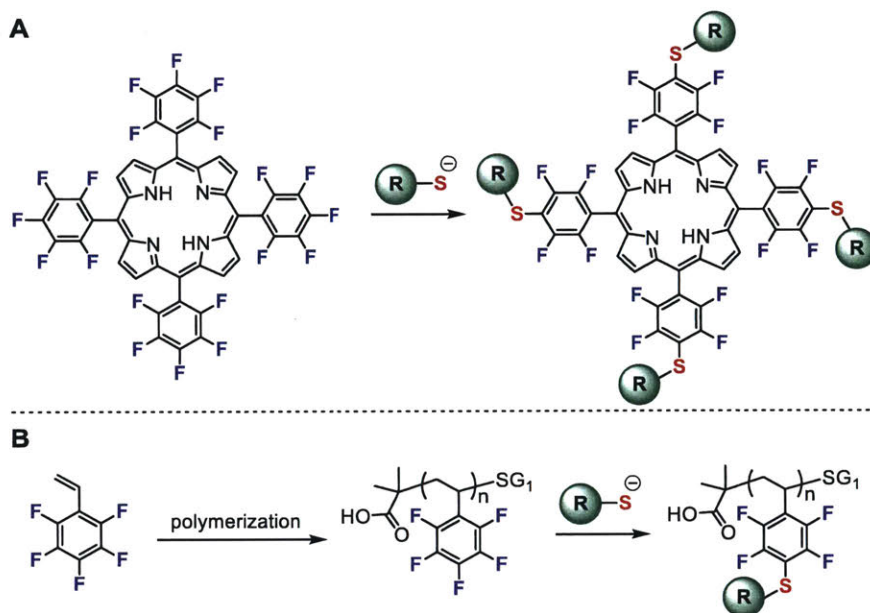
Figure 1.11. Perfluoroaryl-thiol S$_N$Ar reaction for modification of porphyrins (A) and polymers (B).

In a series of manuscripts, Wooley was first to demonstrate how polymers containing pentafluoroaryl groups can be functionalized and cross-linked via S$_N$Ar modification chemistry.[97–99] Following these seminal contributions, Schubert developed methods to modify perfluoroaromatic polymers with thiol-containing sugars and suggested that the perfluoroaryl-thiol reaction can be classified as a click process due to its high efficiency and functional group tolerance (Figure 1.11B).[89,100]

More recently, this perfluoroaryl-thiol click reaction was used by Spokoyny to generate precisely modified boron clusters.[101] Through peralkylation of the boron cluster, they were able to make rigid and three-dimensionally defined nanomolecules featuring pentafluorophenyl handles (Figure 1.12A). These perfluoroaryl boron clusters then served as templates for further diversification with thiol-containing molecules including alkanes, arenes, alcohols, amines, peptides, and sugars (Figure 1.12B). The conjugation reactions were generally very efficient, quantitatively producing multivalent nanoclusters with defined numbers and orientation of the attached ligands. Nanomolecules grafted with glucose exhibited increased binding affinity to

Concanavaline A (ConA, Figure 1.12B), showcasing how this strategy allows the rapid generation of precise multivalency for molecular recognition. The simplicity of this approach is reminiscent of the assembly of thiol-coated gold nanoparticles (AuNPs) but also possesses additional advantages such as atomic precision and full covalency.
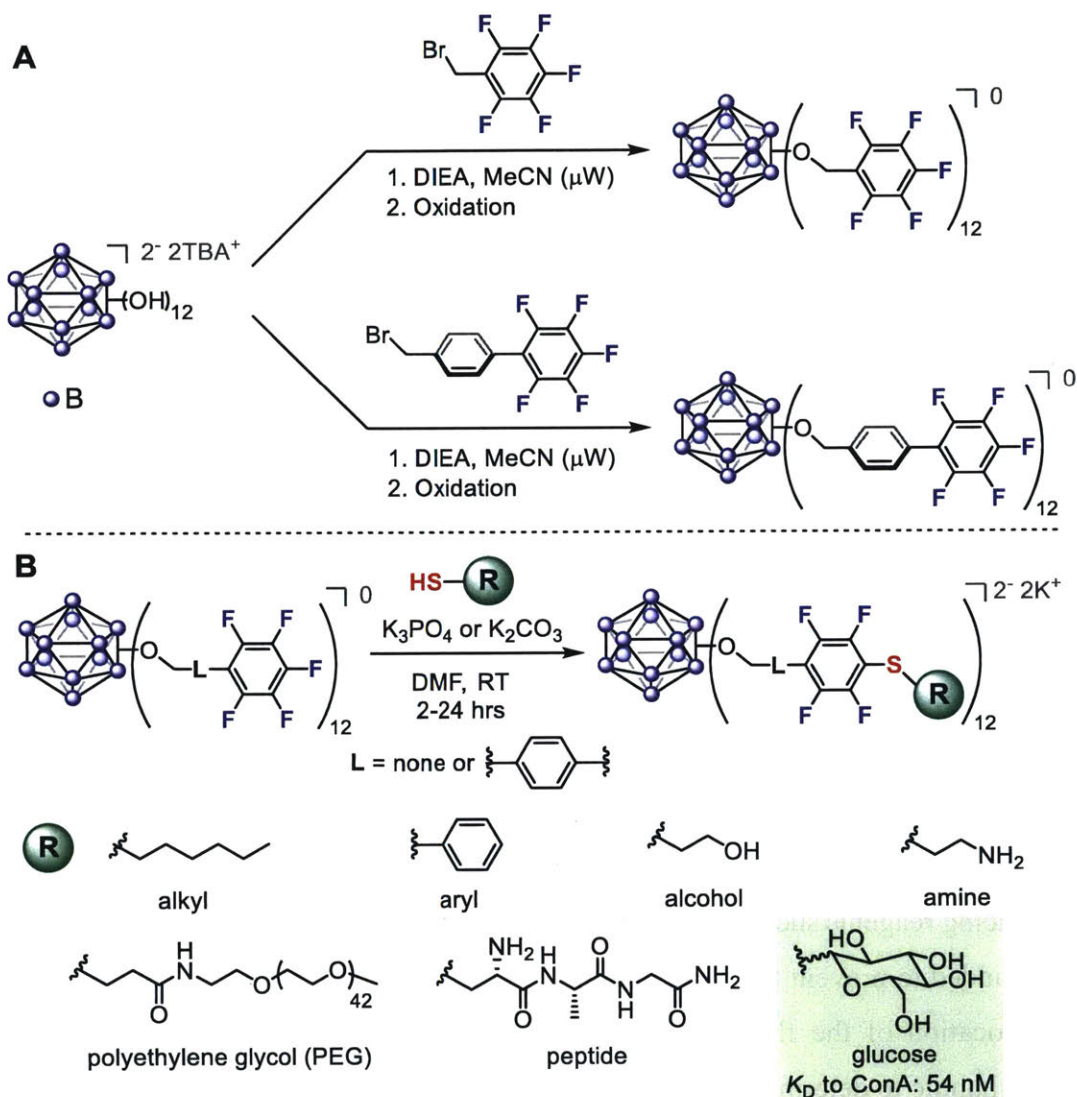
Figure 1.12. Atomically precise functionalization of boron clusters via perfluoroaryl-thiol conjugation reaction.

(A) Boron clusters were modified with perfluoroarenes. (B) Various chemical structures were attached quantitatively to the perfluoroaryl-linked boron cluster. Synthesized sugar modified boron cluster has significantly increased binding affinity to Concanavalin A, presumably because of the multivalency of the modified boron cluster.

Pentelute was able to utilize the same basic principles describe above, to develop perfluoroaryl-thiol SNAr reactions for bioconjugation.[102–108] Although the perfluoroaryl-thiol reactions showed good functional group tolerance in the modification of synthetic molecules (Figure 1.11), their chemoselectivity in a bioconjugation setting had not been extensively studied. As a starting point, reactions between various amino acids and commercially available hexafluorobenzene and decafluorobiphenyl reagents were explored.[102] These simple perfluoroarenes were found to react with the thiol of cysteine but not with other nucleophilic components of amino acid side chains such as the ε–amino group in lysine or the imidazole group in histidine. Consistent with the previously reported activating effect of a *p*-thioether group, the 1,4-disubstituted regioisomer was exclusively observed when allowing hexafluorobenzene to react with an excess of cysteine. The monoarylated product was observed when a large excess (>10-fold) of hexafluorobenzene was used with peptides containing either one or several cysteines. A similar substitution pattern was observed for the reaction between decafluorobiphenyl and cysteine. These results prompted further study of this reaction for the modification of peptides.

Hexafluorobenzene and decafluorobiphenyl both efficiently crosslinked two cysteine thiols in unprotected peptides (Figure 1.13A).[102] When an excess amount of the perfluoroarene was used, monosubstituted product was the major product formed (Figure 1.13B). These arylation reactions were fast and usually complete within hours at room temperature. Organic solvents such as dimethylformamide (DMF), acetonitrile, dimethylsulfoxide (DMSO) can be used. Many mild organic and inorganic bases such as triethylamine and potassium phosphate could be employed. Disulfide reducing reagents such as triphenylphosphine and TCEP are compatible with the reaction and could be added to prevent the oxidation of cysteine thiol. The use of ^{19}F-NMR, which clearly indicate the location of the fluorine atoms on the aromatic ring, provide a convenient NMR spectroscopic handle to determine the result of these transformations.
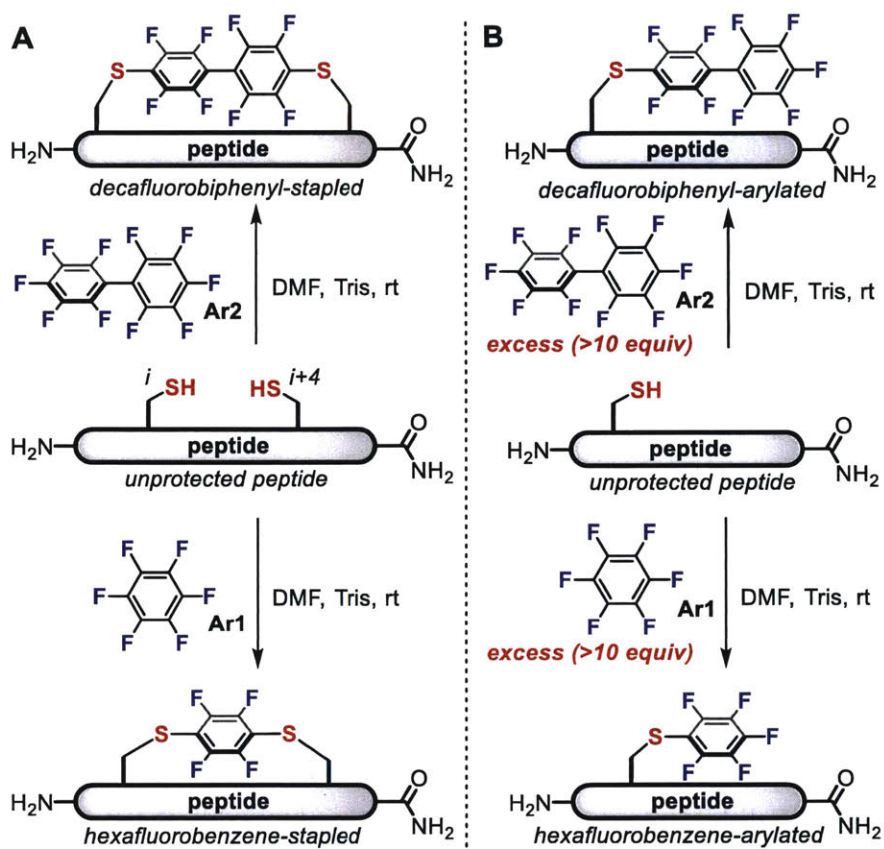
Figure 1.13. First generation perfluoroarenes for peptide modification: hexafluorobenzene and decafluorobiphenyl.

(**A**) Peptide stapling with perfluoroarenes developed by Pentelute. (**B**) Arylation of peptides containing a single cysteine using an excess (>10 equiv.) of perfluoroarenes.

Based on the initial results using commercial perfluoroarenes, Pentelute designed and synthesized four additional perfluoroaryl linkers all of which possess thioether substitutions for activation of p-fluorines.[104] These customized perfluoroaryl linkers were used together with three commercially available perfluoroarenes to cyclize two cysteine thiols positioned on different sites on a peptide chain (Figure 1.14).[104] Fourteen unprotected peptides with cysteine residues positioned from $i, i+1$ to $i, i+14$ sites were cyclized using seven perfluoroaryl linkers (Figure 1.14), producing a total of 98 peptide macrocycles. This convergent diversity-oriented macrocyclization scan highlighted the efficiency, selectivity, and robustness of the perfluoroaryl-cysteine S_NAr reaction for peptide modification.

14 peptides + **7 perfluoroaryl linkers**

98 peptide macrocycles
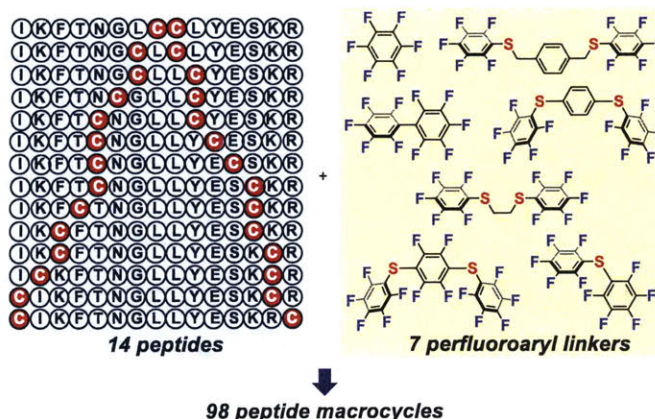
Figure 1.14. Peptide macrocyclization scan with commercially available and customized perfluoroaryl linkers.

The macrocyclic peptide structures availabe were further expanded through a two-step peptide macrocyclization strategy (Figure 1.15).[104] First, using an excess of the simple perfluoroarene, both cysteine thiols in an unprotected peptide were arylated to generate a peptide intermediate with perfluoroarylated cysteine residues. These were further crosslinked using commercially available dithiol reagents. This strategy might enable the synthesis of a range of structurally diverse macrocyclic peptides using commercially available dithiols.
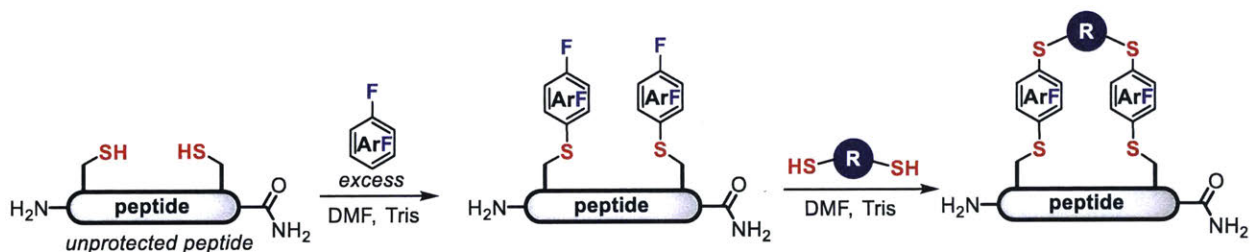
Figure 1.15. Macrocyclization of perfluoroarylated cysteines with dithiol reagents.

Pentelute and several other groups subsequently applied the perfluoroaryl-cysteine S$_N$Ar chemistry to the macrocyclization of bioactive peptides (Table 1.1). Compared to their linear peptide counterparts, macrocyclic or stapled peptides usually have enhanced biological properties such as stability against proteases and binding affinity to the target. Pentelute reported a stapled peptide inhibitor of the C-terminal HIV-1 capsid assembly protein (C-CA) and observed enhanced binding affinity to C-CA and increased stability against trypsin and chymotrypsin digestion (Table 1.1, entry 1).[102] Importantly, enhanced cellular uptake of the stapled peptide was observed, indicating that macrocyclization using perfluoroaryl linkers might render peptides cell permeable.

Pun recently sampled various macrocyclization strategies for a peptide that targets M2 macrophage. Decafluorobiphenyl was found to be the optimum stapling reagent to enhance the binding affinity and the serum stability of the peptide (Table 1.1, entry 2).[109] Schroeder used hexafluorobenzene to synthesize a macrocyclic peptide as tumor-targeting reagent and found that the perfluoroaryl motif was not cytotoxic (Table 1.1, entry 3).[110] Several GLP-1 receptor agonists were cyclized using perfluoroaryl linkers by Craik.[111] However, these macrocyclic peptides showed reduced potency compared to their linear analogs (Table 1.1, entries 4-6). Collectively, perfluoroaryl-cyclized peptides are in most cases more stable than their linear counterparts. The effect of perfluoroaryl linkers on the binding affinity of macrocyclic peptides differs as a function of the peptide structure presumably because the linker itself can directly engage or block the binding of the peptide to its target. This, however, is not an unique property of the perfluoroarene staples, and other carbon-based linkers have shown to exhibit these sequence-dependent variabilities.[112,113]

Table 1.1. Perfluoroaryl-linked macrocyclic bioactive peptides.

Entry	Peptide sequence	Linker	Function	Properties compared to a linear peptide	Ref.
1	ITF**C**DLL**C**YYGKKK	**Ar1** **Ar2**	Bind HIV-1 capsid assembly protein (C-CA)	• Enhanced binding affinity • Increased cellular uptake • Increased proteolysis stability	102
2	**C**GYEQDPWGVRYWY G**C**kkk(K-biotin)	**Ar2**	Bind M2 macrophage	• Increased binding affinity • Increased Serum stability	109
3	**C**GNKRTRG**C**	**Ar1**	Tumor targeting	• Noncytotoxic	110
4	His-aa1-Glu-Gly-**Cys**-aa2-Thr-Ser-**Cys**-aa3-aa4	**Ar1**			
5	His-aa1-Glu-Gly-Thr-aa2-Thr-**Cys**-Asp-aa3-aa4-**Cys**	**Ar1**	GLP-1 receptor agonist	• Reduced potency	111
6	His-aa1-Glu-Gly-**Cys**-aa2-Thr-Ser-Asp-aa3-aa4-**Cys**	**Ar2**			

* aa1: 2-aminoisobutyric acid; aa2: (S)-2-Fluoro-α-methylphenylalanine; aa3: (S)-2-amino-3-(2'ethyl-4-methoxy-[1,1'-biphenyl]4-yl)propanoic acid; aa4: (S)-2-amino-5-phenylpentanoic acid

The perfluoroaryl-cysteine S$_N$Ar reaction is also compatible with the chemical synthesis and folding of proteins. Two small proteins, affibody and human relaxin, were modified with perfluoroaryl linkers by combining perfluoroaryl-cysteine S$_N$Ar reaction with fast flow peptide synthesis and native chemical ligation techniques (Figure 1.16).[102,106] Pentelute installed a perfluoroaryl staple on the C-terminal helix of an affibody targeting human epidermal growth factor receptor 2 (HER-2) (Figure 1.16A).[102] The three segments of affibody were ligated through two sequential native chemical ligation (NCL) reactions. The resulting folded perfluoroaryl-stapled affibody has similar circular dichroism (CD) and binding properties compared to the unmodified affibody. Furthermore, the perfluoroaryl linker could be used as an irreversible disulfide surrogate through replacing the disulfide between Cys-10 and Cys-15 on the A-chain of human relaxin (Figure 1.16B).[106] Upon oxidative folding with the B-chain, a perfluoroaryl relaxin was obtained. Using a cell-based assay, modified relaxin was measured to be active, though with a diminished (100-fold) potency compared to the wild type species.
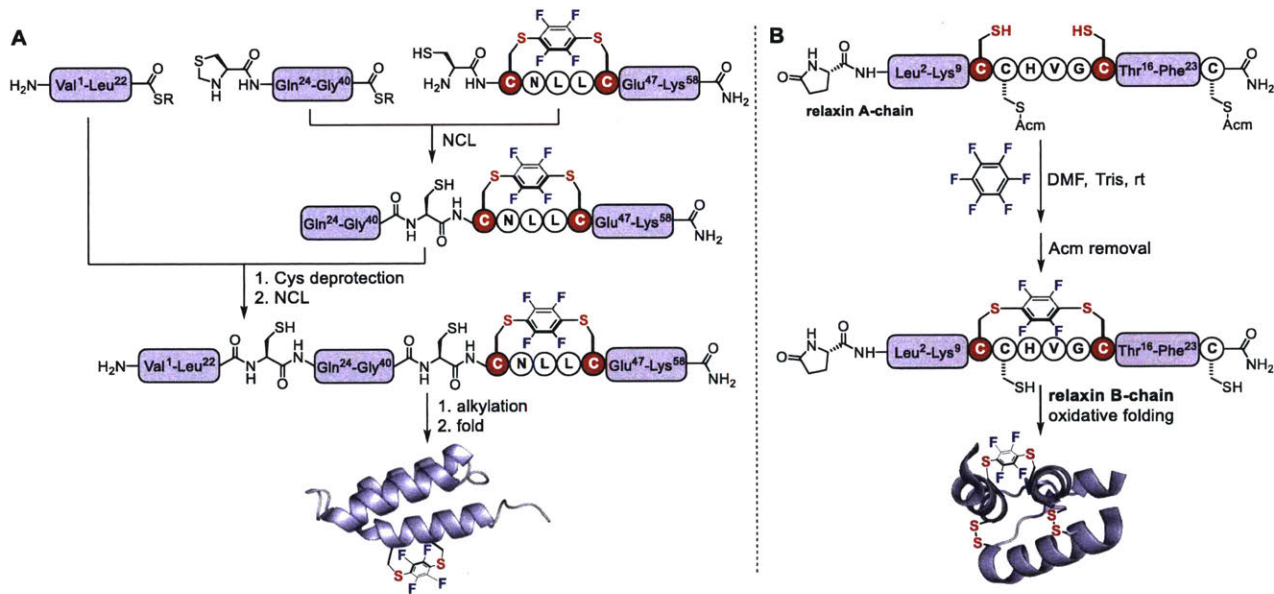
Figure 1.16. Total chemical synthesis of perfluoroaryl-modified affibody (A) and human relaxin (B) proteins.

The perfluoroaryl-cysteine S$_N$Ar reaction is efficient in polar organic solvent but in many cases is very sluggish in aqueous media. Moreover, the low solubility of previously-described perfluoroaryl reagents (Figure 1.14) limits their utility in aqueous media. In an attempt to use perfluoroarenes in aqueous media, Derda identified decafluorobiphenylsulfone as a very reactive electrophile for the arylation of cysteine thiol under aqueous conditions.[114] A reaction rate of 180

M^{-1}s^{-1} was observed when treating a pentapeptide (Ser-Trp-Cys-Arg-Cys) with decafluorobiphenylsulfone in Tris buffer (pH 8.5) with 40% acetonitrile. The increased reactivity of decafluorobiphenylsulfone in aqueous media enabled its application to the modification of biomolecules sensitive to organic solvents. Using this reagent, Derda macrocyclized peptide libraries displayed on the surface of M13 phage with good efficiency (Figure 1.17). This paved the way to further use this peptide stapling reaction and phage display to discover bioactive macrocyclic peptides.
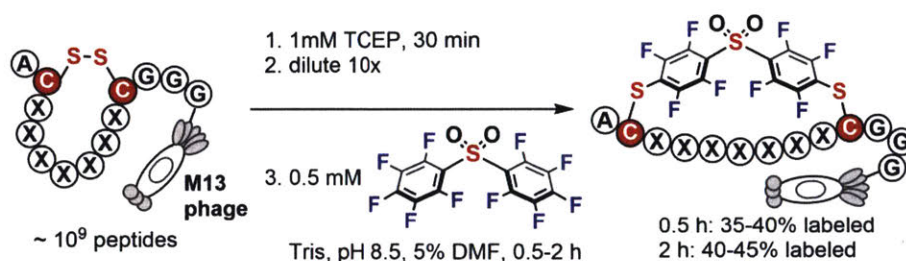
Figure 1.17. Macrocyclization of phage-displayed peptides via cysteine perfluoroarylation with perfluoroarylsulfone reagent.

Another approach to increase the reactivity of perfluoroarenes in water is through proximity-induced reactions. Wang developed a photoswitchable amino acid containing an pentafluorophenyl group for cysteine reaction and an azo bridge for photoswitching (Figure 1.18).[115] This amino acid was introduced into proteins via genetic code expansion, allowing the pentafluorophenyl group to react with a nearby cysteine to form an azo bridge *in situ* during expression of the protein in *E. Coli*. The fluorines changed the photophysical properties of the azo bridge, enabling the photoswitch unit to isomerize under visible light. Applying this strategy to modification of calmodulin enabled the photocontrol of the conformation and binding of calmodulin. The expressed protein was a mixture of desired azo-bridged protein, the unreacted protein, and a side product with the perfluoroaryl reacted with glutathione, indicating that further optimization of the reaction to increase the reaction rate and selectivity is required to generate the desired product with higher yields.
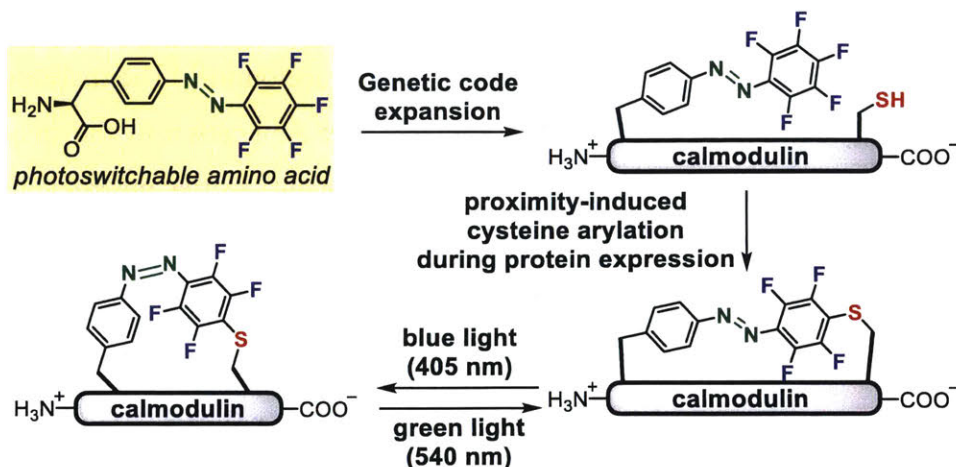
Figure 1.18. Proximity-induced cysteine arylation and genetic code expansion for introduction of photoswitchable azo bridges into proteins.

Cysteine perfluoroarylation was also explored for introducing ^{18}F radiolabels into biomolecules. Chen synthesized perfluoroaryl linked dimers of an integrin binding peptide using ^{18}F-hexafluorobenzene generated from ^{18}F exchange with a non-radiolabeled hexafluorobenzene (Figure 1.19).[116] The ^{18}F-hexafluorobenzene reacted with a thiol containing cyclic peptide (cyclo(RGDfK)) that binds to cell surface integrin. About 50% conversion to the desired product was observed after 20 min reaction at room temperature. The ^{18}F-labeled dimeric peptide was obtained with 40% radiochemical yield after HPLC purification. The perfluoroaryl-linked dimeric peptide showed similar binding affinity to $\alpha_v\beta_3$ integrin as the original monomer. In addition, this dimer showed good tumor uptake and high tumor-to-background contrast in positron emission tomography (PET) imaging of a tumor xenograft mice model.
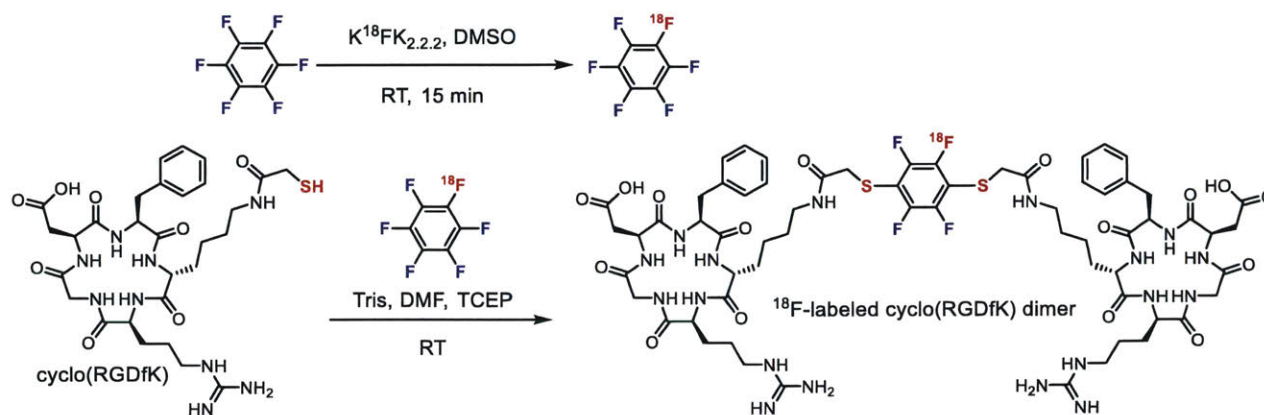
Figure 1.19. Introduction of ^{18}F radiolabel through cysteine perfluoroarylation for PET imaging.

48

(b) Chemo- and regio-selective cysteine S$_N$Ar arylation

Reactions on proteins that contain several solvent-exposed cysteine sites usually lead to mixtures of products with undefined position and number of modifications. For example, maleimide conjugation was previously used to conjugate drug molecules to reduced interchain cysteines in antibodies.[117] The resulting antibody-drug conjugates (ADCs) form mixtures of products with drug-to-antibody ratio (DAR) ranging from 1 to 8. Even for molecules with the same DAR, the position of drug attachment might differ. Each individual member in this ADC mixture might have different therapeutic properties, which presents a significant challenge for systematic studies of ADCs and calls for the development of site-selective chemistry for ADC synthesis.

Although many chemoselective cysteine bioconjugation methods have been developed, regioselective cysteine conjugations are rare. Caddick and co-workers found that one of the two cysteine residues in superfolder green fluorescent protein can selectively react with 2,5-dibromohexanediamide.[118] In certain instances, the N-terminal cysteine could be selectively modified via native chemical ligation, albeit the final ligation product is an amide bond rather than modification on the cysteine thiol.[119–122] Other more sophisticated strategies using protecting groups,[123] enzyme catalysts,[124–126] or multiple chemical steps[127] were also developed for regioselective cysteine conjugation.

Nature uses enzymes to control the chemo- and regio-selectivities of chemical reactions including highly precise post-translational modifications.[128,129] In the case of cysteine arylation, Nature uses glutathione S-transferase (GST) to perform highly selective arylation reactions. GST is a family of promiscuous enzymes that catalyze the conjugation of electrophilic xenobiotic substrates to the thiol of a tripeptide glutathione (γ-Glu-Cys-Gly, GSH) for the purpose of detoxification.[130,131] Aryl halides are a known family of electrophiles accepted by GST. In fact, GST-catalyzed conjugation of 1,4-dinitrochlorobenzene (DNCB) to GSH is a routinely used method for a standard colorimetric assay to assess the enzymatic activity of GSTs (Figure 1.20A).[132–134] Upon arylation of GSH with DNCB under enzyme-catalysis, the arylated product has an increased UV absorbance at 340 nm, allowing the determination of the enzymatic activity of a broad range of GST isozymes.
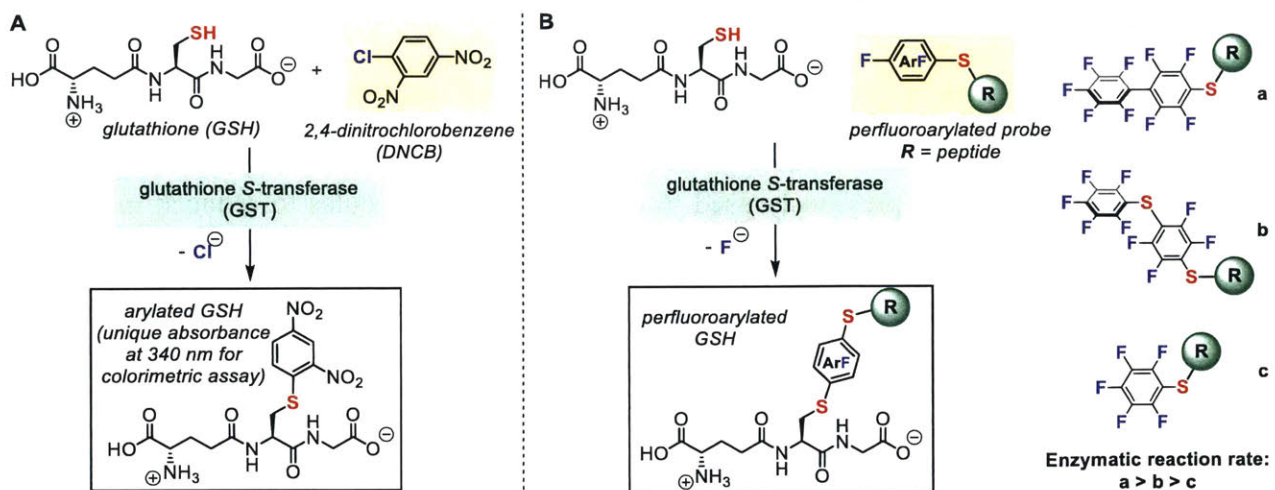
Figure 1.20. Glutathione *S*-transferase-catalyzed arylation of cysteine thiol in glutathione. (**A**) GST catalyzes the arylation of glutathione with 2,4-dinitrochlorobenzene. (**B**) GST catalyzes the perfluoroarylation of glutathione with various peptide-based perfluoroaryl electrophiles.

Because of the substrate promiscuity of natural GSTs, Pentelute hypothesized that these enzymes can be directly hijacked to catalyze the S$_N$Ar reaction between perfluoroarenes and glutathione as a way to improve the rate of previously sluggish perfluoroarylation reactions in aqueous solution (Figure 1.20B). Perfluoroarenes such as hexafluorobenzene and decafluorobiphenyl have very low solubility in water, however one can prepare several water-soluble peptide-based perfluoroaryl electrophiles by treating cysteine-containing peptides with an excess of perfluoroarenes (Figure 1.20B). These peptide-based perfluoroaryl electrophiles were conjugated to GSH using a mixture of commercially available natural GST isozymes.[103] Importantly, no reactions were observed in a control experiment without enzyme. GSTs purified from different sources displayed similar activity and the decafluorobiphenyl-based electrophile was found to be the most reactive electrophile under GST catalysis (Figure 1.20B). It is worth noting that many glutathione *S*-transferases contains cysteine residues,[130] therefore, the GST-catalyzed reaction provides a potential approach for selective cysteine modification in the presence of other thiol species.

To further explore GST-catalyzed reactions for protein modification, GSH was installed as a peptide tag on the protein of interest (Figure 1.21).[103] Inspired by a crystal structure of GSH bound to GST (Figure 1.21A),[135] Pentelute hypothesized that appending the protein of interest to the C-termini of GSH might not significantly affect the recognition of GSH by the enzyme. Such

a design resulted in an *N*-terminal GSH tag for protein modification. When GSH was incorporated into the peptide or protein of interest that containd endogenous cysteines, the GST enzyme might catalyze the regioselective conjugation of perfluoroaryl electrophiles to the thiol of GSH in the presence of other competing cysteine residues on the same molecule (Figure 1.21B). Indeed, it was found that several model peptides with the *N*-terminal GSH tag efficiently conjugated to perfluoroaryl probes under GST catalysis irrespective of the structure of amino acid directly linked to the glycine of the GSH tag.
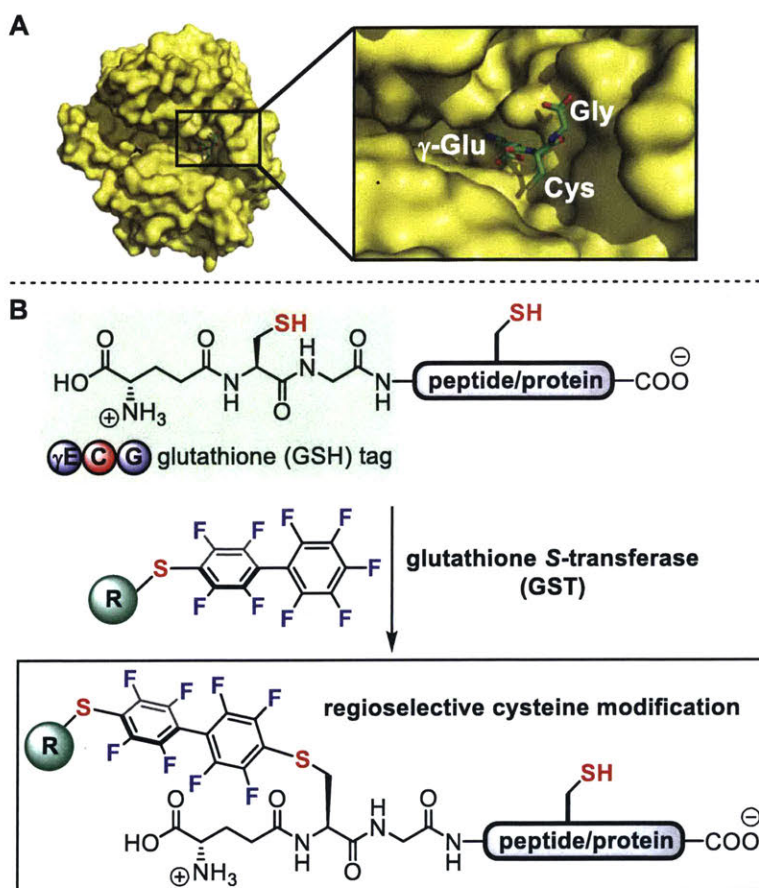
Figure 1.21. GST-catalyzed regioselective cysteine perfluoroarylation.

(**A**) A crystal structure of GSH bound to GST dimer shows that the Gly moiety of GSH is exposed as a possible site for attaching peptide/protein without compromising the recognition of GSH by GST. (**B**) Attaching a three-amino acid glutathione (GSH) tag at the *N*-terminus of the peptide or protein of interest enables recognition by GST and selective conjugation to perfluoroaryl-linked reagents.

The GST-catalyzed regioselective cysteine conjugation enables applications to protecting group-free dual protein modification (Figure 1.22A) and regioselective macrocyclization of long unprotected peptides (Figure 1.22B).[103,105] The GSH tag can be installed on the *N*-terminus of a model protein through native chemical ligation. The resulting protein was dual-labeled with biotin and fluorescein through two sequential reactions. The GSH cysteine was first modified using a perfluoroaryl biotin probe under GST catalysis, then the other unprotected cysteine was labeled with maleimide fluorescein without purification of the previous product (Figure 1.22A). The GSH tag was also installed on the *N*-terminus of peptides that already possess *C*-terminal perfluoroarylated cysteines. Under GST catalysis, macrocyclic peptides with lengths up to 40-residues were readily synthesized (Figure 1.22B). In contrast, cyclization reactions without enzyme produced polymers and other side products.
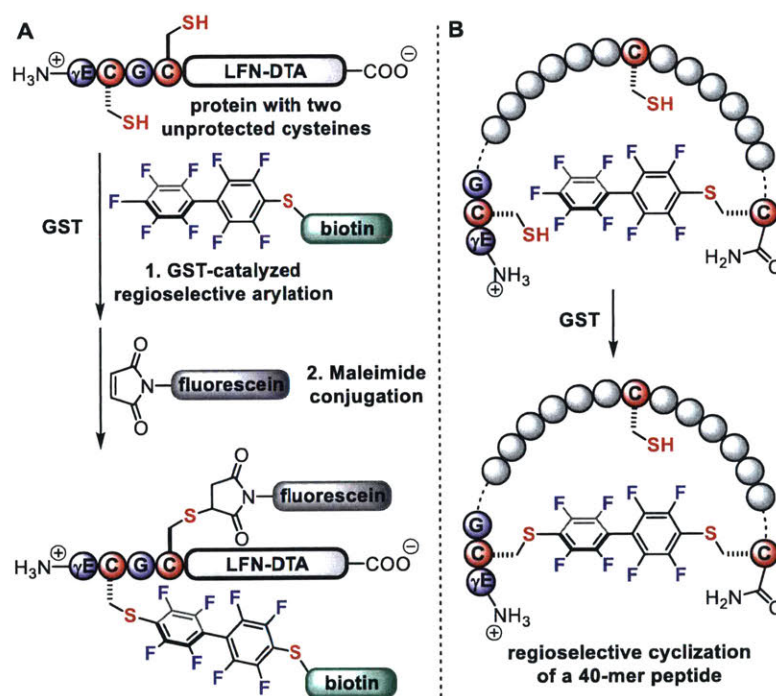
Figure 1.22. Applications of GST-catalyzed regioselective cysteine perfluoroarylation. (**A**) Protecting-group-free selective dual-labeling of a protein. (**B**) Enzyme-catalyzed regioselective macrocyclization of a long peptide.

One limitation of the GST-catalyzed cysteine conjugation is the requirement for a GSH tag which has a γ-glutamic acid that is not genetically encoded. To overcome this limitation, a fully genetically encodable peptide tag that promotes cysteine arylation in water would be highly

desirable. Pioneering work by Tsien described peptide tags containing multiple cysteines to entropically promote selective conjugation to biarsenic reagents.[136,137] More recently, arsenic-based reagents were used to modify two cysteines generated from the reduction of a disulfide.[138] Because of previous observation that different protein microenvironments led to different cysteine reactivity,[58,118] a new strategy was envisioned for site-selective cysteine arylation though fine-tuning the arylation reactivity of a cysteine within a small peptide sequence.

Pentelute discovered that a four-residue peptide sequence (Phe-Cys-Pro-Phe), which is referred to as the "π-clamp" greatly promotes cysteine perfluoroarylation in water (Figure 1.23A).[107] The reaction rate between a π-clamp peptide and a perfluoroaryl probe was $0.76 \ M^{-1}s^{-1}$ which was more than 1000-fold faster than the reaction of a control cysteine-containing peptide. Molecular dynamics (MD) simulation and density function theory (DFT) calculations showed that the π-clamp sequence promoted the reaction both kinetically (lowering the activation energy) and thermodynamically (generating a more stable product) as compared to the control.

The π-clamp was applied to site-specific modification of antibodies (Figure 1.23B).[107] The π-clamp is small and composed entirely of genetically encoded amino acids, enabling the convenient insertion of π-clamp into the C-terminus of antibody heavy chains. The expressed π-clamp antibodies readily conjugated to perfluoroaryl linked probes in a single step under reducing conditions. Only π-clamp cysteine thiols were labeled and other reduced thiols from the antibody interchain disulfides were intact under the reaction conditions. The resulting site-specific antibody conjugates retained their target binding affinity, indicating the insertion of π-clamp and the reaction conditions didn't significantly affect the structure and function of the antibody. A site-specific π-clamp antibody-drug conjugates showed receptor-dependent cell killing. Compared to other site-selective antibody modification methods that require either unnatural amino acids,[139] enzymes,[140–142] or multiple chemical steps,[143,144] the π-clamp provides a single-step method for site-selective antibody modification and shows great promise for novel site-specific antibody-drug conjugates.
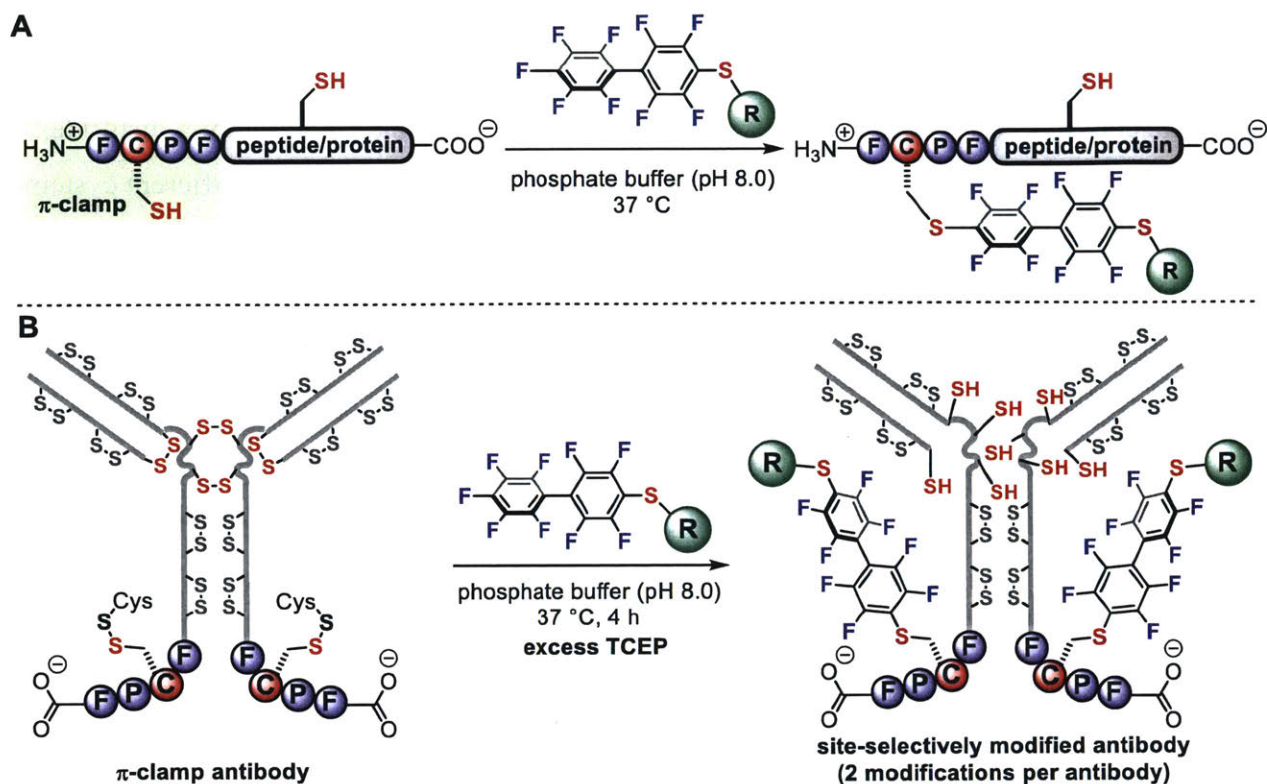
Figure 1.23. π-Clamp-mediated regioselective cysteine modification.

(A) π-Clamp mediates selective arylation on its cysteine residues in the presence of other competing cysteine residues. (B) π-Clamp-mediated site-specific antibody modification.

During the investigation of the π-clamp, Pentelute discovered that adding simple inorganic salts significantly changed the reaction rate of the bioconjugation.[108] Strikingly, different salts can change the rate constant over four orders of magnitude for the π-clamp mediated arylation reaction (Figure 1.24). The observed trend of salt effect on the rate constant followed the Hofmeister series,[145,146] a phenomenon that had existed for more than a hundred years but has never been used for bioconjugation. Protein-compatible salts including ammonium sulfate were used to greatly enhance bioconjugation reactions such as π-clamp-mediated arylation, enabling the fast synthesis of functional antibody-drug conjugates. Importantly, the salt effect can be expanded to reactions beyond arylation; Pentelute showed that ammonium sulfate can also enhance the rate of an *alkylation* reaction, providing the first example of site-specific cysteine alkylation mediated by π-clamp. The advantage of salt effect is significant as well as practical for the modification of biomolecules, considering heating and concentrating are generally not compatible with delicate or easy to aggregate proteins.

condition	rate constant $(M^{-1}s^{-1})$
no salt	0.76
+ 1M sodium chloride	1.03
+ 3M guanidinium thiocyanate	0.002
+ 3M ammonium sulfate	62
+ 2M ammonium citrate	74
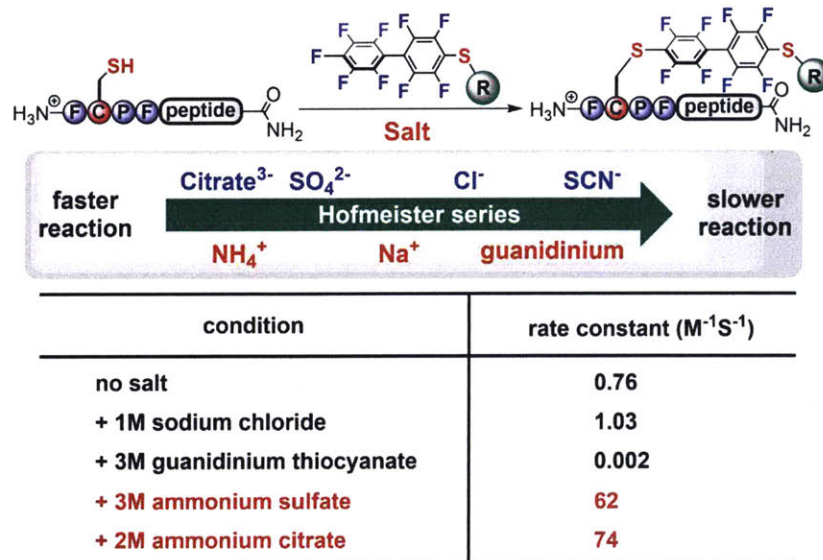
Figure 1.24. Salt effect on π-clamp mediated cysteine arylation.

1.2.2. Metal-mediated cysteine arylation

In the past two decades, metal-mediated bioconjugation reactions have become common tools in the field of chemical biology.[44,49,147–151] Metal-mediated arylation provides a complementary approach to nucleophilic aromatic substitution (S$_N$Ar), enabling an expanded scope of aryl modifications for peptides and proteins. A representative selection of the existing organometallic approaches toward the arylation of biomolecules is presented in different sections of this review below.

(a) Chemoselective metal-mediated cysteine arylation

Selective cysteine arylation using organometallic reagents represents an important example of metal-mediated arylative bioconjugation reactions. Metal-catalyzed C–S cross-coupling reactions have been used for small molecule synthesis by chemists for over three decades.[152–154] Yet, the first report on metal-mediated *S*-arylation of biomolecules appeared very recently, when in 2014 Wong used gold(III) complexes to transfer aryl groups to thiols on short peptides and bovine serum albumin (BSA, Figure 1.25).[155] The cyclometalated structure of the gold(III) complex required the use of arylpyridine units, which limited the scope of aryl structures that could be introduced using this method. Moreover, the slow reaction kinetics of the reductive elimination from gold(III) led to undesired and prolonged heating of the reaction which can be incompatible with certain delicate proteins.
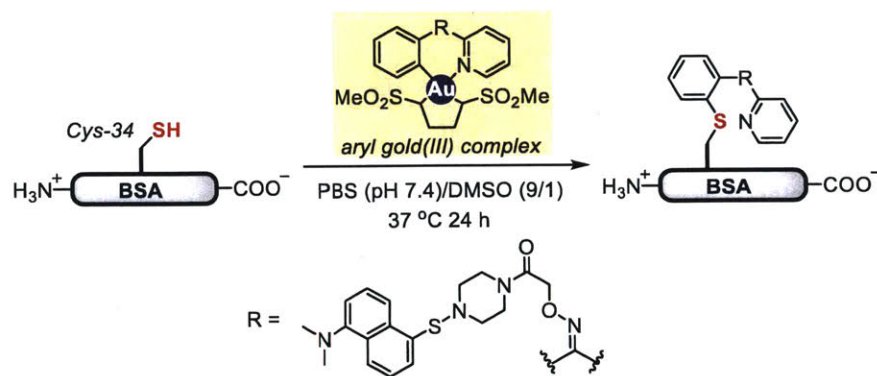
Figure 1.25. Gold(III)-mediated cysteine arylation.

Shortly after the work of Wong, Buchwald and Pentelute reported palladium-based organometallic complexes featuring biarylphosphine ligands as efficient and versatile reagents for chemoselective cysteine arylation (Figure 1.26).[156] The reaction proceeded at low micromolar concentrations of both the organometallic complex and the biomolecule, requiring no heating, and only 5% of common organic co-solvent (CH₃CN, DMF, DMSO) was necessary for quantitative conversions at room temperature within minutes. Uniquely, desired products could be formed under a wide range of pH (2-10), enabling labeling of peptide cysteine residues even in water containing 0.1% of trifluoroacetic acid (TFA).
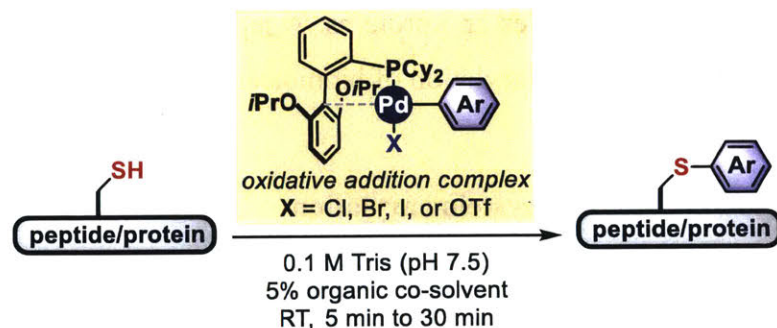
Figure 1.26. Organometallic palladium(II) reagents for cysteine arylation.

A variety of oxidative addition complexes could be accessed in high yields in one step from commercially available aryl halide or from trifluoromethanesulfonate precursors (Figure 1.27A).[156] These storable and air-stable arylation reagents were successfully applied to the cysteine arylation of unprotected peptides, small antibody mimetic proteins, bacterial toxins, and antibodies (Figure 1.27B). Diverse S-aryl and S-heteroaryl bioconjugates were synthesized to introduce fluorescent dyes, affinity tags, and bioconjugation handles into peptides and proteins (Figure 1.27). A comparative study of the resulting conjugates against maleimide and acetamide analogues

showed that *S*-aryl conjugates are more stable under basic, acidic, and oxidative conditions, as well as in the presence of high concentrations of glutathione. The reaction rate of palladium-mediated *S*-arylation was comparable to or slightly faster than maleimide at neutral pH and is significantly faster in acidic solutions in which maleimide reaction generates no product.
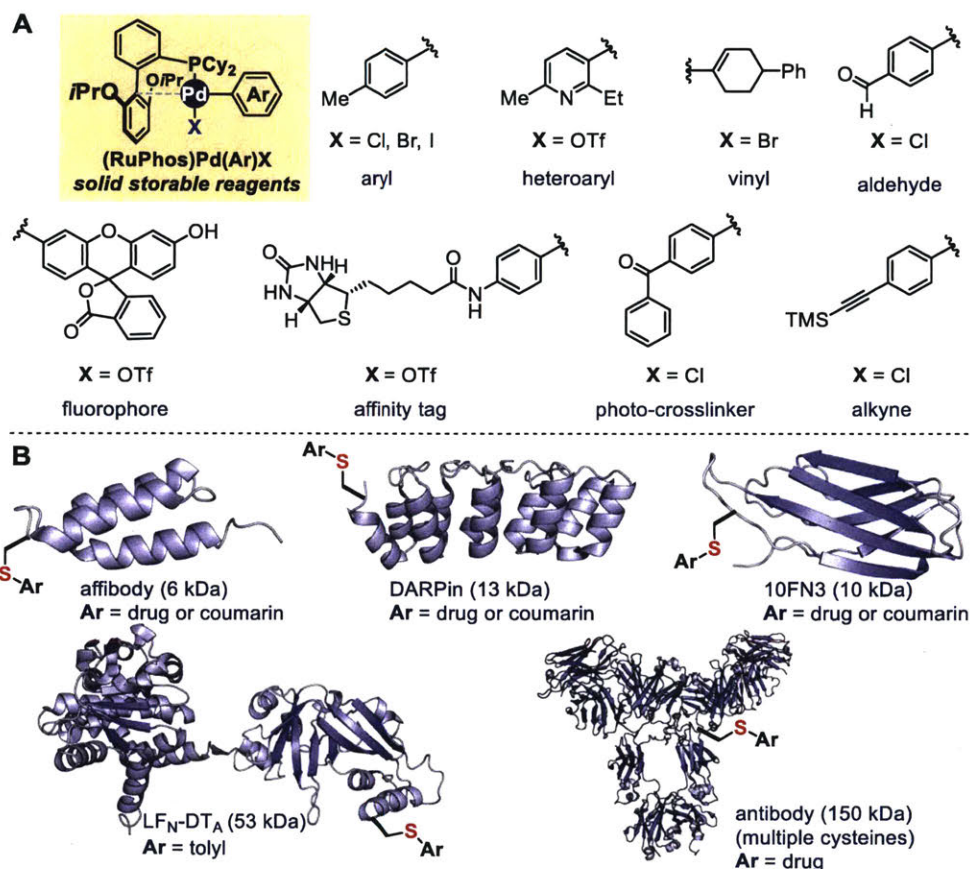
Figure 1.27. Organometallic palladium complexes (A) and protein targets (B) studied in the palladium-mediated cysteine arylation reaction.

Palladium reagents bearing two metal centers were used for peptide stapling without the formation of peptide oligomers (Figure 1.28).[156,157] The high efficiency of this transformation allowed access to a number of aryl-stapled peptides for a comprehensive study on the role of the linker hydrophilicity, rigidity, and overall substitution pattern on various peptide properties including lipophilicity, phospholipid affinity, unique volume of distribution, and target binding affinity (Figure 1.28).
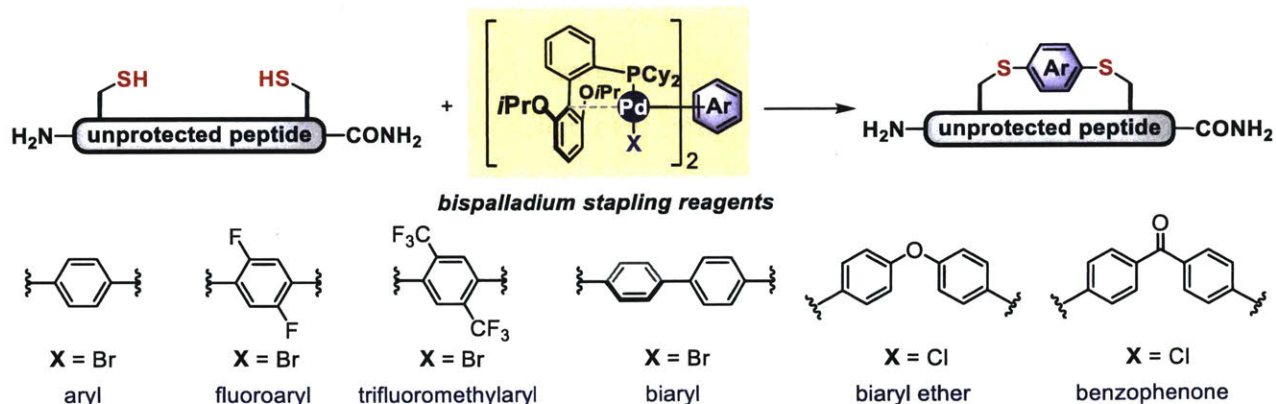
Figure 1.28. Organometallic palladium reagents for peptide stapling.

One of the advantages of palladium-mediated arylation is its compatibility with TCEP commonly used for the reduction of disulfides in proteins and antibodies.[158] This, combined with the mild reaction conditions allowed the application of this transformation for the efficient synthesis of "linker-free" antibody drug conjugates where the drug molecules were directly attached to cysteine thiols in antibodies (Figure 1.29).[156] The resulting ADCs has similar target binding affinity compared to the native antibody, indicating that the reaction conditions and the use of palladium reagents did not significantly affect the antibody function.
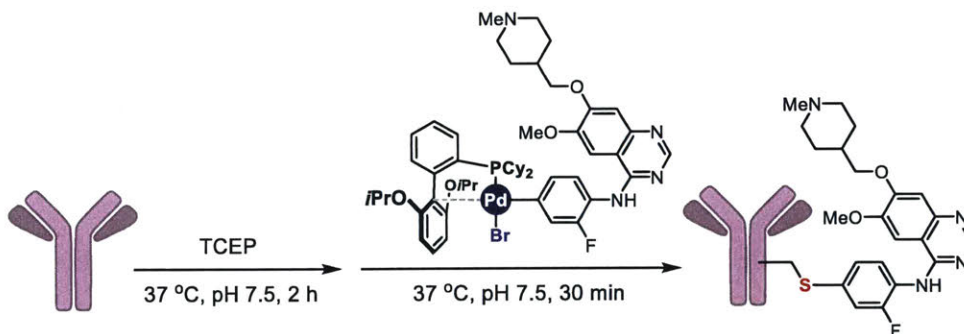
Figure 1.29. Organometallic palladium reagents for antibody-drug conjugation.

Messaoudi showed that medal-mediated *S*-arylative bioconjugation reactions could be performed using catalytic amounts of XantPhos-based aminobiphenyl mesylate palladium pre-catalysts (Figure 1.30).[159] Short unprotected peptides could be converted into the corresponding *S*-aryl bioconjugates using only 2% of the palladium precatalyst in greater than 80% yields. Trastuzumab antibody could also be labeled following pre-reduction with TCEP. However, the efficiency of this catalytic process and the resulting drug-to-antibody ratio were not investigated and potentially warrant further optimization of the catalyst system.

58
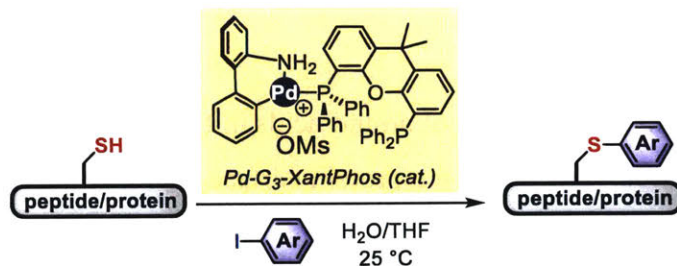
Figure 1.30. Palladium-catalyzed cysteine arylation.

(b) Chemo- and regio-selective metal site-guided cysteine arylation

An important challenge in the field of bioconjugation is to develop efficient reactions with tunable site-selectivity.[52] Organometallic complexes provide a number of promising entry points into the development of site-specific bioconjugation approaches, including the variation of biomolecule microenvironment, metal/ligand combinations, and leaving groups.

Davis used the protein microenvironment for the site-specific arylation of cysteine residues proximal to endogenous metal-binding sites in metalloenzymes.[160] The authors used a known metalloenzyme – mannosyl-glycerate synthase – to showcase that the arylation reaction could be directed with high regioselectivity to the Cys-233 proximal to the endogenous DxD metal-binding motif without affecting other three cysteine residues on the same protein molecule (Figure 1.31). While this initial report highlighted the potential of organometallic catalysts for site-specific bioconjugation, the relatively high reaction temperature (65 °C) can lead to protein denaturation and therefore limits the breadth of potential applications of this method.
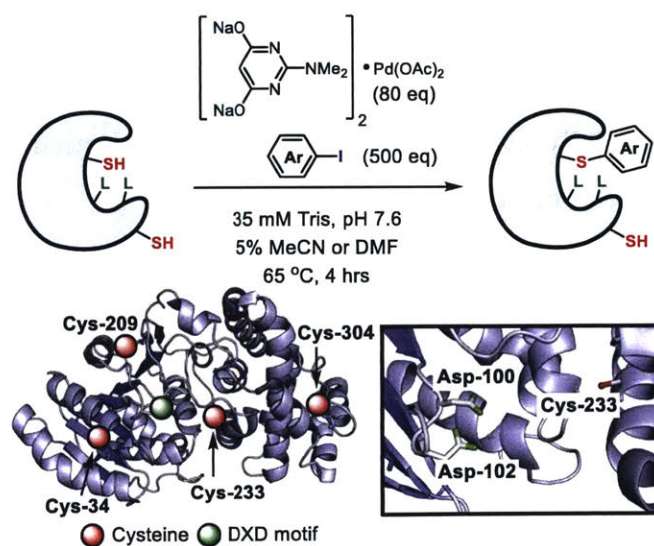
Figure 1.31. Metal-site guided cysteine arylation.

59

1.3. Lysine arylation

Lysine is one of the most abundant amino acids in the human proteome. There is usually more than one solvent-exposed lysines on a protein surface, making it difficult to site-selectively modify one lysine side chain among many. The lysine ε-amino group is more nucleophilic than the N-terminal amine. However, the protonated form of lyseine ε-amino group (pK_a ≈ 10) is less acidic than that of the N-terminal amine (pK_a ≈ 8). When both amines are neutral at higher pH, the lysine ε-amino group might be selectively modified without reaction on the protein N-terminus. At a lower pH, however, the lysine ε-amines are protonated while the free N-terminal amino group can be modified.

Bioconjugation on the ε-amine of lysine is a routine strategy to chemically modify proteins when site-selectivity does not pose a strict requirement.[23,42] Classic methods use electrophiles such as activated esters, isocyanates, and sulfonyl chlorides. Many lysine-based bioconjugates are used as vaccines and biotherapeutics. For example, the first FDA-approved PEGylated protein therapeutic ADAGEN® was made from conjugation of PEG polymers to lysine ε-amines in adenosine deaminase using succinic esters.[161]

N-arylation is a powerful approach to construct N–C(sp^2) bonds in numerous drug molecules and complex natural products. A wide range of electron-deficient aryl halide electrophiles are used to modify amines through nucleophilic aromatic substitution (S$_N$Ar) reactions. More routinely used methods for the arylation of amines are transition-metal catalyzed cross-coupling reactions such as the Buchwald-Hartwig reaction,[36,162,163] the Chan-Lam reaction,[164–166] and Ullmann-Goldberg coupling reactions.[167–170] Many aryl halides and aryl boronic acids are commercially available, providing a wide selection of structures for amine modification using these metal-catalyzed reactions.

Compared to its extensive development for small molecule synthesis, N-arylation is much less explored for the modification of large biomolecules. One of the early examples of the N-arylation of biomolecules is Sanger's reagent for protein sequencing (Figure 1.32).[171] In 1945, Sanger reported the reaction between 2,4-dinitrofluorobenzene (DNFB) and the protein N-terminal amino groups. The DNFB-modified proteins were hydrolyzed under acidic conditions, generating dinitrophenyl-amino acids that were identified through either chromatographic or colorimetric

assays.[171] The selectivity of 2,4-dinitrofluorobenzene for amino group is moderate, as cysteine thiol, histidine imidazole, and tyrosine phenol groups can also react with DNFB.
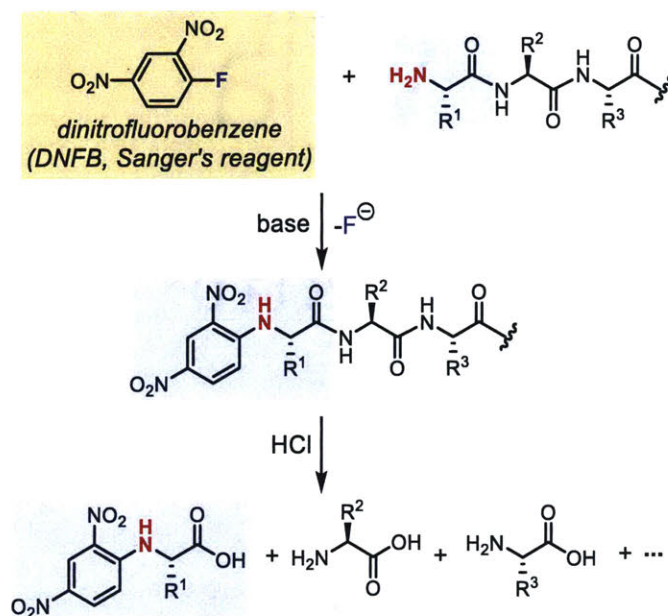
Figure 1.32. Sanger's reagent for peptide sequencing.

1.3.1. Chemoselective lysine arylation via S_NAr chemistry

Fluoronitrobenzene reagents that are structurally similar to Sanger's reagent were developed to label lysine ε-amino groups in proteins (Figure 1.33). These reagents did not distinguish lysine ε-amino groups from the *N*-terminal amino group, resulting in non-selective labeling of both set of amines in proteins. Sutton used a trimethylammonium nitrofluorobenzene reagent for arylation of bovine insulin to introduce positive charges into the protein to increase its solubility and crystallizability (Figure 1.33A).[172] The arylation reagent is not specific to amines, as two of the four tyrosines in bovine insulin were also labeled. A similar approach was developed for protein PEGylation by Ladd (Figure 1.33B).[173] A PEG derivative of fluoronitrobenzene was conjugated to bovine superoxide dismutase. The resulting PEGylated protein has a PEG to protein ratio of 8.8 and retained full enzyme activity. These early examples of fluoronitrobenzene-type electrophiles showed promise for the arylation of proteins under mild conditions albeit with moderate chemoselectivity and poor regioselectivity.
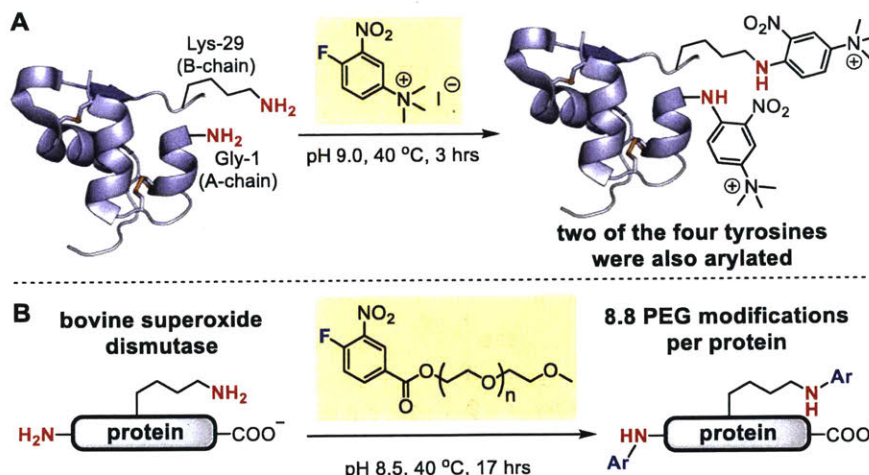
Figure 1.33. Fluoronitrobenzene reagents for protein lysine modification.
(**A**) Labeling of bovine insulin with 1-fluoro-2-nitro-4-trimethylammoniobenzene iodide improved the protein's hydrophilicity and made it readily crystallizable. (**B**) PEGylation of proteins using a PEGylated ester of 4-fluoro-3-nitrobenzoic acid.

Pentelute recently reported a series of aryl halides for peptide lysine modification.[174] The S$_N$Ar reactivity of five fluorinated aromatic electrophiles and two triazine-based electrophiles (Figure 1.34) was evaluated for arylation of lysine ε-amino groups in unprotected peptides. Because of the lower nucleophilicity of amines compared to thiols, a slightly elevated temperature (37 °C) and prolonged reaction time up to 24 hours were needed to generate the arylated product. The decafluorobiphenylsulfone (**Ar8**) was found to be the most reactive reagent for lysine arylation. With it, the desired arylated product was formed in high yields at room temperature within hours. The chemoselectivity of **Ar8** for *N*-arylation is good as nucleophiles including those in tyrosine, arginine, and histidine were not reactive under the developed reaction conditions. Cysteine was not compatible with the reaction conditons because thiol reacted faster than lysine amine with electrophile **Ar8**.

Unprotected peptides containing a range of spacing between two lysines were efficiently stapled by **Ar8**.[174] The lysine-stapled peptides showed improved stability against oxidation and bases compared to the cysteine-stapled counterparts. Furthermore, the stapled peptide variant of a

p53-derived peptide inhibitor of MDM2 showed improved stability against proteases, increased binding to MDM2, and improved cellular uptake compared to a linear peptide analog.
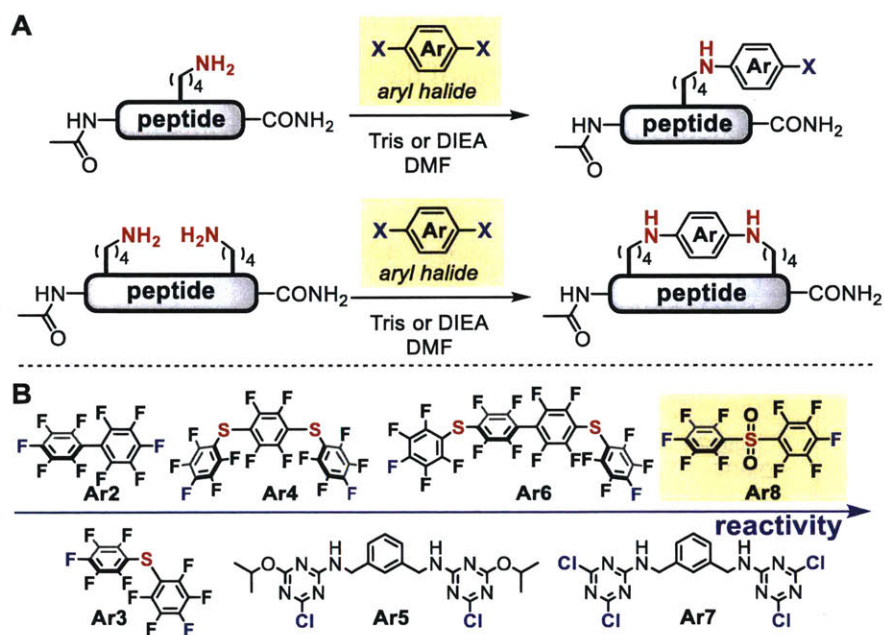
Figure 1.34. Lysine S$_N$Ar arylation for peptide modification.

(**A**) Arylation of a single lysine (top) and cyclization of two lysines (bottom) of unprotected peptides using aryl halides. (**B**) Trend of the arylation reactivity of the aryl halides studied.

1.3.2. Metal-mediated chemoselective lysine arylation

Transition metal-mediated cross-coupling reactions are highly useful for constructing *C-N* bonds in small molecules.[39,40] Importantly, reactivity and selectivity in such metal-mediated processes can be tuned through rational ligand design. Buchwald and Pentelute recently developed new organometallic palladium complexes for lysine arylation in unprotected peptides (Figure 1.35) in organic solvents.[175] Comparing oxidative addition compleses derived from several biarylphosphine ligands showed that the *t*-BuBretPhos (Figure 1.35C) was the optimum ligand for the efficient arylation of lysine ε-amino groups at room temperature in polar organic solvents. The chemoselectivity of the reaction was good as most oxygen and nitrogen protein nucleophiles including those contained in Ser, Tyr, Met, His, Trp and Asn were unreactive while cysteine reacted much faster than lysine under the reported conditions. Some competitive reactivity was seen when with peptides containing an Arg or a *C*-terminal amide. The *N*-terminal amine was more

63

nucleophilic than lysine side chain and selective arylation was observed for the *N*-terminal amine over the lysine ε-amine. Various small molecules were conjugated to peptide lysine using the developed *N*-arylation approach (Figure 1.35A) including stapled peptides from bispalladium complexes (Figure 1.35B). Further optimization of the ligand should potentially enable palladium-mediated *N*-arylation in aqueous media and expand the scope of this reaction to more complex protein targets.
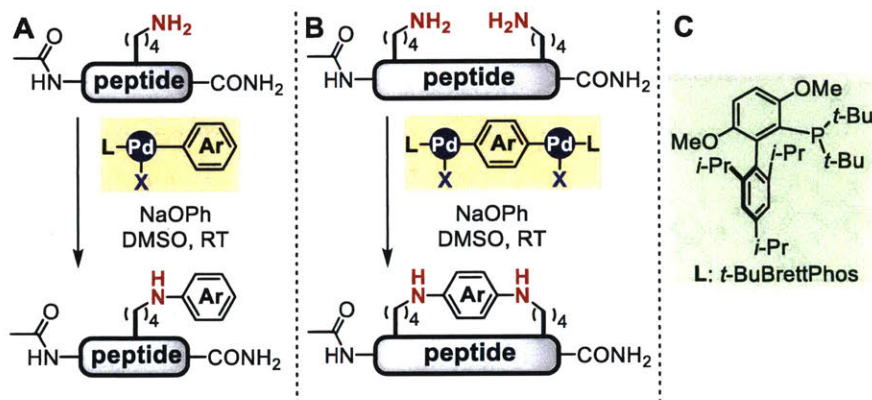
Figure 1.35. Palladium-mediated lysine arylation.

The challenge of site-selective *N*-arylation can also be tackled through exploiting the recognition and directing effects of protein microenvironments. Ball recently reported a histidine-directed method to site-selectively arylate amide in the protein backbone (Figure 1.36).[176] Histidine residues in peptides and proteins were found to direct the oxidative coupling of aryl boronic acids and boronates to a proximal backbone amide nitrogen through a Chan-Lam-type cross-coupling reaction. The reaction proceeded under mild aqueous conditions, enabling backbone modification of histidine-containing proteins. Lysozyme was site-selectively modified with several bioconjugation and affinity handles utilizing its native His-15 as a directing group to promote the arylation of backbone amide of Arg-14 (Figure 1.36). This work demonstrated how the native residues in proteins and peptides were harnessed to direct site-selective chemistry in metal-mediated processes.
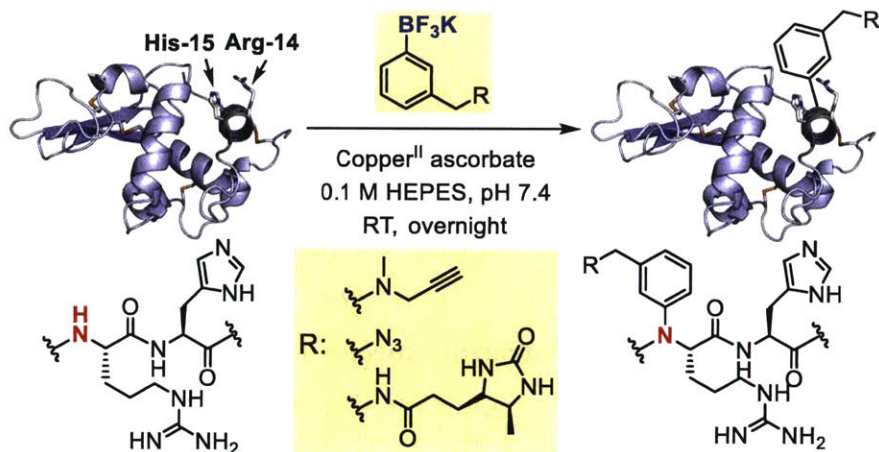
Figure 1.36. Copper-mediated N-arylation for peptide backbone modification.

1.4. Arylation of other amino acids

Bioconjugation reactions have been developed to modify amino acids other than cysteine and lysine. The nucleophilic side chains of tyrosine and tryptophan are major targets for arylation reactions on canonical amino acids. Genetic code expansion enables incorporation of various unnatural amino acids into proteins, facilitating various methods for efficient and robust arylation chemistry for site-selective protein modification.

1.4.1. Arylation of other canonical amino acids

Tyrosine

Weerapana recently identified an aryl halide that reacted in a tyrosine-selective manner, from profiling the reactivity and selectivity of a small library of aryl electrophiles (Figure 1.37).[177] A dichlorotriazine compound bearing an L-leucine methyl ester moiety was identified to selectively arylate the Tyr-108 of glutathione *S*-transferase P1 (GSTP1). This compound, termed LAS17, selectively inhibited GSTP1 in a concentration-dependent manner in cell lysate. The site of covalent modification was confirmed through extensive mutation studies and mass spectrometry analysis. It was proposed that the covalent arylation on Tyr-108 blocked a proximal glutathione (GSH) binding site of GSTP1. Because GSTP1 was found to promote tumorigenesis and drug resistance in cancer, the use of LAS17 as a covalent inhibitor of GSTP1 showed great promise as chemical probe to further interrogate the biological role of GSTP1 and as the starting point to develop potential cancer therapeutics.
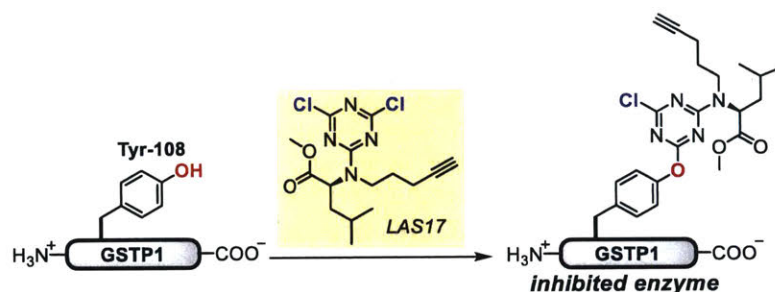
Figure 1.37. Tyrosine-selective dichlorotriazine species as covalent protein inhibitors.

Besides the hydroxyl group, the *o*-carbons (C-3 and C-5) on the phenol ring of tyrosine could also be used for bioconjugation. Heinrich developed a radical-mediated approach to generate 3-aryltyrosine residues in peptides (Figure 1.38).[178] Using an aryl diazonium salt and titanium (III) chloride under acidic conditions, the tyrosine residue of neurotensin peptide fragment could be regioselectively arylated. The proposed radical-mediated mechanism is unique as previous non-transition metal-mediated approaches for using aryl diazonium salt to modify tyrosine residues generated diazo linkages.[179] This method provides a way to quickly generate 3-aryltyrosine analogues of neurotensin peptide to test their activity against neurotensin receptor subtype 2.
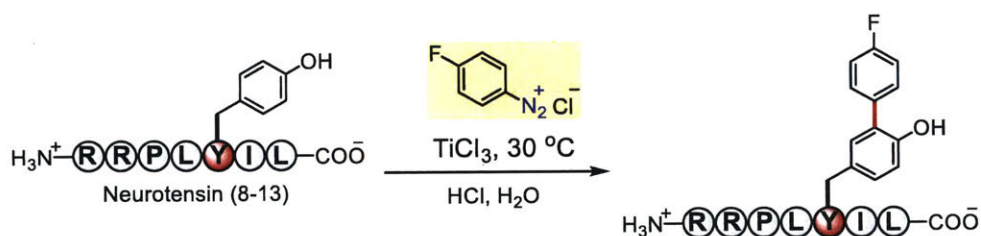
Figure 1.38. Radical arylation of tyrosine residues in peptides.

Other metal-mediated approaches were also developed to arylate tyrosine residues. The versatility of transition-metal mediated arylation reactions was illustrated in the site-selective arylation of a peptide that contains tryptophan, dehydroalanine (Dha), and tyrosine (Figure 1.39).[180] Willis and Frost reported using a combination of palladium and rhodium catalysis to site-selectively label each individual residue in the tripeptide (Trp-Dha-Tyr). Each process showed great chemoselectivity and functional group tolerance. First, using Xantphos (**L1**) as the supporting ligand, selective *N*-arylation of the indole of the tryptophan was observed. With JohnPhos (**L2**), *O*-arylation was seen indicative of the fact that the fine-tuning of ligands provides the means to selectively arylate different residues in peptides and possibly proteins. Moreover, rhodium-catalyzed arylation, with racemic 2,2'-Bis(diphenylphosphino)-1,1'-binaphthalene (*rac-*

66

BINAP, **L3**) as the ligand, was chemoselective for reaction at the dehydroalanine unit, generating a tripeptide product as a single diastereomer.[181]
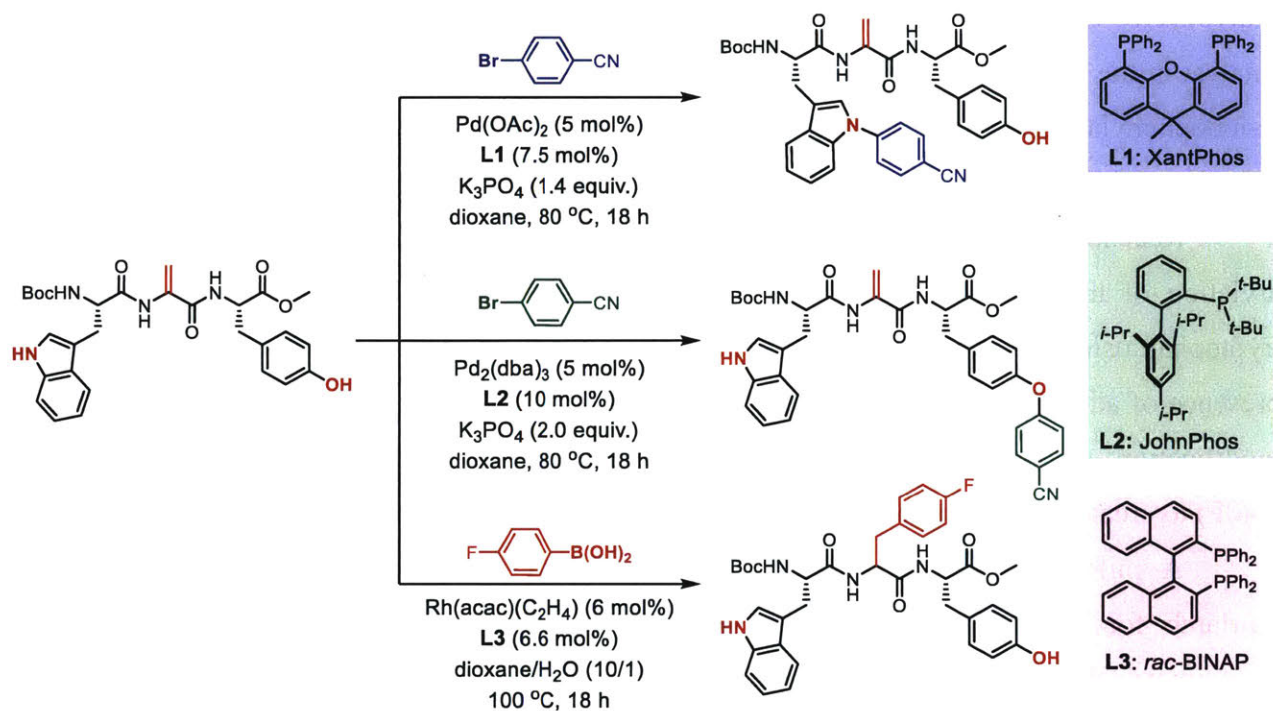
Figure 1.39. Metal-mediated site-selective arylation of dehydroalanine, tyrosine, and tryptophan.

Tryptophan

Transition metal-mediated C–H activation has become an important tool for construction of various carbon-carbon and carbon-heteroatom bonds in complex small molecules.[182] Built upon previous C–H activation strategies for indole-arylation,[183] several palladium- and ruthenium-catalyzed C–H activation processes were recently developed to modify the C-2 position of tryptophan residues.

In 2010, Albericio and Lavilla groups first reported the conjugation of a variety of aryl iodides to tryptophan residues in peptides via palladium-catalyzed C–H activation reactions (Figure 1.40A).[184] The reaction was chemoselective for tryptophan in the presence of other amino acid nucleophiles including Tyr, Arg, His, Lys, Ser, Met, and Gln. This C–H activation strategy was also used to prepare stapled peptides containing Phe-Trp or Tyr-Trp linkages using 3-iodophenylalanine or 2-iodotyrosine (Figure 1.40E).[185,186] The stapling reactions were performed either on-resin or in solution, providing access to a number of structurally diverse stapled peptides.

A similar strategy was used by James for the palladium-catalyzed macrocyclization reaction through C–H arylation of tryptophan side chain with iodophenylalanine (Figure 1.40D).[187] Macrocycles containing simple aryl and alkyl linkers were synthesized using Pd(OAc)$_2$ as the catalyst in the presence of AgBF$_4$ and 2-nitrobenzoic acid (2-NO$_2$-BzOH). Variants with *p*- and *m*-iodophenylalanine were readily crosslinked with tryptophan residue. Substrates with *ortho*-iodophenylalanine were unreactive under the developed conditions presumably for steric reasons.

Recently, Ackermann developed a ruthenium (II)- C–H activation process for the C-2 arylation of tryptophan (Figure 1.40C).[188] A wide range of aryl bromides were coupled to tryptophan using a 2-pyridyl directing group. The reaction showed good chemoselectivity in the presence of added histidine, arginine, tyrosine, aspartic acid, and glutamine. Besides tryptophan modification, this ruthenium (II)-catalyzed reaction was also used for peptide ligation (Figure 1.40F) to create more complex peptide structures.

A mild and selective palladium-mediated C–H activation reaction was reported by Fairlamb for C-2-arylation of tryptophan residues (Figure 1.40B).[189,190] The authors used unsymmetrical diaryliodonium salts as the arylation reagent and Pd(OAc)$_2$ as the catalyst to efficiently and selectively generate the C-2 arylated tryptophan products at 25 °C. The unusually mild conditions are noteworthy as almost all other C–H activation strategies for the arylation of tryptophan require the use of elevated temperatures.
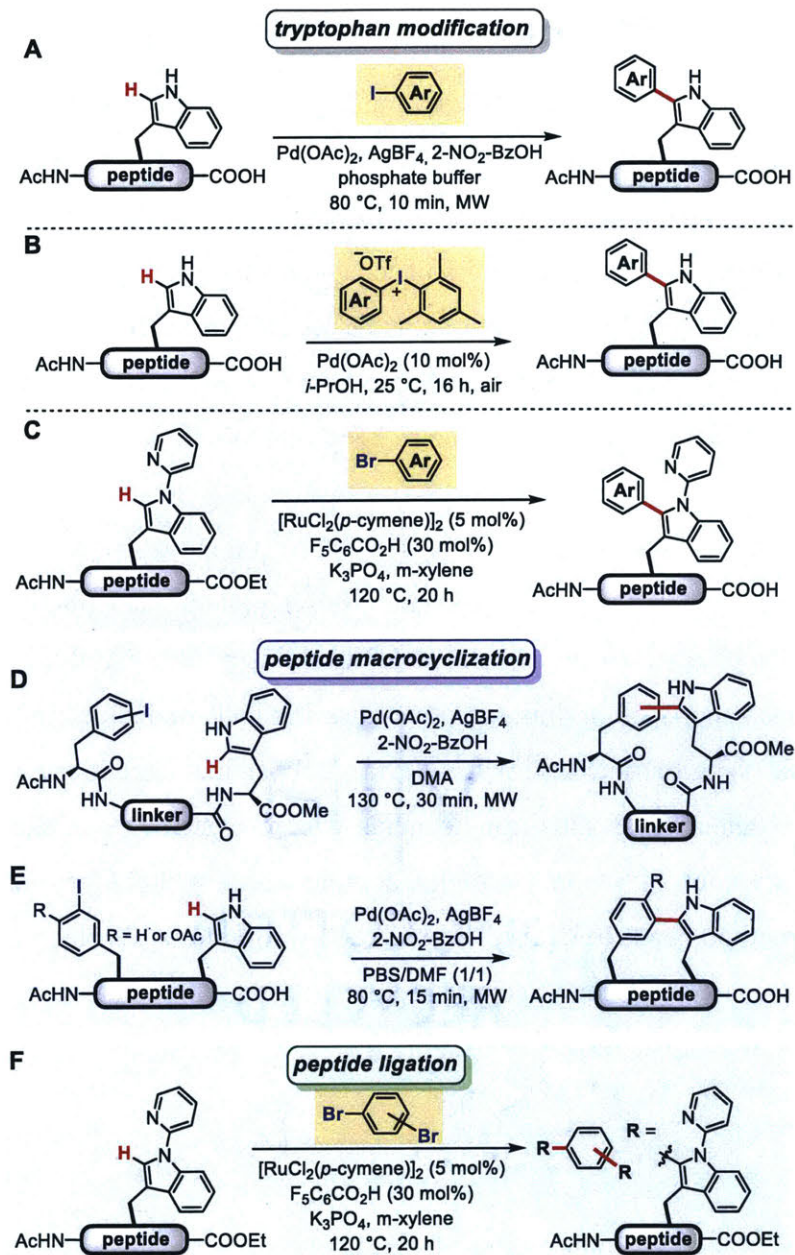
Figure 1.40. C–H activation for arylation of tryptophan residues in peptides.

(**A**) Palladium-mediated C–H activation for the C-2-arylation of a tryptophan unit using aryl iodides. (**B**) Diaryliodonium salts as the aryl source for palladium-mediated arylation of tryptophan. (**C**) Ruthenium-catalyzed C–H activation for the modification of tryptophan in protected peptides using 2-pyridyl as a directing group. (**D**) Palladium-mediated C–H activation for the synthesis of peptidic macrocycles. (**E**) C–H activation for the formation of phenylalanine-tryptophan or tyrosine-tryptophan crosslinks. (**F**) Ruthenium-catalyzed peptide ligation using dibromoaryl linkers.

1.4.2. Umpolung arylation of oxidized selenocysteine electrophiles

Selenocysteine (Sec) is a close structural analogue of cysteine with selenium in place of the sulfur in cysteine.[191] Selenols are much more nucleophilic than thiols mainly because of the greater polarizability of selenium than sulfur.[192] Moreover, alkyl selenols are quite acidic ($pK_a \approx$ 5),[192] making the selenolate the dominant form of selenocysteine residue at physiological pH. Both factors make Sec an appealing bioconjugation handle for site-selective protein modification with electrophiles such as maleimides.[193] Because Sec is prone to oxidation to form a diselenide, the addition of a strong reducing agent is required to keep selenocysteine reduced during traditional Sec-based bioconjugation reactions.

Pentelute and Buchwald recently reported an umpolung approach to the arylation of selenocysteine in unprotected peptides.[194] Instead of relying on the nucleophilic character of Sec, they explored the conjugation of aryl nucleophiles to oxidized Sec electrophiles (Figure 1.41). Inspired by previous reports on arylation of selenides in small molecules,[195,196] an approach to arylate oxidized Sec using boronic acid reagents via a copper-mediated process was devised. The reaction was chemoselective as most amino acids except cysteine were compatible with the reaction, enabling arylation of Sec in a complex peptide with a wide range of aryl boronic acids. The utility of this method needs to be further explored by combining this umpolung reaction with established techniques to incorporate Sec into proteins.
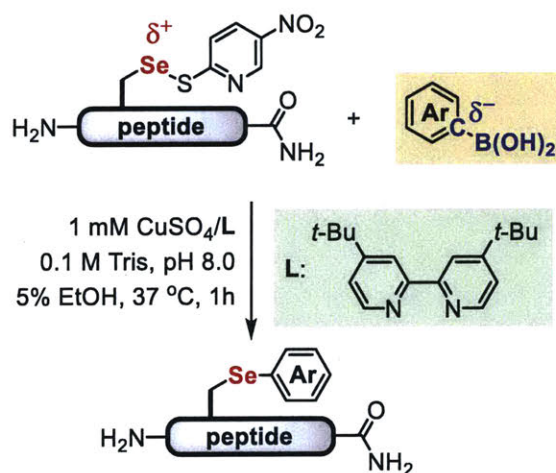
Figure 1.41. Copper-mediated umpolung arylation of oxidized selenocysteine.

1.4.3. Arylation of unnatural amino acids

Transition metal-catalyzed coupling reactions are key methods for arylation reactions on unnatural amino acids (Figure 1.42).[148,149] Palladium-catalyzed Suzuki-Miyaura, Sonogashira, and Heck reactions (Figure 1.42B) have been developed to modify proteins both in the test tube and in living cells. Such metal-mediated bioconjugation processes usually consist of two steps (Figure 1.42A). First, either one of the coupling partners is installed on the protein of interest as an unnatural amino acid through chemical[197–200] or enzymatic[201] tagging, solid-phase peptide synthesis (SPPS),[202,203] or genetic code expansion[204–209] (Figure 1.43). Then the other coupling partner is conjugated to the protein via transition-metal-catalyzed coupling reactions.
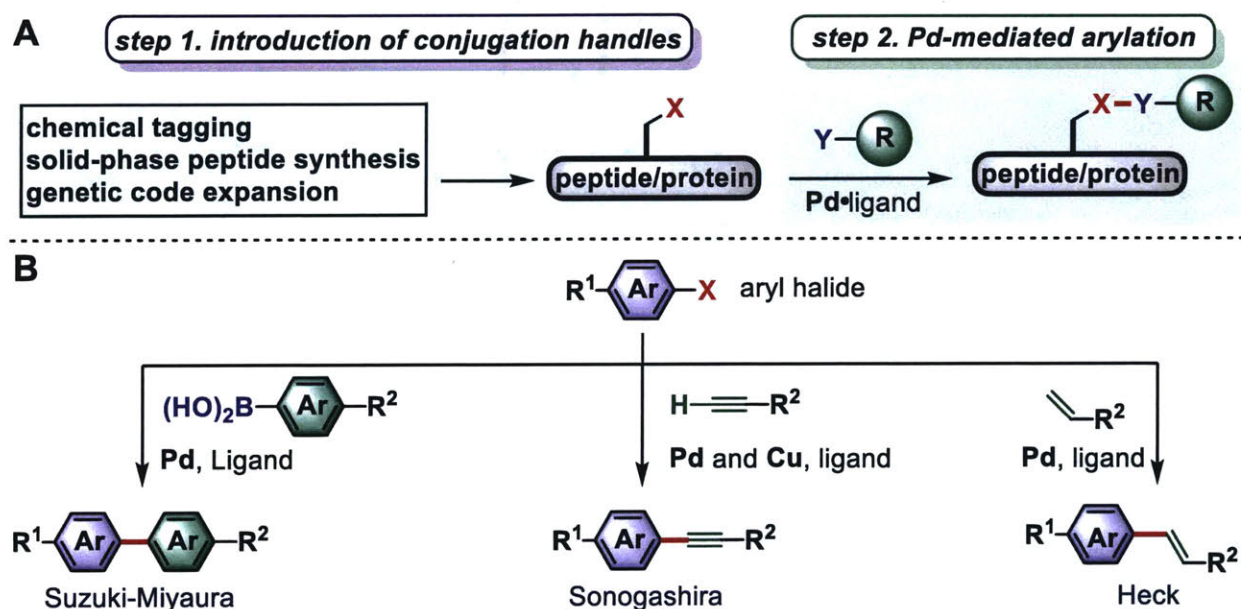
Figure 1.42. Metal-mediated arylation of unnatural amino acids.
(**A**) A two-step process for palladium-mediated arylation of unnatural amino acids. (**B**) Key examples of palladium-mediated coupling strategies applied to bioconjugation.

These metal-mediated arylation reactions require the incorporation of unnatural handles (aryl halides, aryl boronic acids, alkynes, and alkenes) into the protein of interest (Figure 1.43). Chemical tagging and solid-phase synthesis were early approaches used to introduce these handles into proteins. For small peptides that are amenable to solid-phase synthesis, multiple conjugation handles such as iodophenylalanine and iodotyrosine can be directly incorporated into the peptide

chain.[202,203] Canonical amino acid residues in proteins can also be tagged with desired conjugation handles through other conjugation chemistries (e.g., alkylation,[197] iodination,[199] and acylation[198]). Recently, the use of genetic code expansion has enabled the incorporation of these handles as unnatural amino acids.[204–209] Although precise site-selectivity is ensured in genetic code expansion, the number of unnatural amino acids that could be incorporated into a protein is still limited.
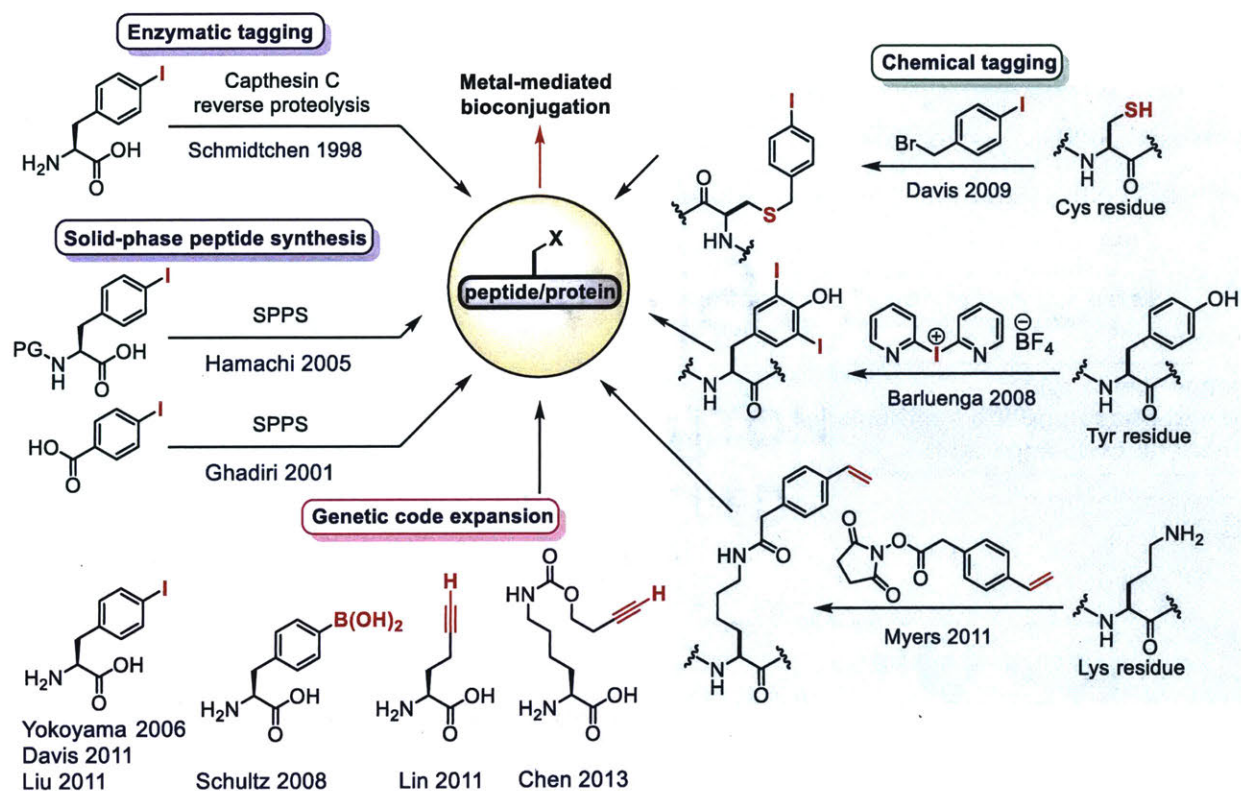
Figure 1.43. Methods for incorporating unnatural arylation handles into peptides or proteins.

Palladium-ligand combinations determine the reactivity and selectivity of these metal-catalyzed arylation reactions (Figure 1.44). Initial reports using phosphine-based ligands (Figure 1.44, **L1** and **L2**)[199,201], Na$_2$PdCl$_4$,[202] or Pd-DBA[205] met with limited success because of the low efficiency of these catalyst and the harsh reaction conditions employed. The utilization of non-phosphine ligands (Figure 1.44, **L3-L8**)[208,210,211] or Pd(NO$_3$)$_2$ [209] enabled highly efficient arylation of unnatural amino acids on proteins both *in vitro* and *in vivo*.
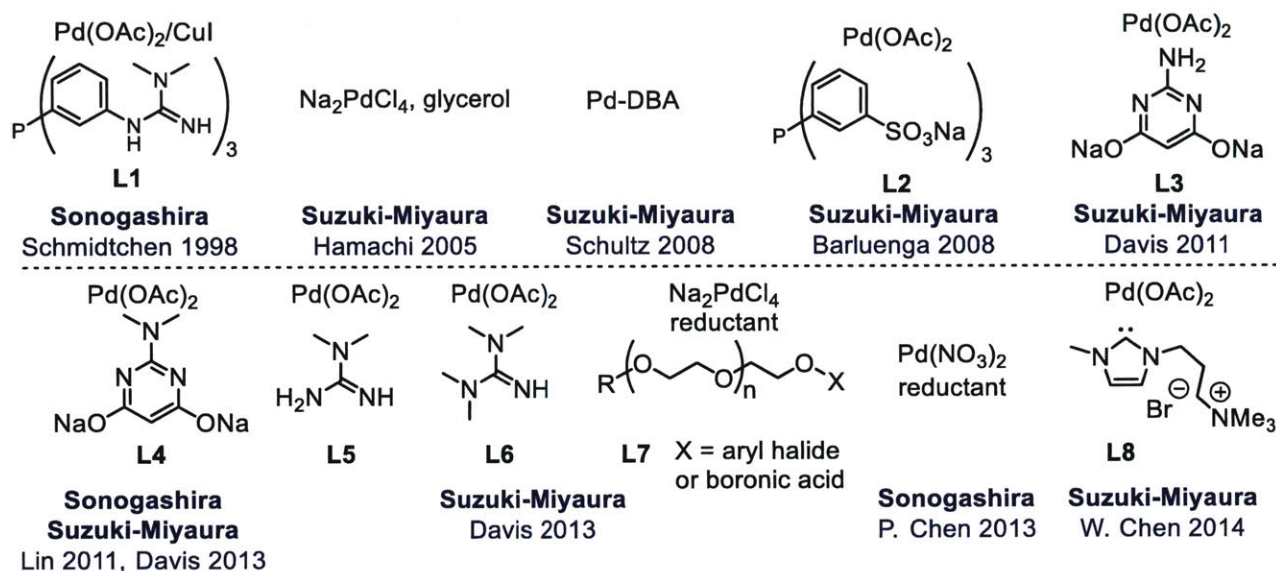
Figure 1.44. Palladium sources and ligands used for the metal-mediated arylation of unnatural amino acids.

In 1998, Schmidtchen first reported the palladium-catalyzed Sonogashira reaction to modify peptides in aqueous media.[201] A water-solubable guanidinoaryl-phosphine ligand (Figure 1.44, **L1**) was used in combination with palladium and copper co-catalysts to modify 4-iodophenylanaline enzymatically installed on a small peptide. Later in 2005, Hamachi reported the Suzuki-Miyaura reaction of small chemically synthesized proteins in aqueous buffers using Na_2PdCl_4 as the catalyst.[202] A significant amount of glycerol (50% v/v) and slightly elevated temperature (40 °C) were required to achieve > 90% arylation yields. Yokoyama and Tachibana first reported the Sonogashira and Heck reactions between terminal alkyne or alkene probes with 4-iodophenylalanine on the protein surface.[200] However, the use 10% of DMSO as a co-solvent was required and low conversions were observed using a $Pd(OAc)_2$-3,3',3''-phosphanetriyltris trisodium salt (TPPTS) catalyst system. Schultz reported the first example of Suzuki-Miyaura reaction of expressed proteins with 4-boronophenylalanine incorporated via genetic code expansion.[205] Although the ability to genetically encode such arylation handles greatly broadened the scope of the reaction, the high temperature (70 °C) and low efficiency of the Pd-DBA catalyst restricted the application of this method. These early examples demonstrated the viability of palladium-mediated bioconjugation reactions but also highlighted the challenge of applying such reactions to complex protein targets.

73

Davis reported a greatly improved Suzuki-Miyaura bioconjugation using palladium acetate in combination with the sodium salt of 2-amino-4,6-dihydroxypyrimidine (Figure 1.44, **L3**) as the ligand (Figure 1.45). The aryl idodide was introduced into the protein through alkylation reaction of a cysteine residue. Full conversion (> 95%) within 30 minutes was achieved at 37 °C in phosphate buffer (pH 8.0) with a number of aryl, heteroaryl, and vinyl boronic acids.[197] However, the use of a large excess of palladium (50 equiv) combined with the ability of proteins to ligate metals resulted in poor mass spectrometry data for the conjugates. This problem was later solved through surveying a number of potential metal scavengers for palladium removal, with 3-mercaptopropionic acid identified as the best candidate.[206] The aryl idodide handle can also be introduced through genetic encoding of L-4-iodophenylalaine, which set the stage for applying this reaction to the labeling of proteins in living systems.[206]
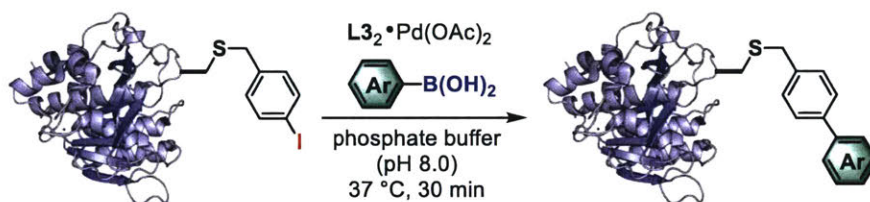
Figure 1.45. Suzuki-Miyaura arylation of a cysteine mutant of subtilisin *Bacillus lentus*.

To this date, a number of ligands has been utilized to improve the palladium-mediated Suzuki-Miyaura reactions for protein modification. Ligands including the *N,N*-dimethyl-2-amino-4,6-dihydroxypyrimidine (**L4**),[208,210] simple guanidines (**L5** and **L6**),[210] polyethylene glycol (PEG) (**L7**),[210] and *N*-heterocyclic carbene (NHC) (**L8**)[211] were shown to promote the Suzuki-Miyaura arylation in water. Unlike other ligands that assist the cross-coupling reaction, the PEG ligand is unique as a self-ligating reagent for protein PEGylation. It is also worth noting that **L4** has been shown by Lin to efficiently catalyze the copper-free Sonogashira reaction (Heck alkynylation) between alkynes on a protein surface and aryl iodide probes.[208]

Combining the high efficiency and selectivity of the Suzuki-Miyaura and Sonogashira reactions with genetic code expansion enabled their applications to labeling proteins in living systems. In 2011, Lin applied the copper-free Sonogashira reaction to labeling of ubiquitin in *E. coli* cells.[208] Around the same time, Davis showed that their Suzuki-Miyaura reactions could be

used to couple a fluorescent boronic acid to the cell surface aryl iodides.[212] N-heterocyclic carbenes (NHCs) were also shown by Ma to be viable catalysts for the labeling of membrane proteins through Suzuki-Miyaura coupling with low catalyst loadings.[211] An advance was made by Chen who developed a "ligand-free" approach using $Pd(NO_3)_2$ to catalyze the arylation of alkyne-containing proteins inside live bacterial cells.[209]
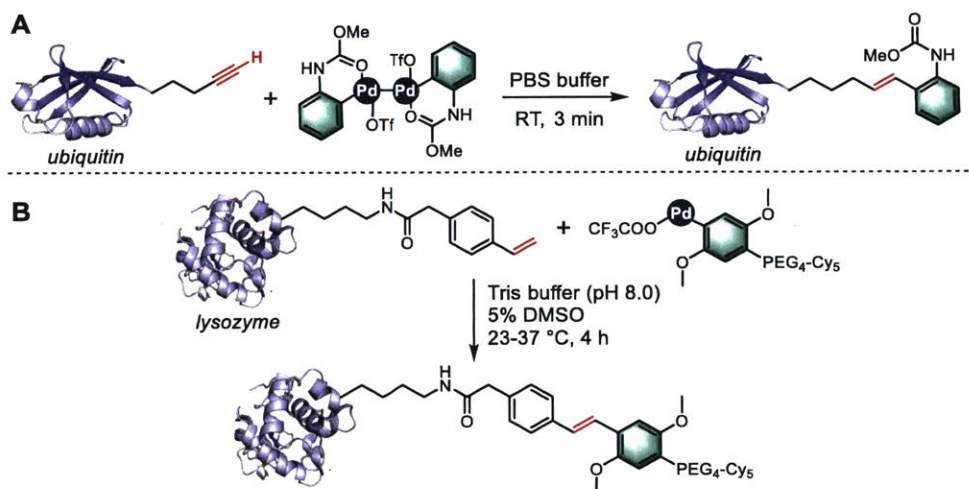
Figure 1.46. Isolated palladium complexes for the arylation of protein alkynes (A) and alkenes (B).

Isolated palladium complexes have also been explored to arylate alkynes and alkenes on proteins. Lin reported a series of N-phenylcarbamate palladacycles for conjugation to an alkyne residue in ubiquitin (Figure 1.46A).[213,214] The reaction rate of this conjugation is very rapid, with the fastest having a rate constant of 19,770 $M^{-1}s^{-1}$. Mass spectrometry analysis of the trypsin digestion of the labeled ubiquitin showed a modified peptide fragment with mass consistent with a styrene product (Figure 1.46A). However, reactions of the palladacycles with a model peptide produced complex reaction mixtures. This disparity between the results of ubiquitin labeling and peptide modification suggests that more detailed studies on the reaction mechanism is required before further applications of the reaction to other protein targets.

Heck-type reactions of proteins containing vinyl aryl groups have been reported (Figure 1.46B) by Myers.[198] They developed storable palladium complexes formed via decarboxylative palladation of aryl carboxylic acids and showed their utility in bioconjugation of functionalized lysozyme under mild conditions.

1.5. Arylation of nucleic acids

Chemically modified oligonucleotides are widely used for applications including biosensing,[215,216] nanotechnology,[217] medicine,[218] and catalysis[219]. Oligonucleotides have fewer monomeric building blocks compared to proteins. This simplifies the range of chemistry necessary for bioconjugation. Traditional chemical methods have utilized either solid phase synthesis[220] or enzyme-catalyzed reactions[221] to modify nucleic acid oligomers.

Recently, Suzuki-Miyaura coupling has emerged as a new strategy for post-synthetic modification of nucleic acids.[48] These methods use synthetically incorporated aryl halides as the handle for site-selective conjugation of aryl boronic acids to both single- and double-stranded DNAs. The required aryl halides were incorporated either through unnatural phosphoramidites during solid-phase synthesis (Figure 1.47) or through direct coupling of aryl halides to nucleic acids through amide bond formation (Figure 1.48).

1.5.1. Post-synthetic arylation of single-stranded DNA

Manderville first applied Suzuki-Miyaura cross-coupling to the modification of 2'-deoxyguanosine (dG) in single-stranded oligodeoxynucleotides (Figure 1.47A).[222] The aryl halide conjugation handle was introduced into DNA using 8-Br-dG phosphoramidite. Using palladium catalysis, a range of aryl boronic acids could be site-selectively coupled to the C-8 site of dG. The dG-arylated product were utilized to study DNA damages caused by radical addition reactions at C-8 site of dG.

A milder Suzuki-Miyaura arylation procedure was developed by Davis to modify 5-iodo-2'-deoxyuridine (5-I-dU) in single-stranded DNA (ssDNA) (Figure 1.47B).[169] The use of 2-aminopyrimidine-type ligands enabled the Suzuki-Miyaura arylation at 37 °C and under buffered aqueous conditions. Phenyl boronic acid and 3-furyl boronic pinacol ester were used to site-selectively arylate the 2'-deoxyuridine in the ssDNA. Importantly, vinyl boronate esters were also conjugated to the ssDNA under the developed reaction conditions, enabling several handles include a benzophenone, a diazirine, a pyrene, a sugar, or an azobenzene to be appended. The method was used to synthesize an ssDNA probe that contained biotin, diazirine, and 5-hydroxylmethylcytosine (5-hmC). This probe was used to identify 5-hmC-binding proteins in the HeLa cell lysate through photo-crosslinking, affinity enrichment, and mass spectrometry analysis.
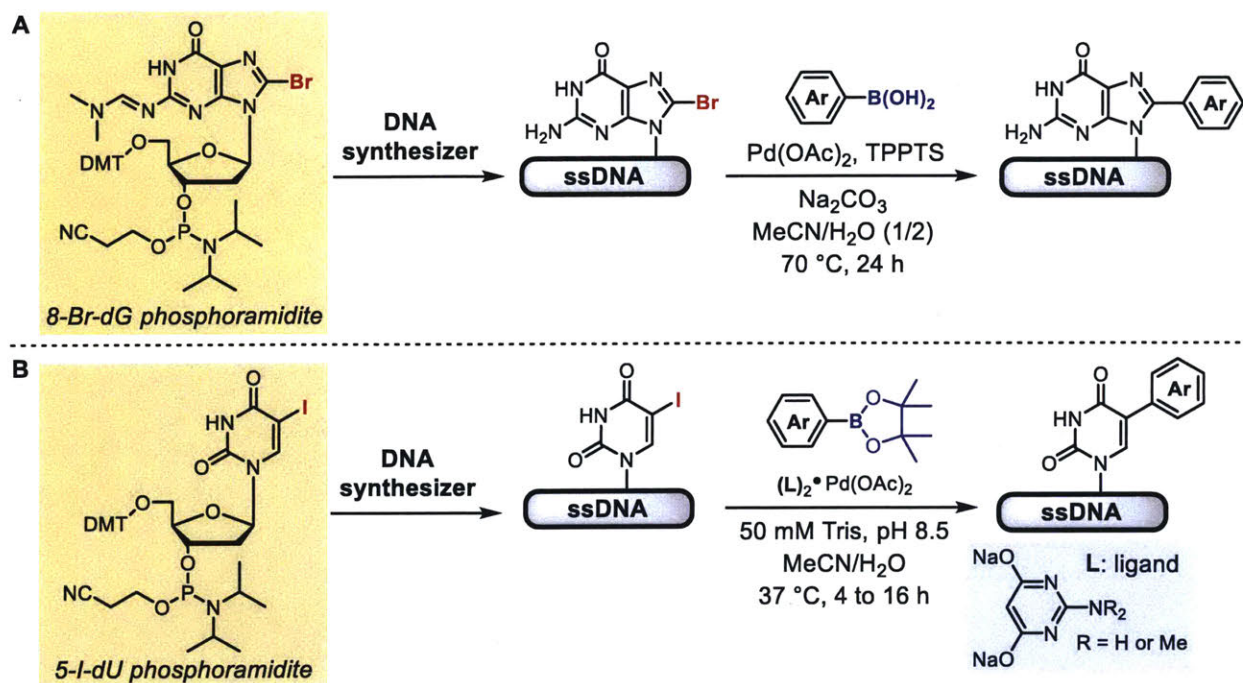
Figure 1.47. Post-synthetic arylation of single-stranded DNA (ssDNA).
8-bromoguanidine (**A**) and 5-iodouracil (**B**) were introduced into single-strand DNA (ssDNA) through solid-phase DNA synthesis using designed phosphoramidites.

1.5.2. Post-synthetic arylation of double-stranded DNA

The Suzuki-Miyaura cross-coupling was also used to synthesize DNA-encoded small molecule libraries. Ding and Clark installed aryl halides on dsDNA through coupling of aryl halide carboxylic acids to the terminal amines in dsDNA (Figure 1.48A).[223] These aryl halide handles were further conjugated to aryl or heteroaryl boronic acids and esters using Pd(Ph3)4 as the catalyst. This catalyst system was applied to the synthesis of a 3.5-million-member DNA-encoded library from which potent hits for phosphoinositide 3-kinase α (PI3Kα) was identified.[224] Recently, they reported a new catalyst system using [(t-Bu)2P(OH)2]PdCl2 (POPd) as the palladium source and sodium 2'-(dicyclohexylphosphino)-2,6-dimethoxy-[1,1'-biphenyl]-3-sulfonate (sSPhos) as the ligand for on-DNA Suzuki-Miyaura coupling of challenging phenyl chlorides and pyrimidinyl chlorides (Figure 1.48B).[225] Increased conjugation yields were observed with this catalyst system for various aryl and heteroaryl boronic acids and esters.

77
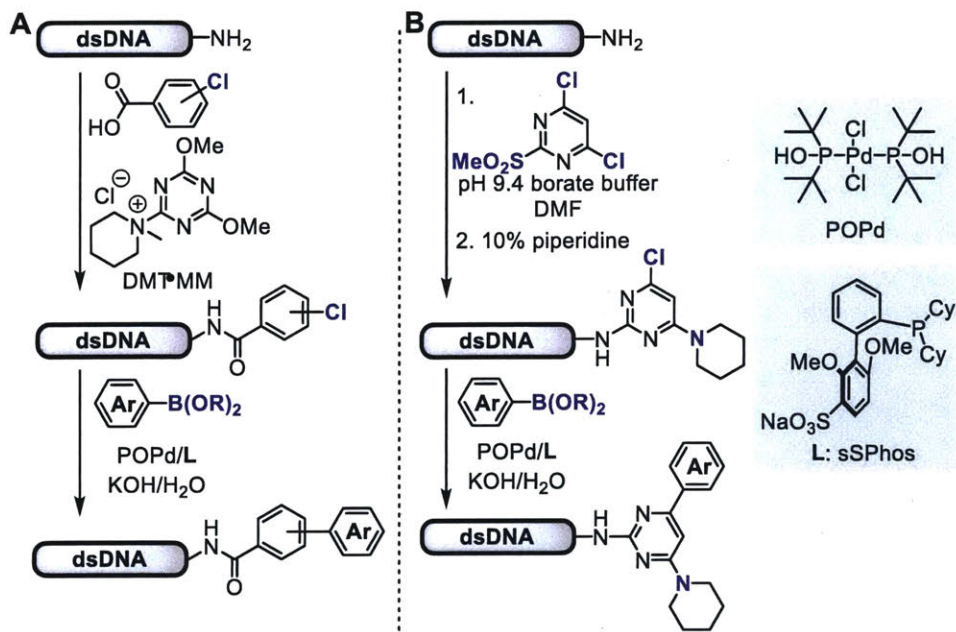
Figure 1.48. Post-synthetic arylation of double-strand DNA (dsDNA).

Aryl or heteroaryl halides were installed on double-strand DNA (dsDNA) through either amide bond formation (**A**) or S$_N$Ar reaction (**B**). The aryl halide-dsDNA conjugates were then arylated using a Suzuki-Miyaura coupling reaction.

1.6. Summary and Outlook

The emerging arylative bioconjugation reactions have begun to allow formation of nucleophile-sp^2 carbon bonds in biomolecules. However, compared to the robust and diverse arylation strategies available for the synthesis of small molecules, arylation chemistry for bioconjugation is still in its infancy.

For the existing arylative bioconjugation reactions, further developments are needed to improve their selectivity, reaction rate, and substrate scope. In particular, site-specific arylation reactions are highly desirable to generate homogeneous bioconjugates. The π-clamp[107] shows how the microenvironment within a small peptide sequence can promote site-selective chemistry. We envision extending the same concept to other perfluoroaryl linkers and arylation chemistry beyond perfluoroarylation. This will provide a set of small genetically encodable peptide tags for site-specific protein labeling with different aryl modifications.

Arylation reactions are desired for natural amino acid residues other than lysine and cysteine. Targeting other nitrogen and oxygen-based nucleophiles (e.g., residues of His, Ser, Thr, and Tyr) is highly valuable for generating structurally diverse protein conjugates. Such nucleophiles are, in most cases, much less reactive compared to the cysteine thiol group or the lysine ε-amine, which presents significant challenge for developing arylation chemistry targeting these sites. A possible solution is to discover new ligands for copper- or palladium-catalyzed arylation reactions that form C(sp^2)-O bonds (for Ser, Thr, and Tyr) and C(sp^2)-N(sp^2) (for His) bonds on proteins. Another strategy is to utilize protein environments and recognition elements to promote otherwise difficult arylation chemistry. The viability of the second strategy has been demonstrated in Ball's method for arylation of amides in protein backbones.

One obvious difficulty in using transition metal-catalyzed or -mediated processes is the removal of the potentially toxic metals at the end of the reaction. While a number of reagents and resin-based techniques have been developed for small molecule applications, much less is known about their application for the reactions of proteins. Developing catalytic reactions to reduce the amount of metal used should aid the metal removal processes.

The highly selective nature of the developed arylation reactions should enable their applications beyond protein and nucleic acid labeling. The modified boron clusters developed by

Spokoyny highlighted the utility of selective arylation reactions for creating new types of nanocomposites. We envision that arylation reactions can be applied to post-synthetic modifications of nanoparticles, surfaces, and metal organic frameworks.[226,227]

The biological benefits (or drawbacks) of aryl modifications remain to be fully explored. Initial studies have shown that certain aryl linkages are chemically more stable than alkyl-based conjugates.[156] However, it is still unclear how these affect the properties of biomolecules in a cellular context. Systematic studies of the biological properties (e.g., stability, toxicity, immunogenicity) of aryl modifications both *in vitro* and in *vivo* will be key to providing guidelines to select aryl linkages for biomolecule modification.

1.7. References

(1)   Jeon, S.; Heinz, U. Introduction to Hydrodynamics. In *Bioconjugate Techniques (Third edition)*; Hermanson, G. T., Ed.; Academic Press: Boston, 2015; pp. 1–60.

(2)   Surís, A.; Smith, J.; Powell, C.; North, C. S. Interfering with the Reconsolidation of Traumatic Memory: Sirolimus as a Novel Agent for Treating Veterans with Posttraumatic Stress Disorder. *Ann. Clin. Psychiatry* **2013**, *25*, 33–40.

(3)   Chari, R. V. J.; Miller, M. L.; Widdison, W. C. Antibody-Drug Conjugates: An Emerging Concept in Cancer Therapy. *Angew. Chemie - Int. Ed.* **2014**, *53*, 3796–3827.

(4)   Heinis, C. Drug Discovery: Tools and Rules for Macrocycles. *Nat. Chem. Biol.* **2014**, *10*, 696–698.

(5)   Chudasama, V.; Maruani, A.; Caddick, S. Recent Advances in the Construction of Antibody-Drug Conjugates. *Nat. Chem.* **2016**, *8*, 114–119.

(6)   Xue, L.; Karpenko, I. A.; Hiblot, J.; Johnsson, K. Imaging and Manipulating Proteins in Live Cells through Covalent Labeling. *Nat. Chem. Biol.* **2015**, *11*, 1–7.

(7)   Cobo, I.; Li, M.; Sumerlin, B. S.; Perrier, S. Smart Hybrid Materials by Conjugation of Responsive Polymers to Biomacromolecules. *Nat. Mater.* **2015**, *14*, 143–159.

(8)   Lutz, J. F.; Zarafshani, Z. Efficient Construction of Therapeutics, Bioconjugates, Biomaterials and Bioactive Surfaces Using Azide-Alkyne "click" chemistry. *Adv. Drug Deliv. Rev.* **2008**, *60*, 958–970.

(9)   Kairdolf, B. A.; Qian, X.; Nie, S. Bioconjugated Nanoparticles for Biosensing, in Vivo Imaging, and Medical Diagnostics. *Anal. Chem.* **2017**, DOI: acs.analchem.6b04873.

(10)   Lang, K.; Chin, J. W. Bioorthogonal Reactions for Labeling Proteins. *ACS Chem. Biol.* **2014**, *9*, 16–20.

(11)   Saito, F.; Noda, H.; Bode, J. W. Critical Evaluation and Rate Constants of Chemoselective Ligation Reactions for Stoichiometric Conjugations in Water. *ACS Chem. Biol.* **2015**, *10*, 1026–1033.

(12)   Meanst, G. E.; Feeney, R. E. Chemical Modifications of Proteins: History and Applications.

Bioconjugate Chem **1990**, *1*, 2–12.

(13) Olcott, H. S.; Fraenkel-Conrat, H. Specific Group Reagents for Proteins. *Chem. Rev.* **1947**, *41*, 151–197.

(14) Herriott, R. M. Reactions of Native Proteins with Chemical Reagents. *Adv. Protein Chem.* **1947**, *3*, 169–225.

(15) Fraenkel-Conrat, H.; Cooper, M.; Olcott, H. S. The Reaction of Formaldehyde with Proteins. *J. Am. Chem. Soc.* **1945**, *67*, 950–954.

(16) Hoare, D. G.; Koshland, D. E. A Method for the Quantitative Modification and Estimation of Carboxylic Acid Groups in Proteins. *J. Biol. Chem.* **1967**, *242*, 2447–2453.

(17) Barman, T. E.; Koshland, D. E. A Colorimetric Procedure for the Quantitative Determination of Tryptophan Residues in Proteins. *J. Biol. Chem.* **1967**, *242*, 5771–5776.

(18) Fields, R. The Rapid Determination of Amino Groups with TNBS. In *Methods in enzymology*; 1972; Vol. 25, pp. 464–468.

(19) Ellman, G. L. Tissue Sulfhydryl Groups. *Arch. Biochem. Biophys.* **1959**, *82*, 70–77.

(20) Baker, B. R. Specific Irreversible Enzyme Inhibitors. *Annu. Rev. Pharmacol.* **1970**, *10*, 35–50.

(21) Edman, P.; Begg, G. A Protein Sequenator. *Eur. J. Biochem.* **1967**, *1*, 80–91.

(22) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662–666.

(23) Hermanson, G. T. The Reactions of Bioconjugation. In *Bioconjugate Techniques*; Hermanson, G. T., Ed.; Academic Press: Boston, 2013; pp. 229–258.

(24) Afagh, N. A.; Yudin, A. K. Chemoselectivity and the Curious Reactivity Preferences of Functional Groups. *Angew. Chemie - Int. Ed.* **2010**, *49*, 262–310.

(25) Lin, C. W.; Ting, A. Y. Transglutaminase-Catalyzed Site-Specific Conjugation of Small-Molecule Probes to Proteins in Vitro and on the Surface of Living Cells. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.

(26)   Wollack, J. W.; Silverman, J. M.; Petzold, C. J.; Mougous, J. D.; Distefano, M. D. A Minimalist Substrate for Enzymatic Peptide and Protein Conjugation. *ChemBioChem* **2009**, *10*, 2934–2943.

(27)   Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically Encoded Short Peptide Tag for Versatile Protein Labeling by Sfp Phosphopantetheinyl Transferase. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15815–15820.

(28)   Cull, M. G.; Schatz, P. J. Biotinylation of Proteins in Vivo and in Vitro Using Small Peptide Tags. In *Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein Purification*; Jeremy Thorner, S. D. E. J. N. A., Ed.; Academic Press, 2000; Vol. 326, pp. 430–440.

(29)   Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. Redirecting Lipoic Acid Ligase for Cell Surface Protein Labeling with Small-Molecule Probes. *Nat. Biotechnol.* **2007**, *25*, 1483–1487.

(30)   Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: A Versatile Method for Protein Labeling. *Nat. Chem. Biol.* **2007**, *3*, 707–708.

(31)   Wang, L.; Xie, J.; Schultz, P. G. Expanding the Genetic Code. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 225–249.

(32)   Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

(33)   Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114*, 4764–4806.

(34)   Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chemie - Int. Ed.* **2001**, *40*, 2004–2021.

(35)   Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chemie - Int. Ed.* **2009**, *48*, 6974–6998.

(36)   Ruiz-Castillo, P.; Buchwald, S. L. Applications of Palladium-Catalyzed C-N Cross-Coupling Reactions. *Chem. Rev.* **2016**, *116*, 12564–12649.

(37) Martin, R.; Buchwald, S. L. Palladium-Catalyzed Suzuki– Miyaura Cross-Coupling Reactions Employing Dialkylbiaryl Phosphine Ligands. *Acc. Chem. Res.* **2008**, *41*, 1461– 1473.

(38) Burnett, J. F.; Zahler, R. E. Aromatic Nucleophilic Substitution Reactions. *Chem. Rev.* **1951**, *49*, 273–412.

(39) Crabtree. *The Organometallic Chemistry of the Transition Metals*; Crabtree, R. H., Ed.; 5th Edition; Wiley, 2014.

(40) Diederich, F.; Stang, P. *Metal-Catalyzed Cross-Coupling Reactions*; John Wiley & Sons, 2008.

(41) Zalatan, D. N.; Du Bois, J.; Yu, J. Q.; Shi, Z. *C-H Activation*; Yu, J.-Q.; Shi, Z., Eds.; Topics in Current Chemistry; Springer Berlin Heidelberg: Berlin, Heidelberg, 2010; Vol. 292.

(42) Boutureira, O.; Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **2015**, *115*, 2174–2195.

(43) Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*, 4740.

(44) Jbara, M.; Maity, S. K.; Brik, A. Palladium in Chemical Protein Synthesis and Modifications. *Angew. Chem., Int. Ed.* **2017**.

(45) Bondalapati, S.; Jbara, M.; Brik, A. Expanding the Chemical Toolbox for the Synthesis of Large and Uniquely Modified Proteins. *Nat. Chem.* **2016**, *8*, 407–418.

(46) Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-Selective Protein-Modification Chemistry for Basic Biology and Drug Development. *Nat. Chem.* **2015**, *8*, 1– 11.

(47) Koniev, O.; Wagner, A. Developments and Recent Advancements in the Field of Endogenous Amino Acid Selective Bond Forming Reactions for Bioconjugation. *Chem. Soc. Rev.* **2015**, *44*, 5495–5551.

(48) Messaoudi, S.; Defrancq, E. Pd-Mediated Labeling of Nucleic Acids. *ChemBioChem* **2016**, 426–431.

(49) Malins, L. R. Transition Metal-Promoted Arylation: An Emerging Strategy for Protein

Bioconjugation. *Aust. J. Chem.* **2016**, *69*, 1360.

(50) Gong, Y.; Pan, L. Recent Advances in Bioorthogonal Reactions for Site-Specific Protein Labeling and Engineering. *Tetrahedron Lett.* **2015**, *56*, 2123–2132.

(51) Baslé, E.; Joubert, N.; Pucheault, M. Protein Chemical Modification on Endogenous Amino Acids. *Chem. Biol.* **2010**, *17*, 213–227.

(52) Hackenberger, C. P. R.; Schwarzer, D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chemie - Int. Ed.* **2008**, *47*, 10030–10074.

(53) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - An Asian J.* **2009**, *4*, 630–640.

(54) Gunnoo, S. B.; Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **2016**, *17*, 529–553.

(55) Giles, N. M.; Giles, G. I.; Jacob, C. Multiple Roles of Cysteine in Biocatalysis. *Biochem. Biophys. Res. Commun.* **2003**, *300*, 1–4.

(56) Jacob, C.; Giles, G. I.; Giles, N. M.; Sies, H. Sulfur and Selenium: The Role of Oxidation State in Protein Structure and Function. *Angew. Chemie - Int. Ed.* **2003**, *42*, 4742–4758.

(57) Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B. D.; Bachovchin, D. a; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative Reactivity Profiling Predicts Functional Cysteines in Proteomes. *Nature* **2010**, *468*, 790–795.

(58) Price, N. C.; Cohn, M.; Schirmer, R. H. Fluorescent and Spin Label Probes of the Environments of the Sulfhydryl Groups of Porcine Muscle Adenylate Kinase. *J. Biol. Chem.* **1975**, *250*, 644–652.

(59) Leesnitzer, L. M.; Parks, D. J.; Bledsoe, R. K.; Cobb, J. E.; Collins, J. L.; Consler, T. G.; Davis, R. G.; Hull-Ryde, E. A.; Lenhard, J. M.; Patel, L.; *et al.* Functional Consequences of Cysteine Modification in the Ligand Binding Sites of Peroxisome Proliferator Activated Receptors by GW9662. *Biochemistry* **2002**, *41*, 6640–6650.

(60) Lee, G.; Elwood, F.; McNally, J.; Weiszmann, J.; Lindstrom, M.; Amaral, K.; Nakamura, M.; Miao, S.; Cao, P.; Marc Learned, R.; *et al.* T0070907, a Selective Ligand for

Peroxisome Proliferator-Activated Receptor Γ, Functions as an Antagonist of Biochemical and Cellular Activities. *J. Biol. Chem.* **2002**, *277*, 19649–19657.

(61) Tyagi, S.; Gupta, P.; Saini, A. S.; Kaushal, C.; Sharma, S. The Peroxisome Proliferator-Activated Receptor: A Family of Nuclear Receptors Role in Various Diseases. *J. Adv. Pharm. Technol. Res.* **2011**, *2*, 236–240.

(62) Ahmadian, M.; Suh, J. M.; Hah, N.; Liddle, C.; Atkins, A. R.; Downes, M.; Evans, R. M. PPARγ Signaling and Metabolism: The Good, the Bad and the Future. *Nat. Med.* **2013**, *99*, 557–566.

(63) Hughes, T. S.; Giri, P. K.; de Vera, I. M. S.; Marciano, D. P.; Kuruvilla, D. S.; Shin, Y.; Blayo, A.-L.; Kamenecka, T. M.; Burris, T. P.; Griffin, P. R.; *et al.* An Alternate Binding Site for PPARγ Ligands. *Nat. Commun.* **2014**, *5*, 3571.

(64) Brust, R.; Lin, H.; Fuhrmann, J.; Asteian, A.; Kamenecka, T. M.; Kojetin, D. J. Modification of the Orthosteric PPARγ Covalent Antagonist Scaffold Yields an Improved Dual-Site Allosteric Inhibitor. *ACS Chem. Biol.* **2017**, acschembio.6b01015.

(65) Johnson, C. M.; Linsky, T. W.; Yoon, D. W.; Person, M. D.; Fast, W. Discovery of Halopyridines as Quiescent Affinity Labels: Inactivation of Dimethylarginine Dimethylaminohydrolase. *J. Am. Chem. Soc.* **2011**, *133*, 1553–1562.

(66) Johnson, C. M.; Monzingo, A. F.; Ke, Z.; Yoon, D. W.; Linsky, T. W.; Guo, H.; Robertus, J. D.; Fast, W. On the Mechanism of Dimethylarginine Dimethylaminohydrolase Inactivation by 4-Halopyridines. *J. Am. Chem. Soc.* **2011**, *133*, 10951–10959.

(67) Banerjee, R.; Pace, N. J.; Brown, D. R.; Weerapana, E. 1,3,5-Triazine As a Modular Scaffold for Covalent Inhibitors With Streamlined Target Identification. *J. Am. Chem. Soc.* **2013**, *135*, 2497–2500.

(68) Shannon, D. A.; Banerjee, R.; Webster, E. R.; Bak, D. W.; Wang, C.; Weerapana, E. Investigating the Proteome Reactivity and Selectivity of Aryl Halides. *J. Am. Chem. Soc.* **2014**, *136*, 3330–3333.

(69) Qian, Y.; Weerapana, E. A Quantitative Mass-Spectrometry Platform to Monitor Changes in Cysteine Reactivity. *Methods Mol. Biol.* **2017**, *1491*, 11–22.

(70) Tucker, M. J.; Courter, J. R.; Chen, J.; Atasoylu, O.; Smith, A. B.; Hochstrasser, R. M. Tetrazine Phototriggers: Probes for Peptide Dynamics. *Angew. Chemie - Int. Ed.* **2010**, *49*, 3612–3616.

(71) Tucker, M. J.; Abdo, M.; Courter, J. R.; Chen, J.; Smith, A. B.; Hochstrasser, R. M. Di-Cysteine S,S-Tetrazine: A Potential Ultra-Fast Photochemical Trigger to Explore the Early Events of Peptide/protein Folding. *J. Photochem. Photobiol. A Chem.* **2012**, *234*, 156–163.

(72) Courter, J. R.; Abdo, M.; Brown, S. P.; Tucker, M. J.; Hochstrasser, R. M.; Smith, A. B. The Design and Synthesis of Alanine-Rich a-Helical Peptides Constrained by an S,s-Tetrazine Photochemical Trigger: A Fragment Union Approach. *J. Org. Chem.* **2014**, *79*, 759–768.

(73) Abdo, M.; Brown, S. P.; Courter, J. R.; Tucker, M. J.; Hochstrasser, R. M.; Smith, A. B. Design, Synthesis, and Photochemical Validation of Peptide Linchpins Containing the S,S-Tetrazine Phototrigger. *Org. Lett.* **2012**, *14*, 3518–3521.

(74) Brown, S. P.; Smith, A. B. Peptide/protein Stapling and Unstapling: Introduction of S-Tetrazine, Photochemical Release, and Regeneration of the Peptide/protein. *J. Am. Chem. Soc.* **2015**, *137*, 4034–4037.

(75) Zhang, D.; Devarie-Baez, N. O.; Li, Q.; Lancaster, J. R.; Xian, M. Methylsulfonyl Benzothiazole (MSBT): A Selective Protein Thiol Blocking Reagent. *Org. Lett.* **2012**, *14*, 3396–3399.

(76) Burnett, J. F.; Zahler, R. E. Aromatic Nucleophilic Substitution Reactions. *Chem. Rev.* **1951**, *49*, 273–412.

(77) Toda, N.; Asano, S.; Barbas, C. F. Rapid, Stable, Chemoselective Labeling of Thiols with Julia- Kocie??ski-like Reagents: A Serum-Stable Alternative to Maleimide-Based Protein Conjugation. *Angew. Chemie - Int. Ed.* **2013**, *52*, 12592–12596.

(78) Patterson, J. T.; Asano, S.; Li, X.; Rader, C.; Barbas, C. F. Improving the Serum Stability of Site-Specific Antibody Conjugates with Sulfone Linkers. *Bioconjug. Chem.* **2014**, *25*, 1402–1407.

(79) Aristeidis Chiotellis. Novel chemoselective18 F-Radiolabeling of Thiol-Containing Biomolecules under Mild Aqueous Conditions. *ChemComm* **2016**, *52*, 6083–6086.

(80)   Zhang, Q.; Dall'Angelo, S.; Fleming, I. N.; Schweiger, L. F.; Zanda, M.; O'Hagan, D. Last-Step Enzymatic [18F]-Fluorination of Cysteine-Tethered RGD Peptides Using Modified Barbas Linkers. *Chem. - A Eur. J.* **2016**, *22*, 10998–11004.

(81)   Farrukh, A.; Paez, J. I.; Salierno, M.; Fan, W.; Berninger, B.; del Campo, A. Bifunctional Poly(acrylamide) Hydrogels through Orthogonal Coupling Chemistries. *Biomacromolecules* **2017**, *18*, 906–913.

(82)   Ema, T.; Inoue, H. Chemical Modification of Lipase for Rational Enhancement of Enantioselectivity. *Chem. Lett.* **2015**, *44*, 1374–1376.

(83)   Zhou, P.; Yao, J.; Hu, G.; Fang, J. Naphthalimide Scaffold Provides Versatile Platform for Selective Thiol Sensing and Protein Labeling. *ACS Chem. Biol.* **2016**, *11*, 1098–1105.

(84)   Bauer, M. R.; Joerger, A. C.; Fersht, A. R. 2-Sulfonylpyrimidines: Mild Alkylating Agents with Anticancer Activity toward p53-Compromised Cells. *Proc. Natl. Acad. Sci.* **2016**, *113*, E5271–E5280.

(85)   Robson, P.; Smith, T. A.; Stephens, R.; Tatlow, J. C. 691. Aromatic Polyfluoro-Compounds. Part XIII. Derivatives of Penta- and 2,3,5,6-Tetra-Fluorothiophenol. *J. Chem. Soc.* **1963**, 3692.

(86)   Birchall, J. M.; Green, M.; Haszeldine, R. N.; Pitts, A. D. Mechanism of the Nucleophilic Substitution Reaction of Polyfluoroarenes. *Chem. Commun.* **1967**, 338–339.

(87)   Bhupathiraju, N. V. S. D. K.; Rizvi, W.; Batteas, J. D.; Drain, C. M. Fluorinated Porphyrinoids as Efficient Platforms for New Photonic Materials, Sensors, and Therapeutics. *Org. Biomol. Chem.* **2016**, *14*, 389–408.

(88)   Ott, C.; Hoogenboom, R.; Schubert, U. S. Post-Modification of Poly(pentafluorostyrene): A Versatile "click" method to Create Well-Defined Multifunctional Graft Copolymers. *Chem. Commun. (Camb).* **2008**, *30*, 3516–3518.

(89)   Becer, C. R.; Babiuch, K.; Pilz, D.; Hornig, S.; Heinze, T.; Gottschaldt, M.; Schubert, U. S. Clicking Pentafluorostyrene Copolymers: Synthesis, Nanoprecipitation, and Glycosylation. *Macromolecules* **2009**, *42*, 2387–2394.

(90)   Noy, J.-M.; Koldevitz, M.; Roth, P. J. Thiol-Reactive Functional Poly(meth)acrylates:

Multicomponent Monomer Synthesis, RAFT (Co)polymerization and Highly Efficient Thiol– Para-Fluoro Postpolymerization Modification. *Polym. Chem.* **2014**, *6*, 436–447.

(91)   Battioni, P.; Brigaud, O.; Desvaux, H.; Mansuy, D.; Traylor, T. G. Preparation of Functionalized Polyhalogenated Tetraaryl-Porphyrins by Selective Substitution of the P-Fluorines of Meso-Tetra-(Pentafluorophenyl)porphyrins. *Tetrahedron Lett.* **1991**, *32*, 2893–2896.

(92)   Yang, J.; Gabriele, B.; Belvedere, S.; Huang, Y.; Breslow, R. Catalytic Oxidations of Steroid Substrates by Artificial Cytochrome P-450 Enzymes. *J. Org. Chem.* **2002**, *67*, 5057–5067.

(93)   Varotto, A.; Todaro, L.; Vinodu, M.; Koehne, J.; Liu, G.; Drain, C. M. Self-Organization of a New Fluorous Porphyrin and C60 Films on Indium-Tin-Oxide Electrode. *Chem. Commun.* **2008**, *1*, 4921.

(94)   Hao, E.; Friso, E.; Miotto, G.; Jori, G.; Soncin, M.; Fabris, C.; Sibrian-Vazquez, M.; Vicente, M. G. H. Synthesis and Biological Investigations of Tetrakis(p-Carboranylthio-Tetrafluorophenyl)chlorin (TPFC). *Org. Biomol. Chem.* **2008**, *6*, 3732–3740.

(95)   Bhupathiraju, N. V. S. D. K.; Vicente, M. G. H. Synthesis and Cellular Studies of Polyamine Conjugates of a Mercaptomethyl-Carboranylporphyrin. *Bioorganic Med. Chem.* **2013**, *21*, 485–495.

(96)   Bhupathiraju, N. V. S. D. K.; Hu, X.; Zhou, Z.; Fronczek, F. R.; Couraud, P. O.; Romero, I. A.; Weksler, B.; Vicente, M. G. H. Synthesis and in Vitro Evaluation of BBB Permeability, Tumor Cell Uptake, and Cytotoxicity of a Series of Carboranylporphyrin Conjugates. *J. Med. Chem.* **2014**, *57*, 6718–6728.

(97)   Gan, D.; Mueller, A.; Wooley, K. L. Amphiphilic and Hydrophobic Surface Patterns Generated from Hyperbranched Fluoropolymer/linear Polymer Networks: Minimally Adhesive Coatings via the Crosslinking of Hyperbranched Fluoropolymers. *J. Polym. Sci. Part A Polym. Chem.* **2003**, *41*, 3531–3540.

(98)   Cheng, C.; Wooley, K. L.; Khoshdel, E. Hyperbranched Fluorocopolymers by Atom Transfer Radical Self-Condensing Vinyl Copolymerization. *J. Polym. Sci. Part A Polym. Chem.* **2005**, *43*, 4754–4770.

89

(99)   Powell, K. T.; Cheng, C.; Gudipati, C. S.; Wooley, K. L.; Wooley, K. L.; Wooley, K. L.; Schaefer, J. Design, Synthesis, and Characterization of Linear Fluorinated Poly(benzyl Ether)s: A Comparison Study with Isomeric Hyperbranched Fluoropolymers. *J. Mater. Chem.* **2005**, *15*, 5128.

(100) Remzi Becer, C.; Hoogenboom, R.; Schubert, U. S. Click Chemistry beyond Metal-Catalyzed Cycloaddition. *Angew. Chemie - Int. Ed.* **2009**, *48*, 4900–4908.

(101) Qian, E. A.; Wixtrom, A. I.; Axtell, J. C.; Saebi, A.; Jung, D.; Rehak, P.; Han, Y.; Moully, E. H.; Mosallaei, D.; Chow, S.; *et al.* Atomically Precise Organomimetic Cluster Nanomolecules Assembled via per Fl Uoroaryl-Thiol S N Ar Chemistry. *Nat. Chem.* **2016**, 1–8.

(102) Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y. S.; Pentelute, B. L. A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.

(103) Zhang, C.; Spokoyny, A. M.; Zou, Y.; Simon, M. D.; Pentelute, B. L. Enzymatic "click" ligation: Selective Cysteine Modification in Polypeptides Enabled by Promiscuous Glutathione S-Transferase. *Angew. Chemie - Int. Ed.* **2013**, *52*, 14001–14005.

(104) Zou, Y.; Spokoyny, A. M.; Zhang, C.; Simon, M. D.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. Convergent Diversity-Oriented Side-Chain Macrocyclization Scan for Unprotected Polypeptides. *Org. Biomol. Chem.* **2013**, *12*, 566–573.

(105) Zhang, C.; Dai, P.; Spokoyny, A. M.; Pentelute, B. L. Enzyme-Catalyzed Macrocyclization of Long Unprotected Peptides. *Org. Lett.* **2014**, *16*, 3652–3655.

(106) Lühmann, T.; Mong, S. K.; Simon, M. D.; Meinel, L.; Pentelute, B. L. A Perfluoroaromatic Abiotic Analog of H2 Relaxin Enabled by Rapid Flow-Based Peptide Synthesis. *Org. Biomol. Chem.* **2016**, *14*, 3345–3349.

(107) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π-Clamp-Mediated Cysteine Conjugation. *Nat. Chem.* **2015**, *8*, 1–9.

(108) Dai, P.; Zhang, C.; Welborn, M.; Shepherd, J. J.; Zhu, T.; Van Voorhis, T.; Pentelute, B. L. Salt Effect Accelerates Site-Selective Cysteine Bioconjugation. *ACS Cent. Sci.* **2016**, *2*, acscentsci.6b00180.

(109) Ngambenjawong, C.; Pineda, J. M. B.; Pun, S. H. Engineering an Affinity-Enhanced Peptide through Optimization of Cyclization Chemistry. *Bioconjug. Chem.* **2016**, acs.bioconjchem.6b00502.

(110) Conibear, A. C.; Chaousis, S.; Durek, T.; Johan Rosengren, K.; Craik, D. J.; Schroeder, C. I. Approaches to the Stabilization of Bioactive Epitopes by Grafting and Peptide Cyclization. *Biopolymers* **2016**, *106*, 89–100.

(111) Swedberg, J. E.; Schroeder, C. I.; Mitchell, J. M.; Durek, T.; Fairlie, D. P.; Edmonds, D. J.; Griffith, D. A.; Ruggeri, R. B.; Derksen, D. R.; Loria, P. M.; *et al.* Cyclic Alpha-Conotoxin Peptidomimetic Chimeras as Potent GLP-1R Agonists. *Eur. J. Med. Chem.* **2015**, *103*, 175–184.

(112) Hilinski, G. J.; Kim, Y.-W.; Hong, J.; Kutchukian, P. S.; Crenshaw, C. M.; Berkovitch, S. S.; Chang, A.; Ham, S.; Verdine, G. L. Stitched α-Helical Peptides via Bis Ring-Closing Metathesis. *J. Am. Chem. Soc.* **2014**, *136*, 12314–12322.

(113) Gunzburg, M. J.; Kulkarni, K.; Watson, G. M.; Ambaye, N. D.; Del Borgo, M. P.; Brandt, R.; Pero, S. C.; Perlmutter, P.; Wilce, M. C. J.; Wilce, J. A. Unexpected Involvement of Staple Leads to Redesign of Selective Bicyclic Peptide Inhibitor of Grb7. *Sci. Rep.* **2016**, *6*, 27060.

(114) Kalhor-Monfared, S.; Jafari, M.-R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. J.; Derda, R. Rapid Biocompatible Macrocyclization of Peptides with Decafluorosulfone. *Chem. Sci.* **2016**, *0*, 1–6.

(115) Hoppmann, C.; Maslennikov, I.; Choe, S.; Wang, L. In Situ Formation of an Azo Bridge on Proteins Controllable by Visible Light. *J. Am. Chem. Soc.* **2015**, *137*, 11218–11221.

(116) Jacobson, O.; Yan, X.; Ma, Y.; Niu, G.; Kiesewetter, D. O.; Chen, X. Novel Method for Radiolabeling and Dimerizing Thiolated Peptides Using 18F-Hexafluorobenzene. *Bioconjug. Chem.* **2015**, *26*, 2016–2020.

(117) Sun, M. M. C.; Beam, K. S.; Cerveny, C. G.; Hamblett, K. J.; Blackmore, R. S.; Torgov, M. Y.; Handley, F. G. M.; Ihle, N. C.; Senter, P. D.; Alley, S. C. Reduction-Alkylation Strategies for the Modification of Specific Monoclonal Antibody Bisulfides. *Bioconjug. Chem.* **2005**, *16*, 1282–1290.

(118) Nathani, R. I.; Moody, P.; Chudasama, V.; Smith, M. E. B.; Fitzmaurice, R. J.; Caddick, S. A Novel Approach to the Site-Selective Dual Labelling of a Protein via Chemoselective Cysteine modification. *Chem. Sci.* **2013**, *4*, 3455–3458.

(119) Daly, N. L.; Love, S.; Alewood, P. F.; Craik, D. J. Chemical Synthesis and Folding Pathways of Large Cyclic Polypeptides: Studies of the Cystine Knot Polypeptide Kalata B1. *Biochemistry* **1999**, *38*, 10606–10614.

(120) Tam, J. P.; Lu, Y. A.; Yu, Q. Thia Zip Reaction for Synthesis of Large Cyclic Peptides: Mechanisms and Applications. *J. Am. Chem. Soc.* **1999**, *121*, 4316–4324.

(121) Aboye, T. L.; Li, Y.; Majumder, S.; Hao, J.; Shekhtman, A.; Camarero, J. A. Efficient One-Pot Cyclization/folding of Rhesus Θ-Defensin-1 (RTD-1). *Bioorganic Med. Chem. Lett.* **2012**, *22*, 2823–2826.

(122) Clark, R. J.; Craik, D. J. Native Chemical Ligation Applied to the Synthesis and Bioengineering of Circular Peptides and Proteins. *Biopolymers* **2010**, *94*, 414–422.

(123) Puljung, M. C.; Zagotta, W. N. Labeling of Specific Cysteines in Proteins Using Reversible Metal Protection. *Biophys. J.* **2011**, *100*, 2513–2521.

(124) Gautier, A.; Juillerat, A.; Heinis, C.; Corria, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **2008**, *15*, 128–136.

(125) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008**, *130*, 12240–12241.

(126) Duckworth, B. P.; Zhang, Z.; Hosokawa, A.; Distefano, M. D. Selective Labeling of Proteins by Using Protein Farnesyltransferase. *ChemBioChem* **2007**, *8*, 98–105.

(127) Kurpiers, T.; Mootz, H. D. Regioselective Cysteine Bioconjugation by Appending a Labeled Cystein Tag to a Protein by Using Protein Splicing in Trans. *Angew. Chemie - Int. Ed.* **2007**, *46*, 5234–5237.

(128) Walsh, C. Enabling the Chemistry of Life. *Nature* **2001**, *409*, 226–231.

(129) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications:

The Chemistry of Proteome Diversifications. *Angew. Chemie - Int. Ed.* **2005**, *44*, 7342–7372.

(130) Hayes, J. D.; Flanagan, J. U.; Jowsey, I. R. Glutathione Transferases. *Rev. Lit. Arts Am.* **2005**, *45*, 51–88.

(131) Armstrong, R. N. Structure, Catalytic Mechanism, and Evolution of the Glutathione Transferases. *Chem. Res. Toxicol.* **1997**, *10*, 2–18.

(132) Buczyński, P.; Dijkstra, K. D. B.; Mauersberger, R.; Moroz, M. D. Review of the Odonata of Belarus. *Odonatologica* **2006**, *35*, 1–13.

(133) Mannervik, B.; Danielson, U. H. Glutathione Transferases--Structure and Catalytic Activity. *CRC Crit. Rev. Biochem.* **1988**, *23*, 283–337.

(134) Wilce, M. C. J.; Parker, M. W. Structure and Function of Glutathione S-Transferases. *Biochim. Biophys. Acta* **1994**, *1205*, 1–18.

(135) Oakley, a J.; Lo Bello, M.; Battistoni, a; Ricci, G.; Rossjohn, J.; Villar, H. O.; Parker, M. W. The Structures of Human Glutathione Transferase P1-1 in Complex with Glutathione and Various Inhibitors at High Resolution. *J. Mol. Biol.* **1997**, *274*, 84–100.

(136) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science.* **1998**, *281*, 269–272.

(137) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124*, 6063–6076.

(138) Wilson, P.; Anastasaki, A.; Owen, M. R.; Kempe, K.; Haddleton, D. M.; Mann, S. K.; Johnston, A. P. R.; Quinn, J. F.; Whittaker, M. R.; Hogg, P. J.; *et al.* Organic Arsenicals as Efficient and Highly Specific Linkers for Protein/peptide-Polymer Conjugation. *J. Am. Chem. Soc.* **2015**, *137*, 4215–4222.

(139) Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; *et al.* Synthesis of Site-Specific Antibody-Drug Conjugates Using Unnatural Amino Acids. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*,

16101–16106.

(140) Zhu, Z.; Ramakrishnan, B.; Li, J.; Wang, Y.; Feng, Y.; Prabakaran, P.; Colantonio, S.; Dyba, M. A.; Qasba, P. K.; Dimitrov, D. S. Site-Specific Antibody-Drug Conjugation through an Engineered Glycotransferase and a Chemically Reactive Sugar. *MAbs* **2014**, *6*, 1190–1200.

(141) Drake, P. M.; Albers, A. E.; Baker, J.; Banas, S.; Barfield, R. M.; Bhat, A. S.; De Hart, G. W.; Garofalo, A. W.; Holder, P.; Jones, L. C.; *et al.* Aldehyde Tag Coupled with HIPS Chemistry Enables the Production of ADCs Conjugated Site-Specifically to Different Antibody Regions with Distinct in Vivo Efficacy and PK Outcomes. *Bioconjug. Chem.* **2014**, *25*, 1331–1341.

(142) Dennler, P.; Chiotellis, A.; Fischer, E.; Brégeon, D.; Belmant, C.; Gauthier, L.; Lhospice, F.; Romagne, F.; Schibli, R. Transglutaminase-Based Chemo-Enzymatic Conjugation Approach Yields Homogeneous Antibody-Drug Conjugates. *Bioconjug. Chem.* **2014**, *25*, 569–578.

(143) Junutula, J. R. J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P. S. P.; Dennis, M. S. M. S.; *et al.* Site-Specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index. *Nat. Biotechnol.* **2008**, *26*, 925–932.

(144) Shen, B.-Q.; Xu, K.; Liu, L.; Raab, H.; Bhakta, S.; Kenrick, M.; Parsons-Reponte, K. L.; Tien, J.; Yu, S.-F.; Mai, E.; *et al.* Conjugation Site Modulates the in Vivo Stability and Therapeutic Activity of Antibody-Drug Conjugates. *Nat. Biotechnol.* **2012**, *30*, 184–189.

(145) Hofmeister, F. Zur Lehre von Der Wirkung Der Salze. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **1888**, *24*, 247-260–260.

(146) Kunz, W.; Henle, J.; Ninham, B. W. "Zur Lehre von Der Wirkung Der Salze" (about the Science of the Effect of Salts): Franz Hofmeister's Historical Papers. *Curr. Opin. Colloid Interface Sci.* **2004**, *9*, 19–37.

(147) Chalker, J. M. Metal-Mediated Bioconjugation. In *Chemoselective and Bioorthogonal Ligation Reactions*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2017; pp. 231–270.

(148) Yang, M.; Li, J.; Chen, P. R. Transition Metal-Mediated Bioorthogonal Protein Chemistry

in Living Cells. *Chem. Soc. Rev.* **2014**, *43*, 6511.

(149) Chankeshwara, S. V.; Indrigo, E.; Bradley, M. Palladium-Mediated Chemistry in Living Cells. *Curr. Opin. Chem. Biol.* **2014**, *21*, 128–135.

(150) Ball, Z. T. Molecular Recognition in Protein Modification with Rhodium Metallopeptides. *Curr. Opin. Chem. Biol.* **2015**, *25*, 98–102.

(151) Antos, J. M.; Francis, M. B. Transition Metal Catalyzed Methods for Site-Selective Protein Modification. *Curr. Opin. Chem. Biol.* **2006**, *10*, 253–262.

(152) Kosugi, M.; Shimizu, T.; Migita, T. Reactions of Aryl Halides With Thiolate Anions in the Presence of Catalytic Amounts of Tetrakis(Triphenylphosphine)Palladium Preparation of Aryl Sulfides. *Chem. Lett.* **1978**, *7*, 13–14.

(153) Migita, T.; Shimizu, T.; Asami, Y.; Shiobara, J.-I.; Kato, Y.; Kosugi, M. The Palladium Catalysed Nucleophilic Substitution of Aryl Halides by Thiolate Anions. *Bull. Chem. Soc. Jpn.* **1980**, *53*, 1385–1389.

(154) Lee, C. F.; Liu, Y. C.; Badsara, S. S. Transition-Metal-Catalyzed C-S Bond Coupling Reaction. *Chem. - An Asian J.* **2014**, *9*, 706–722.

(155) Kung, K. K.-Y.; Ko, H.-M.; Cui, J.-F.; Chong, H.-C.; Leung, Y.-C.; Wong, M.-K. Cyclometalated gold(III) Complexes for Chemoselective Cysteine Modification via Ligand Controlled C-S Bond-Forming Reductive Elimination. *Chem. Commun. (Camb).* **2014**, *50*, 11899–11902.

(156) Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic Palladium Reagents for Cysteine Bioconjugation. *Nature* **2015**, *526*, 687–691.

(157) Rojas, A. J.; Zhang, C.; Vinogradova, E. V.; Buchwald, N.; Reilly, J.; Pentelute, B.; Buchwald, S. Divergent Unprotected Peptide Macrocyclisation by Palladium-Mediated Cysteine Arylation. *Chem. Sci.* **2017**, *0*, 1–7.

(158) Shafer, D. E.; Inman, J. K.; Lees, A. Reaction of Tris(2-Carboxyethyl)phosphine (TCEP) with Maleimide and Alpha-Haloacyl Groups: Anomalous Elution of TCEP by Gel Filtration. *Anal. Biochem.* **2000**, *282*, 161–164.

(159) Al-Shuaeeb, R. A. A.; Kolodych, S.; Koniev, O.; Delacroix, S.; Erb, S.; Nicolaÿ, S.; Cintrat, J.-C.; Brion, J.-D.; Cianférani, S.; Alami, M.; *et al.* Palladium-Catalyzed Chemoselective and Biocompatible Functionalization of Cysteine-Containing Molecules at Room Temperature. *Chem. - A Eur. J.* **2016**, *22*, 11365–11370.

(160) Willwacher, J.; Raj, R.; Mohammed, S.; Davis, B. G. Selective Metal-Site-Guided Arylation of Proteins. *J. Am. Chem. Soc.* **2016**, *138*, 8678–8681.

(161) Booth, C.; Gaspar, H. B. Pegademase Bovine (PEG-ADA) for the Treatment of Infants and Children with Severe Combined Immunodeficiency (SCID). *Biologics* **2009**, *3*, 349–358.

(162) Guram, A.; Buchwald, S. Palladium-Catalyzed Aromatic Aminations with in Situ Generated Aminostannanes. *J. Am. Chem. Soc.* **1994**, *116*, 7901–7902.

(163) Paul, F.; Patt, J.; Hartwig, J. F. Palladium-Catalyzed Formation of Carbon-Nitrogen Bonds. Reaction Intermediates and Catalyst Improvements in the Hetero Cross-Coupling of Aryl Halides and Tin Amides. *J. Am. Chem. Soc.* **1994**, *116*, 5969–5970.

(164) Chan, D. M. T.; Monaco, K. L.; Wang, R. P.; Winters, M. P. New N- and O-Arylations with Phenylboronic Acids and Cupric Acetate. *Tetrahedron Lett.* **1998**, *39*, 2933–2936.

(165) Lam, P. Y. S.; Clark, C. G.; Saubern, S.; Adams, J.; Winters, M. P.; Chan, D. M. T.; Combs, A. New Aryl/heteroaryl C-N Bond Cross-Coupling Reactions via Arylboronic Acid/cupric Acetate Arylation. *Tetrahedron Lett.* **1998**, *39*, 2941–2944.

(166) Lam, P. Y. S. Chan-Lam Coupling Reaction: Copper-Promoted C–Element Bond Oxidative Coupling Reaction with Boronic Acids. In *Synthetic Methods in Drug Discovery*; 2016; pp. 242–273.

(167) Ullmann, F. Ueber Eine Neue Bildungsweise von Diphenylaminderivaten. *Berichte der Dtsch. Chem. Gesellschaft* **1903**, *36*, 2382–2384.

(168) Goldberg, I. Ueber Phenylirungen Bei Gegenwart von Kupfer Als Katalysator. *Berichte der Dtsch. Chem. Gesellschaft* **1906**, *39*, 1691–1692.

(169) Lercher, L.; McGouran, J. F.; Kessler, B. M.; Schofield, C. J.; Davis, B. G. DNA Modification under Mild Conditions by Suzuki-Miyaura Cross-Coupling for the Generation of Functional Probes. *Angew. Chemie - Int. Ed.* **2013**, *52*, 10553–10558.

(170) Sambiagio, C.; Marsden, S. P.; Blacker, A. J.; McGowan, P. C. Copper Catalysed Ullmann Type Chemistry: From Mechanistic Aspects to Modern Development. *Chem. Soc. Rev.* **2014**, *43*, 3525.

(171) Sanger, F. The Free Amino Groups of Insulin. *Biochem. J.* **1945**, *39*, 507–515.

(172) Sutton, B. D. A.; Drewes, S. E.; Welz, U. Evaluation of 1-Fluoro-2-Nitro-4-Trimethyla Unmondobenzene Iodide , a Protein-Solubilizing Reagent. *Biochem. J.* **1972**, *130*, 589–595.

(173) Ladd, D. L.; Snow, R. A. Reagents for the Preparation of Chromophorically Labeled Polyethylene Glycol-Protein Conjugates. *Anal. Biochem.* **1993**, *210*, 258–261.

(174) Lautrette, G.; Touti, F.; Lee, H. G.; Dai, P.; Pentelute, B. L. Nitrogen Arylation for Macrocyclization of Unprotected Peptides. *J. Am. Chem. Soc.* **2016**, *138*, 8340–8343.

(175) Lee, H. G.; Lautrette, G.; Pentelute, B. L.; Buchwald, S. L. Palladium-Mediated Arylation of Lysine in Unprotected Peptides. *Angew. Chemie Int. Ed.* **2017**, 1–6.

(176) Ohata, J.; Minus, M. B.; Abernathy, M. E.; Ball, Z. T. Histidine-Directed Arylation/Alkenylation of Backbone N-H Bonds Mediated by Copper(II). *J. Am. Chem. Soc.* **2016**, *138*, 7472–7475.

(177) Crawford, L. a; Weerapana, E. A Tyrosine-Reactive Irreversible Inhibitor for Glutathione S-Transferase Pi (GSTP1). *Mol. BioSyst.* **2016**, *12*, 1–4.

(178) Fehler, S. K.; Pratsch, G.; Östreicher, C.; Fürst, M. C. D.; Pischetsrieder, M.; Heinrich, M. R. Radical Arylation of Tyrosine Residues in Peptides. *Tetrahedron* **2016**, *72*, 7888–7893.

(179) Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B. Dual-Surface Modification of the Tobacco Mosaic Virus. *J. Am. Chem. Soc.* **2005**, *127*, 3718–3723.

(180) Chapman, C. J.; Matsuno, A.; Frost, C. G.; Willis, M. C. Site-Selective Modification of Peptides Using Rhodium and Palladium Catalysis: Complementary Electrophilic and Nucleophilic Arylation. *Chem. Commun. (Camb).* **2007**, 3903–3905.

(181) Chapman, C. J.; Hargrave, J. D.; Bish, G.; Frost, C. G. Peptide Modification through Site-Selective Residue Interconversion: Application of the Rhodium-Catalysed 1,4-Addition of Aryl Siloxanes and Boronates. *Tetrahedron* **2008**, *64*, 9528–9539.

(182) Chen, X.; Engle, K. M.; Wang, D.-H.; Yu, J.-Q. Palladium(II)-Catalyzed C-H Activation/C-

C Cross-Coupling Reactions: Versatility and Practicality. *Angew. Chemie Int. Ed.* **2009**, *48*, 5094–5115.

(183) Sandtorv, A. H. Transition Metal-Catalyzed C□H Activation of Indoles. *Adv. Synth. Catal.* **2015**, *357*, 2403–2435.

(184) Ruiz-Rodriguez, J.; Albericio, F.; Lavilla, R. Postsynthetic Modification of Peptides: Chemoselective C-Arylation of Tryptophan Residues. *Chem. - A Eur. J.* **2010**, *16*, 1124–1127.

(185) Mendive-Tapia, L.; Preciado, S.; García, J.; Ramón, R.; Kielland, N.; Albericio, F.; Lavilla, R. New Peptide Architectures through C–H Activation Stapling between Tryptophan-phenylalanine/tyrosine Residues. *Nat. Commun.* **2015**, *6*, 7160.

(186) Mendive-Tapia, L.; Bertran, A.; García, J.; Acosta, G.; Albericio, F.; Lavilla, R. Constrained Cyclopeptides: Biaryl Formation through Pd-Catalyzed C–H Activation in Peptides-Structural Control of the Cyclization vs. Cyclodimerization Outcome. *Chem. - A Eur. J.* **2016**, *22*, 13114–13119.

(187) Dong, H.; Limberakis, C.; Liras, S.; Price, D.; James, K. Peptidic Macrocyclization via Palladium-Catalyzed Chemoselective Indole C-2 Arylation. *Chem. Commun.* **2012**, *48*, 11644–11646.

(188) Schischko, A.; Ren, H.; Kaplaneris, N.; Ackermann, L. Bioorthogonal Diversification of Peptides through Selective Ruthenium(II)-Catalyzed C-H Activation. *Angew. Chemie Int. Ed.* **2017**, 1576–1580.

(189) Williams, T. J.; Reay, A. J.; Whitwood, A. C.; Fairlamb, I. J. S. A Mild and Selective Pd-Mediated Methodology for the Synthesis of Highly Fluorescent 2-Arylated Tryptophans and Tryptophan-Containing Peptides: A Catalytic Role for Pd0 Nanoparticles? *Chem. Commun.* **2014**, *50*, 3052.

(190) Reay, A. J.; Williams, T. J.; Fairlamb, I. J. S. Unified Mild Reaction Conditions for C2-Selective Pd-Catalysed Tryptophan Arylation, Including Tryptophan-Containing Peptides. *Org. Biomol. Chem.* **2015**, *13*, 8298–8309.

(191) Hatfield, D. L.; Tsuji, P. A.; Carlson, B. A.; Gladyshev, V. N. Selenium and Selenocysteine: Roles in Cancer, Health, and Development. *Trends Biochem. Sci.* **2014**, *39*, 112–120.

(192) Byun, B. J.; Kang, Y. K. Conformational Preferences and pKa Value of Selenocysteine Residue. *Biopolymers* **2011**, *95*, 345–353.

(193) Hofer, T.; Thomas, J. D.; Burke, T. R.; Rader, C. An Engineered Selenocysteine Defines a Unique Class of Antibody Derivatives. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 12451–12456.

(194) Cohen, D. T.; Zhang, C.; Pentelute, B. L.; Buchwald, S. L. An Umpolung Approach for the Chemoselective Arylation of Selenocysteine in Unprotected Peptides. *J. Am. Chem. Soc.* **2015**, *137*, 9784–9787.

(195) Taniguchi, N. Convenient Synthesis of Unsymmetrical Organochalcogenides Using Organoboronic Acids with Dichalcogenides via Cleavage of the S-S, Se-Se, or Te-Te Bond by a Copper Catalyst. *J. Org. Chem.* **2007**, *72*, 1241–1245.

(196) Zheng, B.; Gong, Y.; Xu, H. J. Copper-Catalyzed C-Se Coupling of Diphenyl Diselenide with Arylboronic Acids at Room Temperature. *Tetrahedron* **2013**, *69*, 5342–5347.

(197) Chalker, J. M.; Wood, C. S. C.; Davis, B. G. A Convenient Catalyst for Aqueous and Protein Suzuki-Miyaura Cross-Coupling. *J. Am. Chem. Soc.* **2009**, *131*, 16346–16347.

(198) Simmons, R. L.; Yu, R. T.; Myers, A. G. Storable arylpalladium(II) Reagents for Alkene Labeling in Aqueous Media. *J. Am. Chem. Soc.* **2011**, *133*, 15870–15873.

(199) Vilaró, M.; Arsequell, G.; Valencia, G.; Ballesteros, A.; Barluenga, J. Arylation of Phe and Tyr Side Chains of Unprotected Peptides by a Suzuki-Miyaura Reaction in Water. *Org. Lett.* **2008**, *10*, 3243–3245.

(200) Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; *et al.* Regioselective Carbon-Carbon Bond Formation in Proteins with Palladium Catalysis; New Protein Chemistry by Organometallic Chemistry. *ChemBioChem* **2006**, *7*, 134–139.

(201) Dibowski, H.; Schmidtchen, F. P. Bioconjugation of Peptides by Palladium-Catalyzed C−C Cross-Coupling in Water. *Angew. Chemie Int. Ed.* **1998**, *37*, 476–478.

(202) Ojida, A.; Tsutsumi, H.; Kasagi, N.; Hamachi, I. Suzuki Coupling for Protein Modification. *Tetrahedron Lett.* **2005**, *46*, 3301–3305.

(203) Bong, D. T.; Ghadiri, M. R. Chemoselective Pd(0)-Catalyzed Peptide Coupling in Water. *Org. Lett.* **2001**, *3*, 2509–2511.

(204) Kodama, K.; Fukuzawa, S.; Nakayama, H.; Sakamoto, K.; Kigawa, T.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Yokoyama, S.; *et al.* Site-Specific Functionalization of Proteins by Organopalladium Reactions. *ChemBioChem* **2007**, *8*, 232–238.

(205) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G. A Genetically Encoded Boronate-Containing Amino Acid. *Angew. Chemie - Int. Ed.* **2008**, *47*, 8220–8223.

(206) Spicer, C. D.; Davis, B. G. Palladium-Mediated Site-Selective Suzuki-Miyaura Protein Modification at Genetically Encoded Aryl Halides. *Chem. Commun.* **2011**, *47*, 1698–1700.

(207) Wang, Y.-S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P.-J.; Russell, D. H.; Liu, W. R. The de Novo Engineering of Pyrrolysyl-tRNA Synthetase for Genetic Incorporation of L-Phenylalanine and Its Derivatives. *Mol. Biosyst.* **2011**, *7*, 714.

(208) Li, N.; Lim, R. K. V; Edwardraja, S.; Lin, Q. Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells. *J. Am. Chem. Soc.* **2011**, *133*, 15316–15319.

(209) Li, J.; Lin, S.; Wang, J.; Jia, S.; Yang, M.; Hao, Z.; Zhang, X.; Chen, P. R. Ligand-Free Palladium-Mediated Site-Specific Protein Labeling inside Gram-Negative Bacterial Pathogens. *J. Am. Chem. Soc.* **2013**, *135*, 7330–7338.

(210) Dumas, A.; Spicer, C. D.; Gao, Z.; Takehana, T.; Lin, Y. A.; Yasukohchi, T.; Davis, B. G. Self-Liganded Suzuki-Miyaura Coupling for Site-Selective Protein PEGylation. *Angew. Chemie - Int. Ed.* **2013**, *52*, 3916–3921.

(211) Ma, X.; Wang, H.; Chen, W. N-Heterocyclic Carbene-Stabilized Palladium Complexes as Organometallic Catalysts for Bioorthogonal Cross-Coupling Reactions. *J. Org. Chem.* **2014**, *79*, 8652–8658.

(212) Spicer, C. D.; Triemer, T.; Davis, B. G. Palladium-Mediated Cell-Surface Labeling. *J. Am. Chem. Soc.* **2012**, *134*, 800–803.

(213) Cheng, G.; Lim, R. K. V.; Li, N.; Lin, Q.; Herman, J. G.; Graff, J. R.; Myohanen, S.; Nelkin, B. D.; Baylin, S. B.; Kaiser, E. T.; *et al.* Storable Palladacycles for Selective

Functionalization of Alkyne-Containing Proteins. *Chem. Commun.* **2013**, *49*, 6809.

(214) Cheng, G.; Lim, R. K. V; Ramil, C. P.; Lin, Q. Storable N-Phenylcarbamate Palladacycles for Rapid Functionalization of an Alkyne-Encoded Protein. *Chem. Commun. (Camb).* **2014**, *50*, 11679–11682.

(215) Liu, J.; Cao, Z.; Lu, Y. Functional Nucleic Acid Sensors. *Chem. Rev.* **2009**, *109*, 1948–1998.

(216) Du, Y.; Dong, S. Nucleic Acid Biosensors: Recent Advances and Perspectives. *Anal. Chem.* **2017**, *89*, 189–215.

(217) *DNA Nanotechnology*; Fan, C., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013.

(218) Lundin, K. E.; Gissberg, O.; Smith, C. I. E. Oligonucleotide Therapies: The Past and the Present. *Hum. Gene Ther.* **2015**, *26*, 475–485.

(219) Silverman, S. K. DNA as a Versatile Chemical Component for Catalysis, Encoding, and Stereocontrol. *Angew. Chemie Int. Ed.* **2010**, *49*, 7180–7201.

(220) Beaucage, S. L.; Iyer, R. P. The Synthesis of Modified Oligonucleotides by the Phosphoramidite Approach and Their Applications. *Tetrahedron* **1993**, *49*, 6123–6194.

(221) Jäger, S.; Rasched, G.; Kornreich-Leshem, H.; Engeser, M.; Thum, O.; Famulok, M. A Versatile Toolbox for Variable DNA Functionalization at High Density. *J. Am. Chem. Soc.* **2005**, *127*, 15071–15082.

(222) Omumi, A.; Beach, D. G.; Baker, M.; Gabryelski, W.; Manderville, R. A. Postsynthetic Guanine Arylation of DNA by Suzuki-Miyaura Cross-Coupling. *J. Am. Chem. Soc.* **2011**, *133*, 42–50.

(223) Ding, Y.; Clark, M. A. Robust Suzuki–Miyaura Cross-Coupling on DNA-Linked Substrates. *ACS Comb. Sci.* **2015**, *17*, 1–4.

(224) Ding, Y.; Franklin, G. J.; DeLorey, J. L.; Centrella, P. A.; Mataruse, S.; Clark, M. A.; Skinner, S. R.; Belyanskaya, S. Design and Synthesis of Biaryl DNA-Encoded Libraries. *ACS Comb. Sci.* **2016**, *18*, 625–629.

(225) Ding, Y.; DeLorey, J. L.; Clark, M. A. Novel Catalyst System for Suzuki-Miyaura Coupling of Challenging DNA-Linked Aryl Chlorides. *Bioconjug. Chem.* **2016**, *27*, 2597–2600.

(226) Wang, Z.; Cohen, S. M.; Stern, C. L.; Hupp, J. T.; Li, J.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M.; Chaudret, B.; *et al.* Postsynthetic Modification of Metal–organic Frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1315.

(227) Kango, S.; Kalia, S.; Celli, A.; Njuguna, J.; Habibi, Y.; Kumar, R. Surface Modification of Inorganic Nanoparticles for Development of Organic–inorganic nanocomposites—A Review. *Prog. Polym. Sci.* **2013**, *38*, 1232–1261.

Chapter 2. Glutathione S-Transferase-Catalysed Regioselective Cysteine Arylation

The work presented in this chapter is a part of the following manuscripts and is reproduced here with permission from John Wiley & Sons and the American Chemical Society.

Zhang, C.; Spokoyny, A. M.; Zou, Y.; Simon, M. D.; Pentelute, B. L.

Enzymatic "Click" Ligation: Selective Cysteine Modification in Polypeptides Enabled by Promiscuous Glutathione S-Transferase

Angew. Chem. Int. Ed. **2013**, *15*, 721–733. DOI: 10.1002/anie.201306430

Zhang, C.; Dai, P.; Spokoyny, A. M.; Pentelute, B. L.

Enzyme-Catalyzed Macrocyclization of Long Unprotected Peptides

Org. Lett. **2014**, *16*, 3652–3655. DOI: 10.1021/ol501609y

2.1. Introduction

Post-translational modifications of biomolecules represent nature's versatile chemical toolbox, allowing for precise and on-demand placement of various functional moieties *in vivo*.[1] Synthetic chemists have strived to develop a similar reaction toolkit with chemical orthogonality, selectivity and reactivity comparable to those of natural systems.[2,3] Recently, "click" chemistry has produced several powerful transformations applicable to a wide range of synthetic settings.[4–9] While the efficiency and chemical orthogonality of such "click" processes are remarkable, regio-control for these transformations has been limited and remains a major challenge.[10] The difficulty of selectively differentiating between two chemically identical sites within a macromolecule is routinely surmounted in natural systems by recognition elements such as binding pockets, directing groups and allosteric components. Although various enzymes have been engineered to catalyze reactions for protein modification and engineering,[11–18] they often require a long recognition sequence in polypeptides, and their substrate scope is relatively restricted. In certain cases, enzymes are capable of recognizing and carrying out efficient transformations on more than one specific substrate, a phenomenon known as enzyme promiscuity. Promiscuous enzymes have been previously engineered to catalyze several synthetically important organic transformations.[19–23]

Both natural and hybrid cyclic peptides represent an important class of molecules being actively explored as potential therapeutics.[24–26] These species are attractive because of their enhanced binding affinity, exo- and endo-peptidase resistance, and in certain cases increased cell penetration compared to their linear counterparts. Although enzymes catalyze versatile transformations in nature, enzyme-catalyzed peptide macrocyclizations remain rare.[27,28] Subtiligase was used to catalyze amide-bond formation between an *N*-terminal amine and a *C*-terminal ester to generate a 31 residue cyclic peptide with 85% yield.[29] Thioesterase was successfully employed to cyclize linear peptides generated from the nonribosomal peptide synthetase system.[30,31] Sortase was used to catalyze the cyclization of peptides, glycopeptides, and proteins.[32–36] With the sortase-catalyzed peptide cyclization method,[32] a 67% yield of cyclic peptide product was produced and the linear oligomer was the major side product.

Synthetic methods for modifying biomolecules provide opportunities for studying protein structure-function relationships as well as creating proteins with new properties and applications.[37–44] Cysteine (Cys) is often chosen for modification because of the unique reactivity

of its thiol group[45,46] as well as the low abundance of Cys residues in the majority of naturally occurring proteins (ca. 1.7%).[47] Various chemical methods for Cys modification have been developed, such as alkylation[48], oxidation[49], and desulfurization[50–55].

Both cysteine-based reactions and other chemical macrocyclization methods have been developed to cross-link two functional groups within a linear peptide substrate through head-to-tail, head-to-side chain, tail-to-side chain, or side chain-to-side chain.[56–60] Cross-linking methods include amide-bond or ester bond formation,[61,62] thiol-based substitution with alkyl and benzyl halide electrophiles (commonly referred to as "alkylation"),[63–67] oxidation at cysteine or selenocysteine residues to form dichalocogenide bond,[24,68] native chemical ligation,[50,69–72] intein-mediated approaches,[73,74] Staudinger ligation,[75] transition metal-catalyzed coupling reactions,[76,77] cycloadditions,[78–81] coordination-based approaches,[82,83] as well as several non-covalent strategies[84,85].

Recently, we developed a perfluoroaryl-cysteine S_NAr chemistry approach to modifying unprotected peptides (Figure 2.1A).[86,87] In particular, two cysteine residues on an unprotected polypeptide were cross-linked with bifunctional perfluoroaromatic reagents. This macrocyclization strategy was versatile enabling the cross-linking of two cysteine moieties positioned at all sites ranging from *i, i+1* to *i, i+14* on a 14-residue unprotected peptide.[87] For longer peptides, however, we found that entropic penalty becomes large enough to render macrocyclizations inefficient. This fundamental limitation is therefore expected to be encountered with any long and unstructured polypeptide irrespective of the cyclization chemistry used. Moreover, while this perfluoroaryl-cysteine conjugation satisfies several requirements of "click" chemistry,[7,88] the insolubility and low reactivity of the perfluoroaromatic reagents in aqueous media restricts the general application of this process to a narrow range of biomolecules.

In our search to render this chemistry suitable for aqueous conditions we turned our attention to glutathione *S*-transferase (GST).[89,90] Known as a class of promiscuous enzymes, GST catalyzes conjugation reactions between the Cys residue of glutathione (GSH, γ-Glu-Cys-Gly) and various electrophiles, thus allowing the cell to detoxify xenobiotics *in vivo* (Figure 2.1B).[89] We reasoned that the substrate promiscuity of GST might be hijacked to catalyze the reaction between perfluroaromatic reagents and GSH tag installed on a polypeptide. Such enzyme-catalyzed

reaction might provide enhanced reaction efficiency, allowing us to overcome the length restrictions previously encountered in non-catalyzed S$_N$Ar macrocyclization reactions.
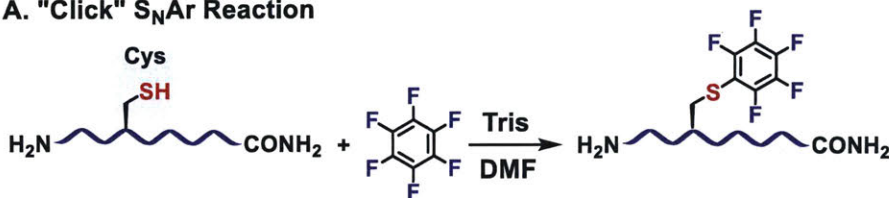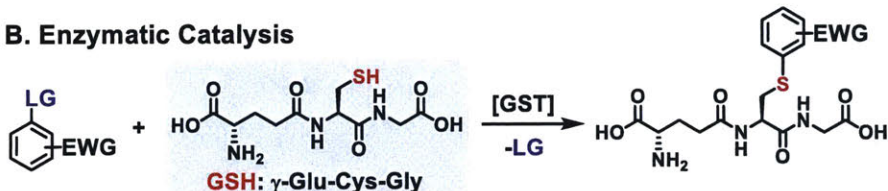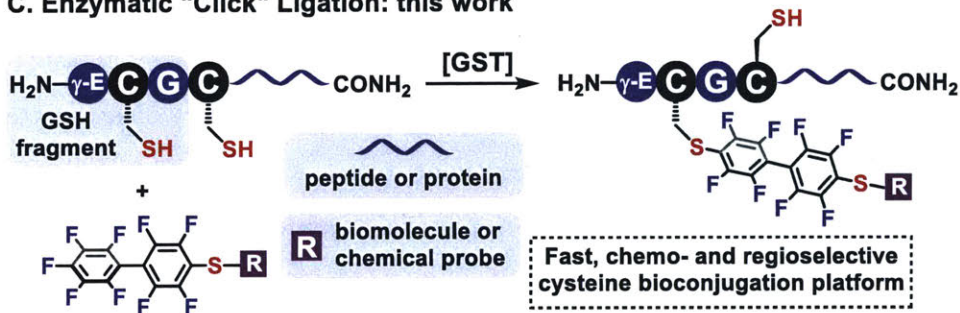
Figure 2.1. Concept of GST-catalyzed cysteine arylation.

(**A**) Peptide "click" modification via perfluoroaryl-cysteine S$_N$Ar chemistry. (**B**) GST-catalyzed S$_N$Ar reaction. GST catalyzes the conjugation of activated aromatic electrophiles to GSH cysteine thiol. EWG: electron withdrawing group, LG: leaving group. (**C**) GST-catalyzed cysteine arylation. GST catalyzes the conjugation of probes bearing 4-mercaptoperfluoro-biphenyl moiety (Cys-**III**) to the *N*-terminal γ-Glu-Cys-Gly sequence of a peptide or a protein. Amino acids are shown in single-letter code. γ-E stands for γ-glutamyl.

2.2. Results

Here we report a novel bioconjugation chemistry that combines a previously developed perfluoroaryl-cysteine S$_N$Ar "click" reaction with GST enzyme catalysis (Figure 2.1C). The promiscuous nature of GST facilitates the bioconjugation with polypeptides other than GSH in aqueous media and provides previously unattainable chemo- and regioselective functionalization of a single cysteine thiol in the presence of other unprotected cysteine residues and reactive functional groups on the same polypeptide chain. This chemistry can be carried out over a broad range of temperatures (4-60 °C) and is compatible with the addition of organic co-solvents (up to 20%). Furthermore, we applied this process to peptide macrocyclization where GST catalyzes the selective S$_N$Ar reaction between an *N*-terminal glutathione (GSH, γ-Glu-Cys-Gly) tag and a *C*-terminal perfluoroaryl-modified cysteine on the same polypeptide chain. Cyclic peptides ranging from 9 to 24 residues were quantitatively produced within 2 hours in aqueous buffer (pH 8.0) at room temperature. The reaction was highly selective for cyclization at the GSH tag, enabling the combination of GST-catalyzed ligation with native chemical ligation to generate a large 40-residue peptide macrocycle.

2.2.1. GST-catalyzed arylation of glutathione with a L-pentafluorophenylalanine electrophile

We hypothesized that the broad scope of electrophiles accepted by GST isozymes might be sufficient for members of this enzyme family to mediate reactions between perfluoroaryl electrophiles and peptides containing GSH in an aqueous environment. To achieve the broadest electrophile scope, a mixture of GST isozymes was chosen for screening (see Experimental section 2.4.2). We first tested the GST-catalyzed conjugation of GSH to model peptides containing L-pentafluorophenylalanine (residue **I**, Figure 2A). Reacting **1** with GSH at 37 °C in aqueous solution at pH 8.0 with 2 mg/mL GST (ca. 5-10 mol% relative to **1**) for two hours generated conjugated product **2**, as confirmed by LC-MS analysis (Figure 2.2), whereas no product was observed without the enzyme. Nucleophilic residues in model peptide **1**, such as Cys and Lys, were unreactive, indicating that the Cys of GSH can be selectively modified with pentafluorophenyl-based electrophiles under GST catalysis.
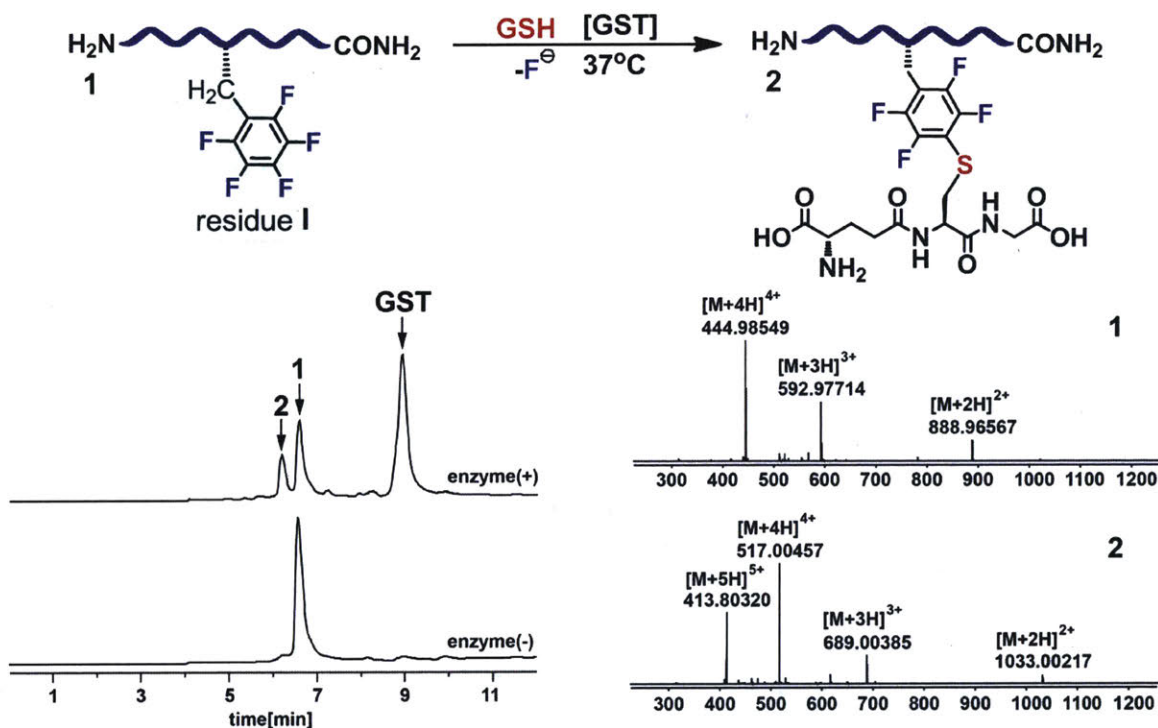
Figure 2.2. GST-catalyzed conjugation of GSH to L-pentafluorophenylalanine residue.
Peptide sequence of **1**: NH2-ITPCNLLF*YYGKKK-CONH2, F* stands for L-pentafluorophenylalanine. Reaction conditions: 1 mM peptide **1**, 1 mM GSH, 2 mg/mL GST, 20 mM *tris*(2-carboxylethyl)phosphine hydrochloride (TCEP·HCl), 0.1 M phosphate buffer, pH 8.0, 37 °C. Total ion current (TIC) traces are shown; mass spectra of LC peaks at highest intensity are shown as an inset.
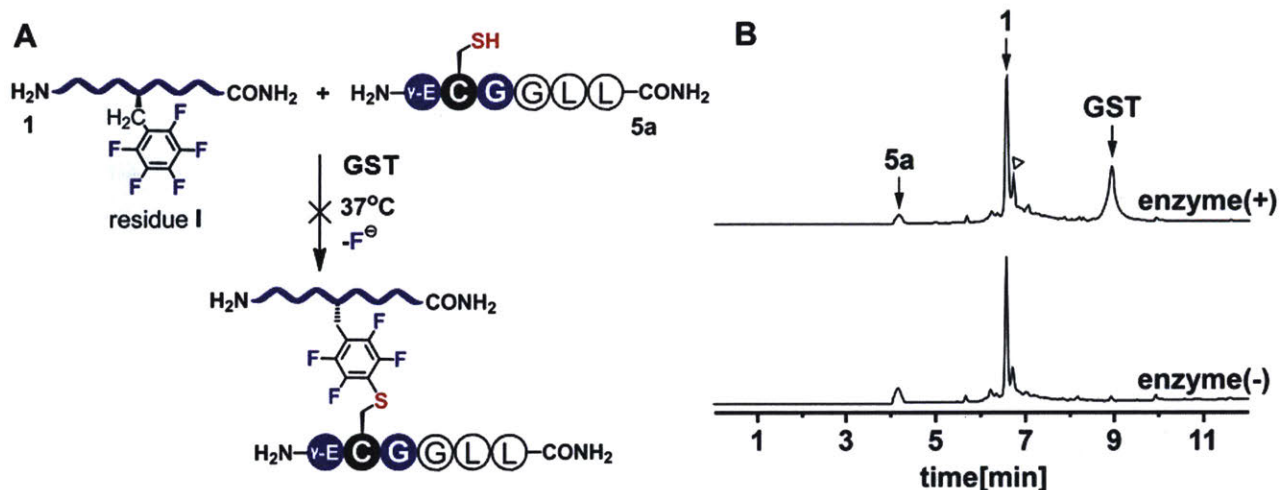
Figure 2.3. GST cannot catalyze the conjugation of L-pentafluorophenylalanine residue to GSH-tagged peptide.

Reactions of peptide **1** containing L-pentafluorophenylalanine residue (residue **I**) with peptide **5a** containing an *N*-terminal GSH (γ-Glu-Cys-Gly) sequence. No product was observed with the addition of GST. (**A**) Reaction conditions: 2 mM **1**, 1 mM **5a**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C. (**B**) LC-MS analysis of crude reactions at 10 hours. Total ion current (TIC) traces are shown. Bottom chromatogram represents a control experiment conducted without an enzyme. The triangle labeled shoulder peak is unreactive impurity contaminant of **1**.

2.2.2. GST-catalyzed arylation of glutathione with perfluoroarylated cysteine electrophiles

To examine the substrate scope of this reaction, we first tested whether the mixture of GST isozymes could catalyze the conjugation of L-pentafluorophenylalanine residue to peptides bearing N-terminal GSH (γ-Glu-Cys-Gly). However, a hexapeptide containing N-terminal glutathione sequence (γ-Glu-Cys-Gly-Gly-Leu-Leu) did not show reactivity towards **1** (Figure 2.3). We then hypothesized that increasing the electrophilicity of the perfluoroaryl moiety might improve the reactivity of the peptide-based substrate sufficiently to allow GST-mediated conjugation with peptides containing N-terminal GSH sequence. Our previous study showed that *para*-thioether substituent on the perfluoroaryl moiety can stabilize the negative charge of the SNAr reaction intermediate thereby increasing the reaction rate.[86] We evaluated the enzymatic reactivity of peptides containing several *para*-thioether substituted electrophiles derived from cysteine. Importantly, reactions with these peptides showed enhanced reaction rates as compared to peptide containing L-pentafluorophenylalanine (residue **I**, Figure 2.2). Specifically, peptide containing cysteine modified with perfluorophenyl residue (Cys-**II**) (Figure 2.4A and Figure 2.4C, **3a**; see Experimental section 2.4.8 for synthesis) reacted with GSH in the presence of GST at a significantly higher rate as compared to **1**, yielding 93% of GSH-conjugated product in less than four hours. Reactions with peptides containing Cys moiety functionalized with perfluorobiphenyl species (Cys-**III**) (Figure 2.4A, **3b**) and perfluorobiphenyl sulfide (Cys-**IV**) (Figure 2.4A, **3c**) proceeded with quantitative conversions in less than 30 minutes (Figure 2.4C).
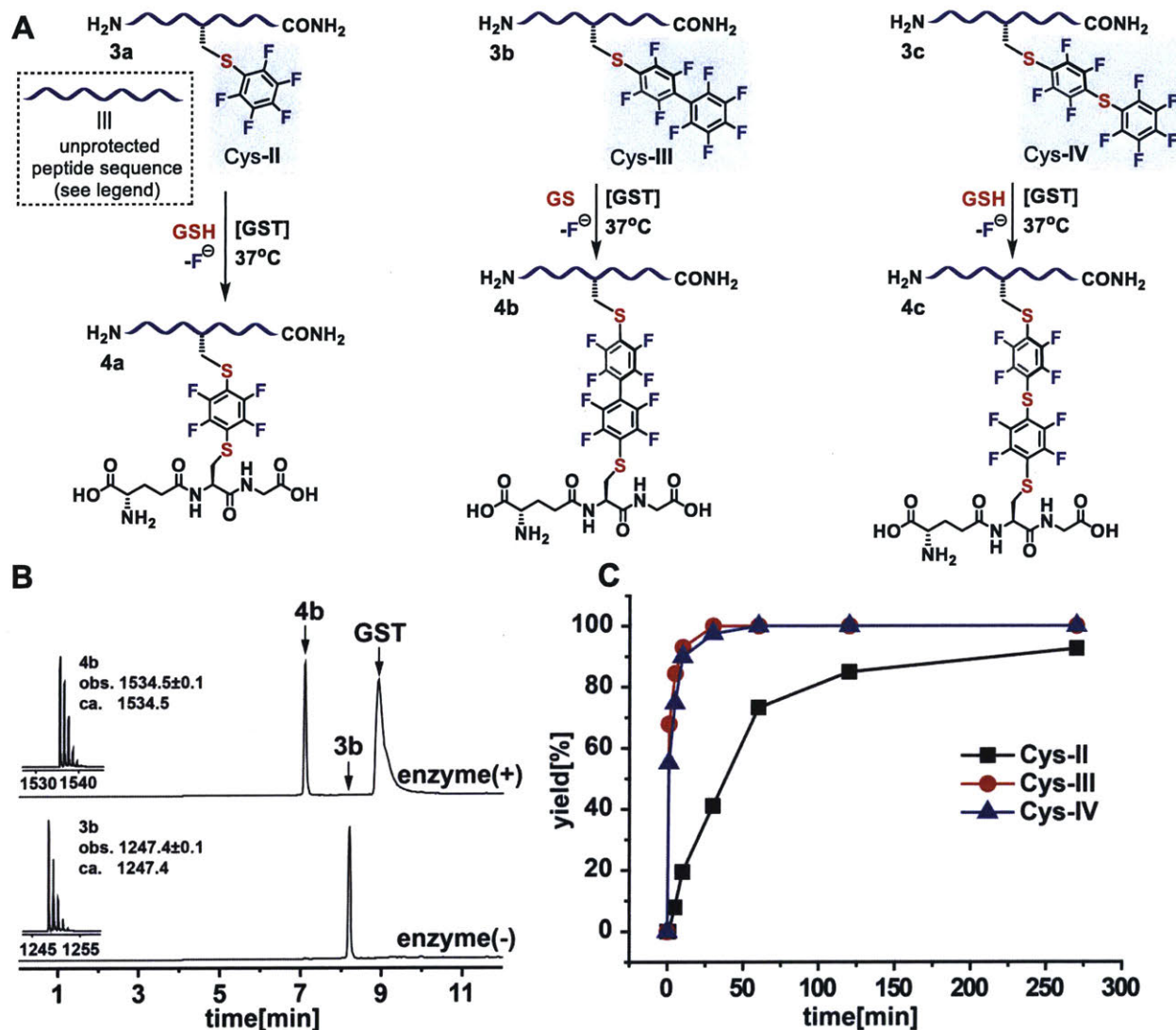
Figure 2.4. GST-catalyzed arylation of GSH with perfluoroarylated cysteine electrophiles.
(**A**) GST-catalyzed conjugation of GSH with peptides containing perfluoroaromatic electrophilic residues. peptide sequences of **3a-c**: H₂N-VTLPSTC*GAS-CONH₂, C* refers to the modified cysteine. Reaction conditions: 1 mM **1** or **3a-c**, 1 mM GSH, 2 mg/mL GST, 20 mM *tris*(2-carboxylethyl)phosphine hydrochloride (TCEP·HCl), 0.1 M phosphate buffer, pH 8.0, 37 °C. (**B**) LC-MS analysis of crude reaction with peptide containing Cys-**III** residue after 30 minutes. (**C**) Rates of formation of GSH-conjugated product with different electrophiles (see Figure 2.5 for complete LC-MS analysis). Yields were determined by LC-MS analysis (see Experimental section 2.4.6 for the method for calculating yields).
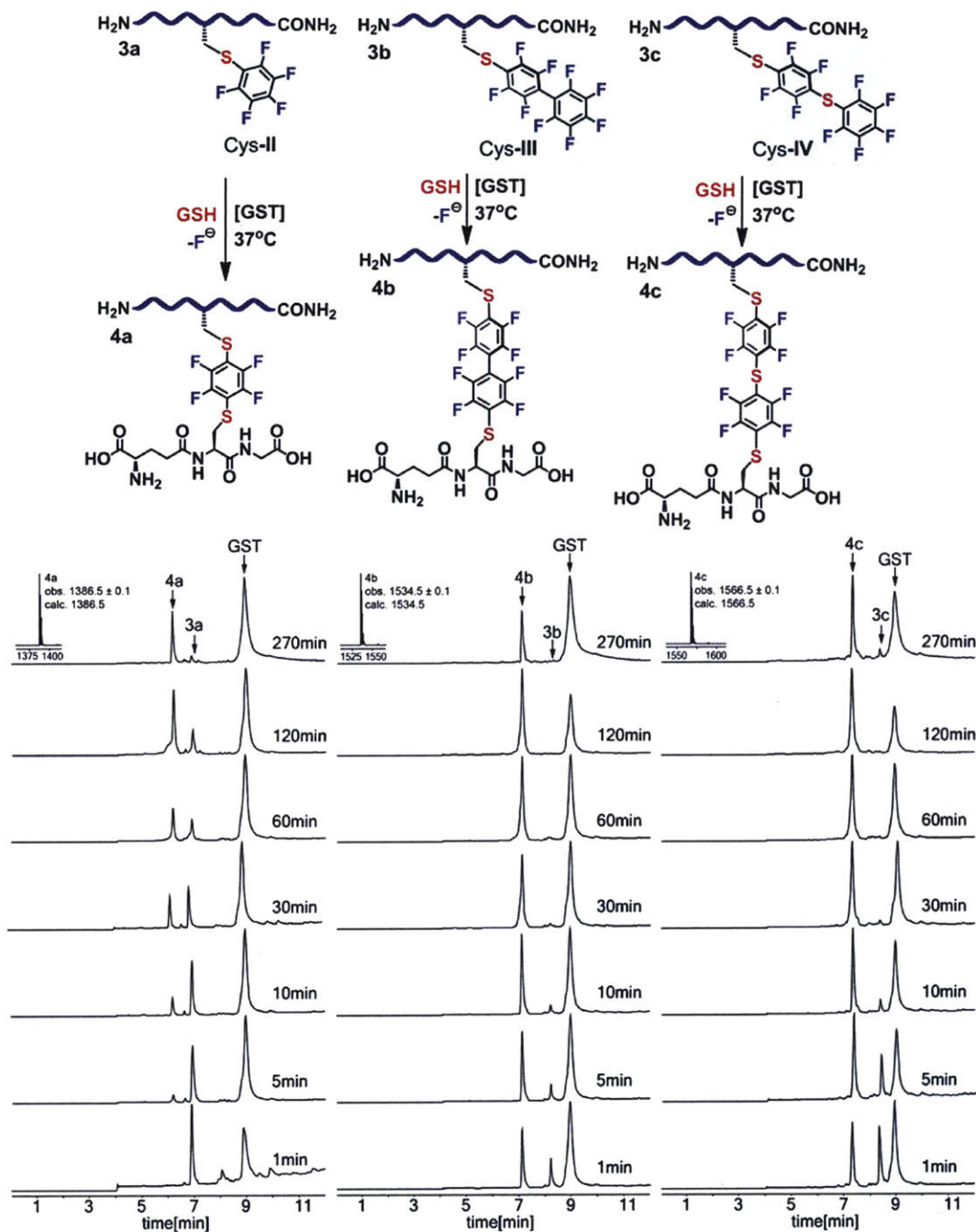
Figure 2.5. LC-MS analysis of GST-catalyzed arylation of GSH with perfluoroarylated cysteine electrophiles.

LC traces of GST-catalyzed conjugation of GSH to peptides 3a-c acquired at multiple timepoints during the reaction. Reaction conditions are described in Figure 2.4. Total ion current (TIC) traces are shown; mass spectra of TIC peaks at highest intensity are shown as an inset.

2.2.3. GST-catalyzed arylation of GSH-tagged peptides

Using more reactive electrophiles, we conducted studies beyond the GSH substrate. We focused on ligations between peptides containing Cys-**III** residue and hexamer peptides with an N-terminal γ-Glu-Cys-Gly. Our study commenced with the synthesis of glycyl-modified GSH-based peptides, where the first amino acid directly after GSH was varied (Figure 2.6, **5a-e**). All reactions proceeded quantitatively within two hours (Figure 2.6, entries 1-5), and reaction with **5a** showed high conversion within 30 minutes (Figure 2.6, entry 1). Decreased reaction rates with **5b** and **5e** as compared to **5a** were observed, suggesting this site may be important for interacting with GST. Changing the second amino acid in the sequence linked to the Gly site to a less bulky residue (Figure 2.7) had no effect on the conjugation rate. Nucleophilic residues (His and Lys, Figure 2.6, entries 3 and 4) are compatible with the reaction. Surprisingly, the reaction with **5c** produced 98% conversion within 30 minutes, suggesting that His residue may favor the reaction.

While the residue C-terminal to GSH can affect the relative rates of the GST-catalyzed reaction, the observed product yield for reactions with Cys-**III** has thus far been independent of the peptide sequence employed, further suggesting that the electrophilicity of the perfluoroaryl substituents can dominate the GST-catalyzed S$_N$Ar reaction with modified GSH (*vide supra*). These results show the promiscuity associated with GST is adequate to catalyze the selective bioconjugation of two unprotected polypeptide fragments and is unique, as it demonstrates the possibility of enhancing a click reaction via enzymatic catalysis.[91]
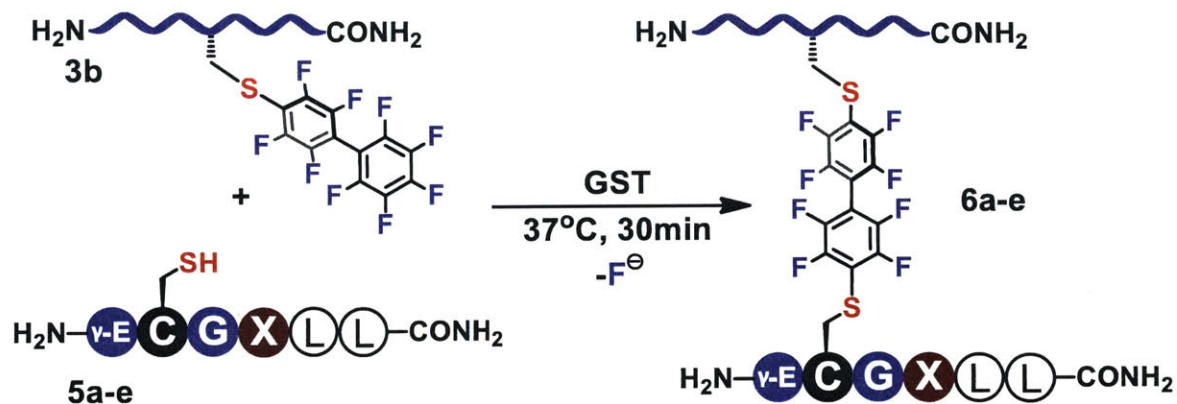
Entry	X		Yield (%)
1	Gly	5a	99
2	Leu	5b	86 (99*)
3	His	5c	98 (99*)
4	Lys	5d	68 (99*)
5	Pro	5e	67 (99*)

Figure 2.6. GST-catalyzed arylation of GSH-tagged peptides with a perfluoroarylated cysteine electrophile.

Conjugation of peptide **3b** containing Cys-**III** residue (H2N-VTLPSTC*GAS-CONH2, C* refers to the modified cysteine) to peptides with N-terminal GSH (γ-Glu-Cys-Gly) sequence featuring varied neighboring amino acid residues under GST-catalysis. Yields were determined by LC-MS analysis at λ=280 nm after 30 minutes (see Experimental section 2.4.6 for the method for yield calculation). Reaction conditions: 2 mM **3b**, 1 mM **5a-e**, 2 mg/mL GST, 20 mM TCEP·HCl, 0.1 M phosphate, pH 8.0, 37 °C. [*] Yields for reactions at 120 minutes.

114
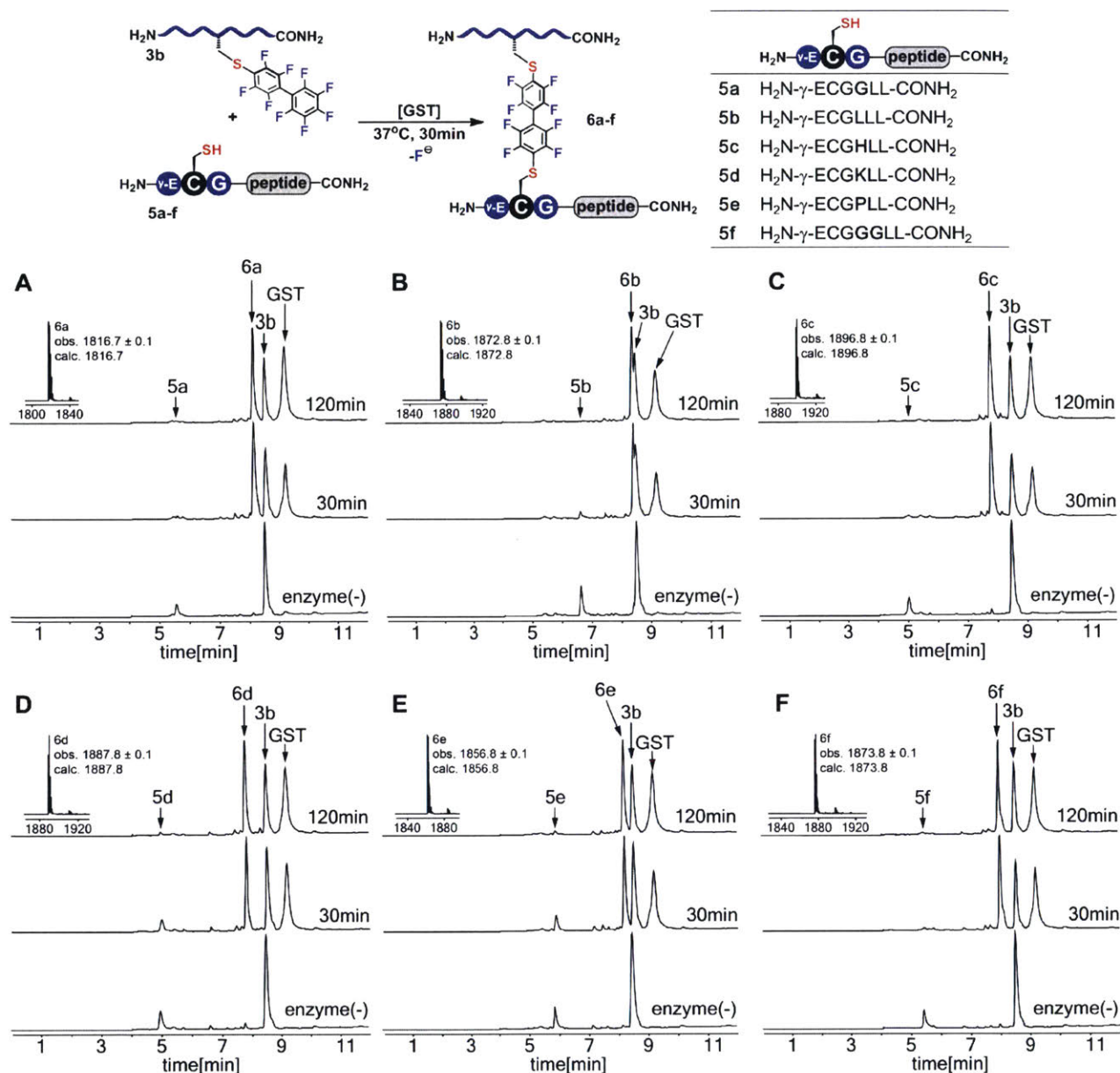
Figure 2.7. LC-MS analysis of GST-catalyzed arylation of GSH-tagged peptides with a perfluoroarylated cysteine electrophile.

Reactions of peptide **3b** with peptides containing *N*-terminal γ-Glu-Cys-Gly sequence under GST catalysis. Reaction conditions are described in Figure 3. Panels (A)-(F) show the LC-MS analysis of each crude reaction with **5a-f** at 30 minutes (center chrotogram) and 120 minutes (top chromatogram). Bottom chromatogram in each panel represents a control experiment conducted without enzyme at 120 minutes. Total ion current (TIC) traces are shown; mass spectra of TIC peaks at highest intensity are shown as an inset.

2.2.4. GST-catalyzed selective arylation of GSH in the presence of competing nucleophiles

This reaction is highly selective for arylation of Cys thiols, as evidenced by the competition experiment with a large excess of exogenous thiol (4-mercaptophenylacetic acid - MPAA, 100-fold, Figure 2.8). No MPAA-arylated product was generated, while the GST-catalyzed product was exclusively produced. In addition, reacting peptide **3b** with another peptide **5j** containing C-terminal hydrazide moiety and N-terminal γ-Glu-Cys-Gly sequence gave solely S-arylated product **6j**, which could be further modified by hydrazone ligation (Figure 2.9).[92]
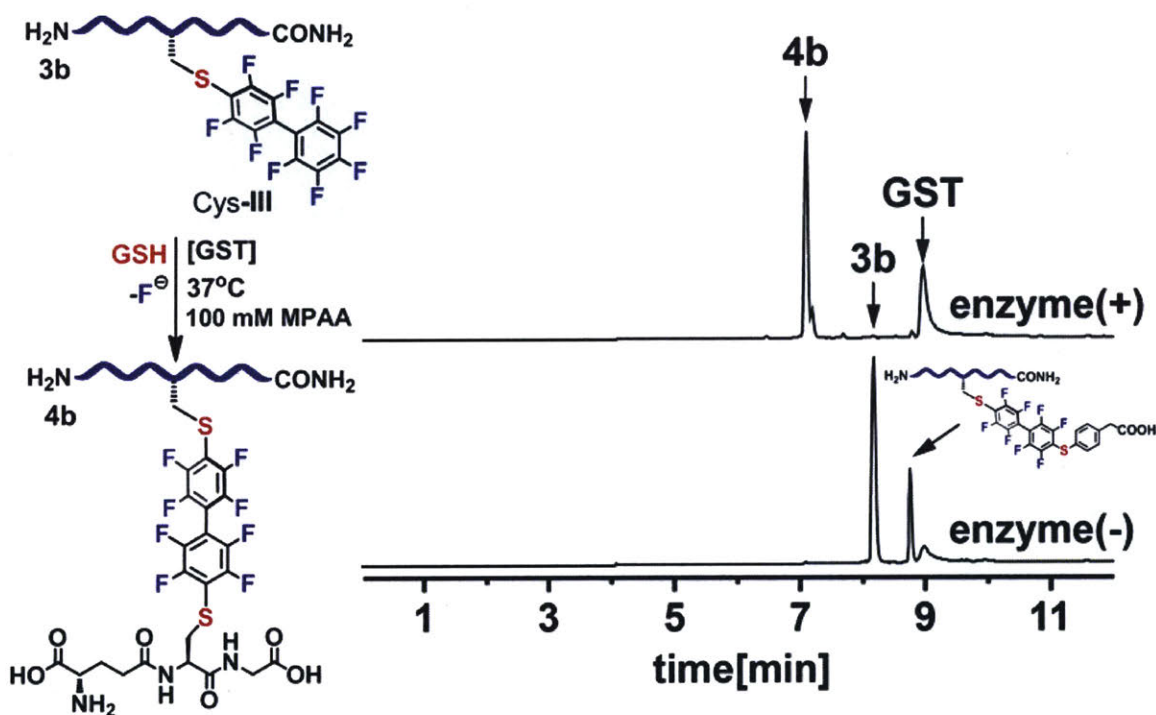
Figure 2.8. GST-catalyzed arylation of GSH with a perfluoroarylated cysteine electrophile in the presence of a competing thiol species.

GST-catalyzed conjugation of GSH to peptide **3b** in the presence of competing 4-mercaptophenylacetic acid (MPAA). Reaction conditions: 1 mM **3b**, 1 mM GSH, 100 mM MPAA, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C, 30min. Total ion current (TIC) traces are shown. While in non-enzymatic reaction MPAA could be arylated, in GST-catalyzed reaction, the GSH-arylated product was exclusively generated.
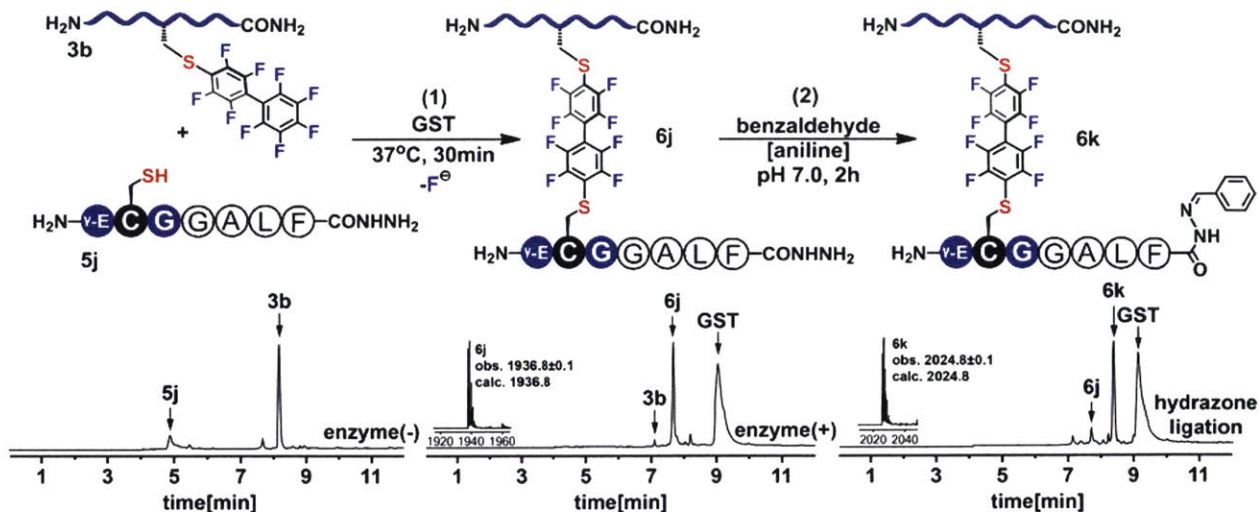
Figure 2.9. GST-catalyzed arylation of GSH-tagged peptide with a perfluoroarylated cysteine electrophile in the presence of a competing hydrazide species.

GST-catalyzed conjugation between peptide hydrazide **5j** and peptide **3b**. The peptide hydrazide was further functionalized by aniline-catalyzed hydrazone ligation with benzaldehyde to give peptide **6k**. Reaction conditions: (1) 1 mM **3b**, 1 mM **5j**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C, 30 min. (2) 2 mM benzaldehyde dissolved in 0.1 M phosphate (pH 7.0) and 200 mM aniline dissolved in 0.1 M phosphate (pH 7.0) were added in situ to the crude reaction mixture resulting from (1). The final reaction conditions were: 100 μM **6j**, 400 μM benzaldehyde, 100 mM aniline, 0.1 M phosphate, pH 7.0, 2 hours. Total ion current (TIC) traces are shown. Left chromatogram represents a control reaction (1) conducted without enzyme.

2.2.5. GST-catalyzed regioselective arylation for protecting group-free dual cysteine modification

The unique chemo- and regioselectivity of the GST-catalyzed arylation reaction could be exploited to label one Cys residue in the presence of another on a fully unprotected peptide or protein. We first labeled the *N*-terminal GSH Cys with biotin and subsequently modified a separate Cys with a fluorophore maleimide. This provides the first example of orthogonal site-specific labeling of two unprotected Cys residues within the same unprotected peptide or protein. Biotin probe containing Cys-**III** moiety (Figure 2.10, **11**, see Experimental section 2.4.8 for synthesis) was conjugated to the Cys position of *N*-terminal γ-Glu-Cys-Gly fragment in protein **7** (see Experimental section 2.4.8 for synthesis). The enzymatic reaction produced the mono-labeled product **8**, and the other cysteine was labeled with fluorescein-5-maleimide **10**, thereby producing the site-specific dual-labeled protein species **9** (Figure 2.10). Using similar protocol, the biotin and fluorescein probes were regiospecifically attached to peptide **14** to produce the dual labeled species **16** (Figure 2.11). The authentic sample was prepared using conventional protecting-group chemistry to confirm the identity of mono-labeled peptide species **15** (Figure 2.12). Importantly, regiospecific and orthogonal modification of two chemically identical Cys sites was enabled by GST selective recognition of the *N*-terminal GSH moiety. This result indicates that the GST-catalyzed arylation could greatly expand the scope of previous cysteine modification methods, which necessitate the use of protecting groups or multiple steps to differentially functionalize two or more cysteine residues.[93,94]
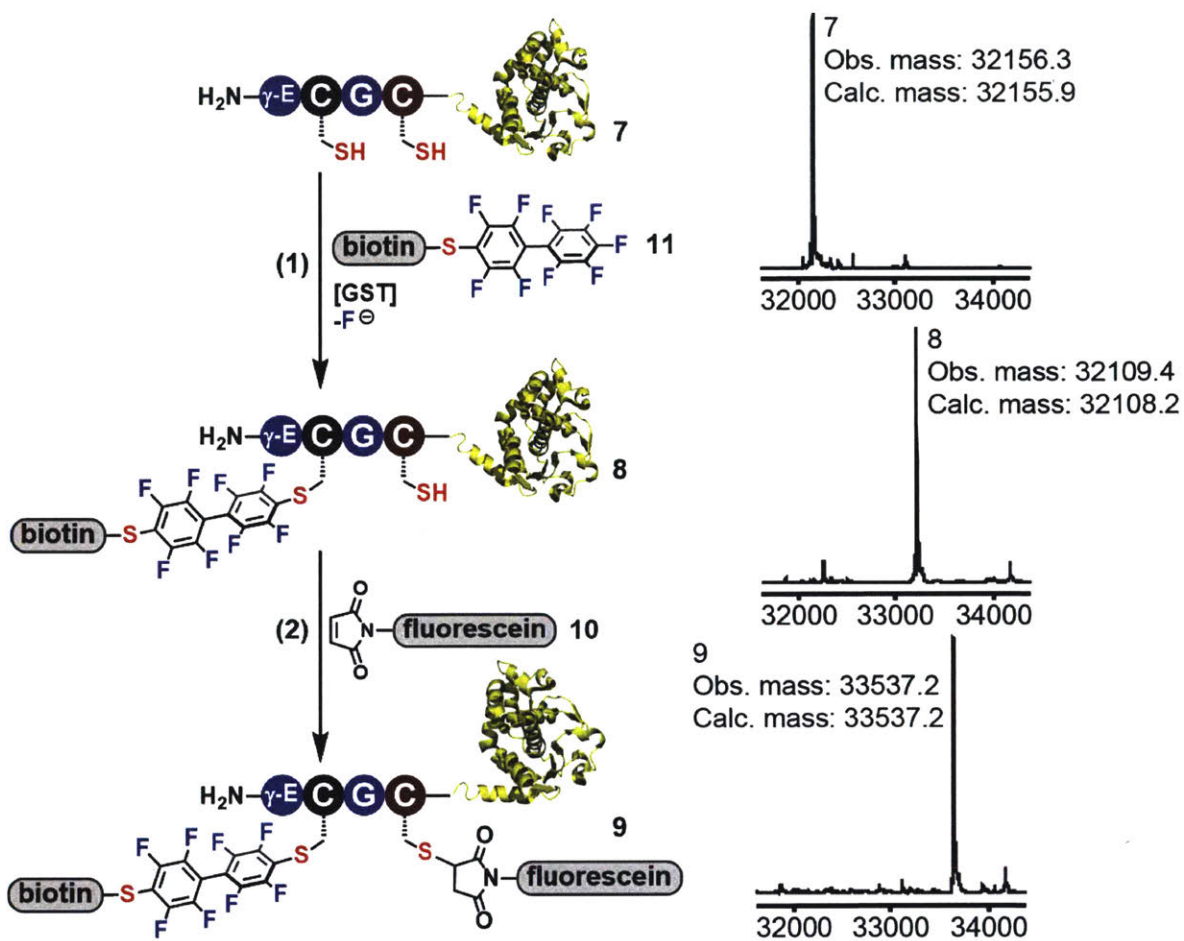
Figure 2.10. GST-catalyzed regioselective cysteine modification enables protecting group-free dual modifiaction of a protein.

Sequential labeling of protein **7** containing two Cys residues with biotin and fluorescein probes. Crystal structure shown is the *N*-terminal domain of anthrax toxin lethal factor 1-263, PDB ID: 1J7N. Reaction conditions: (1) 0.5 mM **11**, 26 µM **7**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C, 2 hours; (2) 13 µM **8**, 100 µM **10**, 0.1 M phosphate, pH 6.0, room temperature, 10 minutes.
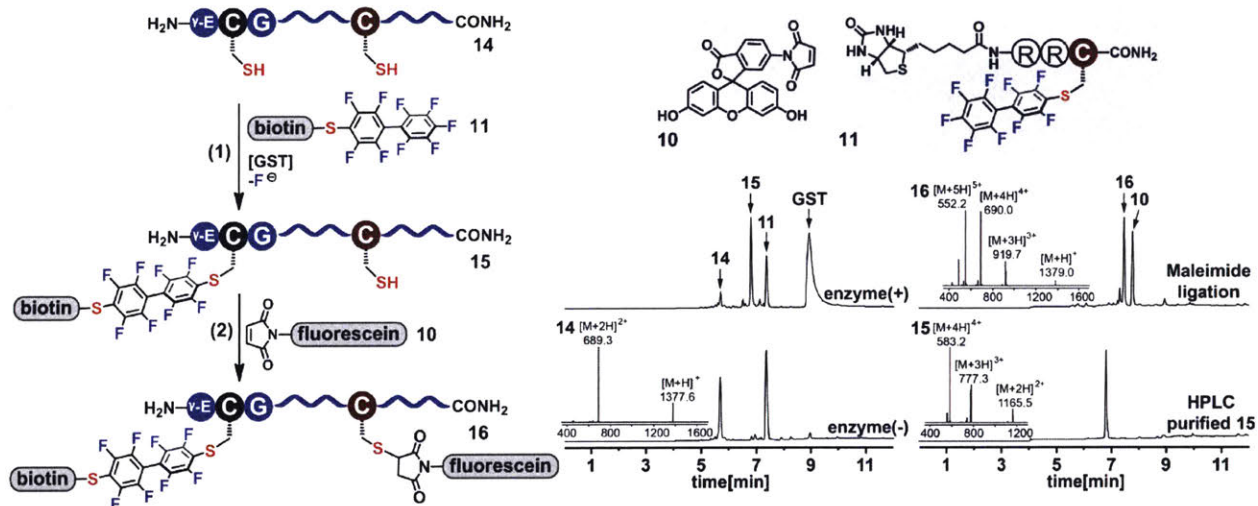
Figure 2.11. GST-catalyzed regioselective cysteine modification enables protecting group-free dual modifiaction of a model peptide.

Sequential labeling of biotin and fluorescein probes to peptide **14** containing two cysteines. Peptide sequence: H2N-γ-ECGPTAAKESCLL-CONH2. Reaction conditions: (1) 0.5 mM **14**, 1 mM **11**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C, 40 minutes; (2) 0.5 mM **15**, 1 mM **10**, 0.1M phosphate, pH 6.0, 10 minutes. Total ion current (TIC) traces are shown; mass spectra of LC peaks at highest intensity are shown as an inset.

120
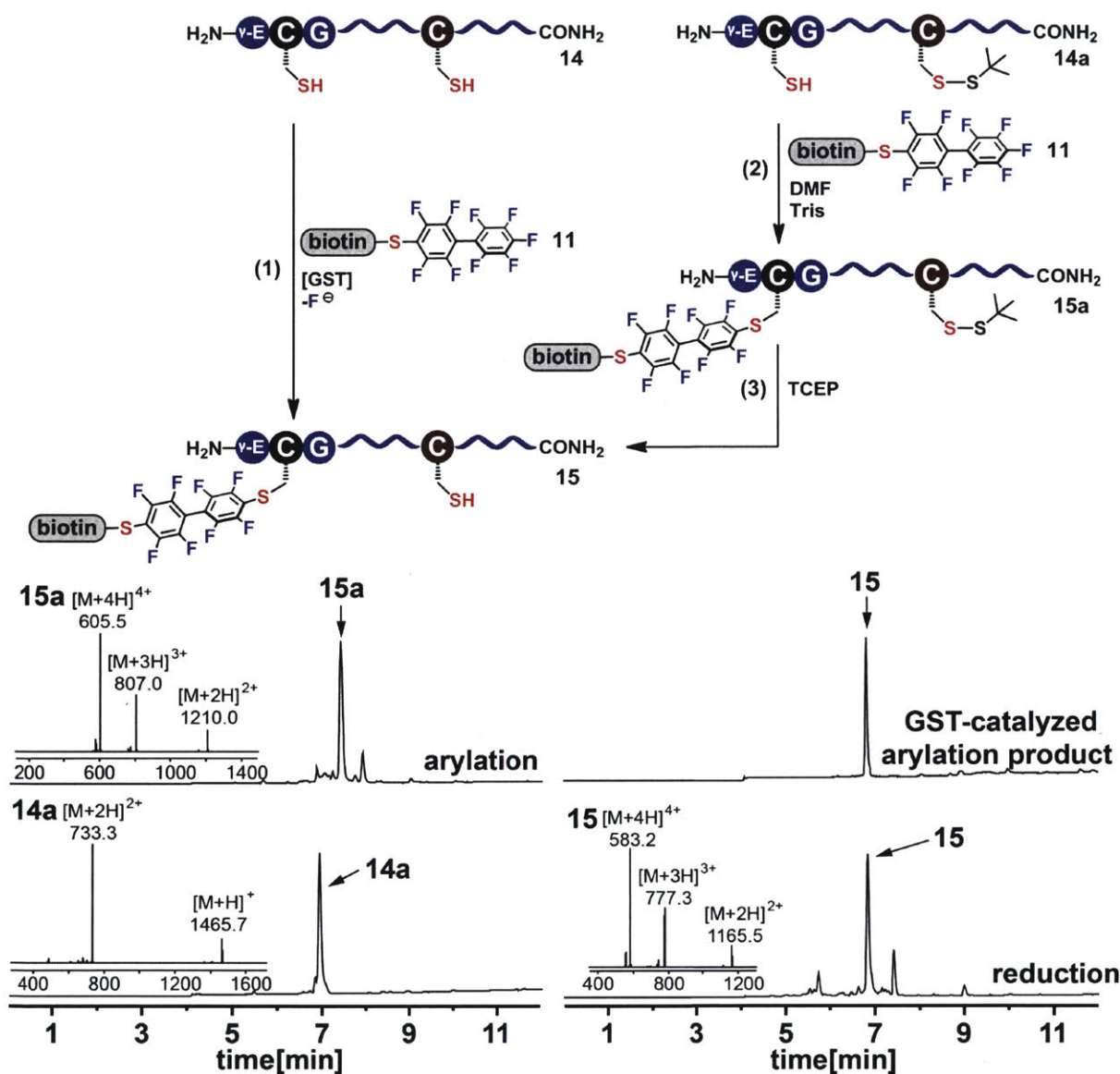
Figure 2.12. Confirming the regioselectivity of GST-catalyzed arylation through preparation of a authentic product using a protecting group-based approach.

GST-catalyzed arylation product and authentic product had same retention time at 6.85 minutes. Reaction conditions: (1) same as those described in Figure 2.11; (2) 1 mM **14a**, 1 mM **11**, 10 mM Tris base, DMF, room temperature, 30 minutes; (3) Equal volume of 1 M phosphate, 200 mM TCEP, pH 8.0 was added to the crude reaction mixture from (2), and the resulting reaction mixture was left at room temperature for 10 minutes before subjected to LCMS analysis. Total ion current (TIC) traces are shown; mass spectra of TIC peaks at highest intensity are shown as an inset.

121

2.2.6. GST-catalyzed peptide macrocyclization with different linkers

We next investigated using the developed GST-catalyzed reaction for macrocyclization of unprotected peptides. Cyclic peptides constitute a very important class of medicinally relevant macrocycles.[95] Although various methods have been previously developed,[56] synthesis of macrocyclic peptide fragments remains challenging. We reasoned that a peptide containing an *N*-terminal GSH tag and a *C*-terminal perfluoroarylated cysteine should be readily cyclized under GST catalysis (Figure 2.13A). The perfluorinated moiety used in this example can potentially enhance the cell-permeability of cyclic peptides as shown previously.[86]

The linear substrates employed in our studies were readily prepared in two steps (Figure 2.13B). Peptides with an *N*-terminal disulfide protected GSH tag and a *C*-terminal free cysteine were synthesized by rapid flow-based Fmoc-SPPS[96,97] followed by *C*-terminal cysteine *S*-arylation with perfluoroaromatic reagents (see Experimental section 2.4.8 for synthesis). The linear substrates were purified by RP-HPLC and then subjected to macrocyclization. *Tris*(carboxylethyl)phosphine hydrochloride (TCEP•HCl) was added to the reaction buffer *in situ* to deprotect the disulfide bond and generate a free cysteine in the GSH tag for GST-catalyzed macrocyclization.
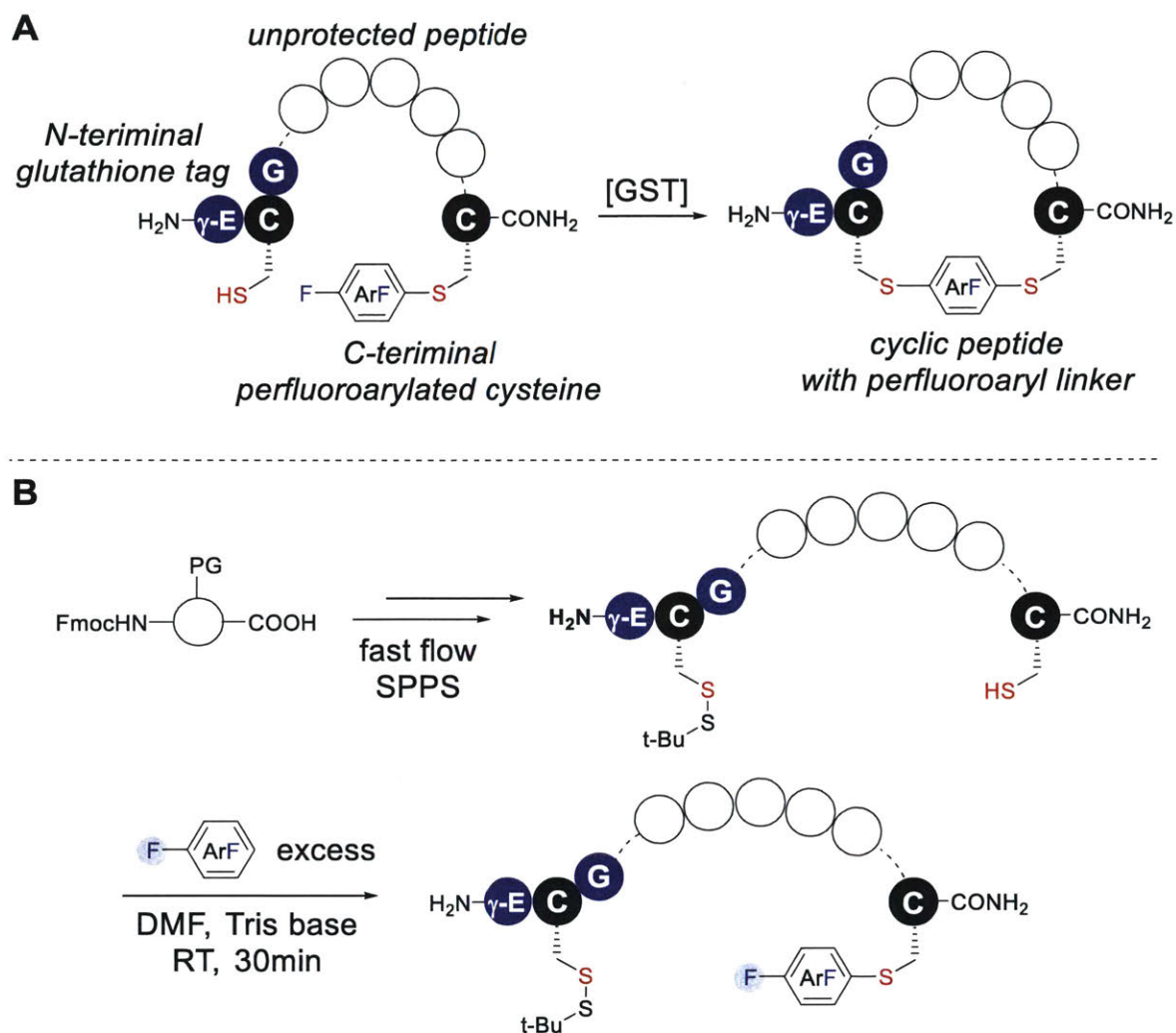
Figure 2.13. GST-catalyzed macrocyclization of unprotected peptides.
(**A**) The macrocyclic peptide is generated by GST-catalyzed S_NAr reaction between the cysteine thiol of an *N*-terminal glutathione (γ-Glu-Cys-Gly) tag and a perfluorayl moiety linked to the *C*-terminal cysteine. Circles represent amino acids. (**B**) Synthesis of perfluoroaryl-linked peptides for GST-catalyzed macrocyclization. Circles represent amino acids; PG: protecting group; SPPS: solid-phase peptide synthesis.

We commenced our study by searching for the best perfluoroaromatic linker for the GST-catalyzed cyclization reaction (Figure 2.14). We prepared model peptides **21a-21c** with *C*-terminal cysteines *S*-arylated with decafluorobiphenyl, hexafluorobenzene, and pentafluorophenyl sulfide, respectively (see Experimental section 2.4.8 for synthesis). Upon incubating 0.1 mM solution of peptide **21a** (Figure 2.14B, top chromatogram) in the presence of 10 mol% of GST (0.2 mg/mL) in pH 8.0 phosphate buffer with 20 mM TCEP, cyclic product **Cyc-21a** was produced in quantitative yield within 5 minutes at room temperature as shown by LC-MS analysis of the crude reaction mixture. We did not observe oligomeric side products (Figure 2.14B, bottom chromatogram). Importantly, the control reaction without added enzyme contained mostly unreacted linear peptide **21a'** with only trace amount of cyclized product (Figure 2.14B, middle chromatogram). Compared with peptide **21a**, peptides **21b** and **21c** showed significantly lower enzyme mediated cyclization yields under the same reaction conditions (Figure 2.14A). At extended reaction time, only 45% and 57% of cyclized products were observed for peptide **21b** and **21c** respectively (Figure 2.15 to Figure 2.18).

The highest cyclization rate under GST-catalysis was achieved with peptide **21a** as determined by LC-MS time course studies (Figure 2.15 and Figure 2.16). The cyclization rate of peptide **21a** under enzyme-catalysis was ~900-fold faster than without enzyme (Figure 2.27). This high cyclization efficiency under enzyme catalysis permitted macrocyclization at increased concentrations of linear peptide: unlike other enzyme-catalyzed macrocyclization methods where increasing the concentrations of linear substrates can lead to the formation of oligomeric side products *via* competing intermolecular reaction pathways.[32] We found that increasing the concentration of peptide **21a** to 10 mM with only 0.1 mol% of GST still produced the cyclized product quantitatively within 3 hours (Figure 2.19); no oligomeric byproducts were observed. The unique preference and high efficiency of peptide **21a** for GST-catalyzed cyclization indicates the perfluorobiphenyl moiety is highly suited for this chemistry.
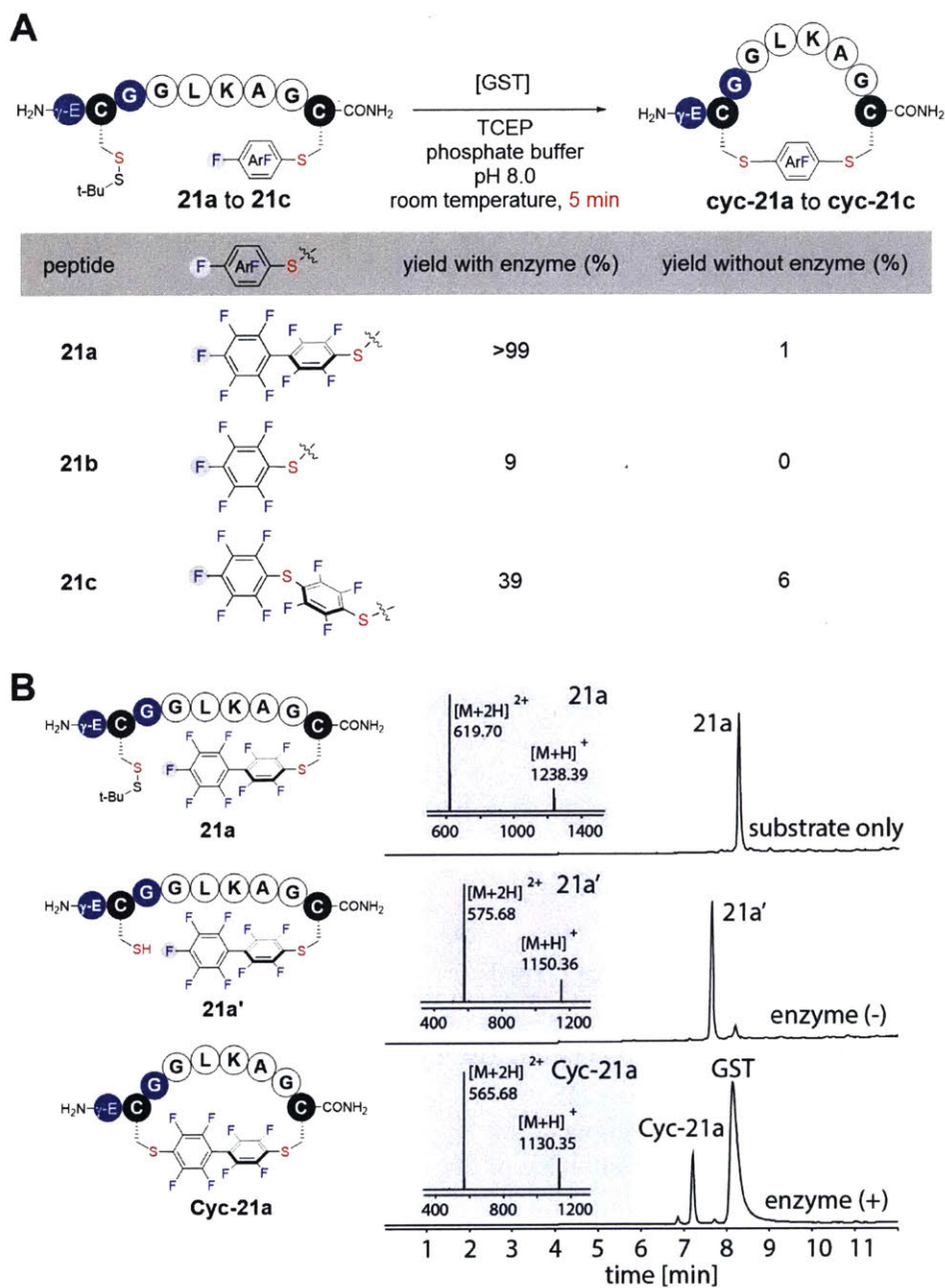
Figure 2.14. Macrocyclization of peptides with different perfluoroaromatic linkers.
Reaction conditions: 0.1 mM peptide **21a-21c**, 0.2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, room temperature (25 °C), 5 minutes. (**A**) Reaction yields with different linkers with or without the addition of GST. (**B**) LC-MS analysis of crude reaction mixtures with peptide **21a**. Amino acids are shown in one-letter code. γ-E: γ-glutamic acid. LC-MS data shown are total ion currents (TIC); mass spectrums at the highest points of the TIC peaks were shown as insets.
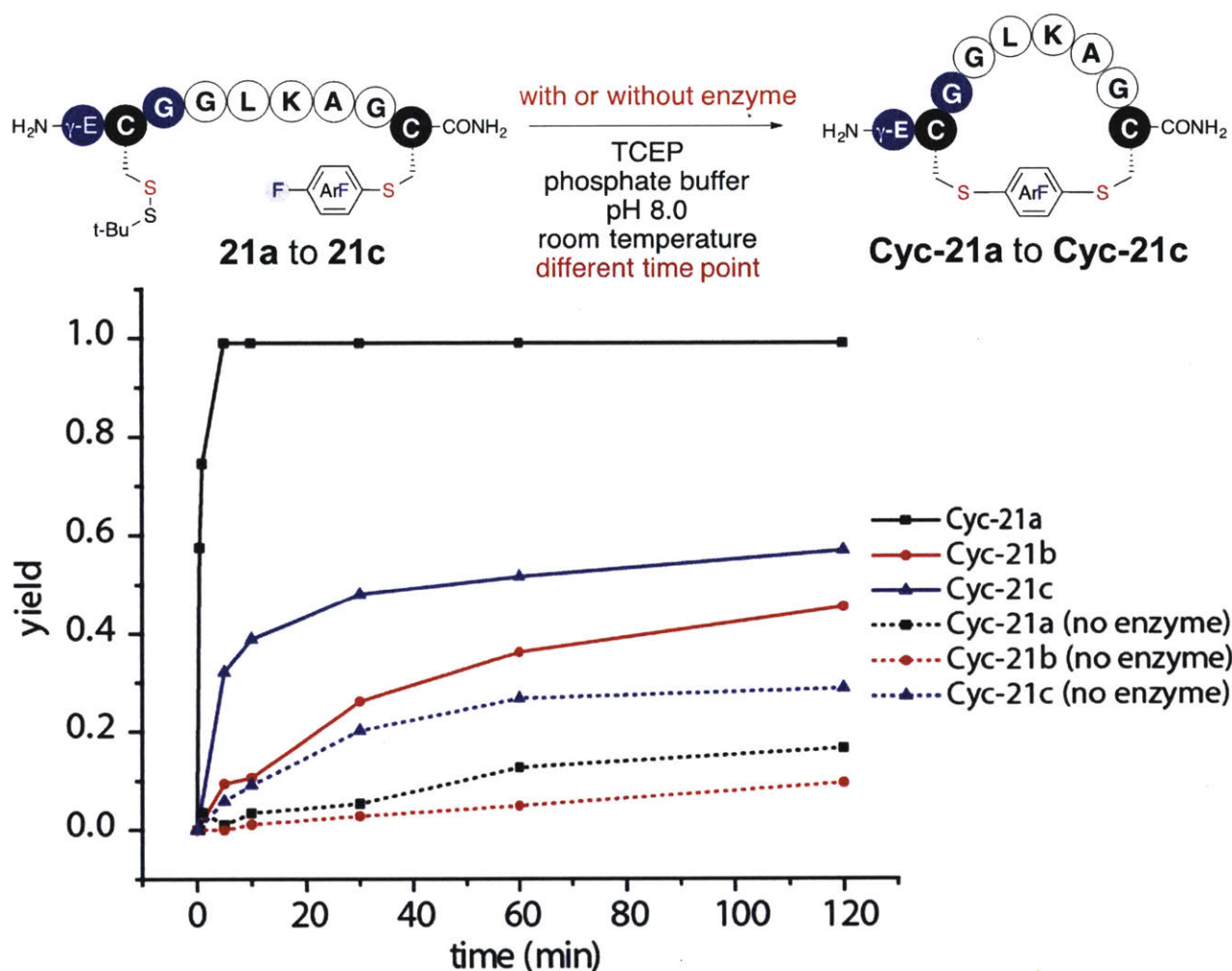
Figure 2.15. Kinetic profiles of macrocyclization of peptides with different perfluoroaromatic linkers.

Yields were calculated from LC-MS analysis of the crude reaction mixture at different time points. All reactions with enzyme showed higher reaction rate than reactions without enzyme. Peptide **21a** showed highest reaction rate under GST catalysis. Reaction conditions are same as those describe in Figure 2.14. LC-MS chromatograms are summarized in Figure 2.16 to Figure 2.18.
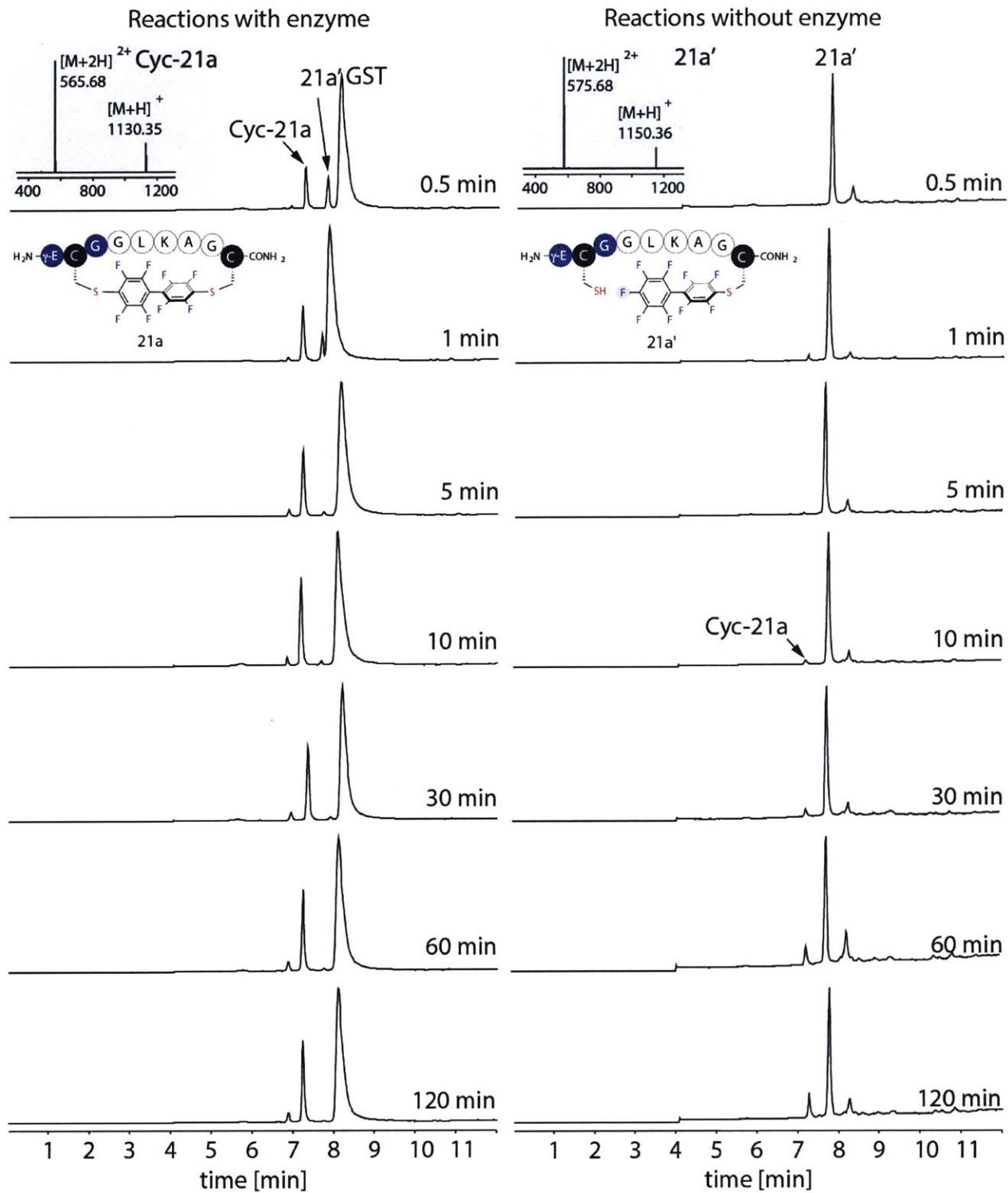
Figure 2.16. Macrocyclization of peptide 21a with and without GST analyzed by LC-MS at different time points.

Reaction conditions are described in Figure 2.14.
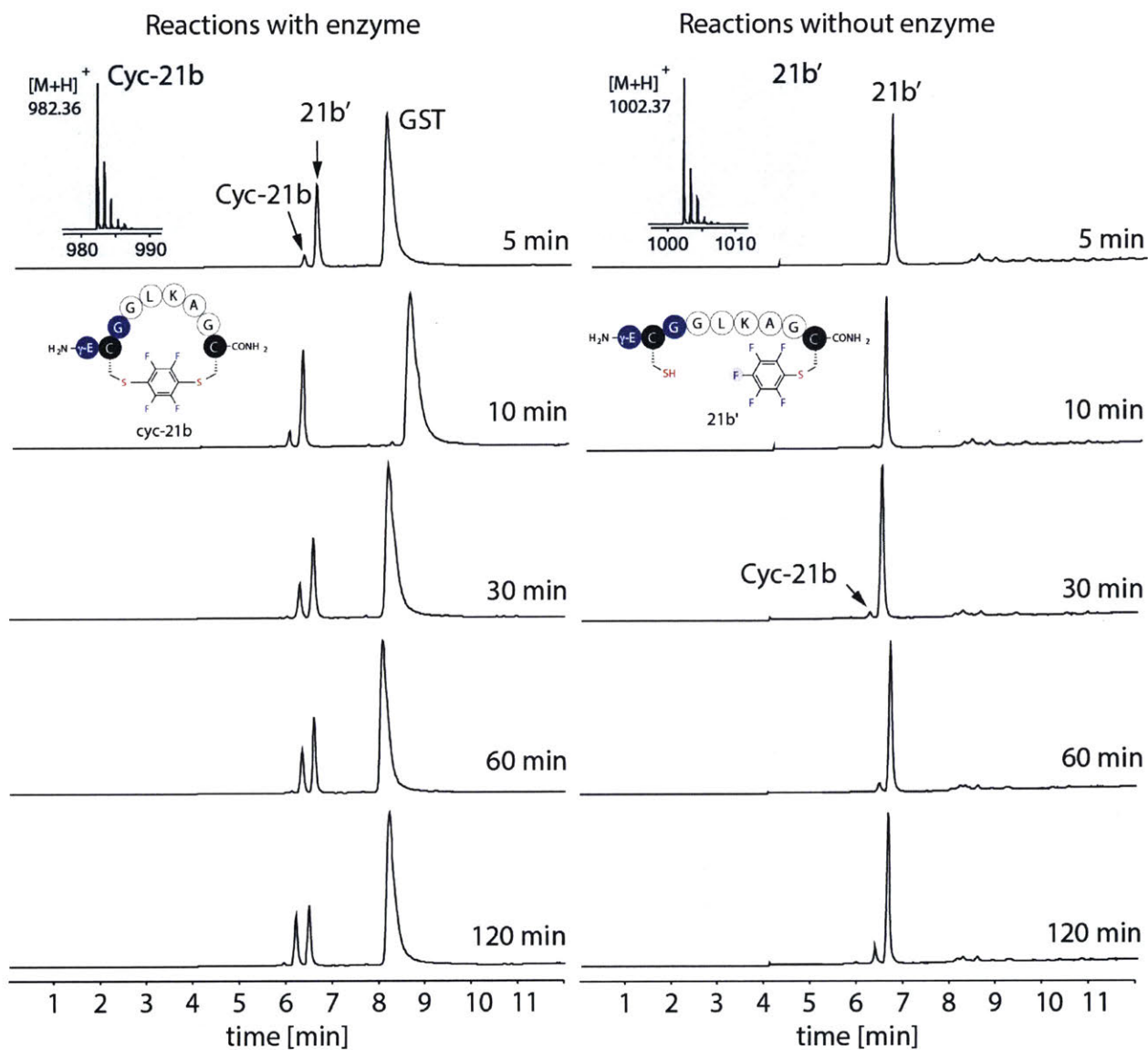
Figure 2.17. Macrocyclization of peptide 21b with and without GST analyzed by LC-MS at different time points.

Reaction conditions are described in Figure 2.14.
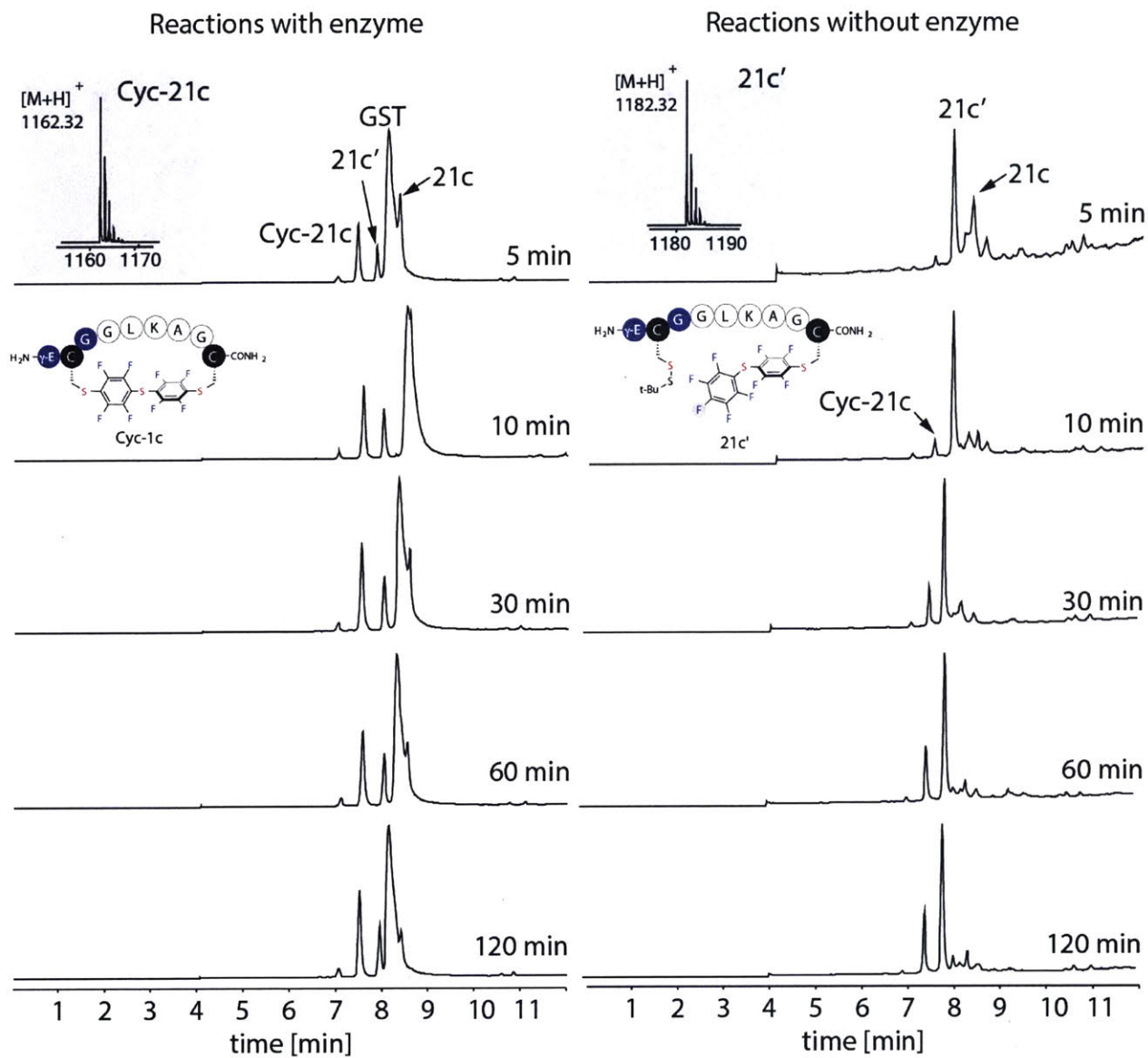
Figure 2.18. Macrocyclization of peptide 21c with and without GST analyzed by LC-MS at different time points.

Reaction conditions are described in Figure 2.14.
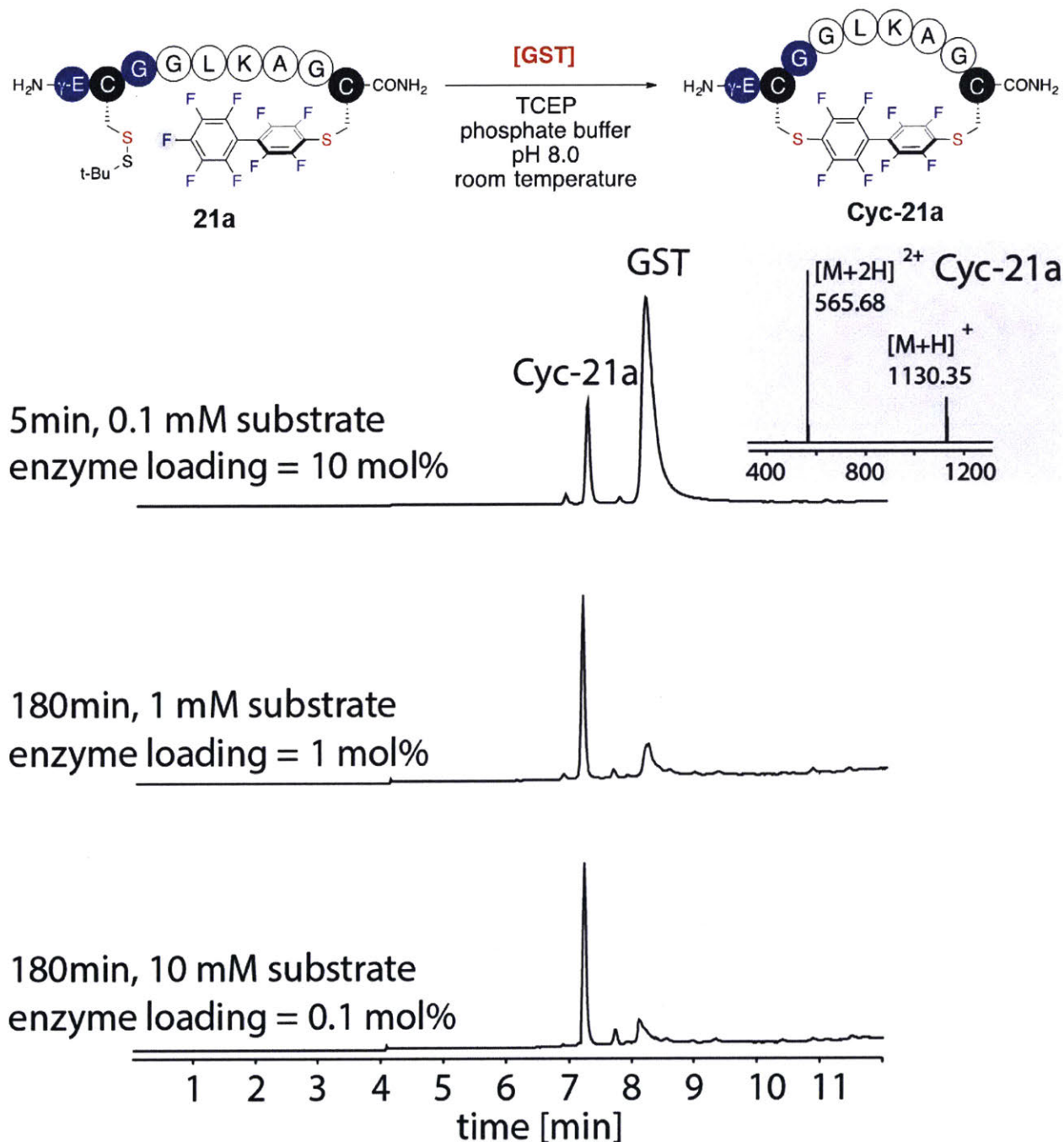
Figure 2.19. Macrocyclization of peptide 21a at increased concentrations.

Concentration of GST was kept constant at 0.2 mg/mL; LC-MS analysis of reaction where concentrations of substrate **21a** is 0.1 mM (top chromatogram), 1 mM (middle chromatogram), and 10 mM (bottom chromatogram). Reactions were performed in 0.1 mM phosphate buffer with 20 mM TCEP•HCl, pH 8.0 at room temperature.

2.2.7. GST-catalyzed peptide macrocyclization with different peptide lengths

To probe macrocyclization of peptides with increasing lengths, we prepared peptides **22a–24a** with variable number (2 to 4) of repeating GLKAG pentapeptide sequences positioned between the GSH tag and the perfluoroaryl-linked *C*-terminal cysteine (see Experimental Section 2.4.8 for synthesis). Although decreased cyclization rates were observed as the peptide lengths increased (Figure 2.21), cyclization of these three peptides under similar reaction conditions as those described in Figure 2.14 yielded cyclized products **Cyc-22a** – **Cyc-24a** nearly quantitatively within 2 hours (Figure 2.20). Only 5% of cyclic products were observed for reactions without GST present (Figure 2.20A and Figure 2.22 to Figure 2.24).

To compare the GST-catalyzed macrocyclization in water with the uncatalyzed macrocyclization in organic solvent side-by-side, we prepared and purified peptides **22a' – 24a'** by reducing the disulfide moiety in peptides **22a – 24a** and purified them by RP-HPLC (Figure 2.25). Adding DMF with Tris base to peptides **22a' – 24a'** initiated their intramolecular cyclization reaction and produced cyclized products **Cyc-22a** – **Cyc-24a**. The highest reaction yield was observed for the shortest peptide **22a'** with 39% of cyclized product generated within 2 hours as measured by LC-MS analysis of the crude reaction mixture (Figure 2.25B). Significant amounts of oligo- and polymeric side products were observed for the macrocyclizations of longer peptides **23a'** and **24a'**; desired cyclic products were minor products identified in the LC-MS chromatograms (Figure 2.25D and F).

To further compare the GST-catalyze cyclization with our previously reported cross-linking chemistry,[17b] we synthesized peptides **2b – 4b** with free cysteines at both the GSH tag and the *C*-terminus. Peptides **2b – 4b** could be cyclized by the reaction with decafluorobiphenyl to produce **Cyc-2a – Cyc-4a**. Similar to peptide **2a' – 4a'**, we observed decreased cyclization yields as the peptide lengths increased. 37% of desired product was formed with the shortest peptide **2b** but no detectable cyclic product resulted with longer peptides **3b** and **4b** (Figure 2.26 and Figure 2.27). We previously found reducing the concentration of linear substrate could improve the reaction yield of the cyclization reaction for small model peptides in DMF, however, cyclization of **23b** and **24b** at diluted conditions still resulted in mixtures of by-products (Figure 2.26F and Figure 2.26I).
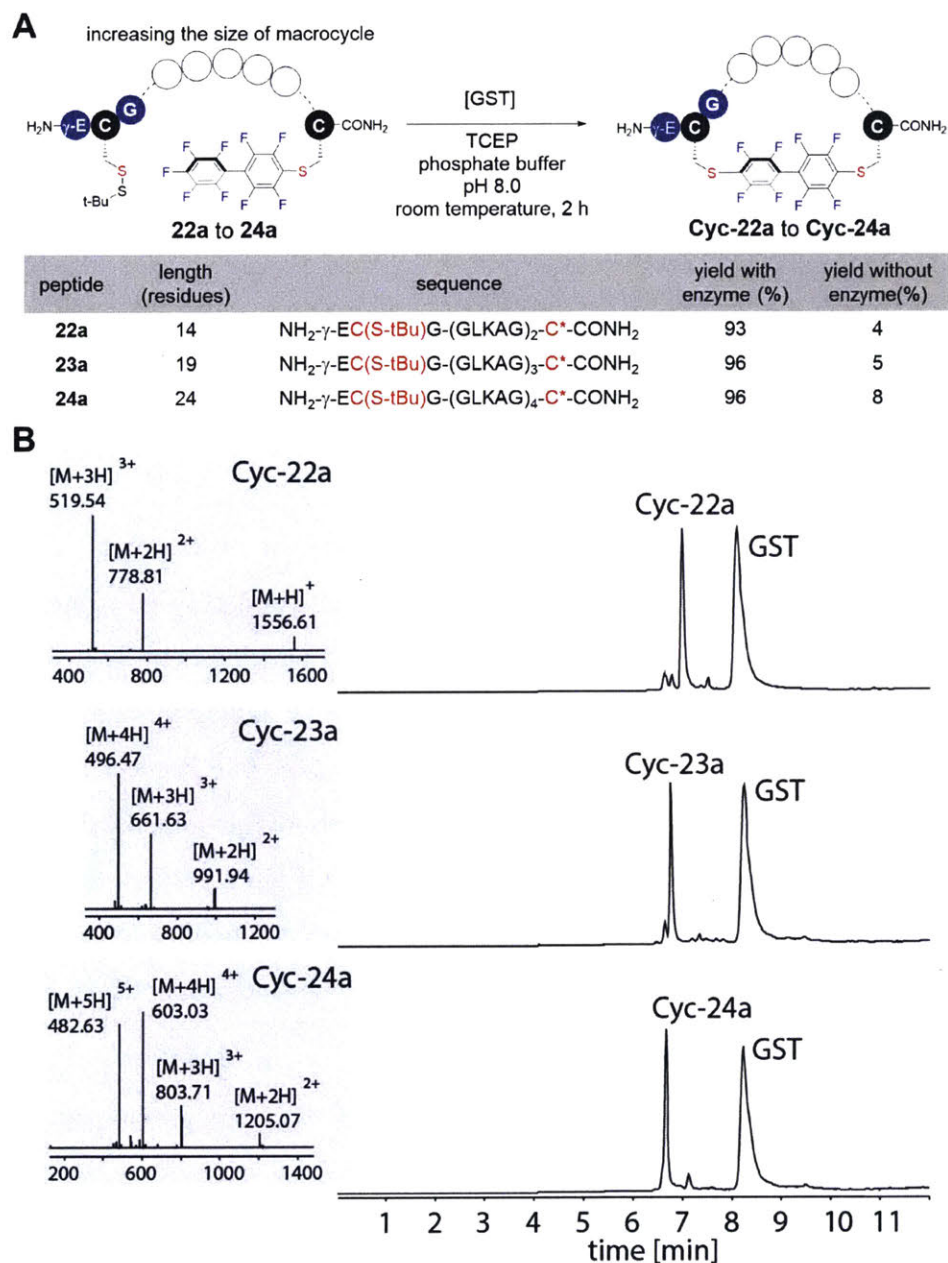
Figure 2.20. Macrocyclization of peptides with increasing lengths.

Reaction conditions: 0.1 mM peptide **22a – 24a**, 0.2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, room temperature (25 °C), 2 hours. (**A**) Yields of cyclization reactions with and without the addition of GST (C(S-*t*Bu): *tert*-butylthio protected cysteine; C*: perfluoroaryl-modified cysteine). The two cysteines that were crosslinked together are highlighted in red. (**B**) LC-MS analysis of reactions of peptide **22a** to **24a** with the addition of GST. LC-MS data shown are total ion currents (TIC); mass spectra at the highest points of the TIC peaks are shown as insets.

Figure 2.21. Kinetic profiles for macrocyclization of peptides with increasing lengths.

Kinetic profiles for macrocyclization of peptides **22a-24a**. Yields are obtained from LC-MS analysis of the crude reaction mixture at different time points. Using Origin software, apparent first-order rate constants of reactions are obtained by fitting the data with first-order kinetics using function: yield $= 1-e^{-Kt}$ where K is the apparent rate constant (min^{-1}) and t is the time (min). Reaction conditions are same as those describe in Figure 2.20. LC-MS chromatograms are summarized in Figure 2.22 to Figure 2.24.
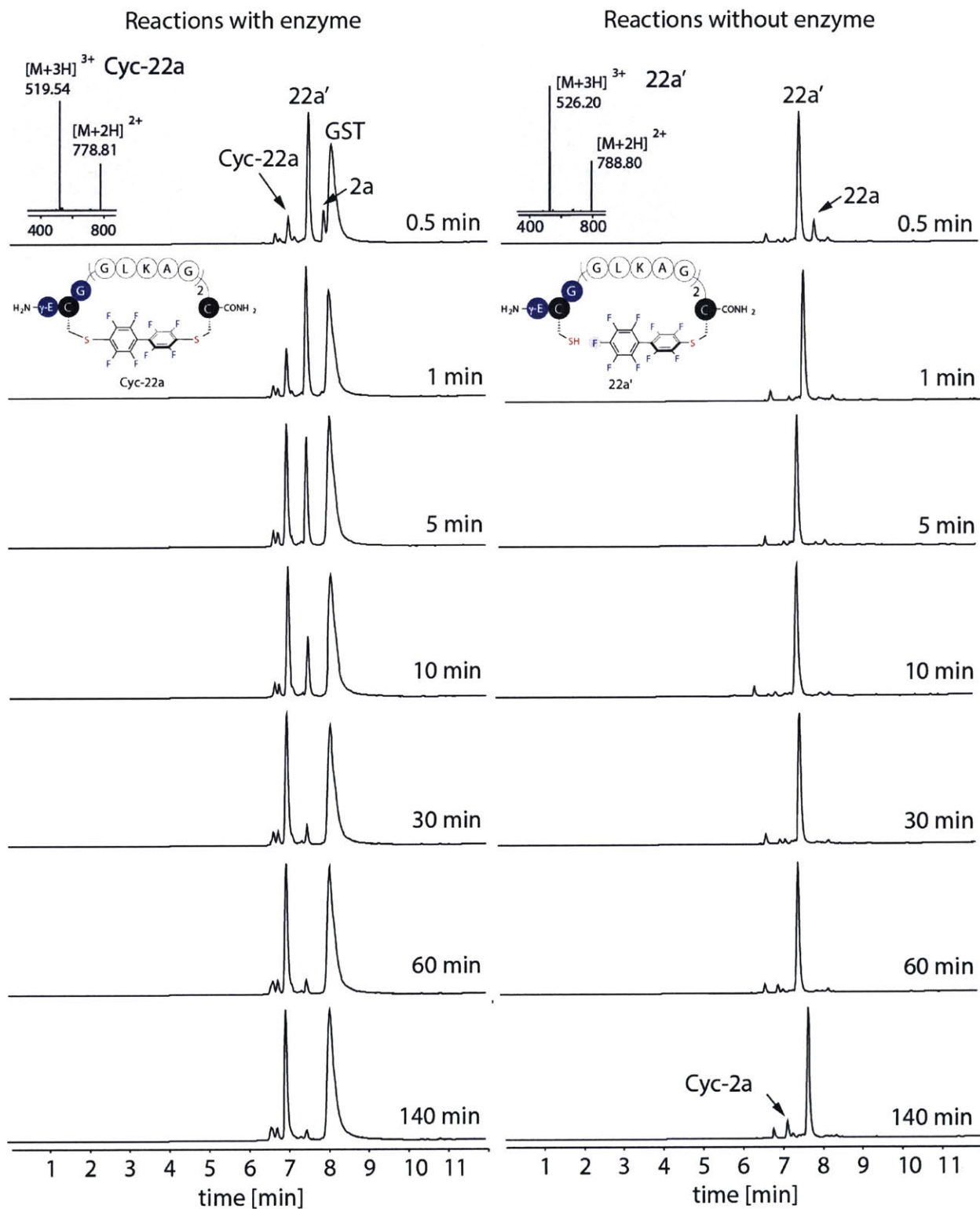
Figure 2.22. LC-MS analysis of macrocyclization of peptide 22a.

Reaction conditions are described in Figure 2.20.
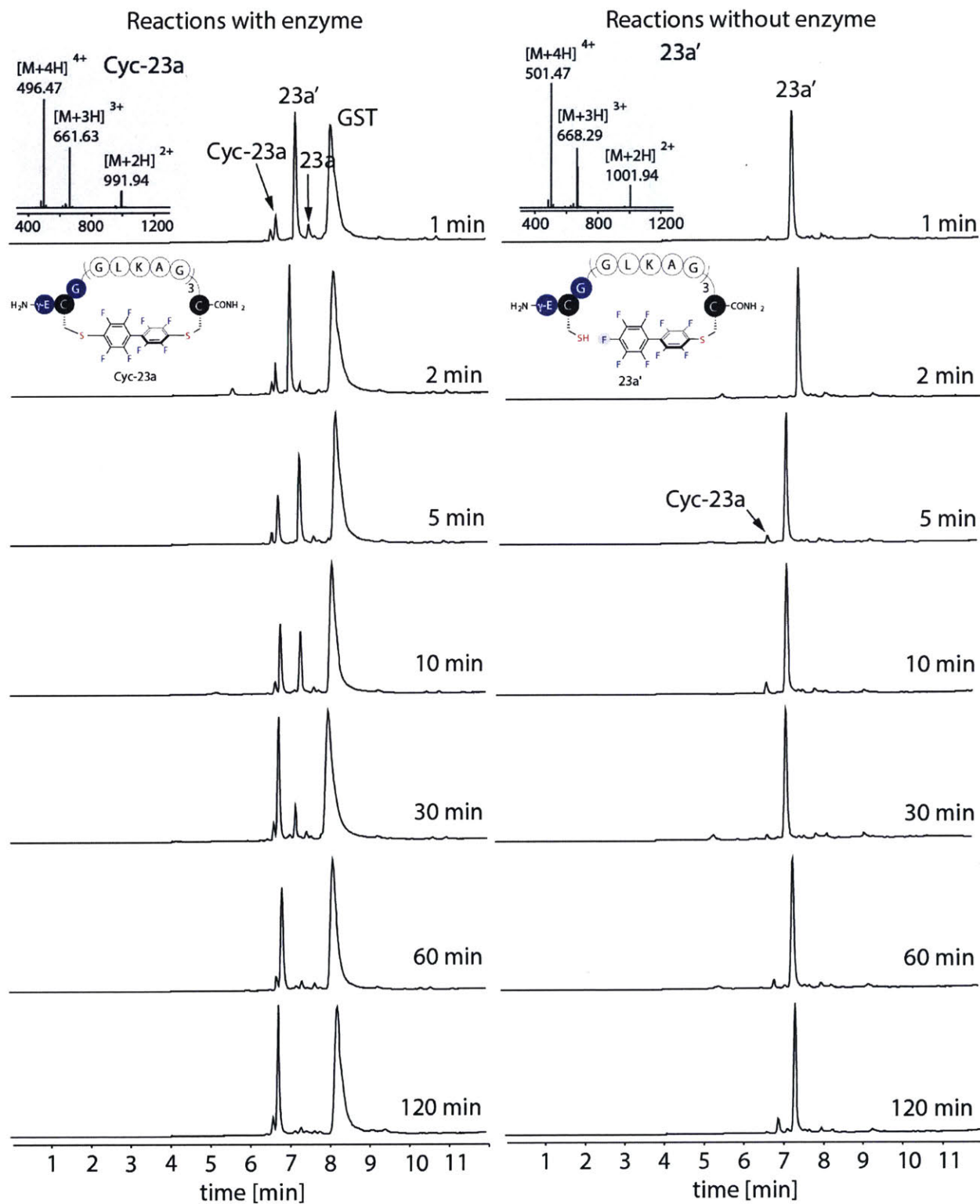
Figure 2.23. LC-MS analysis of macrocyclization of peptide 23a.

Reaction conditions are described in Figure 2.20.
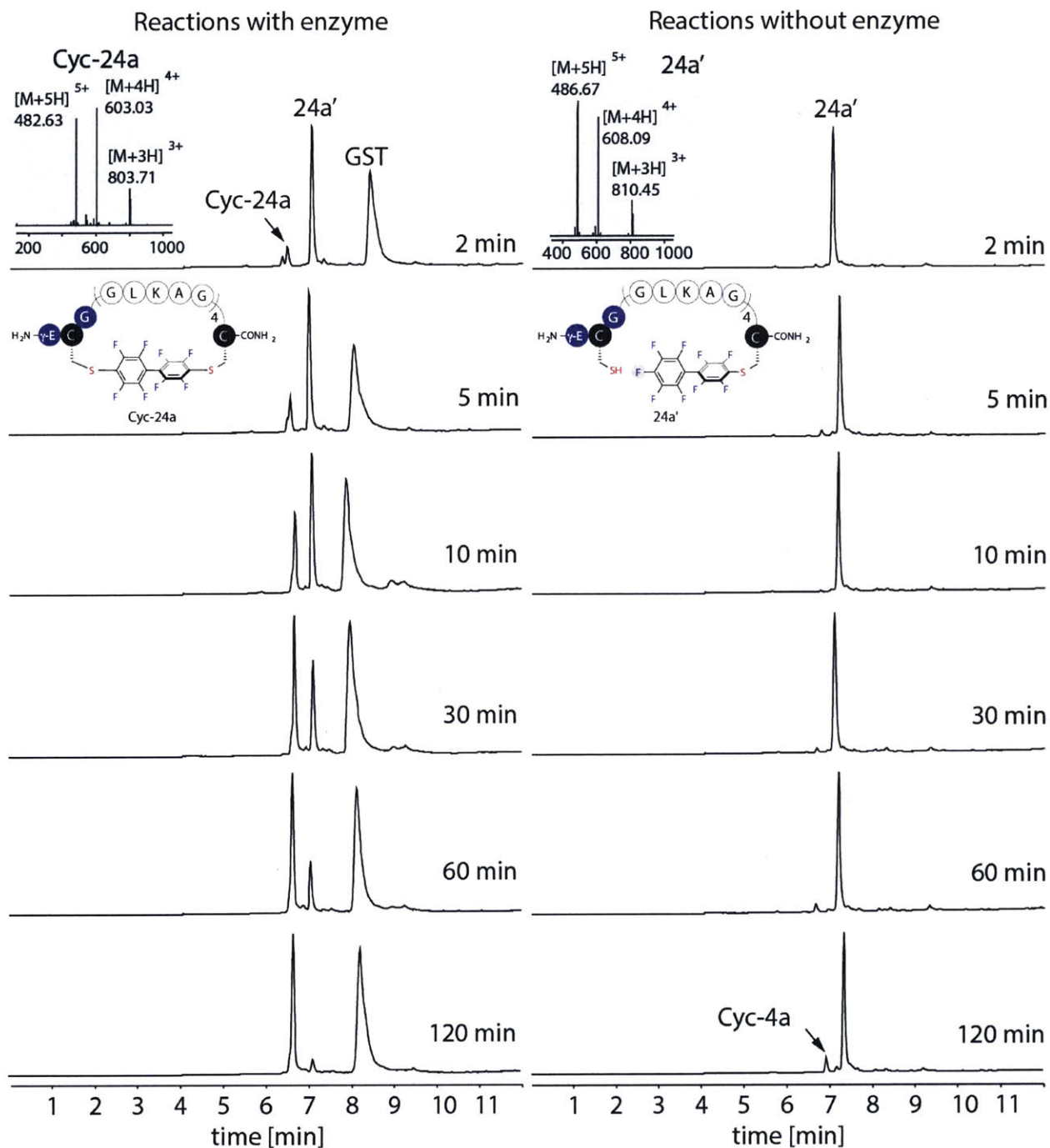
Figure 2.24. LC-MS analysis of macrocyclization of peptide 24a.

Reaction conditions are described in Figure 2.20.
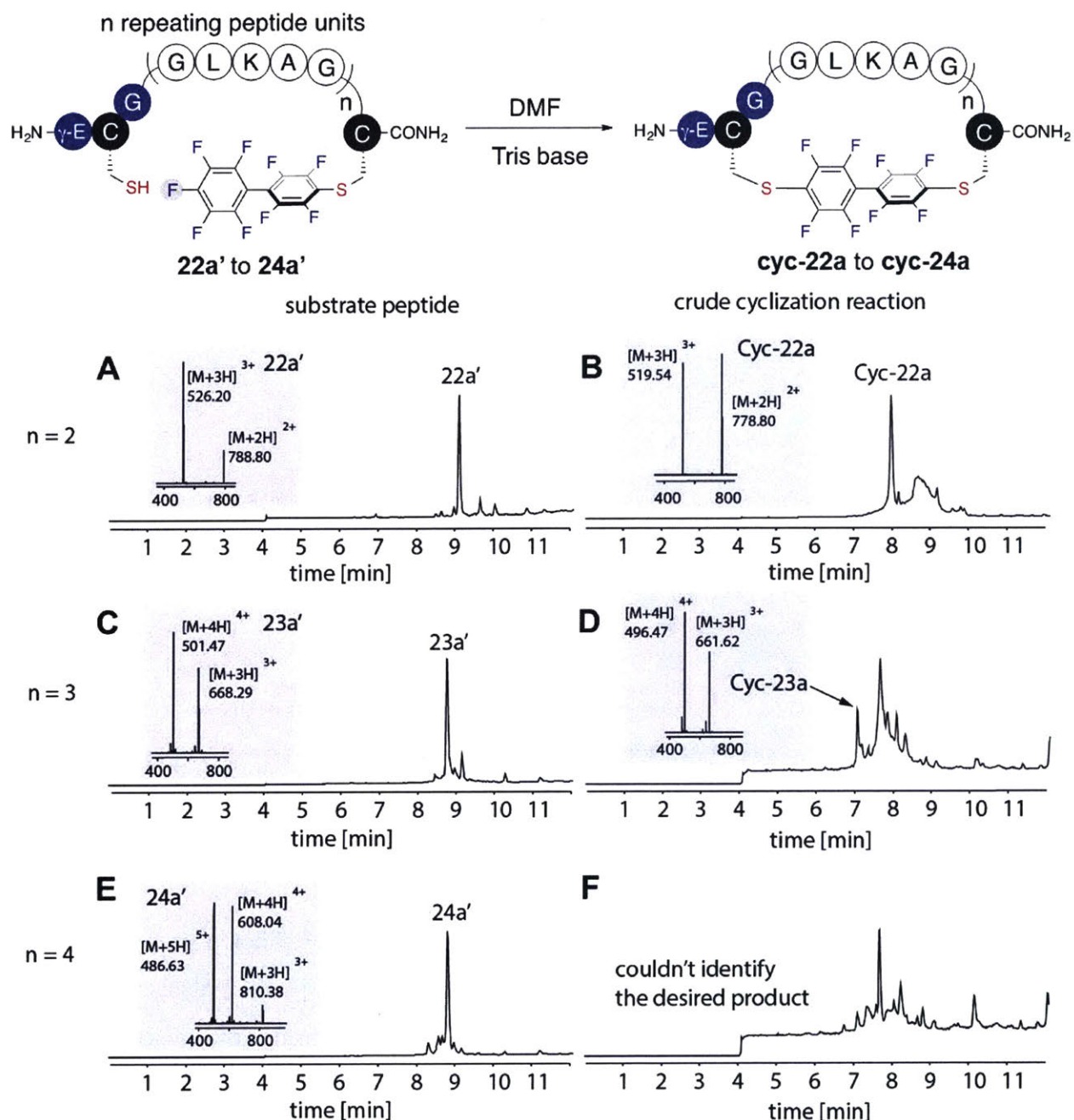
Figure 2.25. Macrocyclization of pre-arylated peptides in organic solvent without enzyme. Macrocyclization of peptide **22a'** to **24a'** in DMF. Shown are two series of chromatograms, each series consist of TIC chromatograms for LC-MS analysis of the pure substrate peptide and crude reaction mixture. Chromatograms **A** and **B** are for peptide **22a**; **C** and **D** are for peptide **23a**; **E** and **F** are for peptide **24a**. Reaction conditions: 0.1 mM peptide **22a** to **24a**, 20 mM *tris*(hydroxymethyl)aminomethane, DMF, 2 hours at room temperature.
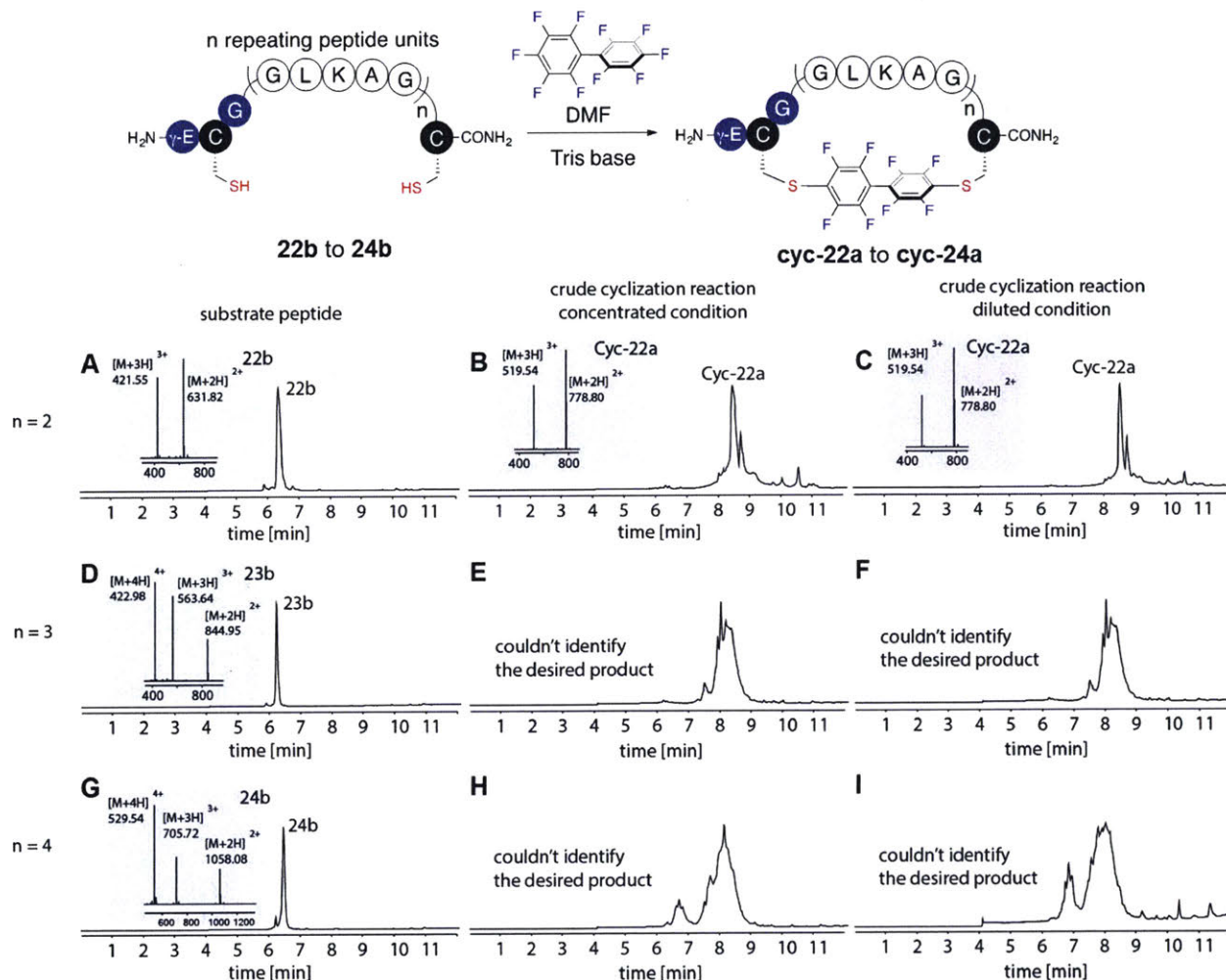
Figure 2.26. Macrocyclization of peptides with two free cysteines using perfluoroaryl linkers in organic solvent.

Macrocyclization of peptide **22b–24b** with decafluorobiphenyl in DMF. Shown are three series of chromatograms, each series consists of TIC chromatograms for LC-MS analysis of the pure substrate peptide, crude reaction mixture under concentrated condition, and crude reaction mixture under diluted condition. Chromatograms **A**, **B**, and **C** correspond to reaction with **22b**; D, E, and F - **23b**; G, H, and I - **24b**. Concentrated conditions: 2 mM substrate peptide, 2.5 mM decafluorobiphenyl, 20 mM tris(hydroxymethyl)aminomethane, 2 hours at room temperature. Diluted conditions: 0.1 mM substrate peptide, 0.125 mM decafluorobiphenyl, 10 mM tris(hydroxymethyl)aminomethane, 2 hours at room temperature.
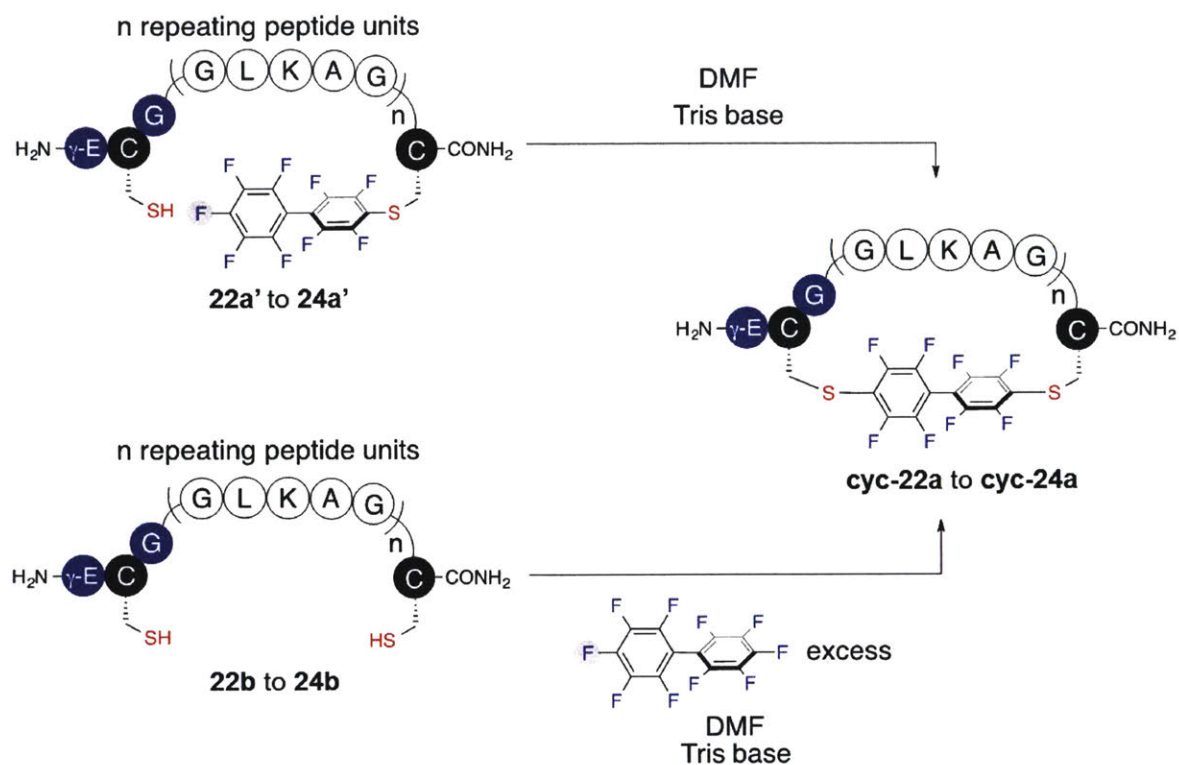
Peptide	Length (residues)	Yield (%)	Yield at diluted condition (%)
22a'	14	39	-
23a'	19	4	-
24a'	24	0	-
22b	14	37	42
23b	19	N/A	N/A
24b	24	N/A	N/A

Figure 2.27. Summary of control macrocyclization reactions in organic solvent.

All reactions showed low yields compared to the corresponding enzyme-catalyzed reactions. For long peptides (**23a'**, **24a'**, **23b**, **24b**), almost no desired cyclization product was observed.

In contrast to the GST-catalyzed cyclization where increasing the peptide length had no effects on the cyclization yield; all cyclization reactions in DMF showed significantly diminished yields as the peptide lengths increased (Figure 2.27). The significantly lowered reaction efficiencies and yields for macrocyclization in DMF compared to macrocyclization under GST catalysis in water highlight the effectiveness of the GST-catalyzed macrocyclization process which compensates the entropic penalty derived from the increased size of the peptide substrate.

2.2.8. GST-catalyzed macrocyclization of a fourty-mer unprotected peptide

The previous regioselective arylation studies (Figure 2.10 and Figure 2.11) have shown that the GST-catalyzed perfluoroarylation reaction was highly selective toward the cysteine residue within the GSH moiety. The desired GSH S-arylated product could be selectively and quantitatively formed in the presence of another cysteine in the same polypeptide chain. We reasoned this highly selective nature of the GST-catalyzed reaction could permit the combination of GST-catalyzed ligation with native chemical ligation (NCL)[42,98] to generate large cyclic peptides from linear substrate synthesized by NCL reaction (Figure 2.28). GST selectively recognizes the GSH tag and only catalyzes the arylation reaction at the GSH thiol in the presence of the free cysteine generated from NCL reaction, thus allowing NCL product to be cyclized by GST-catalyzed ligation (Figure 2.28).

We synthesized and purified a large 40-residue linear peptide **27** obtained from the hydrazide-based NCL reaction[98] (see Experimental section 2.4.11 for synthesis). Cyclization of peptide **27** under GST-catalysis produced desired cyclized product **Cyc-27** with 70% yield as measured by LC-MS analysis of the crude reaction mixture. Trypsin digestion and MS/MS analysis confirmed that the cysteine at the NCL ligation site remained unmodifed after the GST-catalyzed cyclization reaction (Figure 2.30). Reaction without enzyme showed only trace amount of product formation and no regio-selectivity (Figure 2.31). In addition, we further showed that the native chemical ligation and GST-catalyzed ligation could be done in one-pot with the desired macrocyclic peptide as the major product (Figure 2.29). In most cases, the macrocyclic product ring size is restricted to lengths of peptides that can be accessed from SPPS and the efficiency of the cyclization reactions. In our method, combining GST-catalyzed ligation and native chemical ligation offers method to overcome these challenges for the synthesis of large macrocyclic peptides.
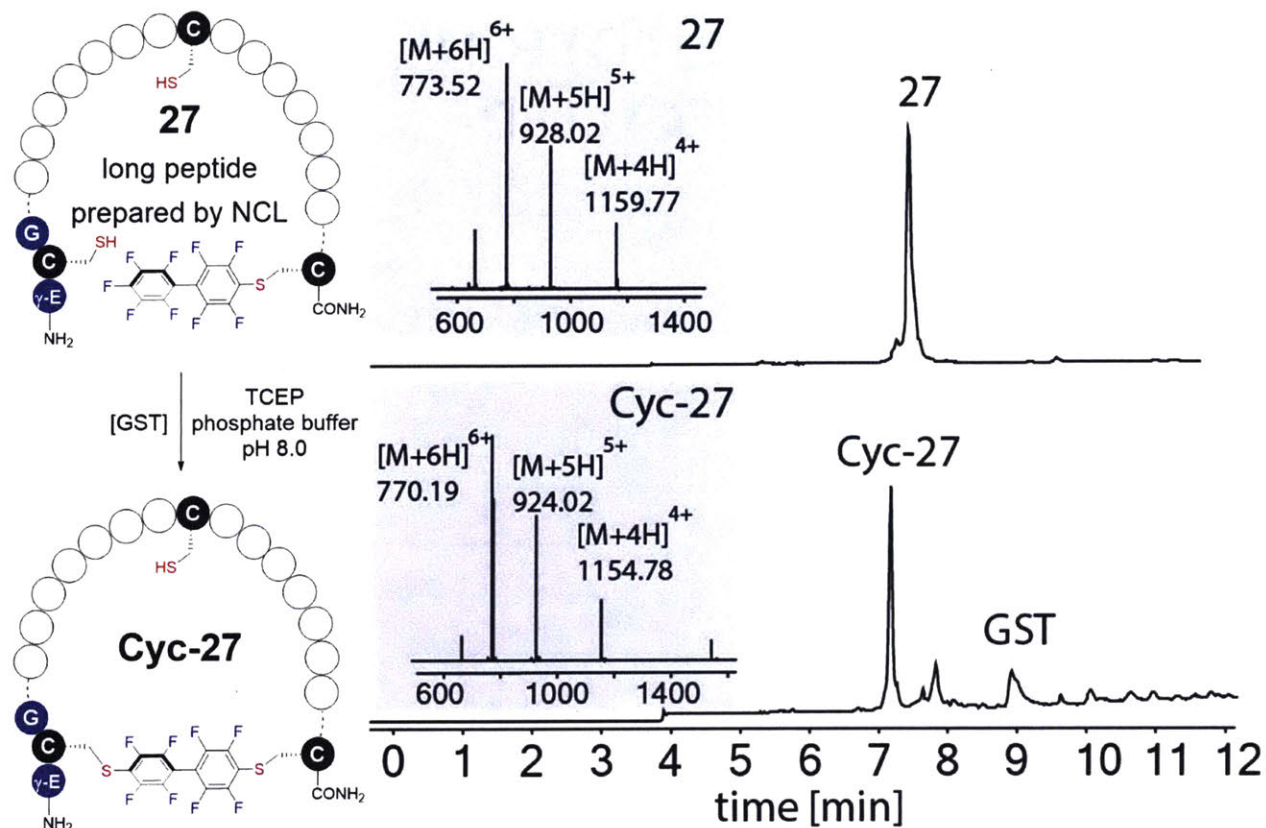
Figure 2.28. Macrocyclization of a 40-residues peptide prepared from native chemical ligation.

Reaction conditions: 0.1 mM peptide **27**, 0.2 mg/mL GST, 0.2 M phosphate, 20 mM TCEP•HCl, pH 8.0, room temperature, 2 hours. LC-MS data shown are total ion currents of starting material (top chromatogram) and crude reaction mixture (bottom chromatogram); mass spectrums at the highest points of the TIC peaks are shown as insets.
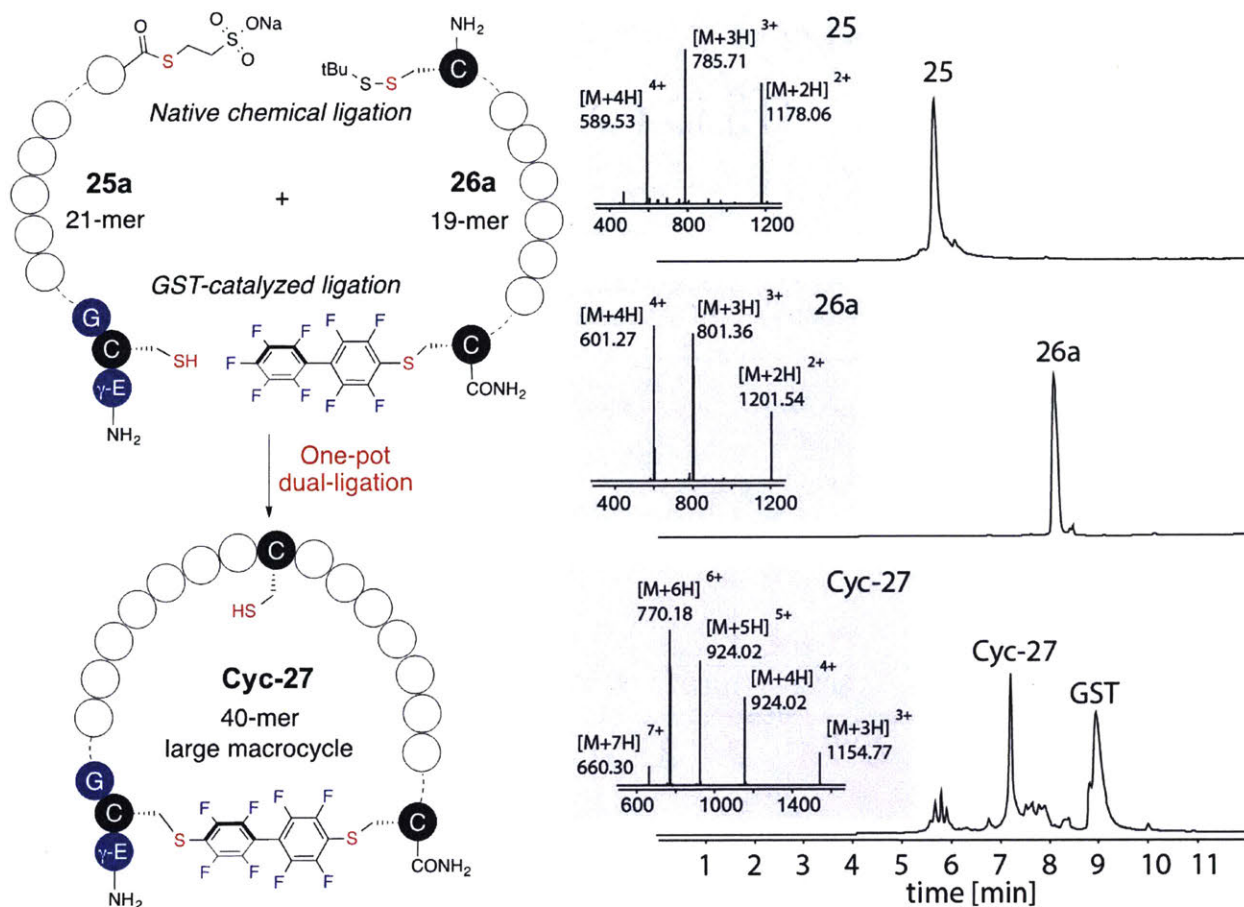
Figure 2.29. One-pot dual-ligation combining NCL and GST-catalyzed ligation for the synthesis of a large macrocyclic peptide.

LC-MS analysis of crude thioester **25a** from oxidation of peptide hydrazide **25** (top chromatogram), purified peptide **26a** (middle chromatogram), and crude dual-ligation reaction mixture (bottom chromatogram). Reaction conditions for one-pot dual ligation: 1 mM peptide **25a**, 1 mM peptide **26a**, 1 mg/mL GST-M5, 0.2 M phosphate, 20 mM TCEP•HCl, pH 8.0.

A

NH$_2$–γ-E**C***GGNQQKRAFIR**SLYDDPSGCANLLAEAK**KLNDAQAPK**C***–CONH$_2$

Cyc-27

Trypsin digestion

NH$_2$–**SLYDDPSGCANLLAEAK**–COOH
fragment(S14-K30)

Cyc-27

GST

before digestion

fragment(S14-K30)

[M+2H]$^{2+}$ 883.9067

[M+3H]$^{3+}$ 589.6048

[M+H]$^{+}$ 1766.8114

after digestion

B

883.9064 parent ion

ion	sequence	m/z
b2	SL	201.1196
b3	SLY	364.1831
b4	SLYD	479.2051
b5	SLYDD	594.2365
y2	AK	218.146
y3	EAK	347.1884
y4	AEAK	418.2241
y5	LAEAK	531.3086
y8	ANLLAEAK	829.4741
y10	GCANLLAEAK	989.506
y11	SGCANLLAEAK	1076.5321
y12	PSGCANLLAEAK	1173.5856
y12	PSGCANLLAEAK	587.2942
y13	DPSGCANLLAEAK	1288.6073
y13	DPSGCANLLAEAK	644.7961
y14	DDPSGCANLLAEAK	1403.6388
y15	YDDPSGCANLLAEAK	783.8491
b5-H2O	SLYDD	576.2226
b14-NH3	SLYDDPSGCANLLA	1403.6388
y2-NH3	AK	201.1196

Figure 2.30. Trypsin digestion and MS/MS analysis of Cyc-27.
Results show that GST regio-selectively catalyzed the cyclization at the GSH cysteine, and the central cysteine remained unmodified after GST-catalyzed cyclization reaction. (**A**) LC-MS analysis of the crude GST-catalyzed ligation product before trypsin digestion (top chromatogram) and after trypsin digestion (bottom chromatogram). Reaction conditions for cyclization reaction were same as those described in Figure 2.28. For trypsin digestion, 20 μL of the crude cyclization product was mixed with 20 μL of trypsin stock (100 μg/mL) and 1 μL of 1.5 M NH$_4$HCO$_3$. The mixture was incubated at 37 °C for 3 hours and was then subjected to LC-MS/MS analysis. The major product after trypsin digestion was identified as fragment (S14-K30), its mass spectrum is shown as the inset in the bottom chromatogram. Cross-linked cysteines are labeled as C* and highlighted in red; arrows indicate the sites of trypsin digestion; sequence of the digested product fragment (S14-K30) is highlighted in blue. (**B**) MS/MS analysis of fragment (S14-K30), masses and sequences of identified ion species are listed in the table. MS/MS result confirmed that the central cysteine C22 was not modified.

143
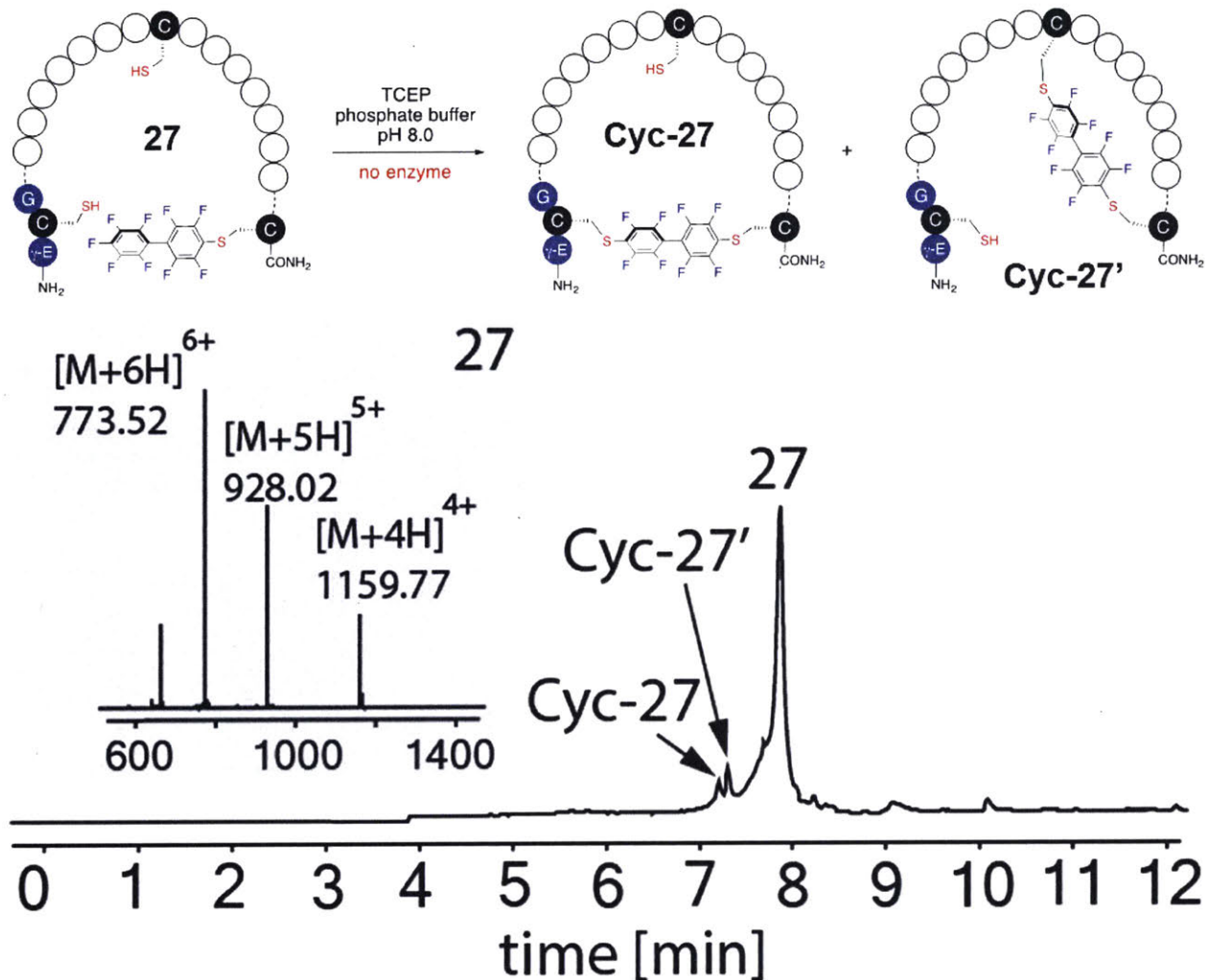
Figure 2.31. Cyclization of peptide 27 without GST catalyst.

Reaction conditions: 0.1 mM peptide **27**, 0.2 M phosphate, 20 mM TCEP•HCl, pH 8.0, room temperature, 2 hours. The reaction lacks regioselectivity producing two isomers **Cyc-27** and **Cys-27'** with arylation at both GSH cysteine and the NCL linkage cysteine, respectively. Peak corresponding to **Cyc-27** was assigned based on retention time identity of the product obtained in the GST-catalyzed cyclization reaction shown in Figure 2.28.

2.2.9. Robustness and enzyme-tag scope of GST-catalyzed arylation

This GST mediated transformation is not limited to specific temperatures, solvent conditions and peptides featuring γ-Glu-Cys-Gly sequence. For example, the reaction between **3b** and GSH under GST catalysis produces the desired arylated product **4b** at temperatures ranging from 4 to 60 °C (Figure 2.32), as well as in the presence of up to 20% organic co-solvent (Figure 2.32).

GST isozyme mixtures from multiple sources showed catalytic activity for the arylation reaction. Different GSTs were used to catalyze the conjugation of peptide **3b** with peptide **5a** (Figure 2.34A). GST from equine liver and GST from human placenta showed similar activity: peptide **5a** was converted nearly quantitatively to product **6a** within 30min under the catalysis of these two isozyme mixtures. However, GST from *Schistosoma Japonicum* showed no significant activity. We also found the activity of GST from *Schistosoma Japonicum* for conjugation of GSH to peptide **3b** (Figure 2.34B) was significantly lower than that of GST from equine liver. Expressed and purified GST-M5 was found to efficiently catalyze the arylation reaction and was used as catalyst for peptide macrocyclization reactions.

Additionally, GSH analogues with the mutation of γ-Glu to a genetically encodable Glu or Asp (Glu-Cys-Gly or Asp-Cys-Gly) undergo facile *S*-arylation with peptide **3b** under the developed GST-catalyzed conditions (Figure 2.35A). Finally, our experiments show that *N*-terminal Glu-Cys-Gly or Asp-Cys-Gly sequences in peptides (Figure 2.35A, peptides **5g** and **5h**) can also be selectively *S*-arylated with peptide **3b** though with lower efficiency as compared to peptides with N-terminal γ-Glu-Cys-Gly moiety (Figure 2.6, **5a**). These results are consistent with previous findings that certain GST isozymes are capable of catalyzing the S$_N$Ar reaction between 1,4-dinitrochlorobenzene and glutathione analogues.[99,100] Together, our observation suggests that the requirement for an *N*-terminal γ-glutamic acid residue could be eliminated with engineered GST.
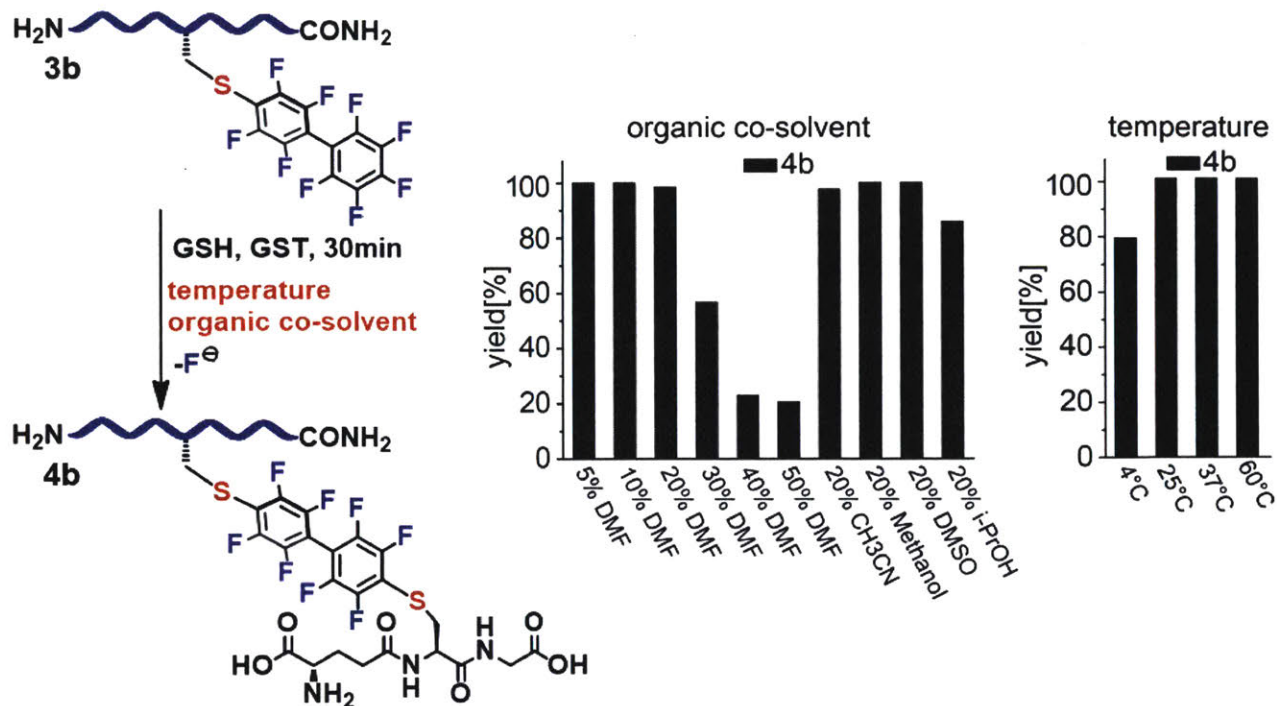
Figure 2.32. GST-catalyzed conjugation of GSH to peptide containing Cys-III residue with mixed solvent at variable conditions and temperatures.

Reaction conditions were the same as those in Figure 2.4 except for solvents (Left panel, percentages shown as volume ratio) and temperatures (right panel). See Figure 2.33 for complete LCMS analysis. DMF: dimethylformamide, DMSO: dimethylsulfoxide, *i*-PrOH: isopropanol.
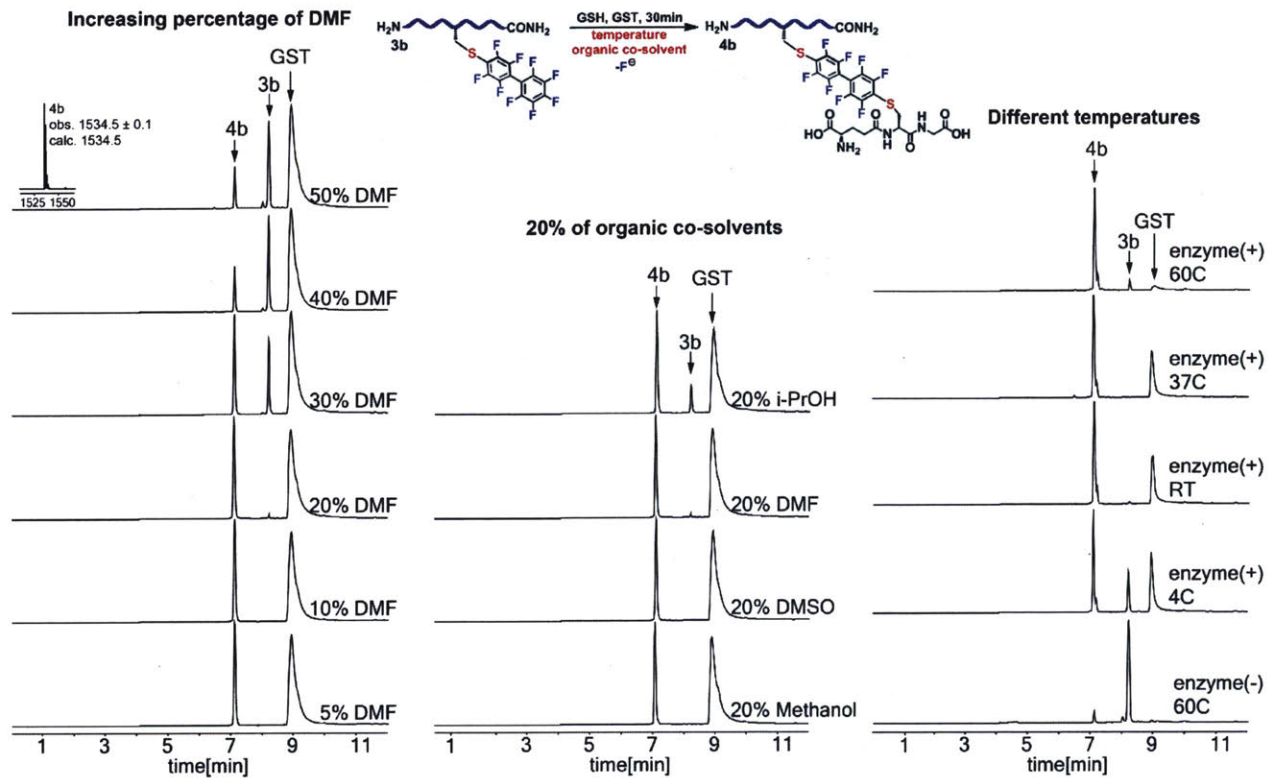
Figure 2.33. LC-MS analysis of GST-catalyzed arylation reactions at different temperature and solvent conditions.

Reaction conditions are described in Figure 2.32. Total ion current (TIC) traces are shown; mass spectra of TIC peaks at highest intensity are shown as an inset.
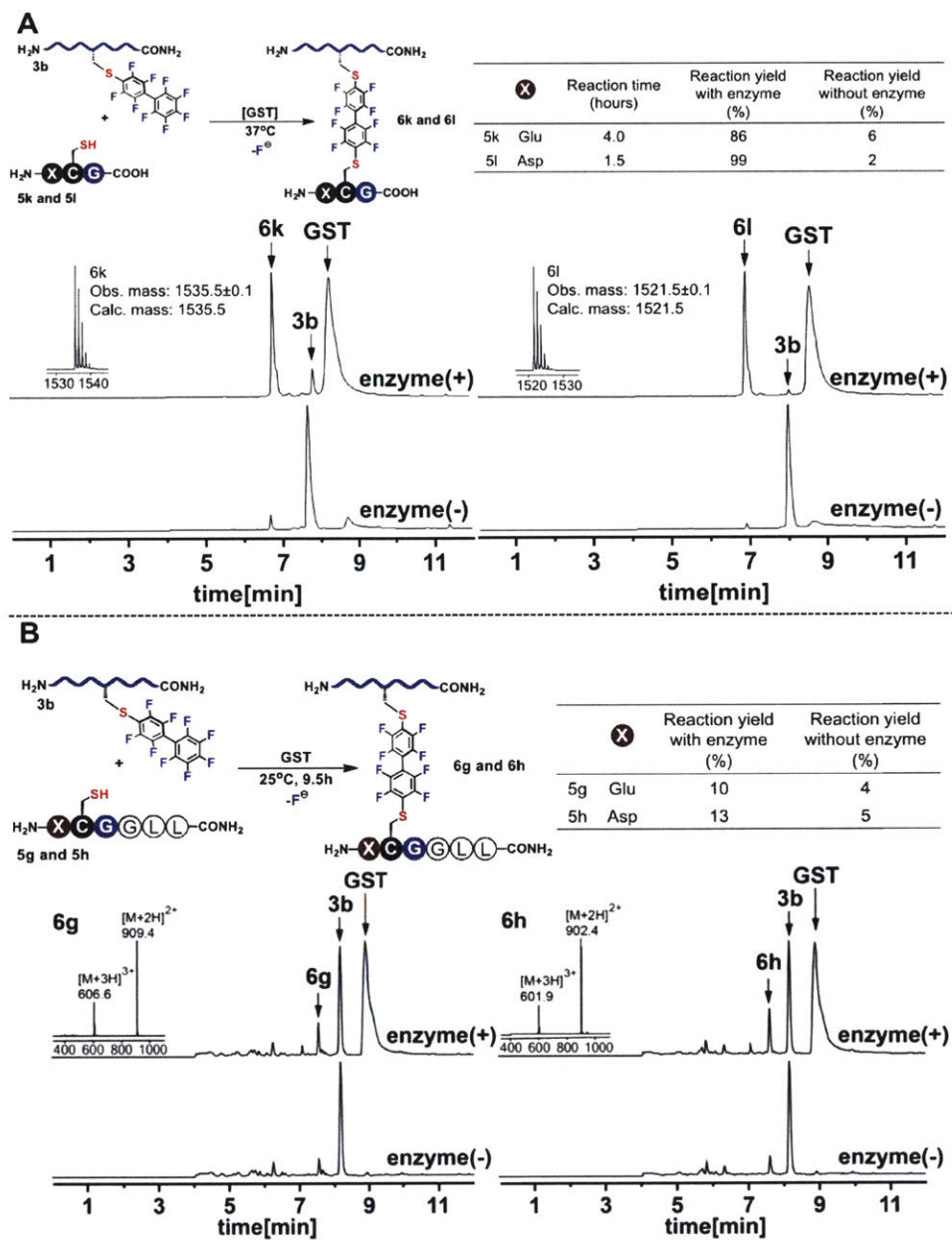
Figure 2.34. Comparison of the activities of various GSTs.

(A) Reaction conditions: 2 mM **3b**, 1 mM **5a**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C. Crude reactions at 30 minutes were subjected to LCMS analysis. (B) Reaction conditions: 1 mM **3b**, 1 mM GSH, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 8.0, 37 °C. Crude reactions at 30 minutes were subjected to LCMS analysis. Total ion current (TIC) LC traces are shown.

148
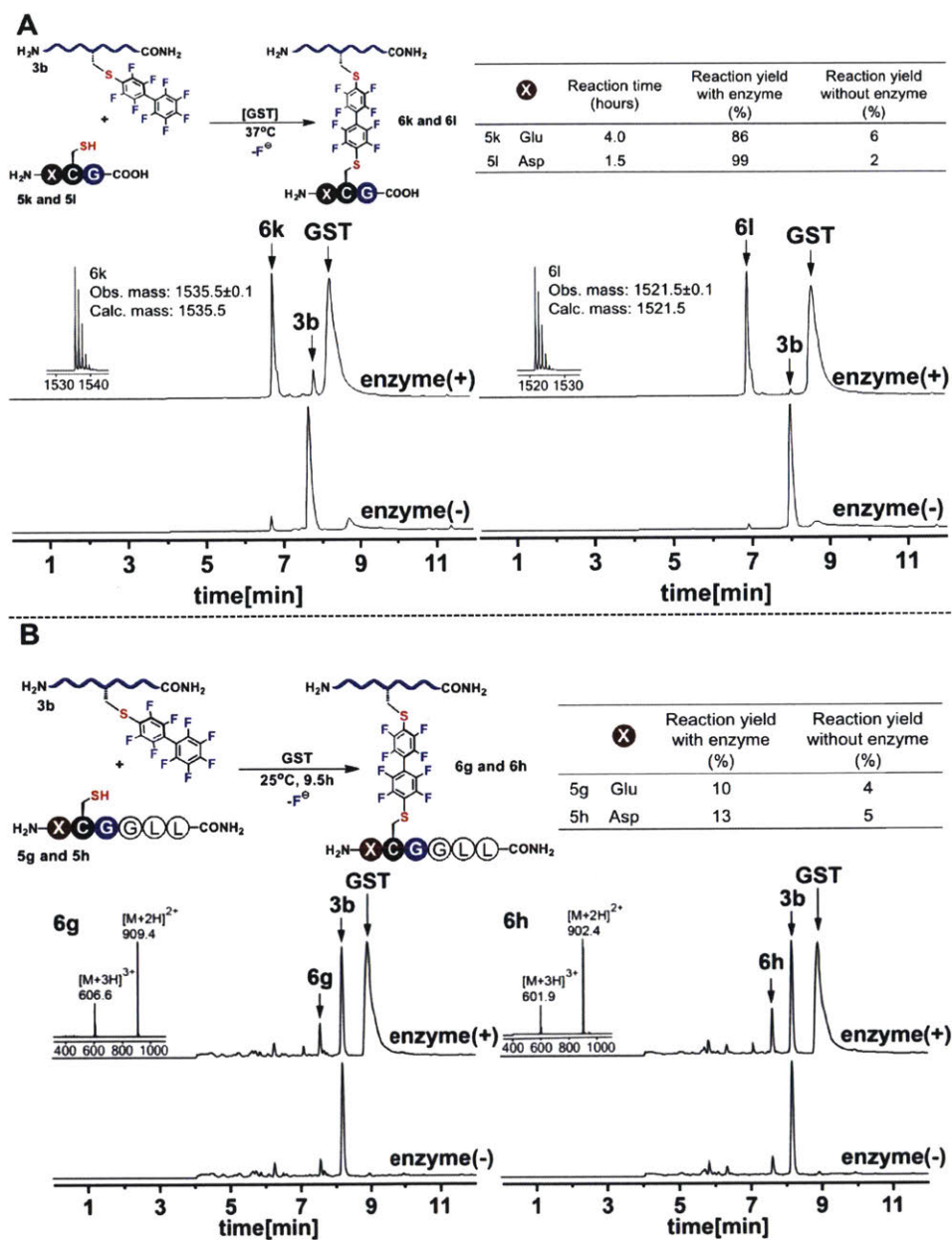
Figure 2.35. GST-catalyzed arylation of genetically encodable analoges of GSH-tag.

(A) Reaction conditions: 1 mM **3b**, 10 mM **5k** or **5l**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 9.0, 25 °C. Tri-peptide Glu-Cys-Gly and Asp-Cys-Gly were synthesized by standard Fmoc-SPPS using 2-chlorotrityl resin from Anaspec (Fermont, CA). (B) GST-catalyzed cysteine arylation of *N*-terminal Glu-Cys-Gly and Asp-Cys-Gly sequences in model peptides. Reaction conditions: 1 mM **3b**, 10 mM **5g** or **5h**, 2 mg/mL GST, 20 mM TCEP, 0.1 M phosphate, pH 9.0, 25 °C, 9.5 hours. Total ion current (TIC) traces are shown; mass spectra of TIC peaks at highest intensity are shown as an inset.

149

2.3. Discussion and Conclusion

In conclusion, we demonstrated a powerful glutathione S-transferase catalyzed S$_N$Ar "click" process for site-specific cysteine modification. This method allowed us to selectively modify cysteine in an N-terminal γ-Glu-Cys-Gly sequence within peptide and protein chains. The unique chemical orthogonality of the discovered arylation enabled by GST provides a route to modifying multiple cysteine sites with different chemical probes or biomolecules while avoiding the use of protecting groups and additional synthetic procedures. By showing how a naturally occurring enzyme catalyst improves and significantly broadens the scope of an established "click" synthetic transformation, this discovery paves a frontier in discovering new enzyme mediated reactivity modes suited for chemoselective modification of biomolecules.

In addition, we have shown how glutathione S-transferase (GST) can efficiently mediate the macrocyclization reaction of long peptides by catalyzing a S$_N$Ar process between cysteine and perfluoroaromatic moieties. This transformation operates under mild aqueous conditions, at ambient temperature, and takes only a few hours to complete. The reaction was shown to be highly efficient and selective, especially for the generation of macrocyclic peptides with large ring sizes. To demonstrate the unique selectivity of our GST-catalyzed reaction, a 40-residue macrocyclic peptide was prepared by combining GST-catalyzed ligation with native chemical ligation.

2.4. Experimental

2.4.1. Chemicals

Hexafluorobenzene and decafluorobiphenyl were purchased from Oakwood Chemicals (West Columbia, SC). Decafluorobiphenyl sulfide was purchased from SynQuest Laboratories (Alachua, FL). *Tris*(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) was purchased from Hampton Research (Aliso Viejo, CA). 2-(1H-Benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), Fmoc-Rink amide linker, Boc-α-Glu(OtBu)-OH, Fmoc-Gly-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Ala-OH, Fmoc-Cys(Trt)-OH, Fmoc-Cys(S-tBu)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Phe-OH, Fmoc-Ile-OH, Fmoc-Ser(tBu)-OH, Fmoc-Tyr(tBu)-OH and Fmoc-Pro-OH were purchased from Chem-Impex International (Wood Dale, IL). 4-methylbenzhydrylamine (MBHA) resin was obtained from Anaspec (Fremont, CA). *N,N*-Dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were obtained from VWR International (Philadelphia, PA). All other reagents were purchased from Sigma-Aldrich and used as received unless otherwise noted.

2.4.2. Glutathione *S*-transferase

GST isozyme mixtures were purchased from Sigma Aldrich. GST from equine liver (Sigma Aldrich catalog number: G6511) and GST from human placenta (Sigma Aldrich catalog number: G8642) were received as lyophilized powders containing Tris and phosphate buffer salts, reduced glutathione, and EDTA. Water was added to dissolve these GST samples to a final concentration of 10 mg/mL. GST from Schistosoma Japonicum (Sigma Aldrich catalog number: G5663) was received as a solution of phosphate buffered saline, pH 7.4, containing 0.02% sodium azide, concentration was ≥ 3 mg/mL and purity was ≥ 85% (SDS-PAGE). All GSTs were dialyzed against 20 mM Tris, 150 mM NaCl, pH 7.5 buffer to remove reduced glutathione, and aliquots of GSTs were stored at -80 °C. All reactions in this work were performed with GST from equine liver unless otherwise noted.

Human GST-M5 was cloned from pcDNA3-HA-GSTM5 (Addgene number: 11986) using ChampionTM pET SUMO expression system (Life Technology) following manufacturer's

recommended protocol. Aliquots of expressed GST-M5 were stored in 20 mM Tris, 150 mM NaCl, pH 7.5 buffer at 6 mg/mL concentration in -80 °C refrigerator.

2.4.3. Solid phase peptide synthesis

All peptides were synthesized on a 0.2 mmol scale using manual fast flow peptide synthesis[96] using 3-minute cycle for each amino acid. Specifically, all reagents and solvents are delivered to a stainless steel reactor containing resins at a constant flow rate using HPLC pump; temperature of the reactor was maintained at 60 °C during the synthesis using water bath. Procedure for amino-acid residue coupling cycle contained 30 second coupling with 1 mmol Fmoc protected amino acids, 1.2 mmol HBTU, and 500 μL of diisopropyl ethyl amine (DIEA) in 2.5 mL of DMF, flow rate was 6 mL/min, note that for coupling of cysteine, 190 μL of DIEA was used to prevent racemization; 1-minute wash with DMF, flow rate was 20 mL/min; 20 seconds deprotection with 50% (v/v) piperidine in DMF, flow rate was 20 mL/min; and 1-minute wash with DMF, flow rate was 20 mL/min. The resin was washed thoroughly with DCM and air dried after completion of the stepwise SPPS. The peptide is then simultaneously cleaved from the resin and side-chain deprotected by treatment with 2.5% (v/v) water, 2.5% (v/v) 1,2-ethanedithiol (EDT), and 1% (v/v) triisoproprylsilane in neat trifluoroacetic acid (TFA) for 2 hours at room temperature. Resulting solution containing peptide was evaporated by blowing a stream of nitrogen gas over its surface for 15 minutes, then triturated and washed with cold diethyl ether three times. Obtained gummy-like solid was dissolved in 50% H_2O: 50% acetonitrile containing 0.1% TFA and lyophilized. *These same solvent compositions were used in most experiments and will be referred to as A: 0.1% TFA in H_2O and B: 0.1% TFA in acetonitrile.* All peptides were synthesized using MBHA resin, unless noted otherwise. Peptide **5** was synthesized using hydrazide resin[98]. All peptides containing glutathione tag at *N*-terminus were synthesized with Boc-α-Glu(OtBu)-OH as the last amino acid residue.

2.4.4. Peptide purification

The crude peptide was dissolved in 95% A: 5% B with 6 M guanidinium hydrochloride and purified by semi-preparative RP-HPLC (Agilent Zorbax SB C_{18} column: 21.2 x 250 mm, 7 μm, linear gradient: 5-50% B over 90 min, flow rate: 5 mL/min). 1 μL of each HPLC fraction was mixed with 1 μL of alpha-cyano-4-hydroxycinnamic acid (CHCA) matrix in 75% A: 25% B,

spotted with MALDI, and checked for fractions with desired molecular mass. The purity of fractions was confirmed by analytical RP-HPLC (Agilent Zorbax SB C_3 column: 2.1 x 150 mm, 5 µm, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min). HPLC fractions containing only product material were confirmed by LC-MS analysis, combined, and then lyophilized. Peptides synthesized using fast flow-based SPPS and purified by RP-HPLC are listed in Table 2.1.

Table 2.1. Peptides synthesized via fast flow peptide synthesis and purified by RP-HPLC.

Peptide	Sequence[#]	Obs. mass	Calc. mass
1	H_2N-ITPCNLLF*YYGKKK-$CONH_2$	1775.9	1775.89
3	H_2N-VTLPSTCGAS-$CONH_2$	933.46	933.46
5a	H_2N-γ-ECGGLL-$CONH_2$	589.29	589.29
5b	H_2N-γ-ECGLLL-$CONH_2$	645.35	645.35
5c	H_2N-γ-ECGHLL-$CONH_2$	669.33	669.33
5d	H_2N-γ-ECGKLL-$CONH_2$	660.36	660.36
5e	H_2N-γ-ECGPLL-$CONH_2$	629.32	629.32
5f	H_2N-γ-ECGGGLL-$CONH_2$	646.31	646.31
5g	H_2N-ECGGLL-$CONH_2$	589.3	589.29
5h	H_2N-DCGGLL-$CONH_2$	575.27	575.27
5i	H_2N-γ-ECG-$CONHNH_2$	321.11	321.11
5j	H_2N-γ-ECGGALF-$CONHNH_2$	709.31	709.32
11a	Biotin-RRC-$CONH_2$	658.31	658.31
14	H_2N-γ-ECGGPTAAKESCLL-$CONH_2$	1376.64	1376.64
14a	H_2N-γ-ECGGPTAAKESC(S-tBu)LL-$CONH_2$	1464.68	1464.68
21	H_2N-γ-EC(S-tBu)G-GLKAG-C-$CONH_2$	923.40	923.40
22	H_2N-γ-EC(S-tBu)G-(GLKAG)$_2$-C-$CONH_2$	1349.66	1349.66
23	H_2N-γ-EC(S-tBu)G-(GLKAG)$_3$-C-$CONH_2$	1775.92	1775.92
24	H_2N-γ-EC(S-tBu)G-(GLKAG)$_4$-C-$CONH_2$	2202.18	2202.18
22b	H_2N-γ-ECG-(GLKAG)$_2$-C-$CONH_2$	1261.63	1261.63
23b	H_2N-γ-ECG-(GLKAG)$_3$-C-$CONH_2$	1687.89	1687.89
24b	H_2N-γ-ECG-(GLKAG)$_4$-C-$CONH_2$	2114.15	2114.15
25	H_2N-γ-ECGGNQQKRAFIRSLYDDPSG-$CONHNH_2$	2354.12	2354.12
26	H_2N-C(S-tBu)ANLLAEAKKLNDAQAPKC-$CONH_2$	2087.07	2087.07

[#]Amino acids are shown in one-letter code. F* is L-pentafluorophenylalanine; γ-E stands for γ - glutamic acid; C(S-tBu) denotes tert-butylthio-protected cysteine.

2.4.5. LC-MS analysis

LC-MS chromatograms and associated mass spectra were acquired using Agilent 6520 ESI-Q-TOF mass spectrometer. Following LC methods were used:

Method A*:* Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min.

Method B*:* Zorbax SB C_{18} column: 2.1 x 250 mm, 5 μm, gradient: 0-2 minutes 1% B, 2-15 minutes 1-70% B, 15-17 minutes 70% B, flow rate: 0.4 mL/min.

All samples were analyzed using ***Method A*** unless otherwise noted. Data was processed using Agilent Mass Hunter software package.

Y-axis in all chromatograms shown in supplementary figures represents total ion current (TIC); mass spectrum insets correspond to the maxima point of the TIC peak.

2.4.6. Determination of reaction yields

For all GST-catalyzed conjugation reactions except peptide macrocyclizations, yields shown were determined by measuring UV absorption at $\lambda = 280$ nm using LC-MS data. First, using Agilent Mass Hunter software package, the peak areas for all relevant species on the chromatogram were integrated. For all reactions, the relevant peaks from UV absorption spectra at $\square = 280$ nm arise from species containing perfluoroaromatic moieties. In Figure 2.4 and Figure 2.32, no products were generated in reactions without enzyme, and clean conversions of substrates to GST-catalyzed products were observed. For these reactions, yields were calculated using the following equation: %yield $= (1-S_E/S_C) \times 100$, where S_E is the peak area of peptide substrates containing perfluoroaromatic moieties in enzymatic reactions and S_C is the peak area of those peptide substrates in control reaction without enzyme. For reactions in Figure 2.35A, both the yields of enzymatic reactions and reactions without enzyme are calculated using equation: %yield $= (1-S_E/S_C) \times 100$, where S_E is the peak area of peptide substrates containing perfluoroacromatic moieties in enzymatic reaction or control reaction without enzyme, and S_C is the peak area of reaction with only peptide **3b**. For reactions in Figure 2.6, yields were calculated using the following equation to account for two equivalents of peptide **3b** used relative to peptides **5a-f**: %yield $= (2-S_E/S_C) \times 100$, where S_E is the peak area of peptide 3b in enzymatic reaction and S_C is the peak area of peptide **3b** in control reaction without enzyme.

All yields for peptide macrocyclization reactions were determined by measuring total ion currents using LC-MS data unless otherwise noted. First, using Agilent Mass Hunter software package, the peak areas for all relevant peptidic species on the chromatogram were integrated. Then yield was calculated as following: %yield = S_{pro}/S_{all} where S_{pro} is the peak area of the desired product and S_{all} is the sum of all TIC peaks.

2.4.7. Protein expression and purification

Expression and Purification of A1C-LFN protein

The A1C-LFN construct was cloned from pET-SUMO-LFN-LPSTGG-H$_6$[44] using QuickChange Lightening Site-Directed Mutagenesis kit according to the manufacturer's protocol. The generated construct encoded for the following protein sequence (the product of SUMO removal is shown in bold, the introduced A1C mutation is highlighted in red):

SUMO-A1C-LFN-H$_6$

MSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGI
RIQADQTPEDLDMEDNDIIEAHREQIGG**C**GGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLKEIMKH
IVKIEVKGEEAVKKEAAEKLLEKVPSDVLEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDIYGKDA
LLHEHYVYAKEGYEPVLVIQSSEDYVENTEKALNVYYEIGKILSRDILSKINQPYQKFLDVLNTIKNASD
SDGQDLLFTNQLKEHPTDFSVEFLEQNSNEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKFNEQEI
NLSLEELKDQRLPSTGGHHHHHH

E. coli BL21(DE3) cells transformed with SUMO-A1C-LFN-LPSTGG-H$_6$ construct were grown in 2 L of LB medium containing ampicillin (100 μg/mL) at 37 °C until OD$_{600}$ = 0.6. Then, expression was induced by addition of 0.5 mM IPTG overnight at 37 °C. After harvesting the cells by centrifugation (6,000 rpm, 30min), the cell pellet was lysed by sonication in 50 mL of 50 mM Tris, 150 mM NaCl, pH 7.5 buffer containing 30 mg lysozyme, 2 mg DNase I, and 1 tablet of protease inhibitor cocktail. The suspension was centrifuged at 17,000 rpm for one hour to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK) and washed with 50 mL of 40 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. The protein was eluted from the column with buffer containing 500 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. Imidazole was removed from protein using HiPrep 26/10 Desalting column (GE Healthcare, UK) into 20 mM Tris, 150 mM NaCl, pH 7.5. Purified protein was

analyzed using an Any kD Mini-PROTEAN TGX Precast Gel (Bio-Rad, CA). In addition, the protein was analyzed by LCMS analysis to confirm its purity and molecular weight analyzed via high-resolution ESI-QTOF MS.

SUMO group on SUMO-A1C-LFN-H$_6$ was cleaved first by incubating 1 µg SUMO protease per mg protein at room temperature for 90 minutes. The crude reaction mixture was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK) and washed with 50 mL of 40 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. The A1C-LFN-H$_6$ protein was eluted from the column with buffer containing 500 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. Imidazole was removed from protein using HiPrep 26/10 Desalting column (GE Healthcare, UK) into 20 mM Tris, 150 mM NaCl, pH 8.5. Purified protein was analyzed using an Any kD Mini-PROTEAN TGX Precast Gel (Bio-Rad, CA). In addition, the protein was analyzed by LCMS to confirm the purity and molecular weight analyzed via high-resolution ESI-QTOF MS. The A1C-LFN-H$_6$ protein was purified as disulfide linked dimer (Figure 2.36, observed mass: 63733.3, calculated mass: 63731.6).

Preparation of protein 7 by hydrazide ligation

γ-Glu-Cys-Gly-CONHNH$_2$ (**5i**) was prepared by Fmoc-SPPS using hydrazide resin.[98] To 4 µmoles of peptide hydrazide dissolved in 0.1 M phosphate, 6 M guanidinium hydrochloride, pH 3.0 was added 100 µL of 100 mM NaNO$_2$ in water at -10 °C and reacted for 20 minutes at the same temperature. Then 500 µL of 200 mM 4-mercaptophenyl acetic acid (MPAA) in 0.1 M phosphate (pH 7.0) was added. The pH of the reaction mixture was adjusted to 7.0 by addition of 5 M sodium hydroxide. The reaction mixture was left at room temperature for 10 minutes and analyzed by LCMS. This crude peptide mixture was then used directly in ligation reaction with protein A1C-LFN. To 50 µL crude thioester mixture was added 10 µL of 40 mg/mL A1C-LFN and 10 µL of 1 M phosphate, 200 mM TCEP·HCl, pH 7.0. The reaction mixture was left at room temperature for 2 hours and characterized by LC-MS analysis (Figure 2.36). Protein **7** was purified by dialysis against 20 mM Tris, 150 mM NaCl, pH 7.5 buffer.
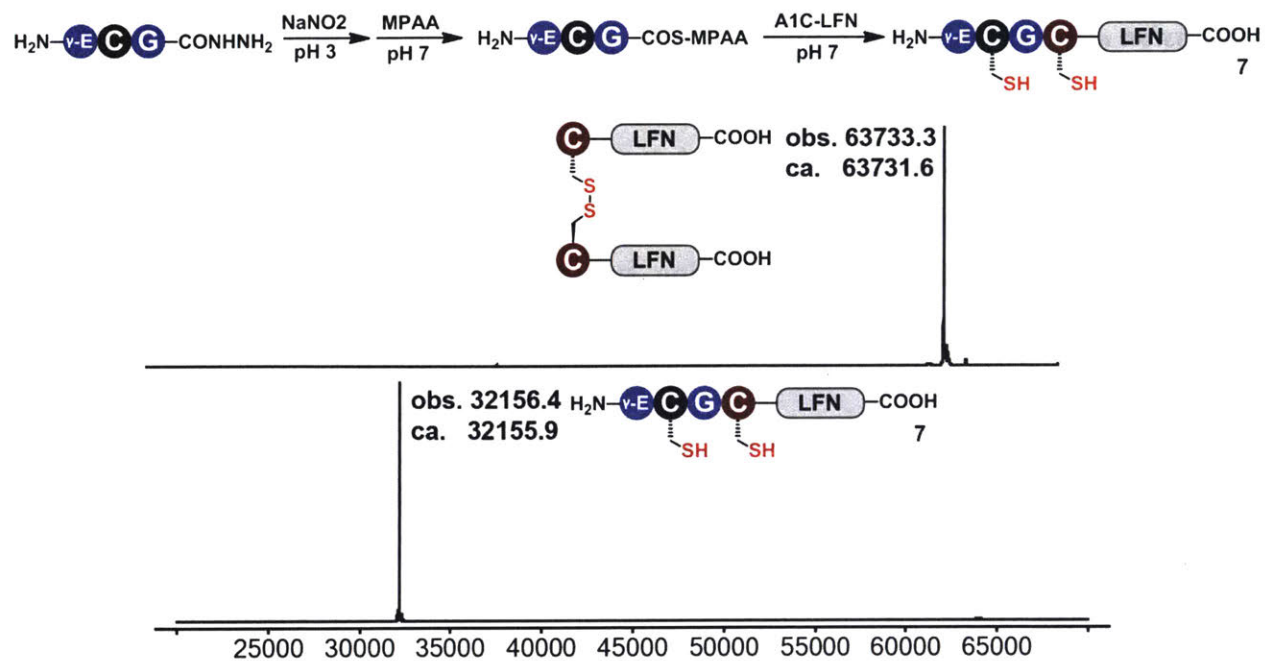
Figure 2.36. Preparation of protein 7 with N-terminal GSH tag by peptide hydrazide ligation.

Deconvoluted mass spectra obtained from LC-MS analysis of these samples are shown.

Expression and Purification of GST-M5

GST-M5 gene was amplified from pcDNA3-HA-GSTM5 (Addgene number: 11986) using primers: 5'-ATGCCCATGACTCTGGG and 5'-CTATTTGCTGTTCCATGTAGCTGAC, after gel purification, the gene was ligated into pET SUMO vector using protocol provided in Champion™ pET SUMO protein expression system manual. The *N*-terminal oligo-glycine sequence was introduced by site-directed mutagenesis using QuickChange Lightning Multi Site-Directed mutagenesis kit (Agilent) with primer: 5'-GGCTCACAGAGAACAGATTGGTGGTggtggtggtggcggcATGCCCATGACTCT-GGGG. The generated pET-SUMO-G5-GST-M5 construct encoded for the following protein sequence (G5-GST-M5 sequence is shown in bold):

SUMO-GST-M5

MSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGI
RIQADQTPEDLDMEDNDIIEAHREQIG**GGGGGGMPMTLGYWDIRGLAHAIRLLLEYTDSSYVEKKYTLGD**
APDYDRSQWLNEKFKLGLDFPNLPYLIDGAHKITQSNAILRYIARKHNLCGETEEEKIRVDILENQVMDN
HMELVRLCYDPDFEKLKPKYLEELPEKLKLYSEFLGKRPWFAGDKITFVDFLAYDVLDMKRIFEPKCLDA
FLNLKDFISRFEGLKKISAYMKSSQFLRGLLFGKSATWNSK

E. coli BL21(DE3) cells transformed with pET-SUMO-G5-GST-M5 plasmid were grown in 1 L of LB medium containing kanamycin (30 μg/mL) at 37 °C until OD_{600} = 0.6. Then, expression was induced by addition of 0.5 mM IPTG overnight at 30 °C. After harvesting the cells by centrifugation (6,000 rpm, 10 min), the cell pellet was lysed by sonication in 25 mL of 50 mM Tris, 150 mM NaCl, pH 7.5 buffer containing 15 mg lysozyme (Calbiochem), 1 mg DNase I (Sigma-Aldrich), and 0.5 tablet of protease inhibitor cocktail (Roche Diagnostics, Germany). The suspension was centrifuged at 17,000 rpm for 30 min to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK), first washed with 40 mL of 20 mM Tris, 150 mM NaCl, pH 8.5, and then washed with 40 mL of 40 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. The protein was eluted from the column with buffer containing 500 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 8.5. Imidazole was removed from protein using HiPrep 26/10 Desalting column (GE Healthcare, UK), protein was eluted into 20 mM Tris, 150 mM NaCl, pH 7.5 buffer. Purified protein was analyzed using an Any kD Mini-PROTEAN

TGX Precast Gel (Bio-Rad, CA). In addition, the protein was analyzed by LC-MS to confirm its purity and molecular weight.

SUMO group on SUMO-G5-GST-M5 was cleaved by incubating 1 μg of SUMO protease per mg protein at room temperature for 60 minutes. The crude reaction mixture was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK) and the flow through containing G5-GST-M5 was collected. The protein was analyzed by LC-MS confirming sample purity and molecular weight (analyzed via high-resolution ESI-Q-TOF MS, observed mass: 25960.2, calculated mass: 25960.5).

2.4.8. Synthesis of *S*-arylated electrophilic probes

Peptides 3a-3c

To a solid sample of peptide **3** (5 μmoles) dissolved in 20mM Tris base in 1mL of DMF in a plastic Eppendorf tube was added 500 μmoles of perfluoroaromatic reagent (hexafluorobenzene for **3a**, decafluorobiphenyl for **3b**, decafluorobiphenyl sulfide for **3c**). The tube was vortexed to ensure complete reagent mixing and left at room temperature for 1 hour. Reaction mixtures were characterized by LC-MS analysis. Resulting reaction mixtures were quenched by addition of 20 mL of 95% A: 5% B. Peptides **3a** and **3c** were purified by solid-phase extraction using Grace C_{18} SPE Maxi-Clean Cartridges (Deerfield, IL) according to the manufacturer's protocol (Figure 2.37, left and right chromatograms), and peptide **3b** was purified by RP-HPLC (Figure 2.37, chromatogram in the center).

Preparation of biotin probe 11

Peptide **11a** was synthesized by Fmoc-SPPS using biotin as the last amino acid. To a solid sample of peptide **11a** (5 μmoles) in a plastic Eppendorf tube dissolved in 20 mM Tris base in 1mL of DMF was added 500 μmoles of decafluorobiphenyl. The tube was vortexed and left at room temperature for 1 hour. Reaction mixtures were characterized by LCMS analysis. Resulting reaction mixtures were quenched by addition of 20 mL of 95% A: 5% B. Peptide **11** was purified by RP-HPLC. Pure fractions were identified by LCMS analysis (Figure 2.38), collected, and lyophilized.

159
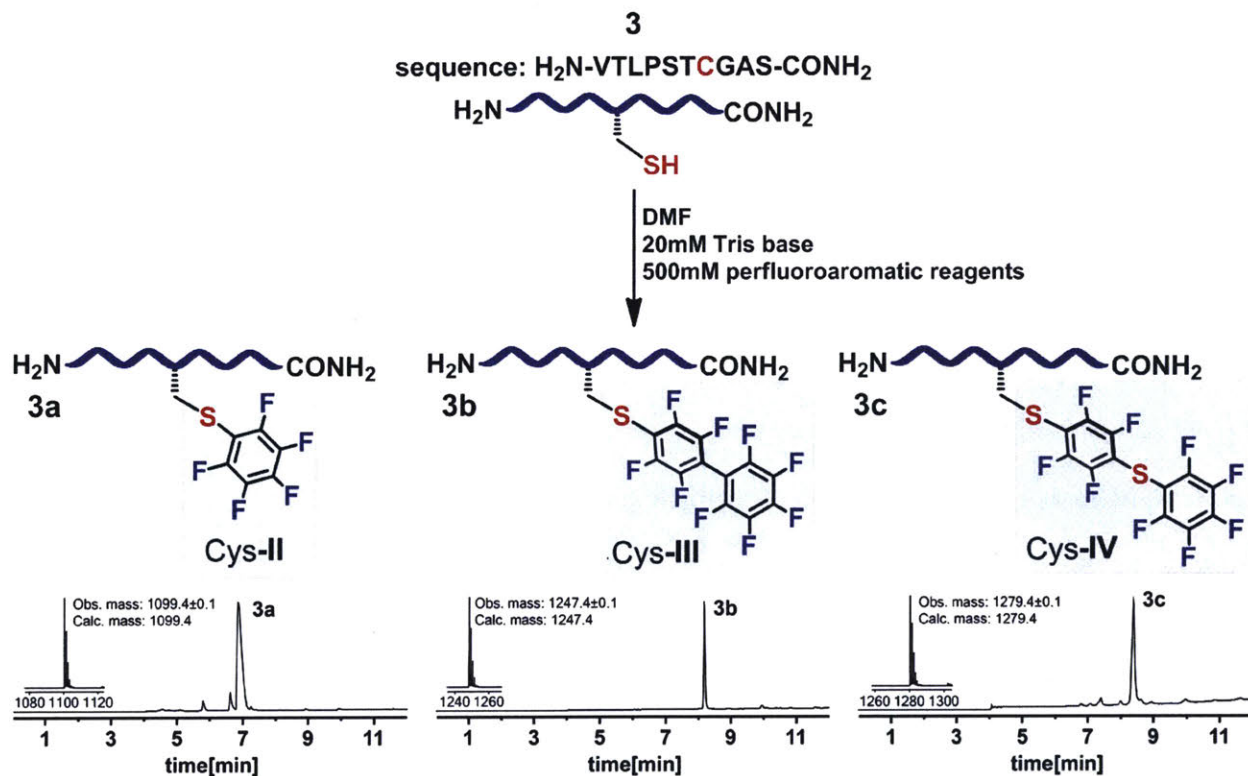
Figure 2.37. Preparation of peptides 3a-3c.

Total ion current (TIC) traces are shown; mass spectra of TIC peaks at highest intensity are shown as an inset.
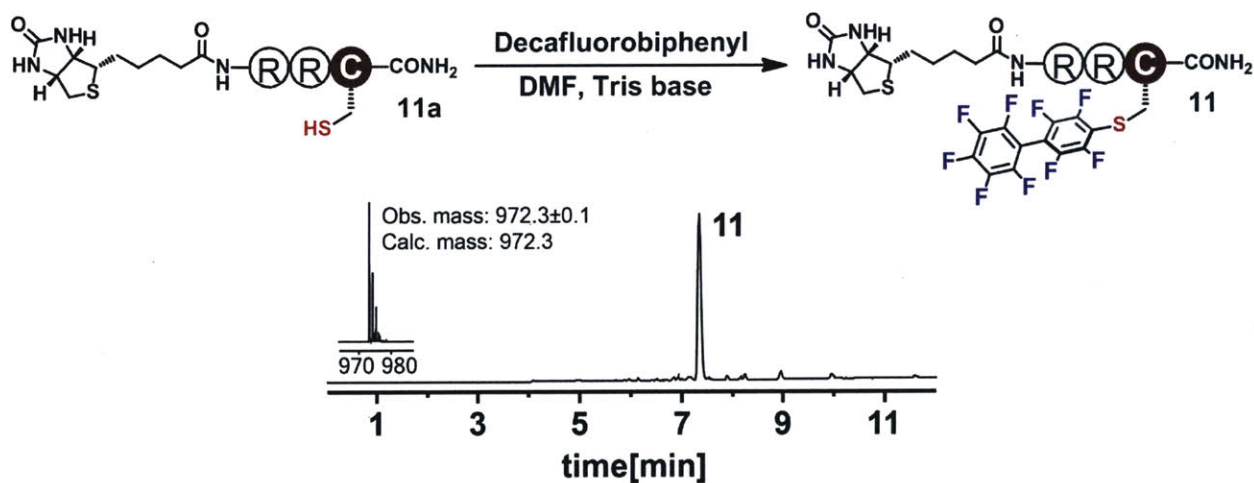
Figure 2.38. Preparation of biotin probe 11.

The total ion current (TIC) trace is shown; mass spectrum of the TIC peak at highest intensity is shown as an inset.

160

Peptides 21a−21c

To a solid sample of peptide 21 (5 μmoles) dissolved in 20 mM Tris base in 1mL of DMF in a plastic Eppendorf tube was added 500 μmoles of perfluoroaromatic reagent (decafluorobiphenyl for product **21a**, hexafluorobenzene for **21b**, decafluorobiphenyl sulfide for **21c**). The tube was vortexed to ensure complete reagent mixing and left at room temperature for 30 minutes. 1 μL of each reaction mixture was quenched by addition of 20 μL of 50% A: 50% B and subjected to LC-MS analysis. Resulting reaction mixtures were quenched by addition of 20 mL of 95% A: 5% B, filtered through 0.22 μm nylon syringe filter, and purified by RP-HPLC.

Peptides 22a-24a

Peptides **22a−24a** were prepared using the same protocol as **21a−21c** by reacting peptide **22−24** with decafluorobiphenyl. Note that for preparation of peptides **23a** and **24a**, 20 mL of 90% A: 10% B was used as a quenching solution.

Peptide *26a* was prepared using the same protocol by reacting peptide **26** with decafluorobiphenyl.

LC-MS data for all reactions obtained *in situ* and RP-HPLC-purified products of peptides **21a** to **21c** and **22a−24a** are summarized in Figure 2.39 and Figure 2.40.
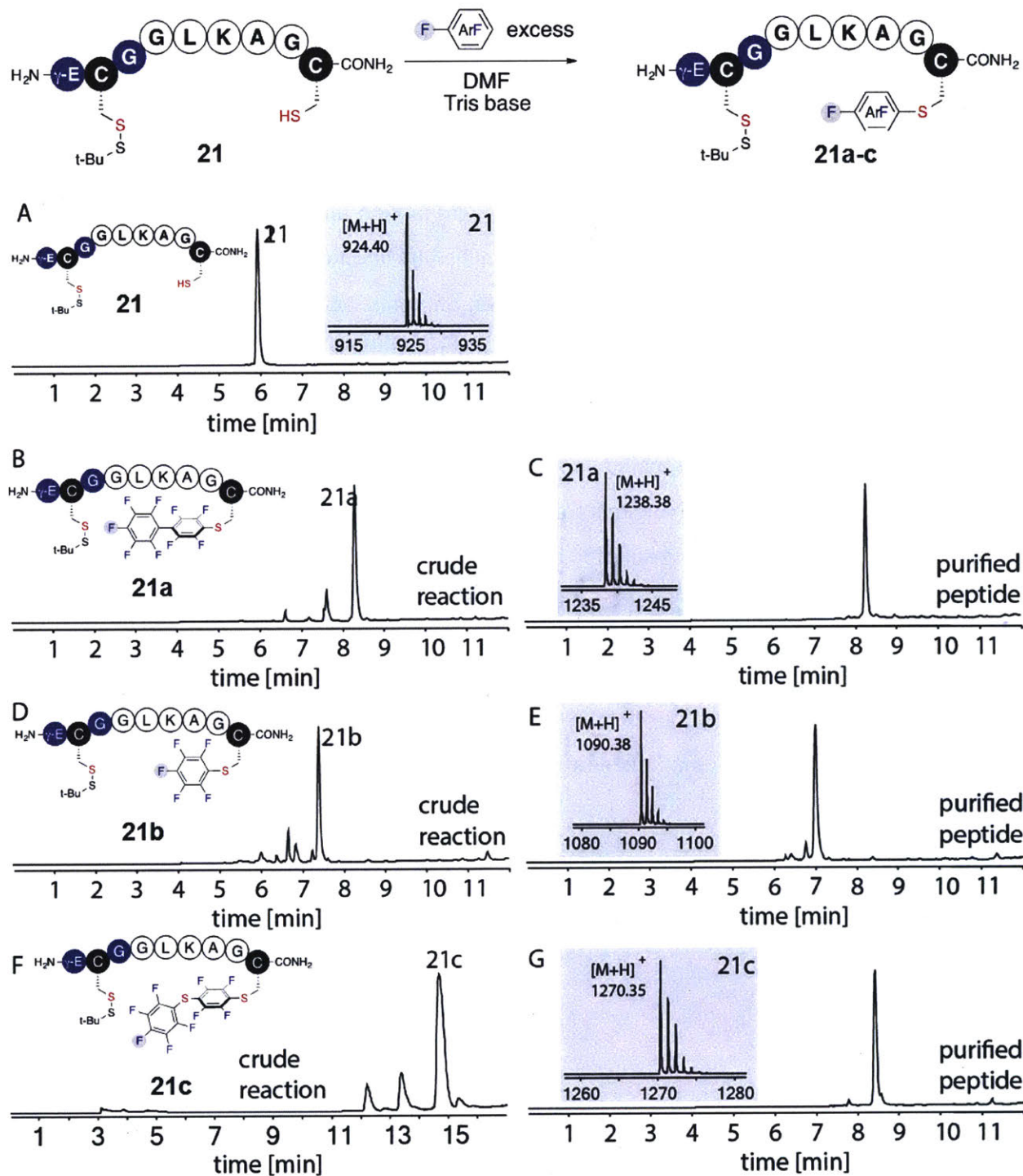
Figure 2.39. Synthesis and purification of peptides 21a-21c.

Reaction conditions are described in the general protocol section. Chromatograms B, D, and F correspond to crude reactions used to prepare peptides **21a**, **21b**, and **21c**, respectively; chromatograms C, E and G represent LC-MS analysis data for RP-HPLC-purified peptides **21a**, **21b**, and **21c**, respectively. Chromatogram F was acquired using LC-MS *method B*.

162
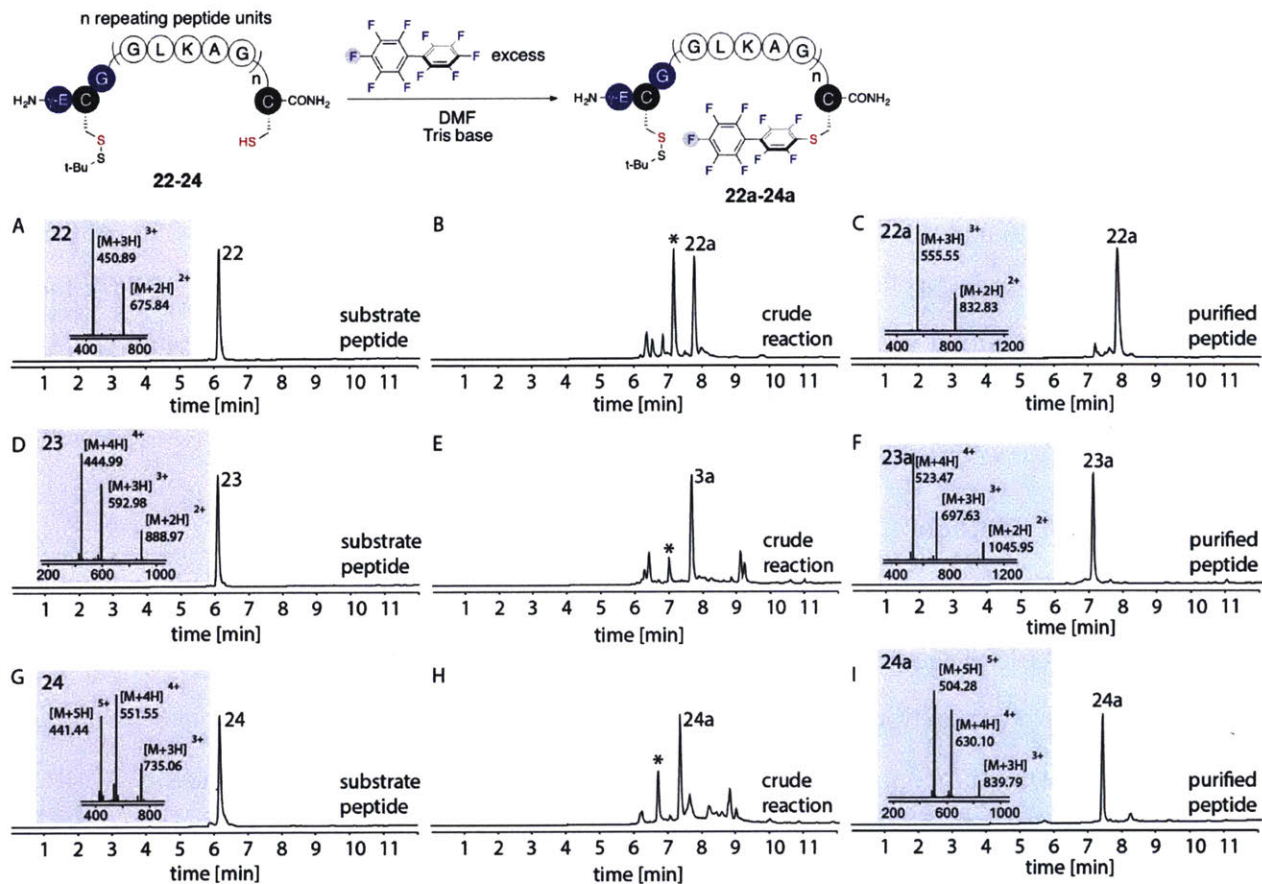
Figure 2.40. Synthesis and purification of peptides 22a-24a.

Shown are three series of chromatograms, each series consists of TIC chromatograms for LC-MS analysis of the pure substrate peptide, crude reaction mixture, and purified perfluoroaryl-linked peptide. Chromatograms A, B, and C are for peptide **22a**; D, E, and F are for peptide **3a**; G, H, and I are for peptide **4a**. Peaks labeled with "*" denote side products with two peptides crosslinked by one decafluorobiphenyl molecule.

2.4.9. General protocol for GST-catalyzed peptide conjugation reactions

Typical GST-catalyzed reactions were performed on a 10 μL scale. Peptide substrates were dissolved in water to make 10 mM stock solutions. Then substrates were mixed with 2 μL of GST stock solution (10 mg/mL), 1 μL reaction buffer (1 M phosphate, 200 mM TCEP, pH 8.0), then water was added to a total volume of 10 μL. Unless otherwise noted, reactions were incubated at 37 °C water bath, quenched by addition of 90 μL of 50% A: 50% B, and subjected to LC-MS analysis.

2.4.10. General protocol for GST-catalyzed peptide macrocyclization

Macrocyclization reactions described in Figure 2.14 and Figure 2.20 were performed on a 100 μL scale. 10 μL of the peptide stock solution (1 mM in water), 2 μL of GST stock solution (10 mg/mL), 10 μL of reaction buffer stock (1M phosphate, 200 mM TCEP•HCl, pH 8.0), and 78 μL of water were combined in a vial (total volume of 100 μL). The reaction mixture was pipetted up and down for 20 times to ensure thorough mixing and left at room temperature (25 °C). 10 μL of the reaction mixture was quenched by addition of 100 μL of 50% A: 50% B, and subjected to LC-MS analysis.

2.4.11. Synthesis and purification of a forty-residue peptide for macrocyclization

Peptide **27** was synthesized *via* hydrazide-based native chemical ligation protocol.[98] To a solution of peptide **25** (2.17 mg, 0.92 μmol) dissolved in 210 μL of 200 mM phosphate buffer, pH 3.0 at -10 °C was slowly added 20 μL of 100 mM $NaNO_2$ in water and reacted for 20 min at the same temperature. 230 μL of 200 mM sodium 2-sulfanylenthanesulfonate in 200 mM phosphate buffer, pH 7.0 was then added to the reaction mixture. The pH of the solution was adjusted to 7.0 by addition of 5 M sodium hydroxide solution. The reaction mixture was left at room temperature for 10 minutes followed by the addition of 460 μL of 2 mM peptide **26a** (*S*-perfluoroarylated peptide **26**) solution and 100 μL of buffer (2 M phosphate, 200 mM TCEP•HCl, pH 7.0). The reaction was left at room temperature and monitored by LC-MS analysis. After ligation for two hours, the product peptide **27** was purified by RP-HPLC (Figure 2.1).
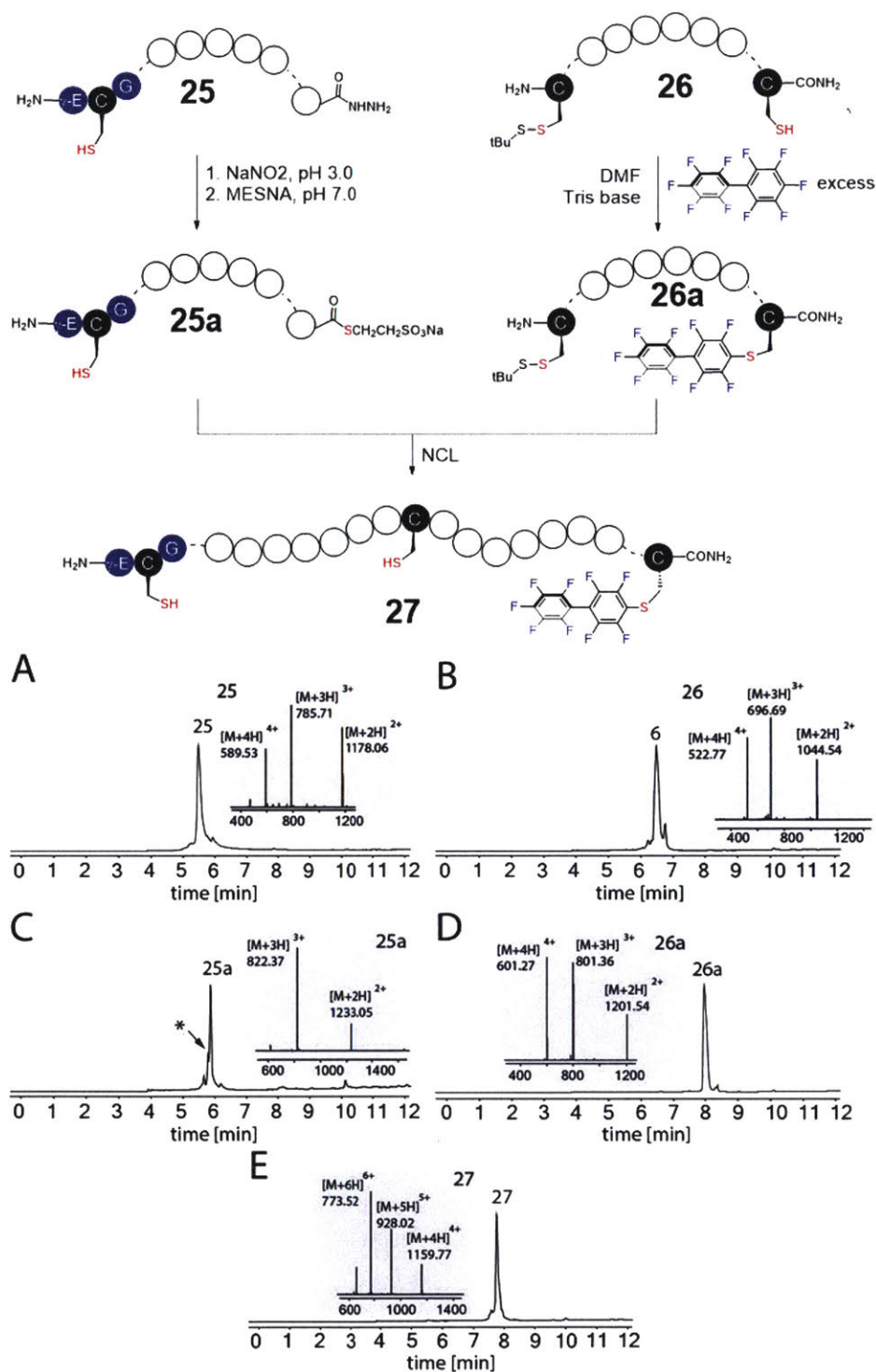
Figure 2.41. Synthesis and purification of peptide 27.

TIC chromatograms for LC-MS analysis of crude peptide thioester **25a** and RP-HPLC purified peptides **25**, **26**, **26a** and **27**. Peak labeled with "*" denotes thiolactone formed at the GSH cysteine from peptide **25a**. Reaction conditions are described in 2.4.10 of the Experimental section.

2.5. Acknowledgements

2.6. References

(1)   Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chemie Int. Ed.* **2005**, *44*, 7342–7372.

(2)   Davis, B. G. Mimicking Posttranslational Modifications of Proteins. *Science.* **2004**, *303*, 480–482.

(3)   Marchetti, L.; Levine, M. Biomimetic Catalysis. *ACS Catal.* **2011**, *1*, 1090–1118.

(4)   Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004–2021.

(5)   Jewett, J. C.; Bertozzi, C. R. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chem. Soc. Rev.* **2010**, *39*, 1272–1279.

(6)   Moses, J. E.; Moorhouse, A. D. The Growing Applications of Click Chemistry. *Chem. Soc. Rev.* **2007**, *36*, 1249–1262.

(7)   Becer, C. R.; Hoogenboom, R.; Schubert, U. S. Click Chemistry beyond Metal-Catalyzed Cycloaddition. *Angew. Chem., Int. Ed.* **2009**, *48*, 4900–4908.

(8)   Blackman, M. L.; Royzen, M.; Fox, J. M. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.

(9)   Hoyle, C. E.; Bowman, C. N. Thiol-Ene Click Chemistry. *Angew. Chem., Int. Ed.* **2010**, *49*, 1540–1573.

(10)   Wong, C.-H.; Zimmerman, S. C. Orthogonality in Organic, Polymer, and Supramolecular Chemistry: From Merrifield to Click Chemistry. *Chem. Commun.* **2013**, *49*, 1679–1695.

(11)   Los, G. V; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3*, 373–382.

(12)   Watanabe, S.; Mizukami, S.; Hori, Y.; Kikuchi, K. Multicolor Protein Labeling in Living Cells Using Mutant Œ≤-Lactamase-Tag Technology. *Bioconjugate Chem.* **2010**, *21*, 2320–2326.

(13)   George, N.; Pick, H.; Vogel, H.; Johnsson, N.; Johnsson, K. Specific Labeling of Cell Surface Proteins with Chemically Diverse Compounds. *J. Am. Chem. Soc.* **2004**, *126*, 8896–8897.

(14)   Chen, I.; Howarth, M.; Lin, W.; Ting, A. Y. Site-Specific Labeling of Cell Surface Proteins with Biophysical Probes Using Biotin Ligase. *Nat. Methods* **2005**, *2*, 99–104.

(15)   Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernández-Suárez, M.; Puthenveetil, S.; Ting, A. Y. A Fluorophore Ligase for Site-Specific Protein Labeling inside Living Cells. *Proc. Natl. Acad. Sci.* **2010**, *107*, 10914–10919.

(16)   Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A General

Method for the Covalent Labeling of Fusion Proteins with Small Molecules in Vivo. *Nat. Biotechnol.* **2003**, *21*, 86–89.

(17)   Witus, L. S.; Francis, M. Site-Specific Protein Bioconjugation via a Pyridoxal 5'-Phosphate-Mediated N-Terminal Transamination Reaction. *Curr Protoc Chem Biol* **2010**, *2*, 125–134.

(18)   Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: A Versatile Method for Protein Labeling. *Nat. Chem. Biol.* **2007**, *3*, 707–708.

(19)   Babtie, A.; Tokuriki, N.; Hollfelder, F. What Makes an Enzyme Promiscuous? *Curr. Opin. Chem. Biol.* **2010**, *14*, 200–207.

(20)   Humble, M. S.; Berglund, P. Biocatalytic Promiscuity. *Eur. J. Org. Chem.* **2011**, *2011*, 3391–3401.

(21)   Khersonsky, O.; Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **2010**, *79*, 471–505.

(22)   Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 138–174.

(23)   Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H. Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013**, *339*, 307–310.

(24)   Craik, D. J.; Cemazar, M.; Wang, C. K. L.; Daly, N. L. The Cyclotide Family of Circular Miniproteins: Nature's Combinatorial Peptide Template. *Biopolymers* **2006**, *84*, 250–266.

(25)   Baeriswyl, V.; Heinis, C. Polycyclic Peptide Therapeutics. *ChemMedChem* **2013**, *8*, 377–384.

(26)   Smith, J. M.; Frost, J. R.; Fasan, R. Emerging Strategies to Access Peptide Macrocycles from Genetically Encoded Polypeptides. *J. Org. Chem.* **2013**, *78*, 3525–3531.

(27)   Aboye, T. L.; Camarero, J. A. Biological Synthesis of Circular Polypeptides. *J. Biol. Chem.* **2012**, *287*, 27026–27032.

(28)   Gruenewald, J.; Marahiel, M. A. Chemoenzymatic and Template-Directed Synthesis of Bioactive Macrocyclic Peptides. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 121–146.

(29)   Jackson, D. Y.; Burnier, J. P.; Wells, J. A. Enzymic Cyclization of Linear Peptide Esters Using Subtiligase. *J. Am. Chem. Soc.* **1995**, *117*, 819–820.

(30)   Kohli, R. M.; Walsh, C. T.; Burkart, M. D. Biomimetic Synthesis and Optimization of Cyclic Peptide Antibiotics. *Nature* **2002**, *418*, 658–661.

(31)   Trauger, J. W.; Kohli, R. M.; Mootz, H. D.; Marahiel, M. A.; Walsh, C. T. Peptide Cyclization Catalysed by the Thioesterase Domain of Tyrocidine Synthetase. *Nature* **2000**, *407*, 215–218.

(32)   Wu, Z.; Guo, X.; Guo, Z. Sortase A-Catalyzed Peptide Cyclization for the Synthesis of Macrocyclic Peptides and Glycopeptides. *Chem. Commun.* **2011**, *47*, 9218–9220.

(33) Bolscher, J. G. M.; Oudhoff, M. J.; Nazmi, K.; Antos, J. M.; Guimaraes, C. P.; Spooner, E.; Haney, E. F.; Garcia Vallejo, J. J.; Vogel, H. J.; van't Hof, W.; *et al.* Sortase A as a Tool for High-Yield Histatin Cyclization. *FASEB J.* **2011**, *25*, 2650–2658, 10-182212.

(34) Antos, J. M.; Popp, M. W.-L.; Ernst, R.; Chew, G.-L.; Spooner, E.; Ploegh, H. L. A Straight Path to Circular Proteins. *J. Biol. Chem.* **2009**, *284*, 16028–16036.

(35) Dasgupta, S.; Samantaray, S.; Sahal, D.; Roy, R. P. Isopeptide Ligation Catalyzed by Quintessential Sortase A: Mechanistic Cues from Cyclic and Branched Oligomers of Indolicidin. *J. Biol. Chem.* **2011**, *286*, 23996–24006.

(36) Jia, X.; Kwon, S.; Wang, C.-I. A.; Huang, Y.-H.; Chan, L. Y.; Tan, C. C.; Rosengren, K. J.; Mulvenna, J. P.; Schroeder, C. I.; Craik, D. J. Semienzymatic Cyclization of Disulfide-Rich Peptides Using Sortase A. *J. Biol. Chem.* **2014**, *289*, 6627–6638.

(37) *Bioconjugate Techniques (Third edition)*; Hermanson, G. T., Academic Press: Boston, **2015**.

(38) Carrico, I. S. Chemoselective Modification of Proteins: Hitting the Target. *Chem. Soc. Rev.* **2008**, *37*, 1423–1431.

(39) Antos, J. M.; Francis, M. B. Transition Metal Catalyzed Methods for Site-Selective Protein Modification. *Curr. Opin. Chem. Biol.* **2006**, *10*, 253–262.

(40) Hackenberger, C. P. R.; Schwarzer, D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chemie Int. Ed.* **2008**, *47*, 10030–10074.

(41) Basle, E.; Joubert, N.; Pucheault, M. Protein Chemical Modification on Endogenous Amino Acids. *Chem. Biol.* **2010**, *17*, 213–227.

(42) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266*, 776–779.

(43) Sato, S.; Nakamura, H. Ligand-Directed Selective Protein Modification Based on Local Single-Electron-Transfer Catalysis. *Angew. Chemie Int. Ed.* **2013**, *52*, 8681–8684.

(44) Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L. Protein Thioester Synthesis Enabled by Sortase. *J. Am. Chem. Soc.* **2012**, *134*, 10749–10752.

(45) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. – An Asian J.* **2009**, *4*, 630–640.

(46) Crankshaw, M. W.; Grant, G. A. *Modification at Cysteine*; Wiley, 1996.

(47) Fodje, M. N.; Al-Karadaghi, S. Occurrence, Conformational Features and Amino Acid Propensities for the π-Helix. *Protein Eng.* **2002**, *15*, 353–358.

(48) Weerapana, E.; Simon, G. M.; Cravatt, B. F. Disparate Proteome Reactivity Profiles of Carbon Electrophiles. *Nat. Chem. Biol.* **2008**, *4*, 405–407.

(49) van Kasteren, S. I.; Kramer, H. B.; Jensen, H. H.; Campbell, S. J.; Kirkpatrick, J.; Oldham, N. J.; Anthony, D. C.; Davis, B. G. Expanding the Diversity of Chemical Protein Modification Allows Post-Translational Mimicry. *Nature* **2007**, *446*, 1105–1109.

(50) Yan, L. Z.; Dawson, P. E. Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001**, *123*, 526–533.

(51) Crich, D.; Banerjee, A. Native Chemical Ligation at Phenylalanine. *J. Am. Chem. Soc.* **2007**, *129*, 10064–10065.

(52) Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem., Int. Ed.* **2007**, *46*, 9248–9252.

(53) Haase, C.; Rohde, H.; Seitz, O. Native Chemical Ligation at Valine. *Angew. Chem., Int. Ed.* **2008**, *47*, 6807–6810.

(54) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angew. Chem., Int. Ed.* **2008**, *47*, 8521–8524.

(55) Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053.

(56) White, C. J.; Yudin, A. K. Contemporary Strategies for Peptide Macrocyclization. *Nat. Chem.* **2011**, *3*, 509–524.

(57) Moretto, A.; Crisma, M.; Formaggio, F.; Toniolo, C. Building a Bridge between Peptide Chemistry and Organic Chemistry: Intramolecular Macrocyclization Reactions and Supramolecular Chemistry with Helical Peptide Substrates. *Biopolymers* **2010**, *94*, 721–732.

(58) Davies, J. S. The Cyclization of Peptides and Depsipeptides. *J. Pept. Sci.* **2003**, *9*, 471–501.

(59) Jiang, S.; Li, Z.; Ding, K.; Roller, P. P. Recent Progress of Synthetic Studies to Peptide and Peptidomimetic Cyclization. *Curr. Org. Chem.* **2008**, *12*, 1502–1542.

(60) Lambert, J. N.; Mitchell, J. P.; Roberts, K. D. The Synthesis of Cyclic Peptides. *J. Chem. Soc., Perkin Trans. 1* **2001**, 471–484.

(61) Montalbetti, C. A. G. N.; Falque, V. Amide Bond Formation and Peptide Coupling. *Tetrahedron* **2005**, *61*, 10827–10852.

(62) Parenty, A.; Moreau, X.; Campagne, J. M. Macrolactonizations in the Total Synthesis of Natural Products. *Chem. Rev.* **2006**, *106*, 911–939.

(63) Timmerman, P.; Beld, J.; Puijk, W. C.; Meloen, R. H. Rapid and Quantitative Cyclization of Multiple Peptide Loops onto Synthetic Scaffolds for Structural Mimicry of Protein Surfaces. *Chembiochem* **2005**, *6*, 821–824.

(64) Smeenk, L. E. J.; Dailly, N.; Hiemstra, H.; van Maarseveen, J. H.; Timmerman, P. Synthesis of Water-Soluble Scaffolds for Peptide Cyclization, Labeling, and Ligation. *Org. Lett.* **2012**, *14*, 1194–1197.

(65) Muppidi, A.; Doi, K.; Edwardraja, S.; Drake, E. J.; Gulick, A. M.; Wang, H.-G.; Lin, Q.

Rational Design of Proteolytically Stable, Cell-Permeable Peptide-Based Selective Mcl-1 Inhibitors. *J. Am. Chem. Soc.* **2012**, *134*, 14734–14737.

(66) Heinis, C.; Rutherford, T.; Freund, S.; Winter, G. Phage-Encoded Combinatorial Chemical Libraries Based on Bicyclic Peptides. *Nat. Chem. Biol.* **2009**, *5*, 502–507.

(67) Jo, H.; Meinhardt, N.; Wu, Y.; Kulkarni, S.; Hu, X.; Low, K. E.; Davies, P. L.; DeGrado, W. F.; Greenbaum, D. C. Development of α-Helical Calpain Probes by Mimicking a Natural Protein-Protein Interaction. *J. Am. Chem. Soc.* **2012**, *134*, 17704–17713.

(68) Besse, D.; Moroder, L. Synthesis of Selenocysteine Peptides and Their Oxidation to Diselenide-Bridged Compounds. *J. Pept. Sci.* **1997**, *3*, 442–453.

(69) Camarero, J. A.; Muir, T. W. Chemoselective Backbone Cyclization of Unprotected Peptides. *Chem. Commun.* **1997**, 1369–1370.

(70) Tulla-Puche, J.; Barany, G. On-Resin Native Chemical Ligation for Cyclic Peptide Synthesis. *J. Org. Chem.* **2004**, *69*, 4101–4107.

(71) Tam, J. P.; Lu, Y.-A.; Yu, Q. Thia Zip Reaction for Synthesis of Large Cyclic Peptides: Mechanisms and Applications. *J. Am. Chem. Soc.* **1999**, *121*, 4316–4324.

(72) Shao, Y.; Lu, W.; Kent, S. B. H. A Novel Method to Synthesize Cyclic Peptides. *Tetrahedron Lett.* **1998**, *39*, 3911–3914.

(73) Tavassoli, A.; Benkovic, S. J. Split-Intein Mediated Circular Ligation Used in the Synthesis of Cyclic Peptide Libraries in E. Coli. *Nat. Protoc.* **2007**, *2*, 1126–1133.

(74) Smith, J. M.; Vitali, F.; Archer, S. A.; Fasan, R. Modular Assembly of Macrocyclic Organo-Peptide Hybrids Using Synthetic and Genetically Encoded Precursors. *Angew. Chem., Int. Ed.* **2011**, *50*, 5075–5080.

(75) Kleineweischede, R.; Hackenberger, C. P. R. Chemoselective Peptide Cyclization by Traceless Staudinger Ligation. *Angew. Chem., Int. Ed.* **2008**, *47*, 5984–5988, S5984/1–S5984/10.

(76) Kim, Y.-W.; Grossmann, T. N.; Verdine, G. L. Synthesis of All-Hydrocarbon Stapled α-Helical Peptides by Ring-Closing Olefin Metathesis. *Nat. Protoc.* **2011**, *6*, 761–771.

(77) Blackwell, H. E.; Grubbs, R. H. Highly Efficient Synthesis of Covalently Cross-Linked Peptide Helices by Ring-Closing Metathesis. *Angew. Chem., Int. Ed.* **1998**, *37*, 3281–3284.

(78) Jewett, J. C.; Bertozzi, C. R.; Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B.; Tornøe, C. W.; Christensen, C.; Meldal, M.; Tornøe, C. W.; et al. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chem. Soc. Rev.* **2010**, *39*, 1272.

(79) Bock, V. D.; Perciaccante, R.; Jansen, T. P.; Hiemstra, H.; Van Maarseveen, J. H. Click Chemistry as a Route to Cyclic Tetrapeptide Analogs: Synthesis of Cyclo-[Pro-Val-Ψ(triazole)-Pro-Tyr]. *Org. Lett.* **2006**, *8*, 919–922.

(80) Turner, R. A.; Oliver, A. G.; Lokey, R. S. Click Chemistry as a Macrocyclization Tool in the Solid-Phase Synthesis of Small Cyclic Peptides. *Org. Lett.* **2007**, *9*, 5011–5014.

(81) Madden, M. M.; Muppidi, A.; Li, Z.; Li, X.; Chen, J.; Lin, Q. Synthesis of Cell-Permeable Stapled Peptide Dual Inhibitors of the p53-Mdm2/Mdmx Interactions via Photoinduced Cycloaddition. *Bioorganic Med. Chem. Lett.* **2011**, *21*, 1472–1475.

(82) Ghadiri, M. R.; Choi, C. Secondary Structure Nucleation in Peptides. Transition Metal Ion Stabilized α-Helices. *J. Am. Chem. Soc.* **1990**, *112*, 1630–1632.

(83) Smith, S. J.; Du, K.; Radford, R. J.; Tezcan, F. A. Functional, Metal-Based Crosslinkers for α-Helix Induction in Short Peptides. *Chem. Sci.* **2013**, *4*, 3740–3747.

(84) Marqusee, S.; Baldwin, R. L. Helix Stabilization by Glu-···Lys+ Salt Bridges in Short Peptides of de Novo Design. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 8898–8902.

(85) Dolain, C.; Hatakeyama, Y.; Sawada, T.; Tashiro, S.; Fujita, M. Inducing α-Helices in Short Oligopeptides through Binding by an Artificial Hydrophobic Cavity. *J. Am. Chem. Soc.* **2010**, *132*, 5564–5565.

(86) Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.

(87) Zou, Y.; Spokoyny, A. M.; Zhang, C.; Simon, M. D.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. Convergent Diversity-Oriented Side-Chain Macrocyclization Scan for Unprotected Polypeptides. *Org. Biomol. Chem.* **2014**, *12*, 566–573.

(88) Becer, C. R.; Babiuch, K.; Pilz, D.; Hornig, S.; Heinze, T.; Gottschaldt, M.; Schubert, U. S. Clicking Pentafluorostyrene Copolymers: Synthesis, Nanoprecipitation, and Glycosylation. *Macromolecules* **2009**, *42*, 2387–2394.

(89) Hayes, J. D.; Flanagan, J. U.; Jowsey, I. R. Glutathione Transferases. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 51–88, 1 plate.

(90) Armstrong, R. N. Structure, Catalytic Mechanism, and Evolution of the Glutathione Transferases. *Chem. Res. Toxicol.* **1997**, *10*, 2–18.

(91) Suzuki, T.; Ota, Y.; Kasuya, Y.; Mutsuga, M.; Kawamura, Y.; Tsumoto, H.; Nakagawa, H.; Finn, M. G.; Miyata, N. An Unexpected Example of Protein-Templated Click Chemistry. *Angew. Chem., Int. Ed.* **2010**, *49*, 6817–6820, S6817/1–S6817/10.

(92) Dirksen, A.; Dawson, P. E. Rapid Oxime and Hydrazone Ligations with Aromatic Aldehydes for Biomolecular Labeling. *Bioconjugate Chem.* **2008**, *19*, 2543–2548.

(93) Kurpiers, T.; Mootz, H. D. Regioselective Cysteine Bioconjugation by Appending a Labeled Cysteine Tag to a Protein by Using Protein Splicing in Trans. *Angew. Chem., Int. Ed.* **2007**, *46*, 5234–5237.

(94) Nathani, R. I.; Moody, P.; Chudasama, V.; Smith, M. E. B.; Fitzmaurice, R. J.; Caddick, S. A Novel Approach to the Site-Selective Dual Labelling of a Protein via Chemoselective Cysteine Modification. *Chem. Sci.* **2013**, *4*, 3455–3458.

(95) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The Exploration of Macrocycles for Drug Discovery - an Underexploited Structural Class. *Nat. Rev. Drug Discov.* **2008**, *7*, 608–624.

(96)   Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; *et al.* Rapid Flow-Based Peptide Synthesis. *Chembiochem* **2014**, *15*, 713–720.

(97)   Mong, S. K.; Vinogradov, A. A.; Simon, M. D.; Pentelute, B. L. Rapid Total Synthesis of DARPin pE59 and Barnase. *Chembiochem* **2014**, *15*, 721–733.

(98)   Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protoc.* **2013**, *8*, 2483–2495.

(99)   Adang, A. E. P.; Brussee, J.; Meyer, D. J.; Coles, B.; Ketterer, B.; Van, der G. A.; Mulder, G. J. Substrate Specificity of Rat Liver Glutathione S-Transferase Isoenzymes for a Series of Glutathione Analogs, Modified at the γ -Glutamyl Moiety. *Biochem. J.* **1988**, *255*, 721–724.

(100) Adang, A. E. P.; Brussee, J.; Van, der G. A.; Mulder, G. J. The Glutathione-Binding Site in Glutathione S-Transferases. Investigation of the Cysteinyl, Glycyl and γ-Glutamyl Domains. *Biochem. J.* **1990**, *269*, 47–54.

Chapter 3. π-Clamp-Mediated Regioselective Cysteine Arylation

The work presented in this chapter is a part of the following manuscripts and is reproduced here with permission from the Nature Publication Group and Macmillan Publishers Limited.

3.1. Introduction

Site-selective chemistry[1–5] is essential for creating homogeneously modified biologics[6,7], studying protein structure and function[8], generating materials with defined composition[9], and on-demand modification of complex small molecules[10,11]. Existing approaches for site-selective chemistry utilize either reaction pairs that are orthogonal to other functional groups on the target of interest (Figure 2.4a, strategy 1)[12,13] or catalysts that mediate selective reactions at a particular site among many competing ones (Figure 2.4A, strategy 2)[14–19]. These strategies have been widely used in protein modification and have led to the development of multiple bio-orthogonal handles[20–25] and enzyme-tag pairs[26–31].

Natural proteins precisely control selective reactions and interactions by building large three-dimensional structures from polypeptides usually much greater than 100 residues.[32] For example, enzymes have folded structures where particular amino acids are placed in a specialized active-site environment.[33] Inspired by this, we envisioned a new strategy for site-selective chemistry on proteins by fine-tuning the local environment around an amino acid residue in a small peptide sequence (Figure 2.4b). This is challenging because peptides are highly dynamic and unstructured thereby presenting a formidable challenge to build defined environments for selective chemical transformations.

Our design efforts leveraged cysteine because Nature has shown its robust catalytic role in enzymes,[34,35] and prior efforts indicate the reactivity of a cysteine residue can vary in different protein environments.[36] Further, cysteine is the first choice in bioconjugation to modify proteins often via maleimide ligation or alkylation.[37,38] However, these traditional cysteine-based bioconjugations are significantly limited because they are not site-specific. When these methods are applied to protein targets with multiple cysteine residues, a mixture of heterogeneous products are generated, as exemplified by recent efforts to conjugate small molecule drugs to antibodies through cysteine-based reactions.[39]

Small peptide tags that contain multiple cysteine residues have been used for bioconjugation. Tsien have developed biarsenic reagents that selectively react with tetra-cysteine motifs in peptides and proteins.[40,41] More recently, organic arsenics have been used to modify two cysteine residues generated from reducing a disulfide bond.[42] These methods can present challenges with thiol selectivity[43] and none report the site-specific modification of one cysteine

residue in the presence of many as enzymes or multiple chemical steps must be used to accomplish this feat.[44,45] An enzyme-free and one-step method for site-selective cysteine conjugation has yet to be developed.

We have previously described a perfluoroaryl-cysteine S_NAr approach for peptide and protein modifications.[46–49] The reactions between perfluoroaryl groups and cysteine residues are fast in organic solvent but extremely sluggish in water unless an enzyme is used.[47,48] This observation inspired us to develop small peptides to promote the S_NAr reaction in an analogous fashion to enzymes.

a. Existing strategies for site-selective chemistry

b. This work: site-selective chemistry by tuning the local environment

c. π-Clamp
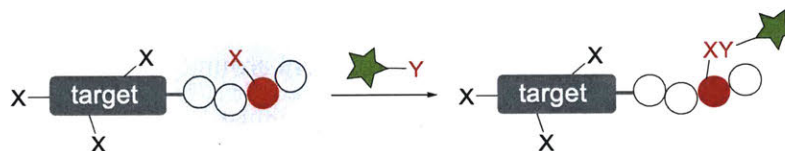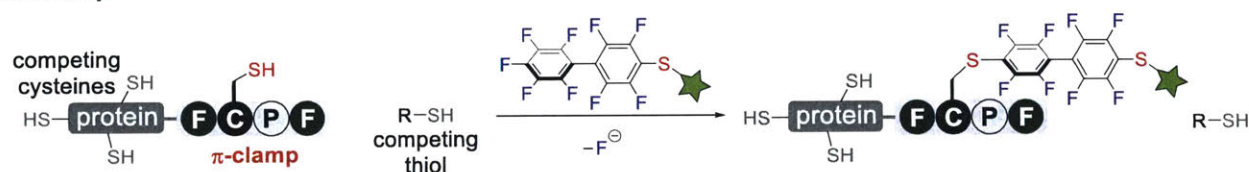
Figure 3.1. π-clamp mediated cysteine conjugation as a new strategy for site-selective chemistry.

(**a**) Existing strategies for site-selective chemistry. Strategy 1: selectivity arises from orthogonal chemistry between site Z and reagent Y. Strategy 2: catalyst mediates the reaction between a particular site X (highlighted in red) and reagent Y. (**b**) This work demonstrates a new strategy for site-selective chemistry by fine-tuning the local chemical environment around the target site. A particular site X (highlighted in red) is tuned to react with reagent Y in the presence of other competing X sites. (**c**) Cysteine residue inside the π-clamp selectively reacts with perfluoroaromatic probes in the presence of other competing cysteine residues and thiol species.

3.2. Results

Here we describe the identification of the π-clamp sequence to mediate site-specific cysteine modification in water without an enzyme, which overcomes the selectivity challenge for cysteine bioconjugation (Figure 2.4c). This offers a new mode for site-specific chemistry by fine-tuning the microenvironment of a four-residue stretch within a complex protein or peptide.

3.1.1. Design and discovery of the π-clamp

We serendipitously discovered the π-clamp via a library selection approach designed to identify peptide sequences that promote arylation reactions in water. To accomplish this, we prepared a peptide library (Zaa-Cys-Zaa-Zaa-Gly-Leu-Leu-Lys, where Zaa is anyone of the 20 natural amino acids except cysteine) and reacted the library with a biotin-perfluoroaryl probe in solution (Figure 3.2). Through streptavidin pull-down and following LC-MS/MS analysis to identify reaction products (see Experimental section 3.4.7), we uncovered the sequence Phe-Cys-Pro-Trp reacted with the perfluoroaryl-cysteine moiety in water (Figure 3.3). This observation is in stark contrast to our prior efforts[47] which showed that cysteine residues and perfluoroaryl moieties do not react in water. Thus the Phe-Cys-Pro-Trp sequence appears to modify the reactivity of the cysteine thiol. Further mutating the Phe and Trp to Gly eliminated the reaction. Based on these findings and a molecular model of Phe-Cys-Pro-Trp, we hypothesize that the Phe and Trp side chains activate the cysteine thiol and interact with the incoming perfluoroaryl group, while the Pro serves to position the Cys, Phe, and Trp residues into a conformation that promotes the reaction. We refer to this distinctive amino acid sequence Xaa-Cys-Pro-Xaa (Xaa = electron-rich aromatic amino acids including Phe, Trp, or Tyr) as a π-clamp.
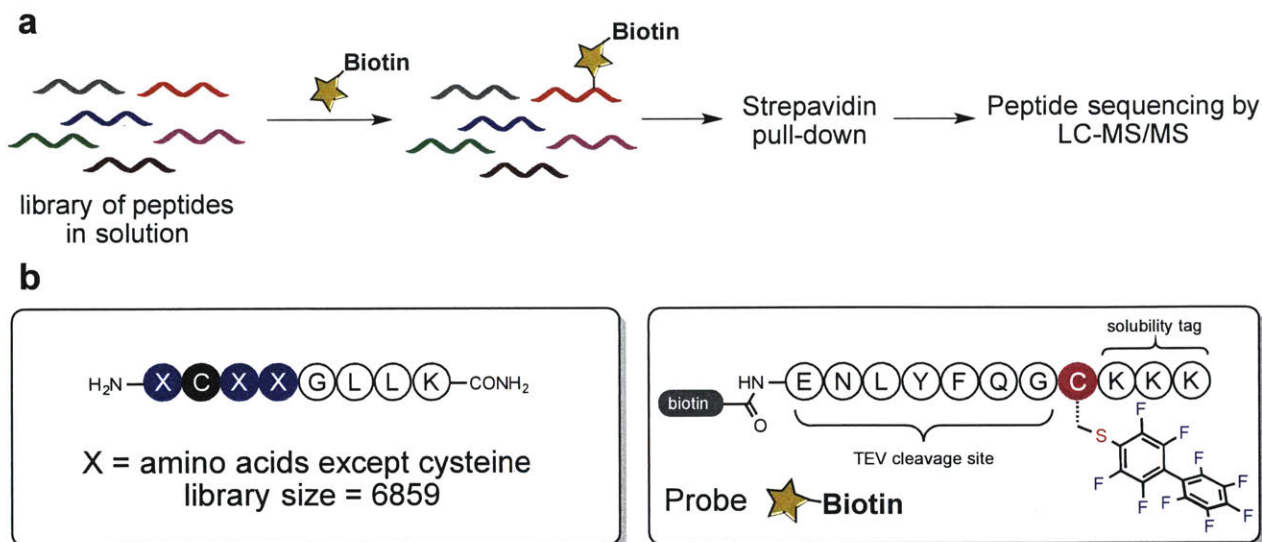
Figure 3.2. A library selection approach for the identification of the π-clamp sequence.

(a) Our strategy combines the streptavidin pull-down and the peptide sequencing by LC-MS to identify reactive peptide sequences for the arylation reaction. (b) The sequence of the peptide library and the structure of the probe used. The peptide library was synthesized using spilt-pool synthesis. The peptide library was cleaved from the resin and the side chain protecting groups were removed using standard cleavage techniques described in the peptide synthesis section.
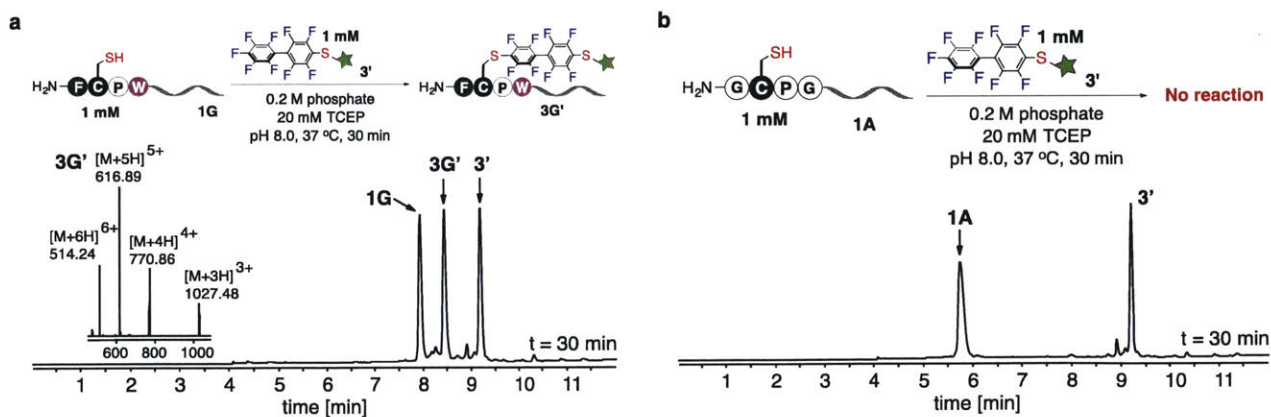
Figure 3.3. Discovery of enhanced arylation reactivity of Phe-Cys-Pro-Trp sequence.

(a) Phe-Cys-Pro-Trp sequence is reactive towards perfluoroaryl probe. (b) Gly-Cys-Pro-Gly sequence showed no product formation under the same reaction conditions. Reaction conditions: 1 mM peptide **1A** or **1G**, 1 mM perfluoroaryl probe **3'**, 0.2 M phosphate and 20 mM TCEP (pH 8.0), at 37 °C for 30 minutes.

178

3.1.2. Mutation studies of the π-clamp sequence

To investigate the π-clamp mediated conjugation, we mutated the aromatic residues. Each of 9 peptides (Xaa-Cys-Pro-Xaa-Gly-Leu-Leu-Lys-Asn-Lys, where Xaa was Phe, Trp, or Tyr) were tested for reaction with a perfluoroaryl-probe (**2**) in 0.2 M phosphate buffer at pH 8.0 and 37 °C with 20 mM TCEP added as the reducing agent. All 9 peptides reacted with probe **2** (rate constants = 0.05 to 0.73 $M^{-1} \cdot S^{-1}$, Figure 3.4, entries 5-13). In contrast, the double glycine mutant (**1A**) formed no product (Figure 3.4, entry 1). The Phe-Phe π-clamp peptide (**1E**) gave quantitative conversion in 30 minutes (rate constants = 0.73 $M^{-1} \cdot S^{-1}$, Figure 3.4, entry 5). Single mutations of each Phe to Gly (**1B** and **1C**, Figure 3.4, entries 2 and 3) or converting the L-Pro to D-Pro (**1D**, Figure 3.4, entry 4) significantly decreased the rate of the arylation reaction. These studies indicate that each amino acid in the π-clamp is essential for product formation.
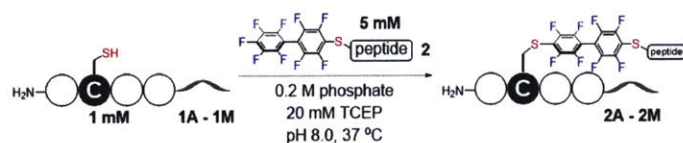
Entry	Peptide	Sequence	$K_2/(M^{-1}s^{-1})$
1	1A	H_2N-GCPGGLLKNK-$CONH_2$	N/A
2	1B	H_2N-FCPGGLLKNK-$CONH_2$	N/A
3	1C	H_2N-GCPFGLLKNK-$CONH_2$	0.089 ± 0.007
4	1D*	H_2N-FC(^{D}P)FGLLKNK-$CONH_2$	0.048 ± 0.002
5	1E	H_2N-FCPFGLLKNK-$CONH_2$	0.73 ± 0.01
6	1F	H_2N-FCPYGLLKNK-$CONH_2$	0.094 ± 0.002
7	1G	H_2N-FCPWGLLKNK-$CONH_2$	0.181 ± 0.005
8	1H	H_2N-YCPFGLLKNK-$CONH_2$	0.315 ± 0.008
9	1I	H_2N-YCPYGLLKNK-$CONH_2$	0.076 ± 0.003
10	1J	H_2N-YCPWGLLKNK-$CONH_2$	0.118 ± 0.005
11	1K	H_2N-WCPFGLLKNK-$CONH_2$	0.475 ± 0.008
12	1L	H_2N-WCPYGLLKNK-$CONH_2$	0.101 ± 0.003
13	1M	H_2N-WCPWGLLKNK-$CONH_2$	0.195 ± 0.002

Figure 3.4. Mutation studies of the π-clamp sequence.

Mutation studies show Phe-Cys-Pro-Phe has the highest reaction rate (entries 5-13) and Phe-1, Pro-3, and Phe-4 are required for its observed reactivity (entries 1 to 5). TCEP: *tris*(2-carboxylethyl)phosphine. Yields shown are from LC-MS analysis of the crude reactions at 30 minutes. Amino acids are shown in one-letter code. *^{D}P stands for D-proline. Kinetics fitting of reactions at different time points are summarized in Figure 3.5. LC-MS chromatograms of these reactions are shown in Figure 3.6 to Figure 3.12.
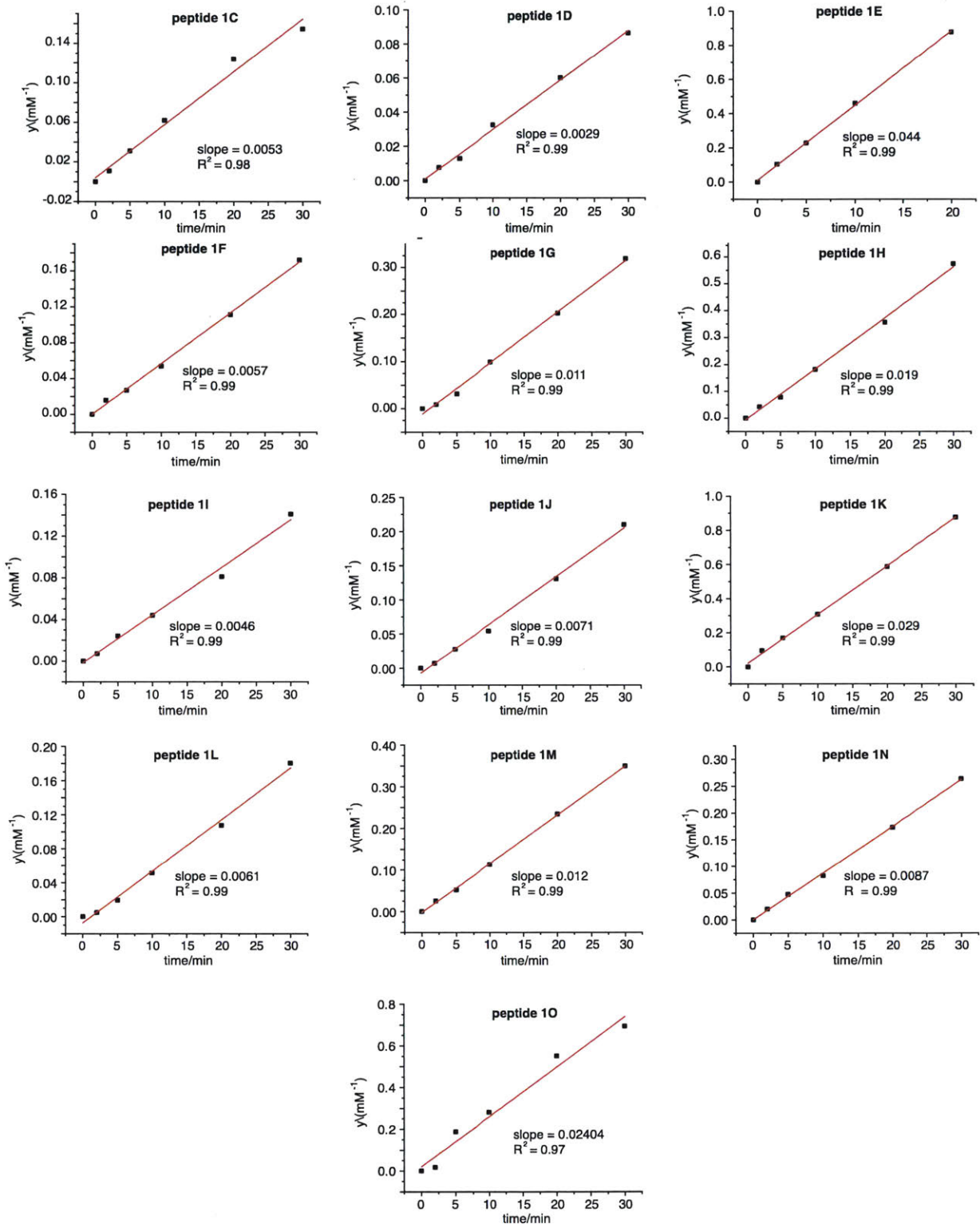
Figure 3.5. Linear fitting of kinetics data for π-clamp peptides and controls to second-order rate equation.

See Experimental section 3.4.11 for detailed method for calculation and fitting.
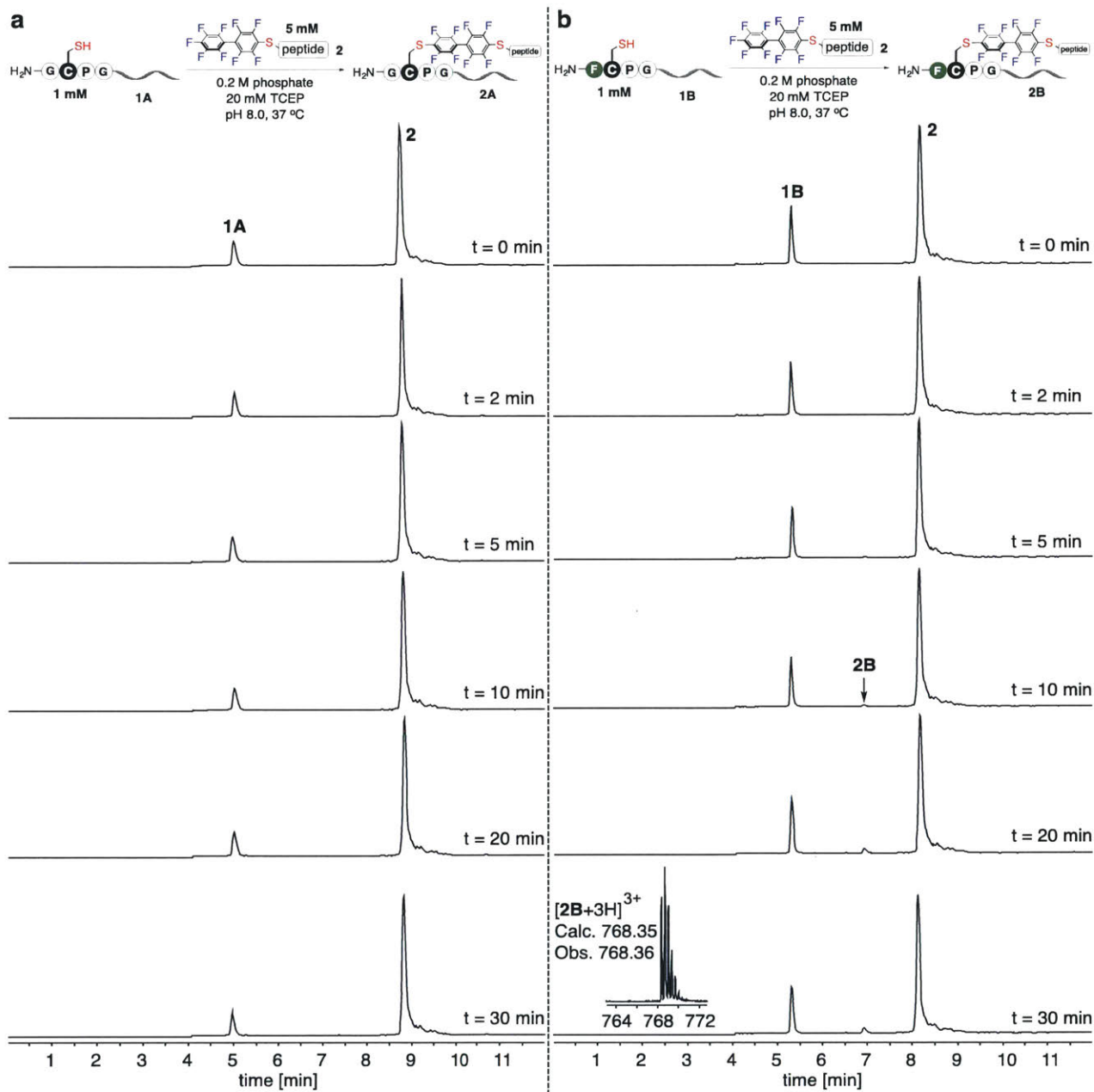
Figure 3.6. LC-MS chromatograms for arylation reactions of peptide 1A (a) and 1B (b) at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
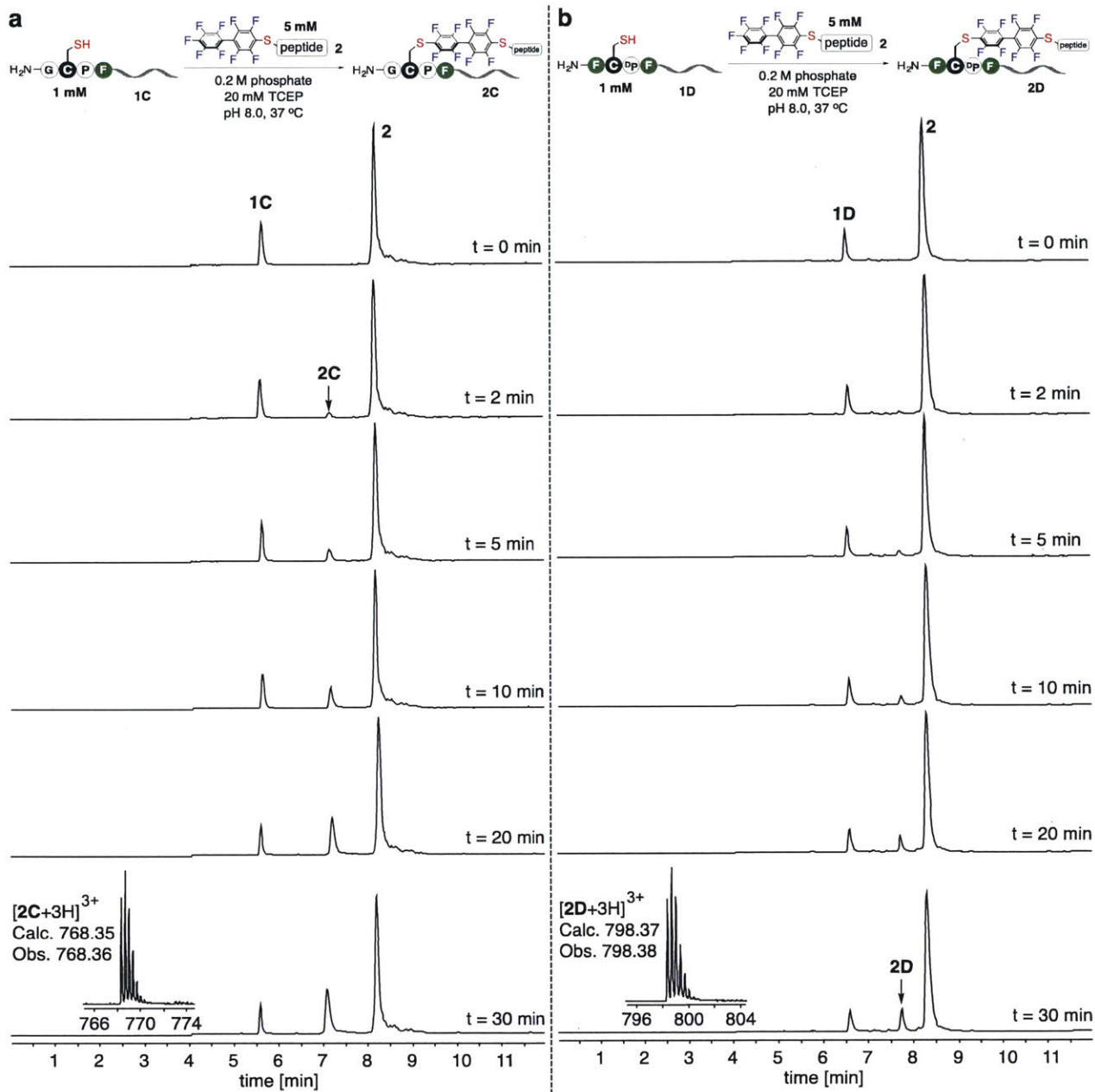
Figure 3.7. LC-MS chromatograms for arylation reactions of peptide 1C (a) and 1D (b) at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
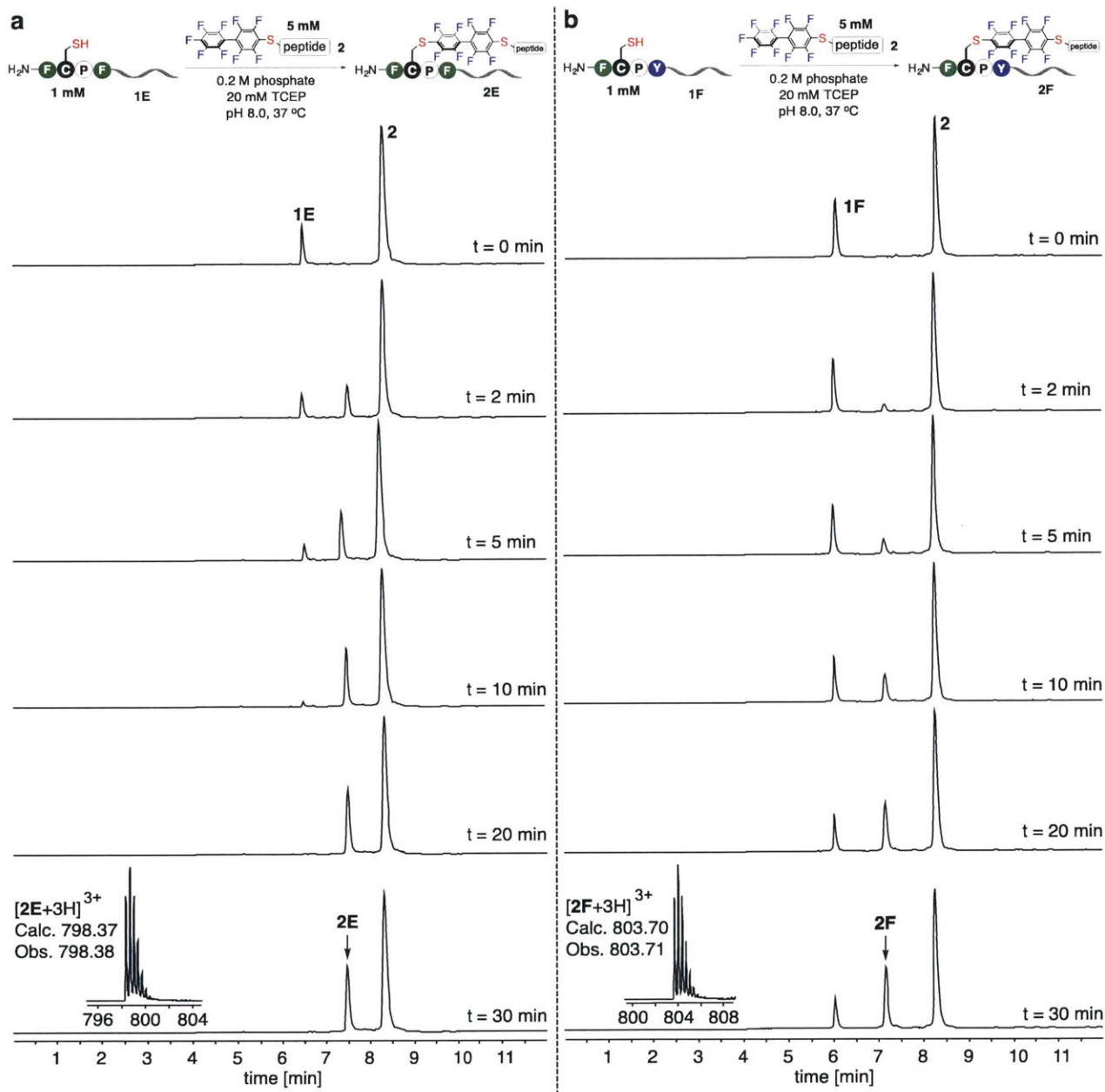
Figure 3.8. LC-MS chromatograms for arylation reactions of peptide 1E (a) and 1F (b) at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
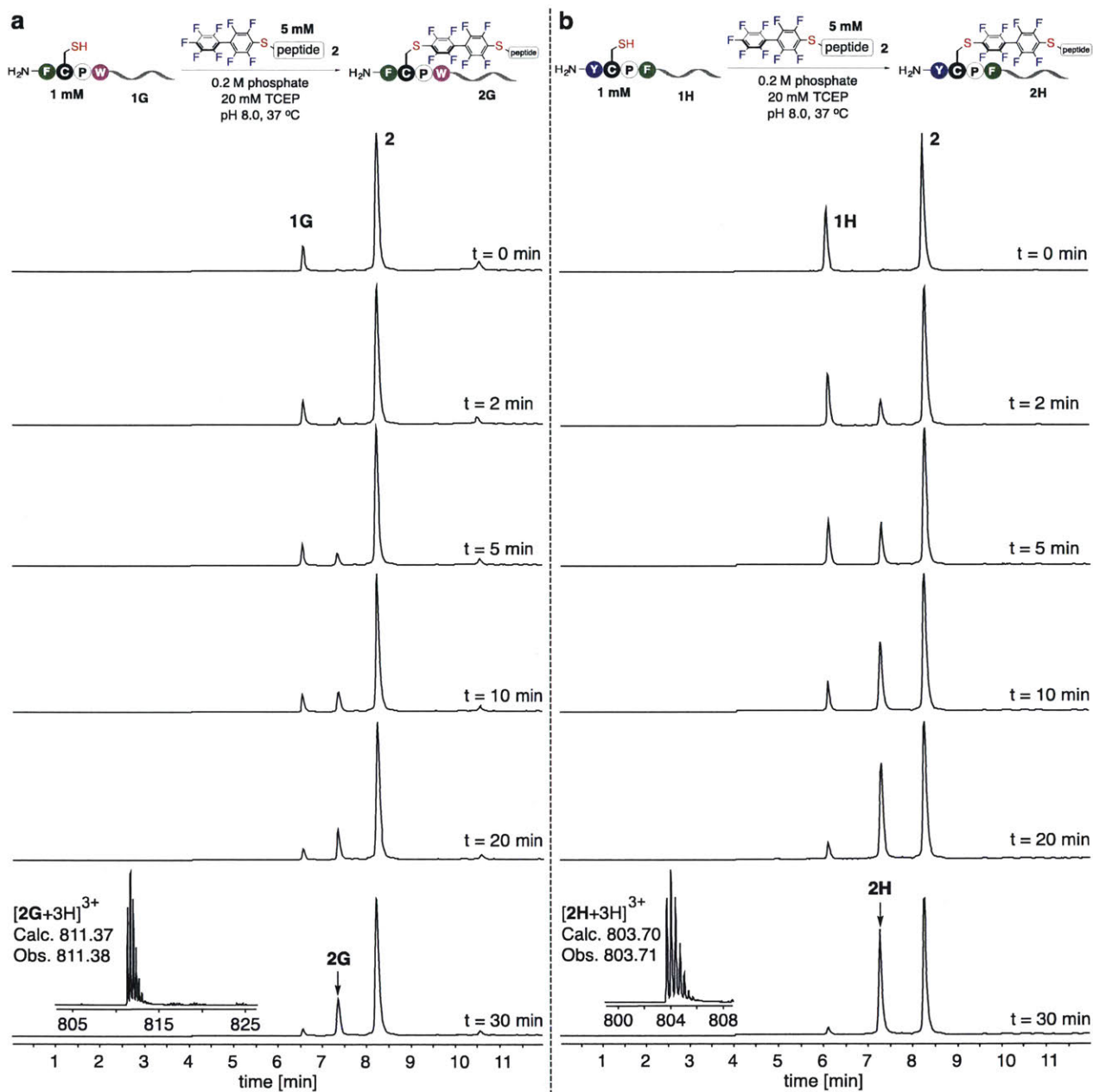
Figure 3.9. LC-MS chromatograms for arylation reactions of peptide 1G (a) and 1H (b) at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.

184
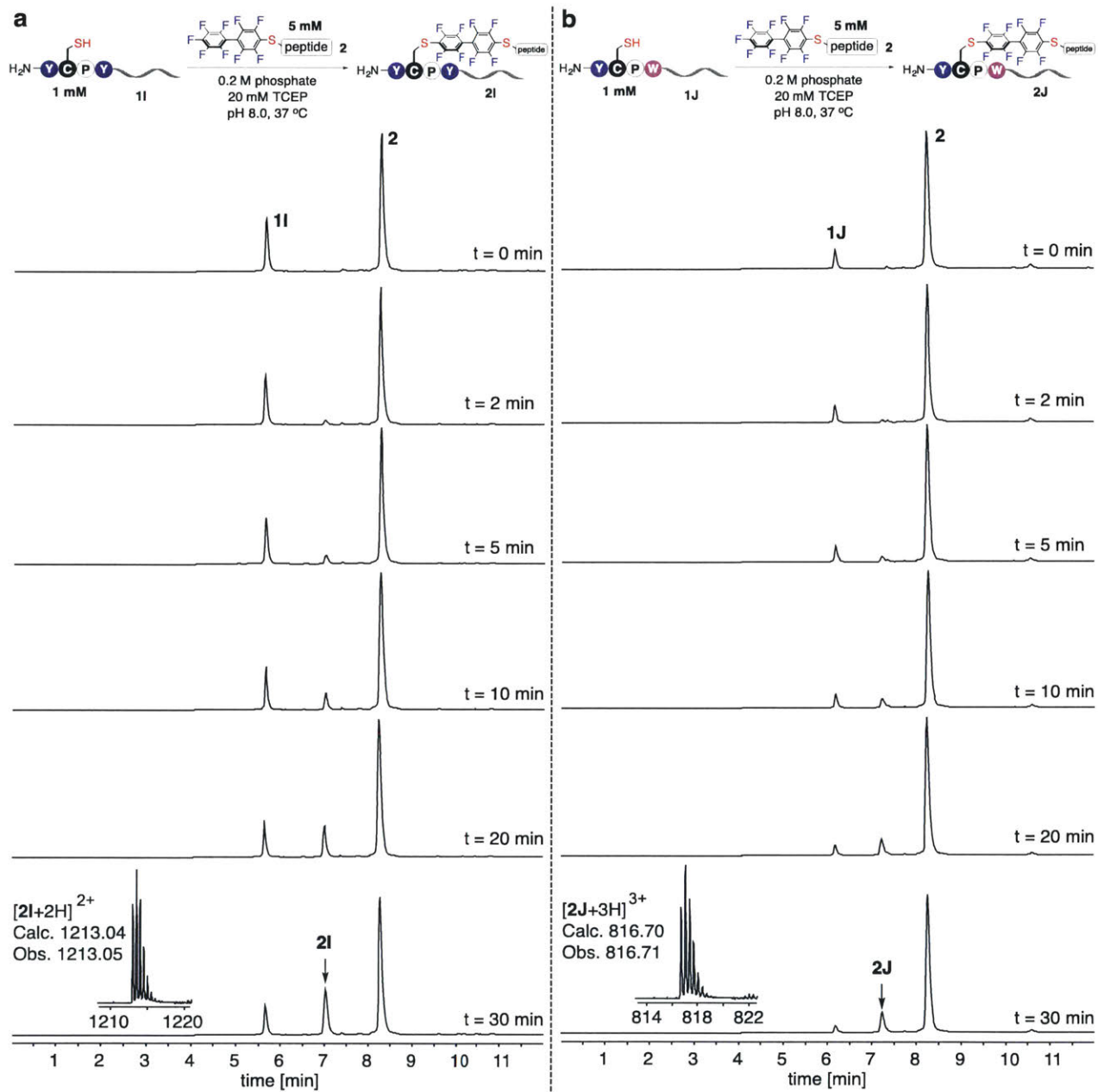
Figure 3.10. LC-MS chromatograms for arylation reactions of peptide 1I (a) and 1J (b) at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
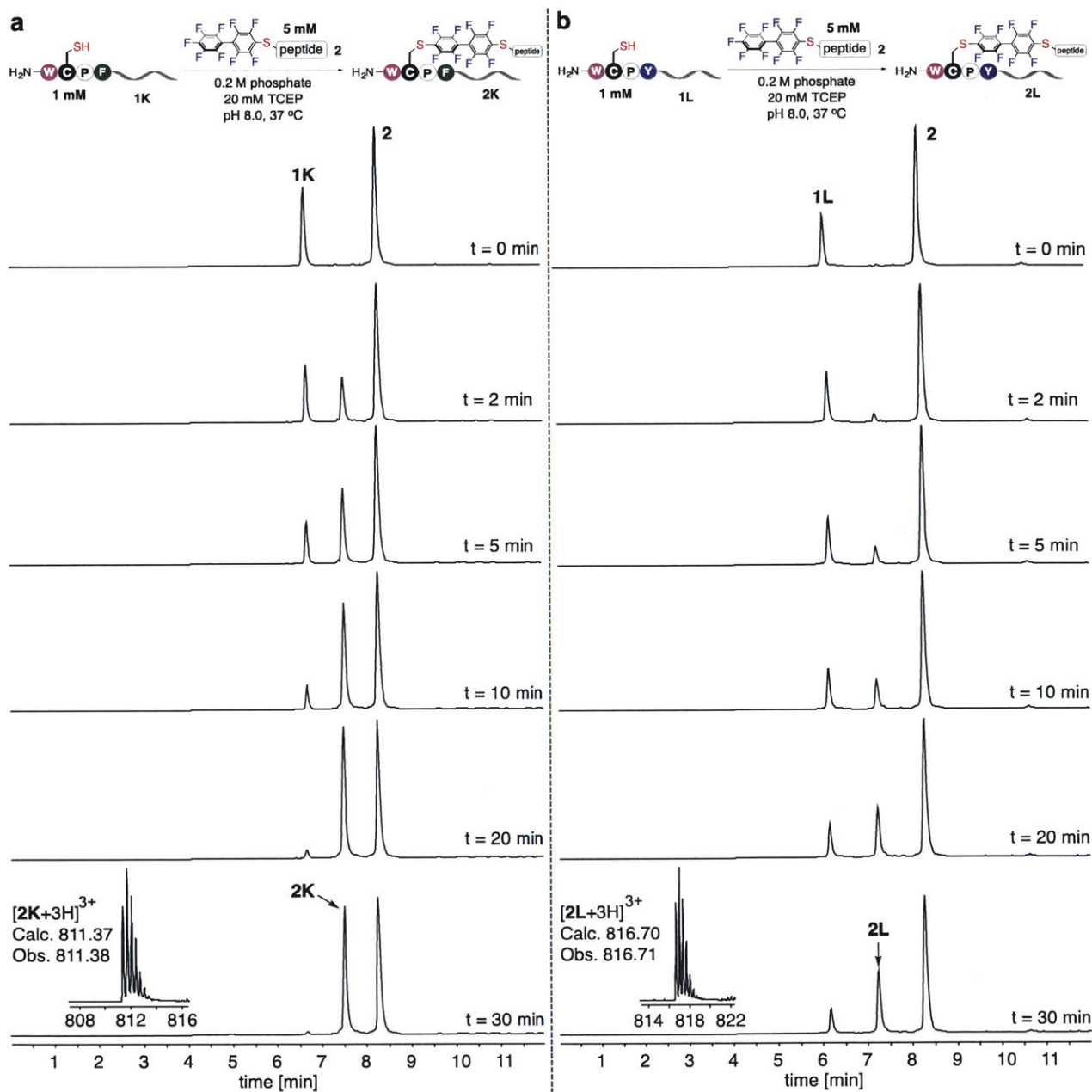
Figure 3.11. LC-MS chromatograms for arylation reactions of peptide 1K (a) and 1L (b) at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
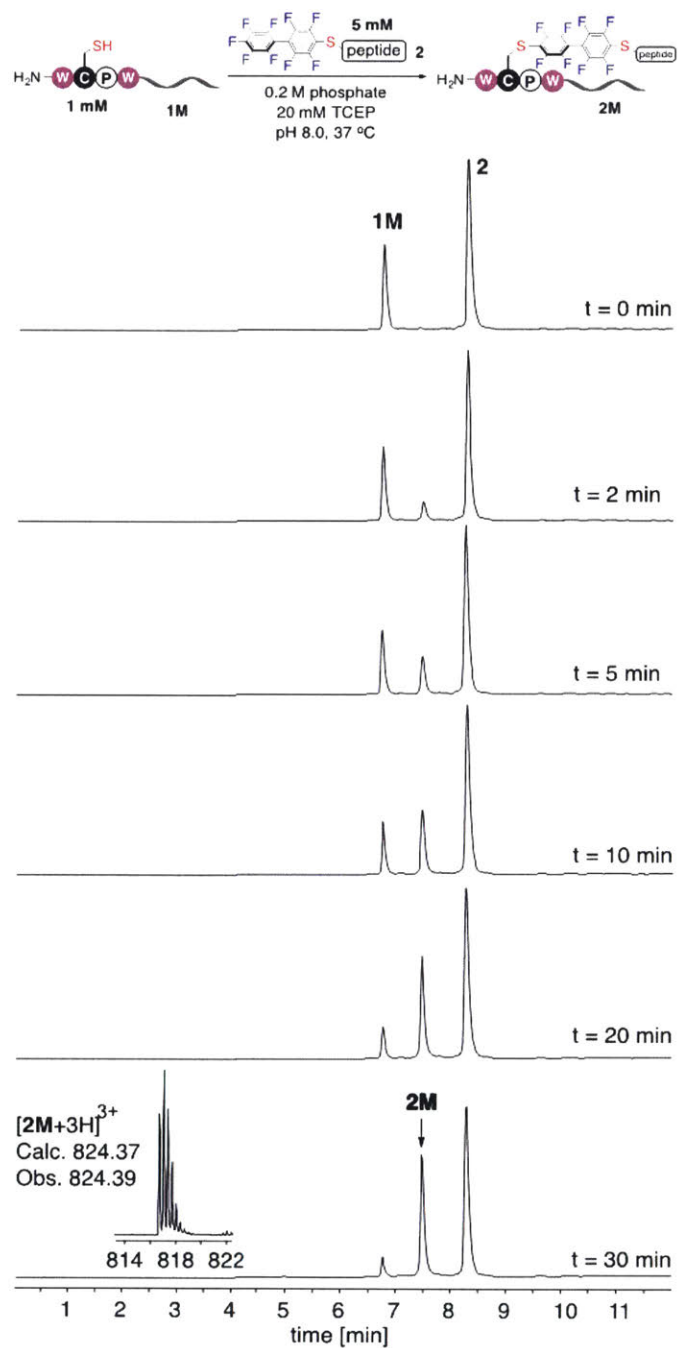
Figure 3.12. LC-MS chromatograms for arylation reactions of peptide 1M at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.

3.1.3. Reactions between π-clamp peptides with perfluoroaryl probes

π-clamp mediated conjugation is highly selective as indicated by our thiol competition experiments. The π-clamp peptide (1E) was found to undergo quantitative conversion with the perfluoroaryl probe (2) in the presence of a double glycine mutant peptide (1A) that served as the competing thiol species. Only the π-clamp peptide reacted quantitatively to form conjugated product in 30 minutes (Figure 3.13).
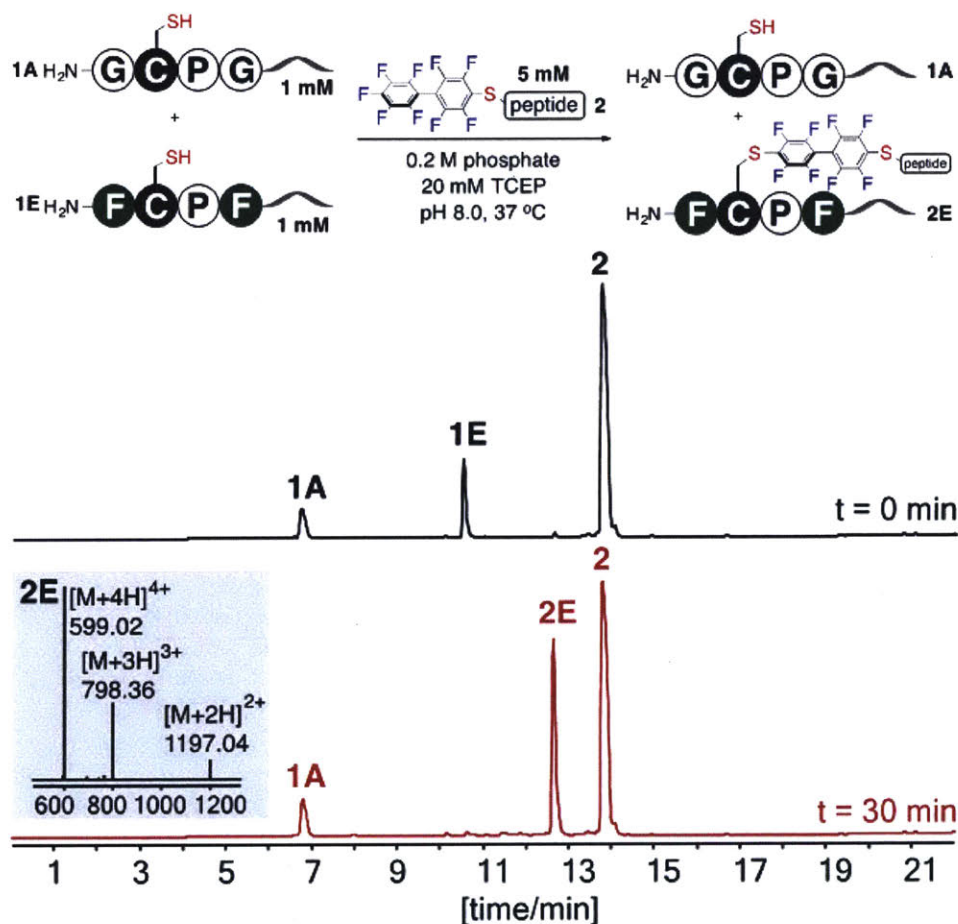
Figure 3.13. π-clamp mediated site-specific cysteine conjugation in the presence of another competing cysteine peptide.

π-Clamp peptide 1E was fully converted to the arylated product 2E while a competing cysteine peptide 1A remain unmodified. Chromatograms shown are total ion currents (TIC) from LC-MS analysis of crude reaction mixtures at 0 minute (black) and 30 minutes (red). The mass spectrum of product 2E is shown as the inset.

188

To further investigate the π-clamp mediated cysteine conjugation, we carried out additional studies to understand if location mattered and the substrate scope. We found that the π-clamp was efficiently modified irrespective of its position on the polypeptide chain (Figure 3.17). π-clamp at the *N*-terminus (**1E**), the *C*-terminus (**1N**), and the middle (**1O**) of the polypeptide chain were readily modified with a diverse set of perfluoroaryl-linked probes including peptide, biotin, fluorescein, alkyne, and polyethylene glycol (**2 – 6**, see Experimental section 3.4.6 for synthesis).
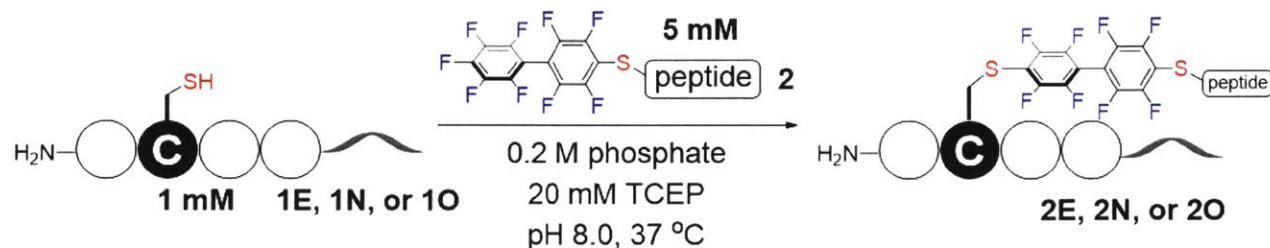
Entry	Peptide	Sequence	$K_2/(M^{-1}s^{-1})$
1	**1E** (N-terminal π-clamp)	NH$_2$-FCPFGLLKNK-CONH$_2$	0.73 ± 0.01
2	**1N** (C-terminal π-clamp)	NH$_2$-KNKLLGFCPF-CONH$_2$	0.094 ± 0.002
3	**1O** (Middle π-clamp)	NH$_2$-KNKLLGFCPFGLLKNK-CONH$_2$	0.181 ± 0.005

Figure 3.14. Effect of the position on peptide chain on the reactivity of π-clamp.

Amino acids are shown in one-letter code. *[D]P stands for D-proline. Kinetics fitting of reactions at different time points are summarized in Figure 3.5. LC-MS chromatograms of these reactions are shown in Figure 3.8a, Figure 3.15, and Figure 3.16.
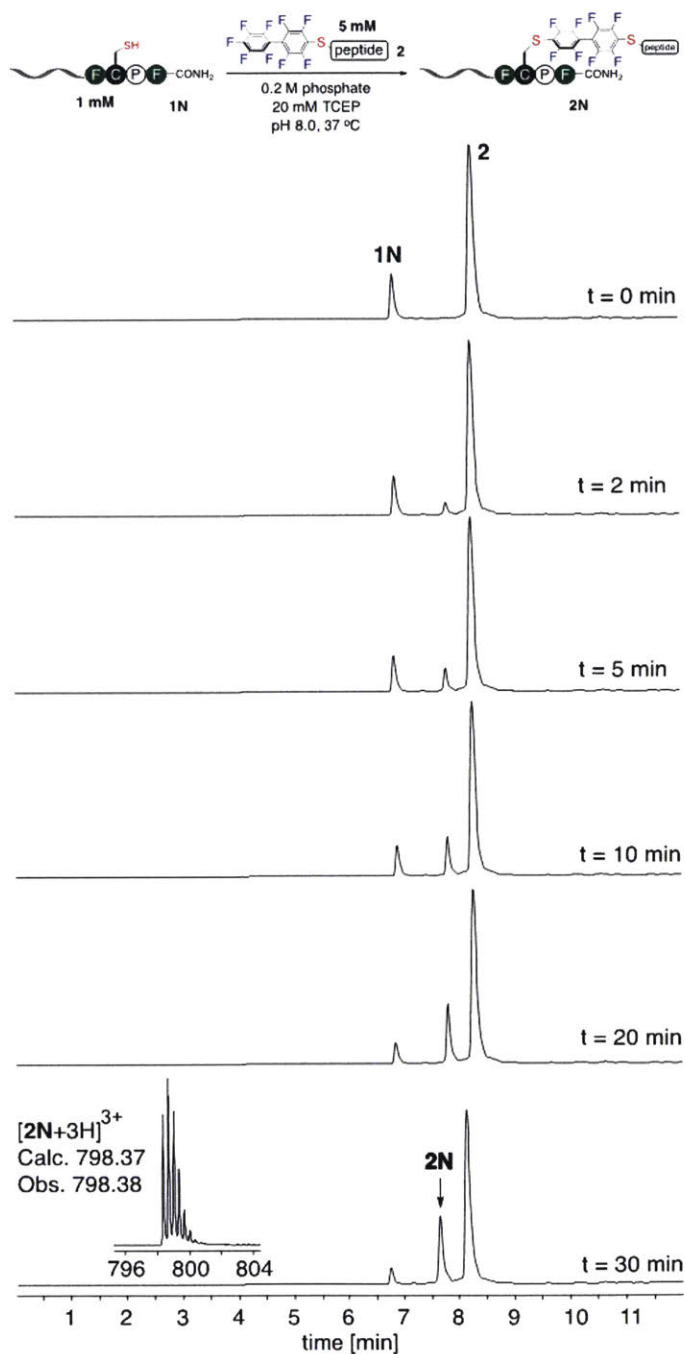
Figure 3.15. LC-MS chromatograms for arylation reactions of peptide 1N at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
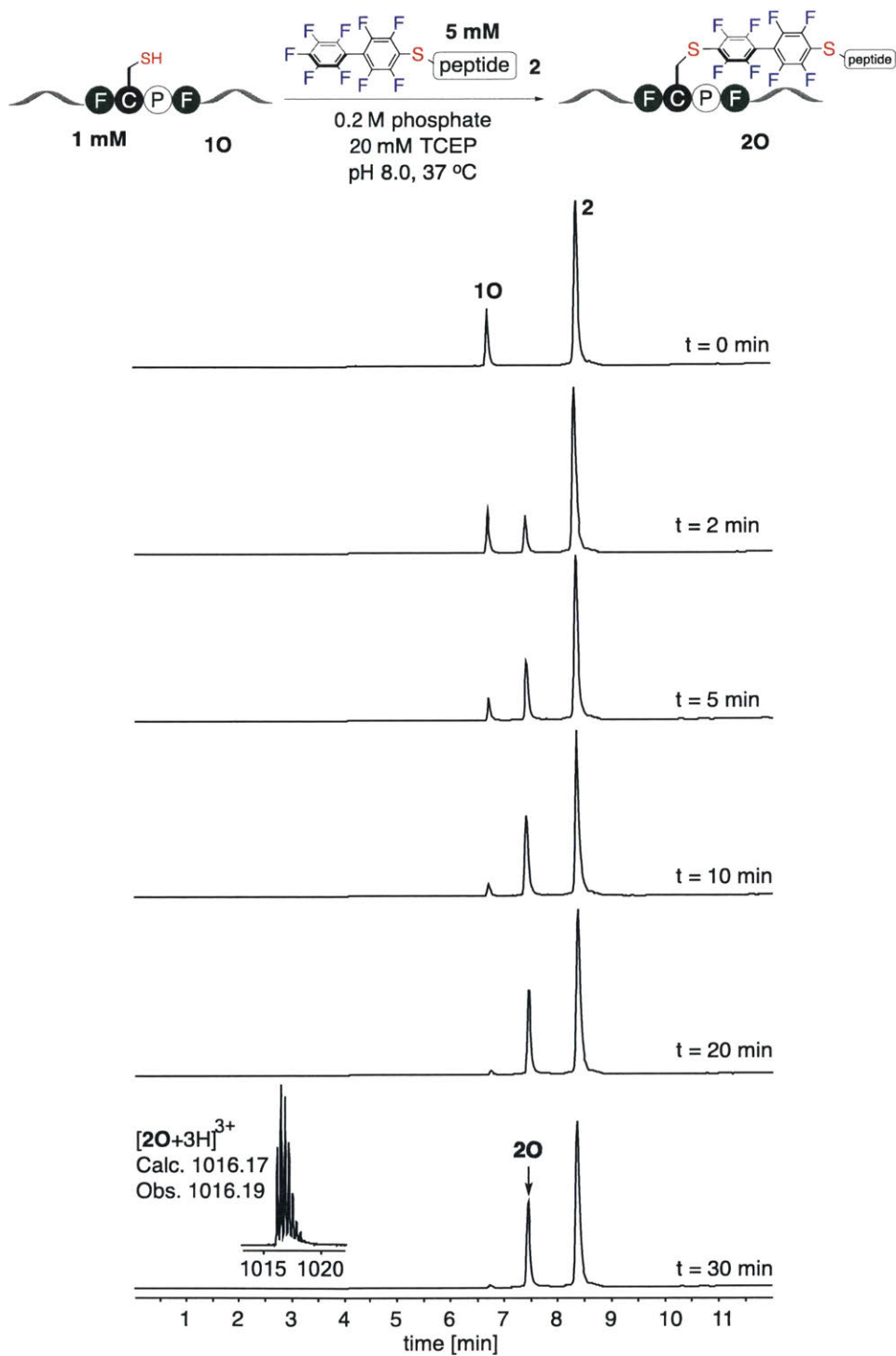
Figure 3.16. LC-MS chromatograms for arylation reactions of peptide 1O at different time points.

Shown are total ion current (TIC) curves. The mass spectrum at the highest intensity of the TIC peak is shown as an inset.
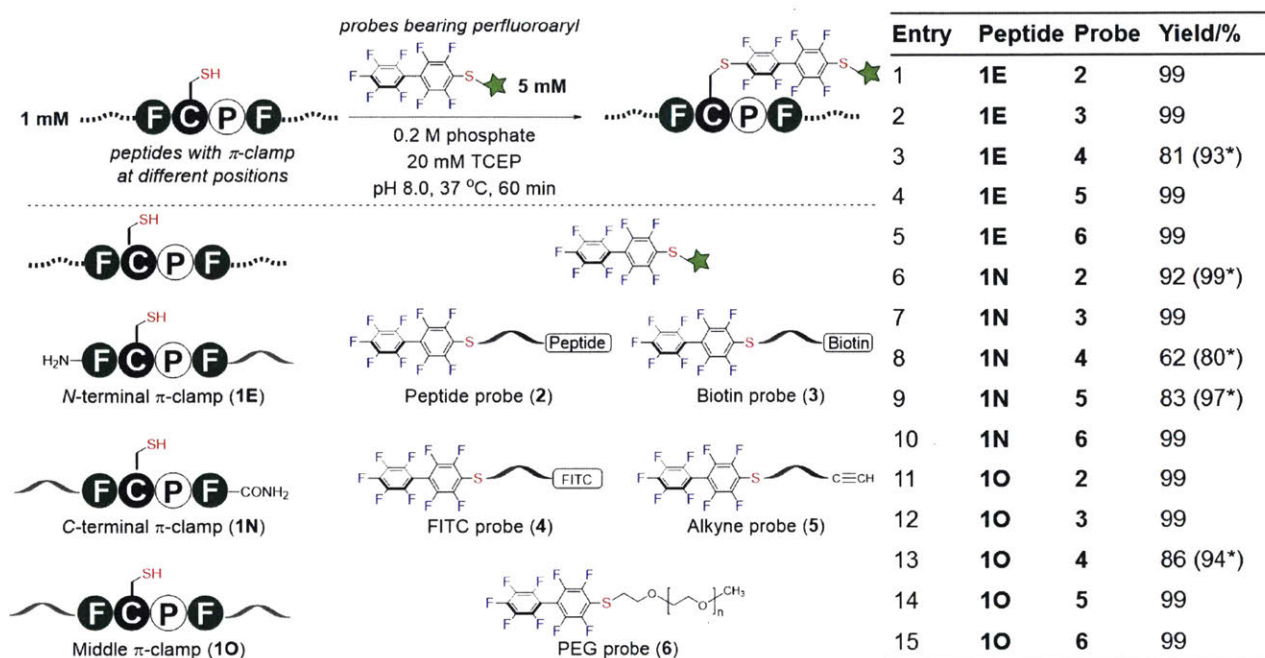
Figure 3.17. Reactions between π-clamp at distinct positions in polypeptides and diverse perfluoroaryl-based probes.

Entry	Peptide	Probe	Yield/%
1	1E	2	99
2	1E	3	99
3	1E	4	81 (93*)
4	1E	5	99
5	1E	6	99
6	1N	2	92 (99*)
7	1N	3	99
8	1N	4	62 (80*)
9	1N	5	83 (97*)
10	1N	6	99
11	1O	2	99
12	1O	3	99
13	1O	4	86 (94*)
14	1O	5	99
15	1O	6	99

π-Clamp at the *N*-terminus, the *C*-terminus, and the middle of peptides were readily reacted with perfluoroaryl probes bearing peptide molecule (**2**), affinity tag (biotin, **3**), fluorescent reporter (fluorescein isothiocyanate, FITC, **4**), click chemistry handle (alkyne, **5**), and polymer (polyethylene glycol, PEG, **6**). Lower reaction yields were observed for the FITC-linked probe **4** potentially because of the low solubility of probe **4** in water. Yields shown are from LC-MS analysis of the crude reactions at 60 minutes. *Yields at 120 minutes. See Figure 3.18 to Figure 3.21 for LC-MS chromatograms. See Experimental section 3.4.5 for the method for calculating yields.
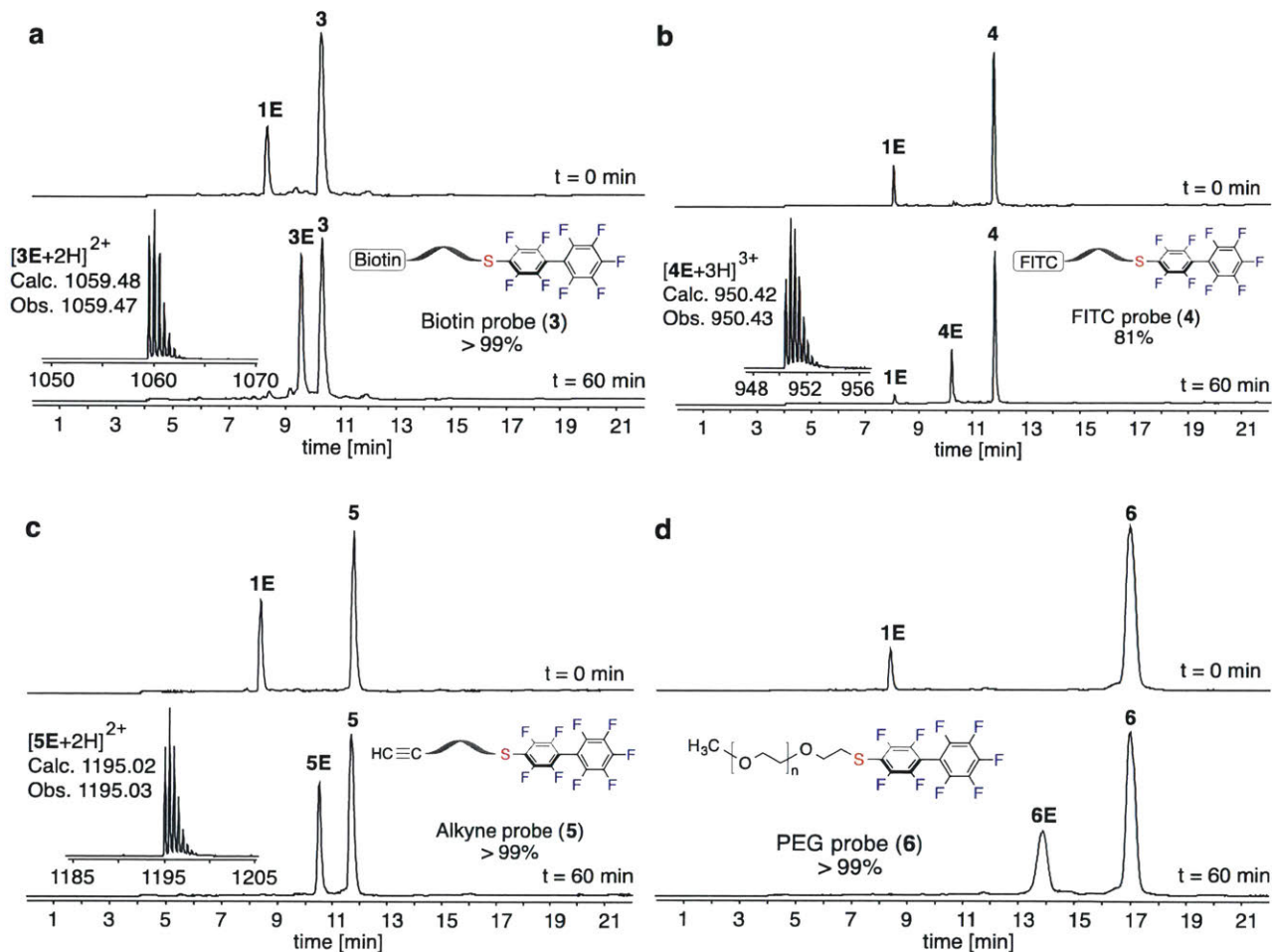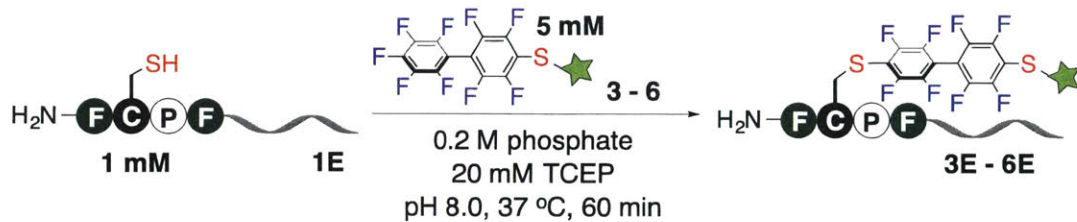
Figure 3.18. Reactions of perfluoroaryl probes with the *N*-terminal π-clamp peptide 1E.

Shown are LC-MS chromatograms for reactions of the π-clamp peptide **1E** with (**a**) biotin probe, (**b**) FITC probe, (**c**) alkyne probe, and (**d**) PEG probe. Reaction conditions: 1 mM peptide **1E**, 5 mM probe, 0.2 M phosphate and 20 mM TCEP (pH 8.0), at 37 °C for 60 minutes.
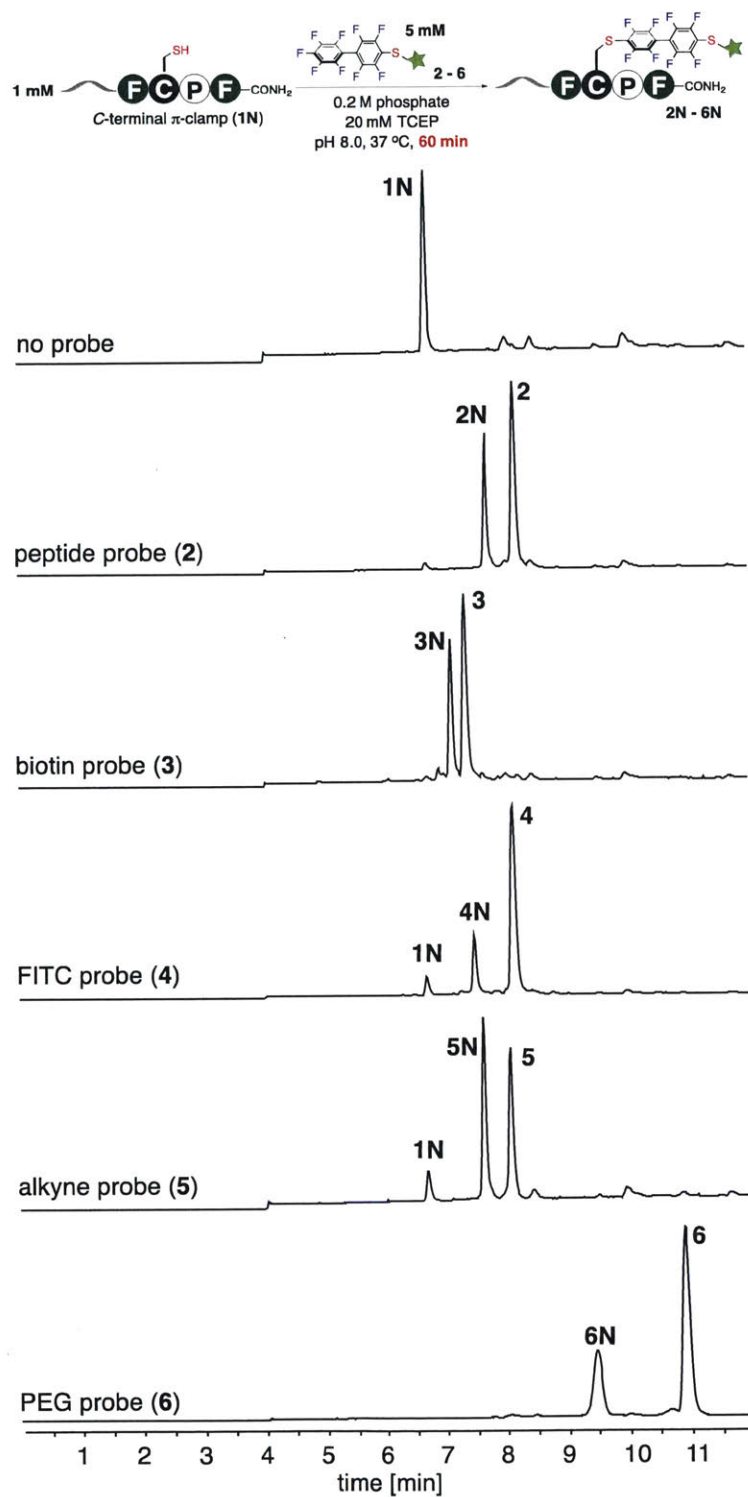
Figure 3.19. Reactions of perfluoroaryl probes with *N*-terminal π-clamp peptide 1O.
Reaction conditions: 1 mM peptide **1O**, 5 mM probe **2 – 6**, 0.2 M phosphate and 20 mM TCEP (pH 8.0), at 37 °C for 60 minutes.
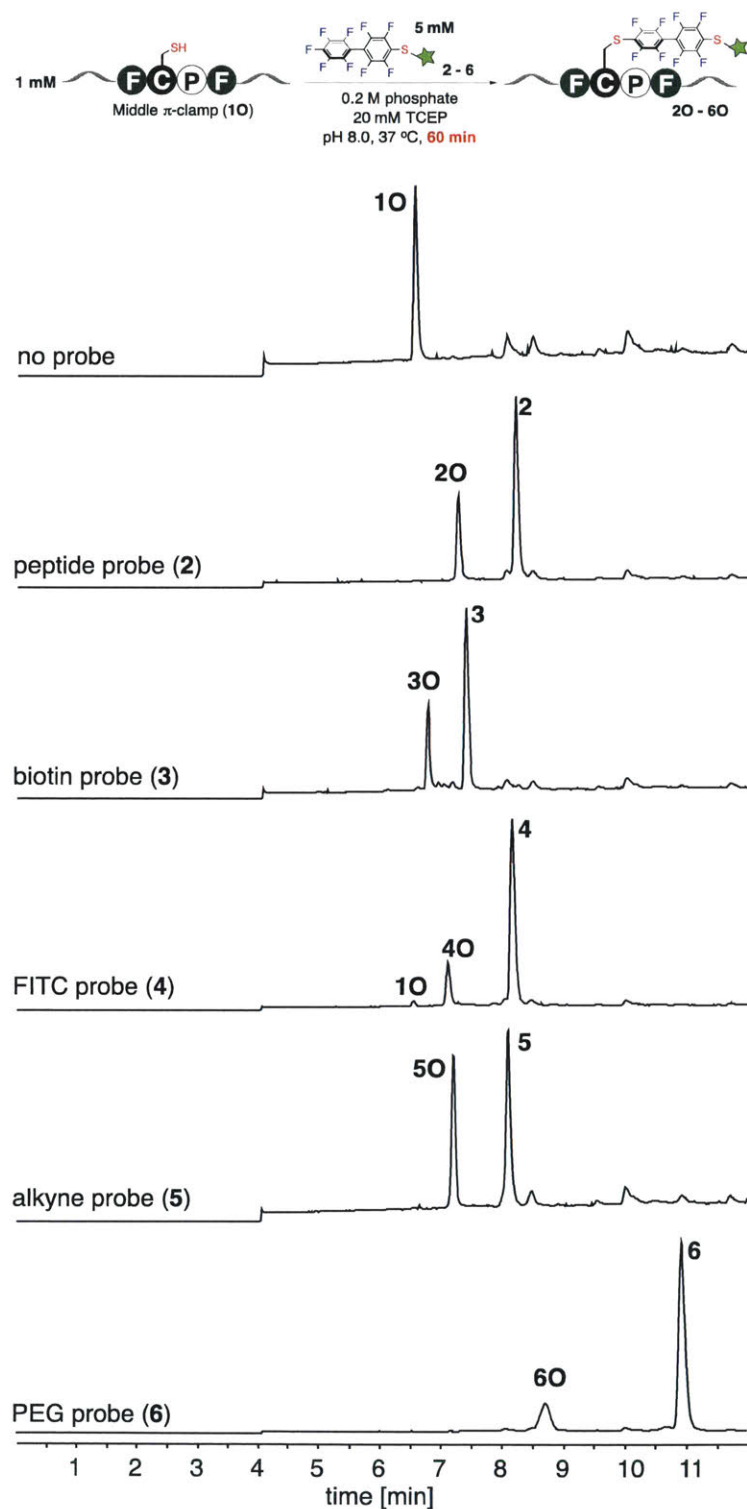
Figure 3.20. Reactions of perfluoroaryl probes with π-clamp at the middle of the peptide chain.

Reaction conditions: 1 mM peptide **1O**, 5 mM probe **2 – 6**, 0.2 M phosphate and 20 mM TCEP (pH 8.0), at 37 °C for 60 minutes.
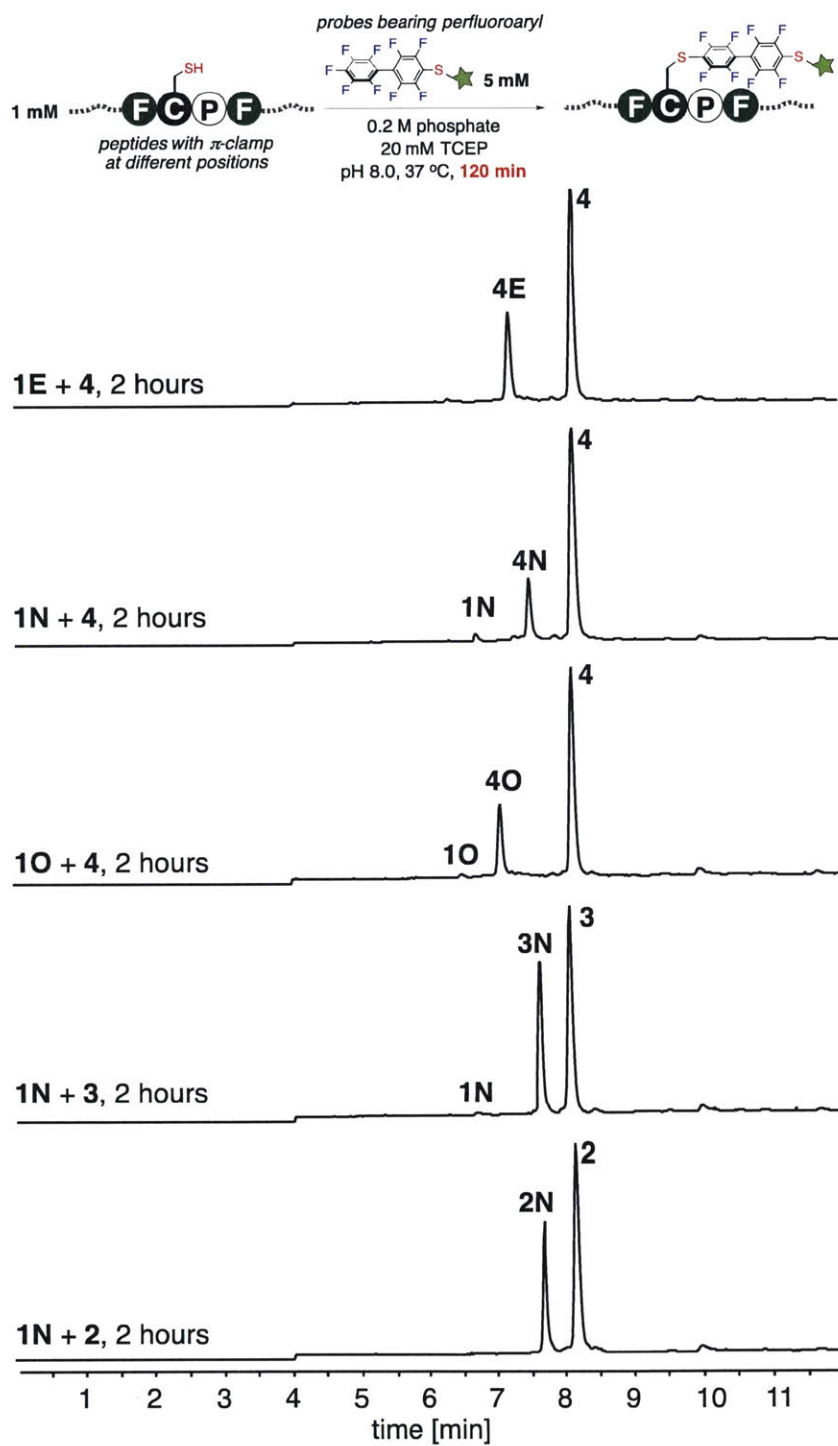
Figure 3.21. Reactions of perfluoroaryl probes with π-clamp peptides for 2 hours.
Reaction conditions: 1 mM π-clamp peptide, 5 mM probe, 0.2 M phosphate and 20 mM TCEP (pH 8.0), at 37 °C for 120 minutes.

3.1.4. π-Clamp-mediated regioselective modification of proteins containing multiple cysteines

We next investigated the regioselectivity on a 55-kDa protein substrate (Figure 3.23). Model protein **7** was designed to contain an *N*-terminal cysteine and a *C*-terminal π-clamp. A protease cleavage site was positioned upstream of the π-clamp thereby allowing for the unequivocal verification of the regioselectivity. Upon reacting the protein (**7**) with probe **2** for 2 hours, we observed > 95% formation of the mono-labeled product (**7A**). The *N*-terminal free cysteine was subsequently labeled with fluorescein-5-maleimide producing the dual-labeled product (**7B**). Upon protease cleavage, only two products were generated: a protein with maleimide-labeled *N*-terminal cysteine (**7C**) and a π-clamp arylated species, confirming the absolute regioselectivity endowed by the π-clamp.
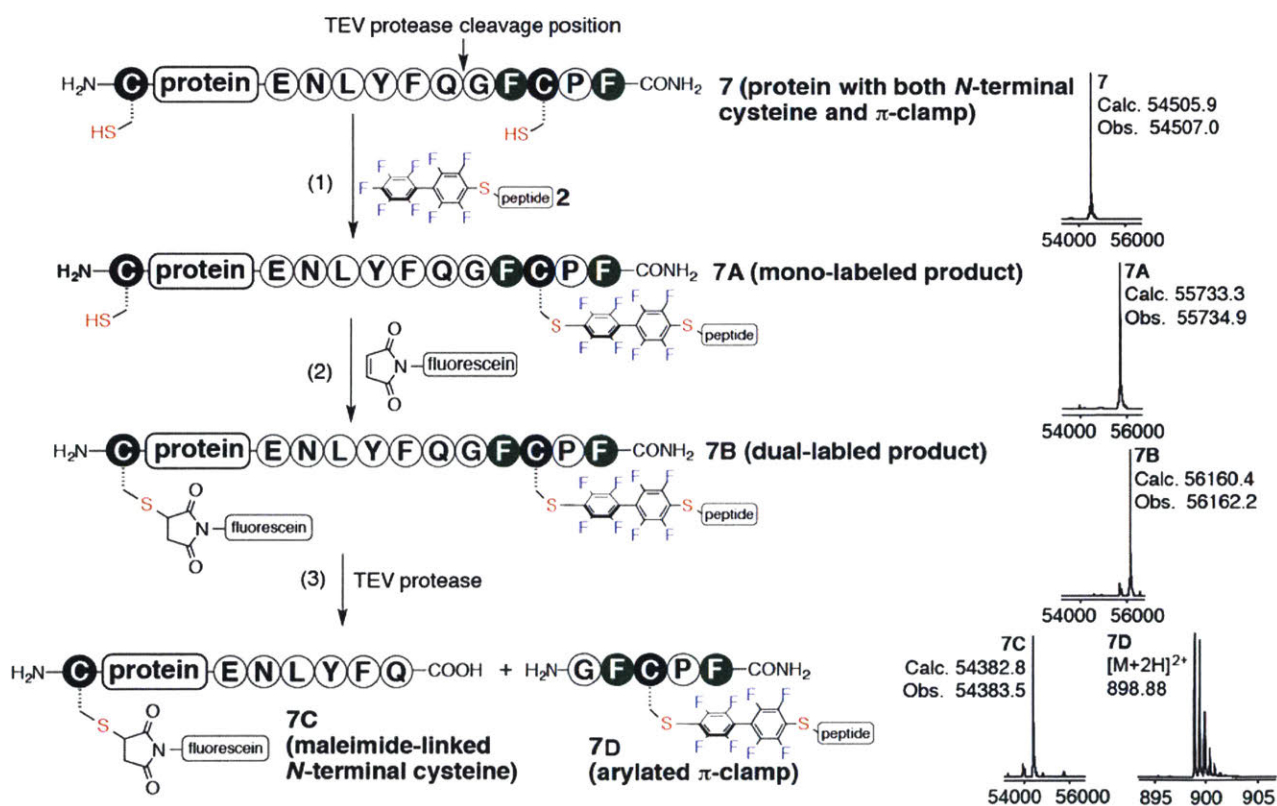
Figure 3.22. π-clamp mediated regioselective cysteine modification enables protecting group-free one-pot dual labeling of a protein.

Protecting group-free one-pot dual labeling of a 55-kDa protein. The protein used was a fusion protein of the anthrax toxin lethal factor 1-263 (LF$_N$) and diphtheria toxin domain A (DTA) with

197

an engineered *N*-terminal cysteine, a *C*-terminal π-clamp, and a protease cleavage site upstream of the π-clamp. After sequential modifications of the two cysteines with perfluoroaryl probe **2** and fluorescein-5-maleimide, the dual-labeled product **7B** was protease-digested to illustrate the selectivity of the π-clamp mediated conjugation. Reaction conditions: (1) 50 μM **7**, 1 mM **2**, 0.2 M phosphate, 20 mM TCEP, 37 °C, 2 hours. (2) 50 μM **7A**, 1 mM fluorescein-5-maleimide, 0.2 M phosphate pH 7.0, room temperature, 10 minutes. (3) 25 μM protein **7B**, 0.1 mg/mL TEV protease, 50 mM Tris, 0.1 mM EDTA, 1 mM DTT, pH 8.0, room temperature, 15 hours. TEV: tobacco itch virus; EDTA: ethylenediaminetetraacetic acid; DTT: dithiothreitol; Tris: 2-amino-2-hydroxylmethyl-propane-1, 3-diol.

Next, we site-specifically modified a cysteine-containing transpeptidase Sortase A (SrtA)[31] (Figure 3.23). An *N*-terminal π-clamp SrtA variant (**8**) reacted with probe **2** to produce > 95% mono-labeled product (**8A**). The modified variant displayed full catalytic activity (Figure 3.24). No reaction took place with SrtA without the π-clamp (**9**). In sharp contrast, when the π-clamp-Sortase (**8**) was reacted with bromoacetamide, a mixture of products was produced with labeling of both cysteine residues (Figure 3.25).
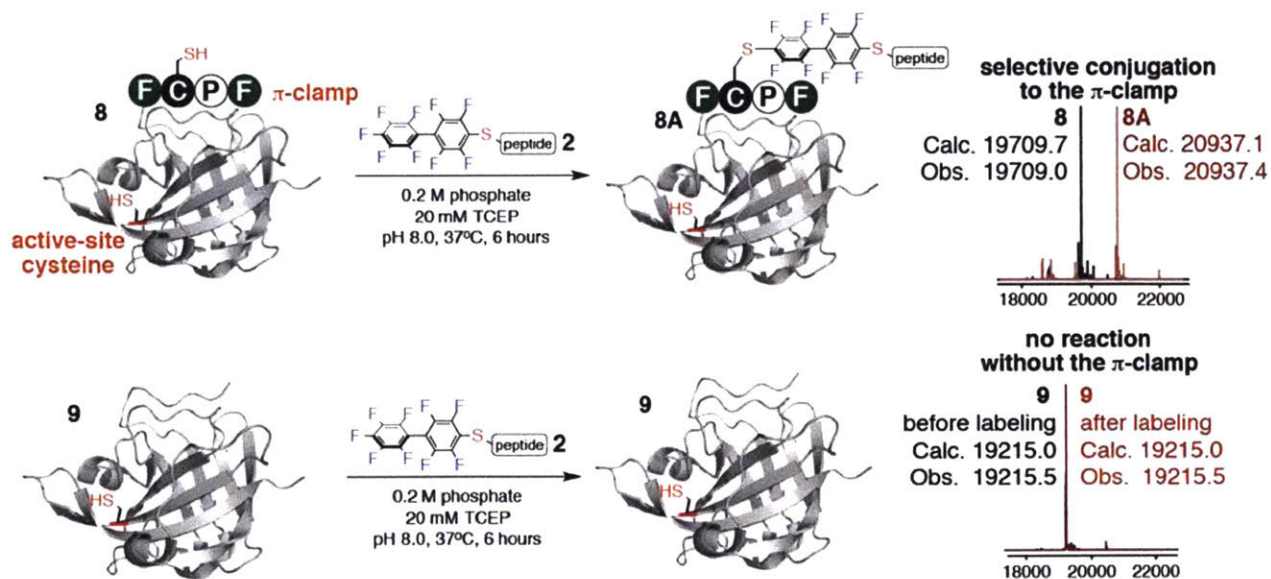
Figure 3.23. π-clamp mediated site-specific conjugation on a cysteine-containing enzyme. Top, quantitative and selective labeling of π-clamp SrtA (PDB entry: 1T2P); bottom, control shows no labeling of SrtA. Reactions conditions: 38 μM **8** or **9**, 1 mM **2**, 0.2 M phosphate, 20 mM TCEP, 37 °C, 6 hours.
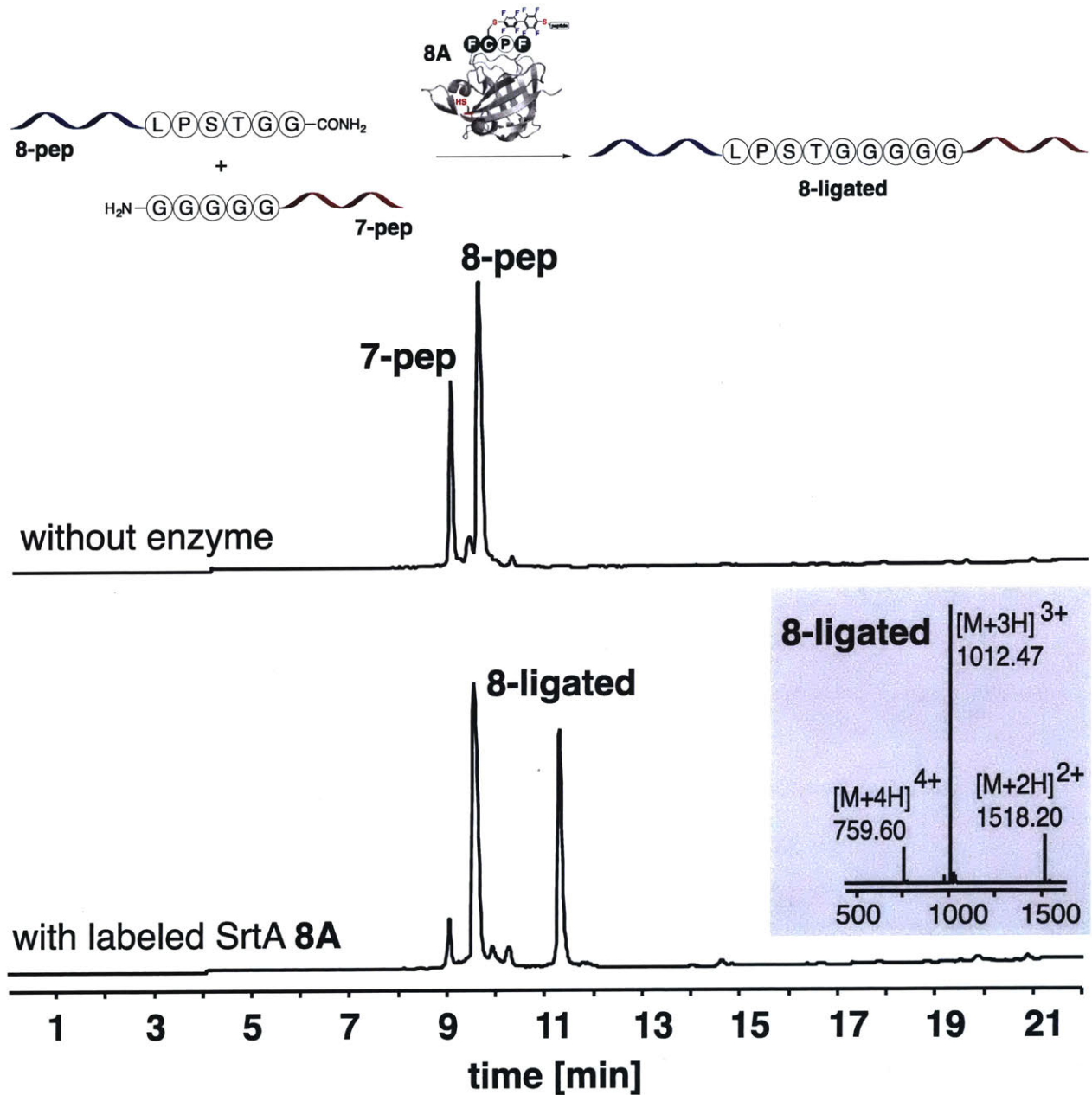
Figure 3.24. Perfluoroaryl-labeled π-clamp SrtA (8A) is able to catalyze the ligation of two peptides.

Reaction conditions: 1 mM **8-pep**, 2 mM **7-pep**, 50 mM Tris, 150 mM NaCl, 10 mM CaCl2, 10 mM TCEP, pH 7.5, 10 μM **8A**, at room temperature for 30 minutes. 10 μL of the reaction mixture was quenched by addition of 100 μL 50% A: 50% B and analyzed by LC-MS.
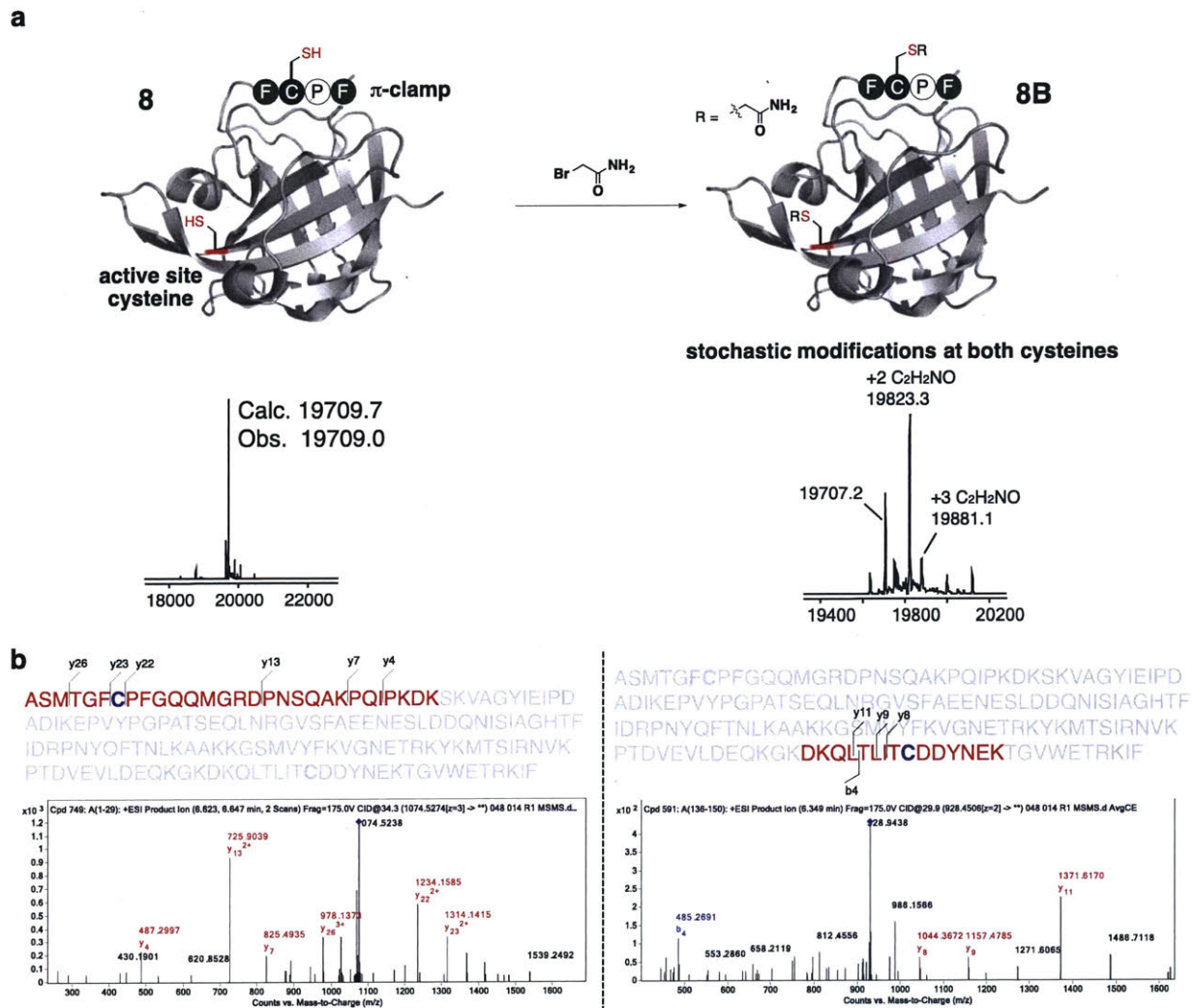
Figure 3.25. Labeling of cysteines in π-clamp SrtA by bromoacetamide..

Reaction conditions: 39 μM π-clamp SrtA **8**, 20 mM bromoacetamide, 0.2 M phosphate and 20 mM TCEP (pH 8.0), at 37 °C for 1 hour. (**b**) Trypsin digestion and LC-MS/MS analysis indicate that both the π-clamp cysteine and the active-site cysteine are labeled by bromoacetamide.

200

3.1.5. π-Clamp-mediated site-specific antibody modification

IgG molecules modified with small molecule drugs (antibody-drug conjugates, ADCs) are currently used as therapeutic agents.[50] However, attaching small molecule agents site-specifically to cysteines in IgGs is as of yet impossible, and thus commercial ADCs are heterogeneous mixtures of conjugates.[50] Approaches to engineer cysteine substitutions in antibodies produce mixed disulfides with cysteine or glutathione, thus a fine-tuned reduction-oxidation protocol must be used to afford the free cysteine thiols for selective drug conjugation in the presence of disulfide bonds.[51,52]

We anticipated that the π-clamp IgG could be used to overcome this specificity problem in ADC synthesis, which is notably challenging because IgGs harbor 32 native cysteine residues. The π-clamp mediated modification on antibodies will be a single-step and site-specific antibody-drug conjugation technology that does not require significant antibody engineering or extra chemical steps[51,52]. To this end, we inserted the Phe-Cys-Pro-Phe sequence into the C-termini of the heavy chains of trastuzuamab[53]. Reacting the π-clamp trastuzumab (protein **10**) with either a biotin-perfluoroaryl probe (**11-Biotin**) or a drug-perfluoroaryl probe (**11-MMAF**) (see Experimental section 3.4.6 for synthesis)under reducing conditions, we observed facile formation of the heavy chain mono-labeled products (**10-Biotin** or **10-MMAF**) by LC-MS analysis (Figure 3.26). Antibodies without the π-clamp showed no desired modification under the same conditions (Figure 3.27), highlighting the specificity of the conjugation. Moreover, this selective conjugation reaction works with other antibodies, reacting a π-clamp C225 antibody[54,55] with **11-Biotin** resulted in only the selective modifications on the π-clamp cysteine residues (Figure 3.28), suggesting that the π-clamp could be a general strategy for site-selective antibody modification.
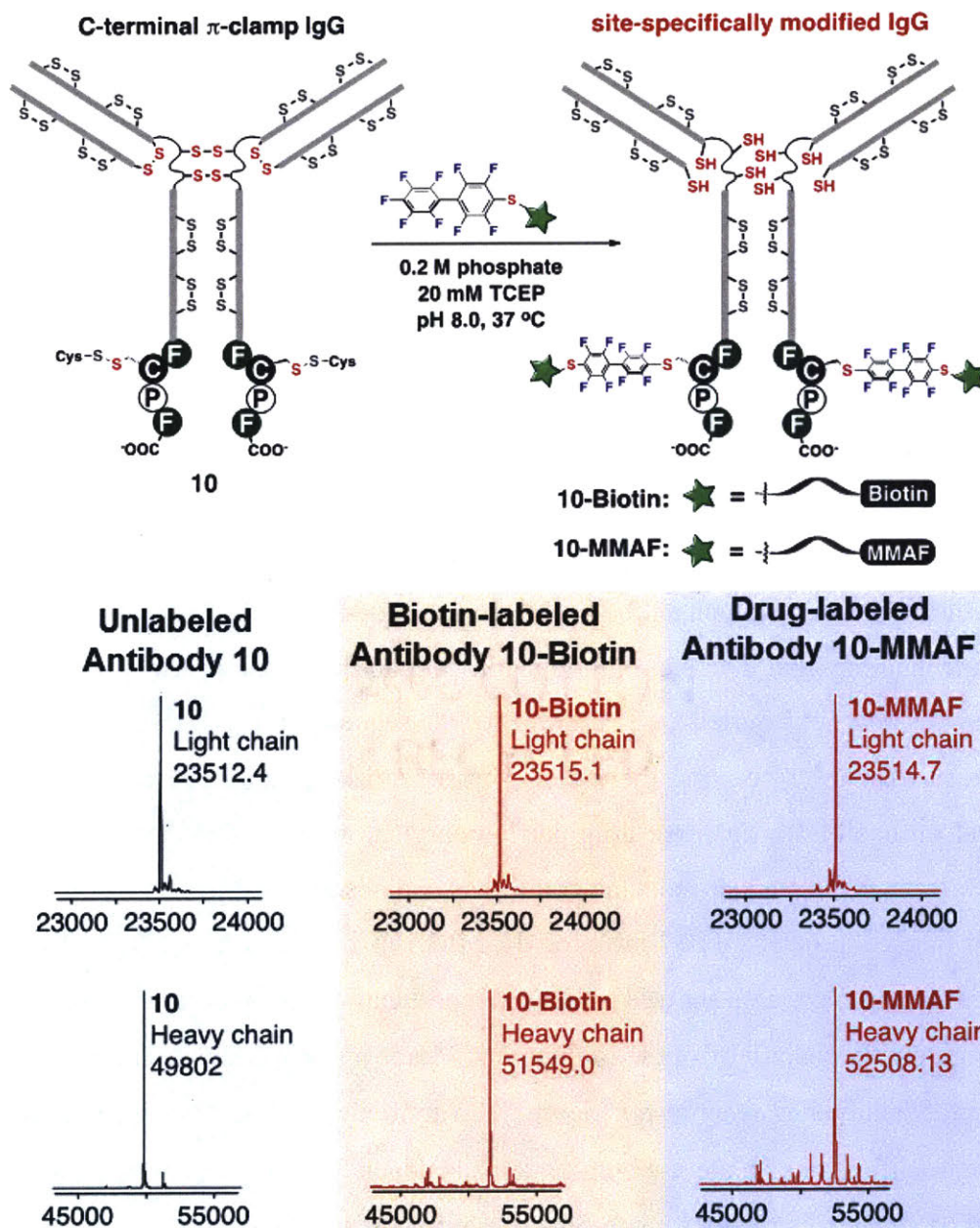
Figure 3.26. Site-specific conjugation of biotin or monomethyl auristatin F (MMAF) to π-clamp trastuzumab 10.

LC-MS analysis showed site-specific labeling of the π-clamp cysteine residues on the trastuzumab heavy chain. The antibodies were treated with PNGase F to remove the *N*-linked glycans before LC-MS analysis. The deconvoluted mass spectrum of the light chain (top) and the deglycosylated heavy chain of π-clamp trastuzumab (**10**, left), the biotin labeled π-clamp trastuzumab (**10-Biotin**, center), and the drug-conjugated π-clamp trastuzumab (**10-MMAF**, right) were shown.
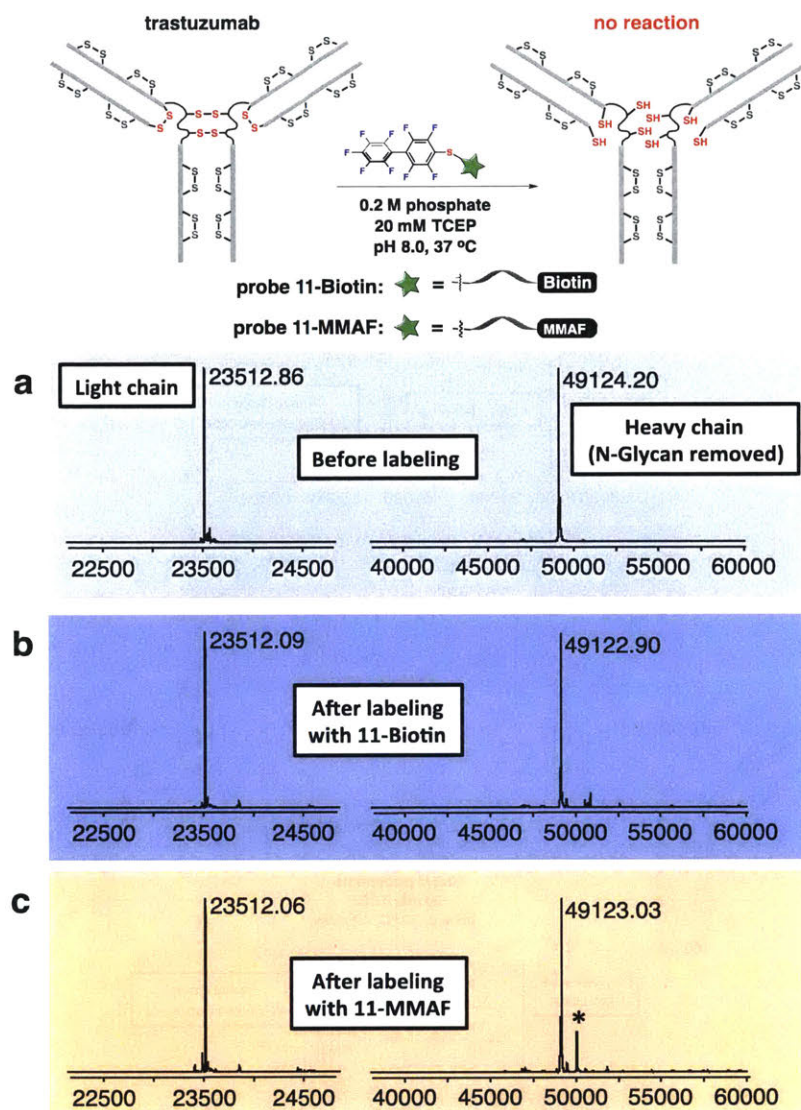
Figure 3.27. Labeling reactions of native trastuzumab with perfluoroaryl probes showed no desired labeling products.

Shown are the mass spectra of the native trastuzumab (**a**), the antibody sample after reaction with **11-Biotin** (**b**), and the antibody sample after reaction with **11-MMAF** (**c**). Reaction conditions for labeling with **11-Biotin** are: 100 μM trastuzumab, 1 mM **11-Biotin**, 0.2 M phosphate, 20 mM TCEP, 37 °C, 4 hours. Reaction conditions for labeling with **11-MMAF**: 100 μM trastuzumab, 1 mM probe **11-Biotin**, 0.2 M phosphate, 20 mM TCEP, 37 °C, 5% DMSO, 16 hours. The peak labeled with "*" indicates the side product from the conjugation of maleimide-MMAF to the heavy chain of the antibody. All antibodies were buffer exchanged with PBS 5 times to remove the probes, then the antibodies were treated with PNGase F to remove the N-linked glycans and were reduced before LC-MS analysis.

203
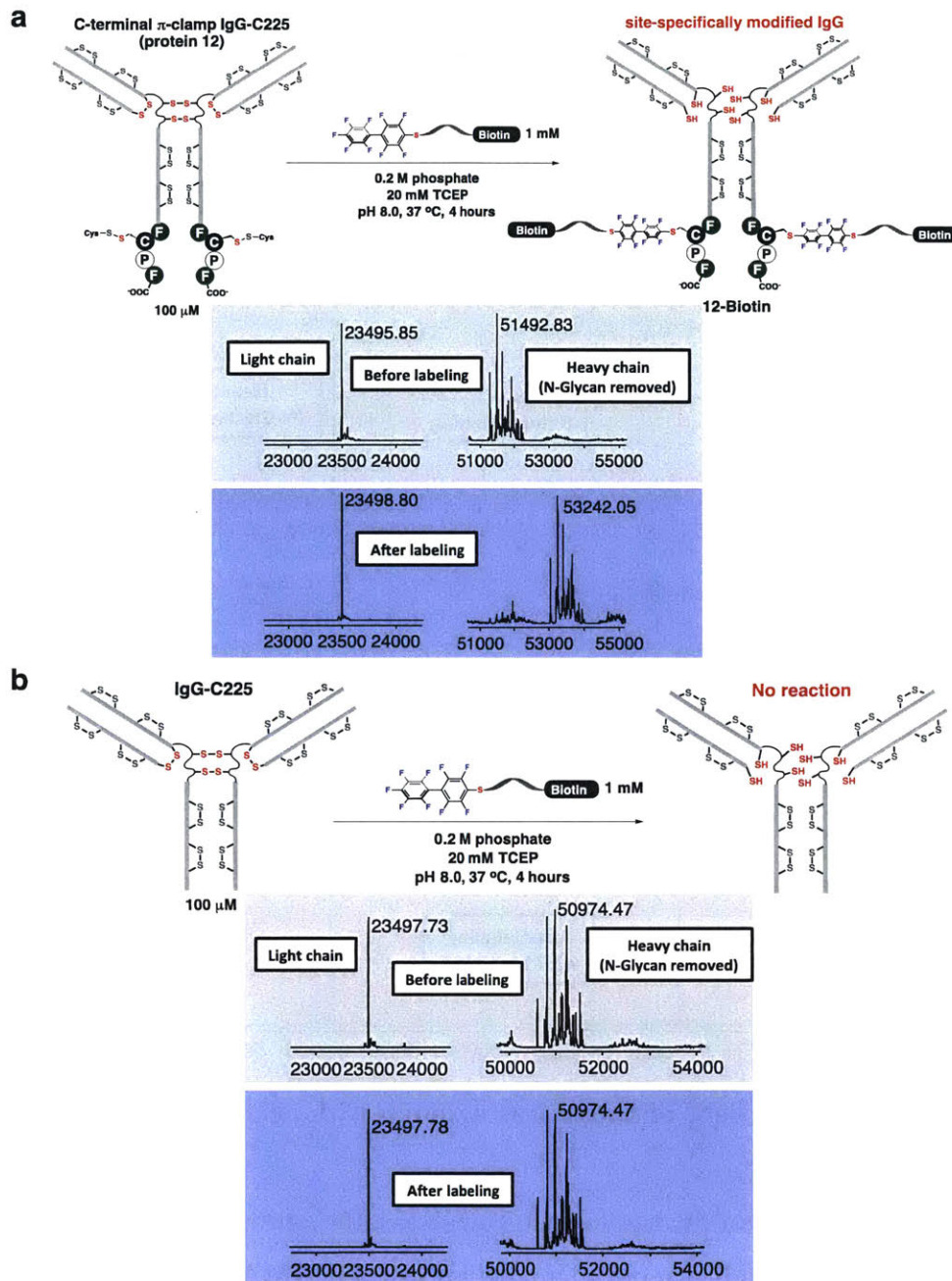
Figure 3.28. Site-selective conjugation of biotin to π-clamp-C225 antibody.

Reaction conditions are: 100 μM protein **12** or C225, 1 mM **11-Biotin**, 0.2 M phosphate, 20 mM TCEP, 37 °C, 4 hours. Mass spectra of the antibodies before labeling and after labeling are shown. (**a**) Conjugation reaction with π-clamp-C225 (**12**) showed complete labeling of the heavy chain. (**b**) Conjugation reaction with C225 showed no reaction. All antibodies were buffer exchanged with PBS 5 times to remove the probes, then the antibodies were treated with PNGase F to remove the *N*-linked glycans and were reduced before LC-MS analysis.

Under the developed reaction conditions (0.2 M phosphate, 20 mM TCEP, pH 8.0, at 37 °C), only the inter-chain disulfides and the π-clamp cysteine residues are reduced (Figure 3.29), and the modified antibodies retained binding affinity to their targets. Biotin modified π-clamp trastuzumab (**10-Biotin**) showed similar binding affinity to HER2 ($K_D = 0.2 \pm 0.2$ nM) compared to native trastuzumab non-selectively modified with a (PEG)4-Biotin (trastuzumab-(PEG)4-Biotin, $K_D = 0.3 \pm 0.1$ nM) (Figure 3.30, see Experimental section 3.4.12 for synthesis of trastuzumab-(PEG)4-Biotin). In addition, both proteins **10** and **10-Biotin** readily bound to BT474 cells (HER2-positive) (Figure 3.31 and Figure 3.32). As another antibody test case, biotin modified C225 antibody (**12-Biotin**) showed similar binding to A431 cells (EGFR-positive) compared to the native C225 antibody (Figure 3.33 and Figure 3.34). Collectively, insertion of the π-clamp into the heavy chains of antibodies and subsequent modification with drugs or probes did not significantly alter the binding properties.

Using the π-clamp mediated cysteine conjugation, we synthesized a site-specific antibody drug conjugate using π-clamp trastuzumab (protein **10**) and a monomethyl auristatin F (MMAF) linked to a perfluoroaryl group (**11-MMAF**, see Experimental section 3.4.6 for synthesis). LC-MS analysis of the conjugation reaction showed selective labeling of the heavy chain π-clamp cysteine residues (Figure 3.26). The prepared ADC selectively killed BT474 cells (HER2 positive) but was not effective for CHO cells (HER2 negative), indicating that the observed toxicity is receptor-dependent (Figure 3.35).
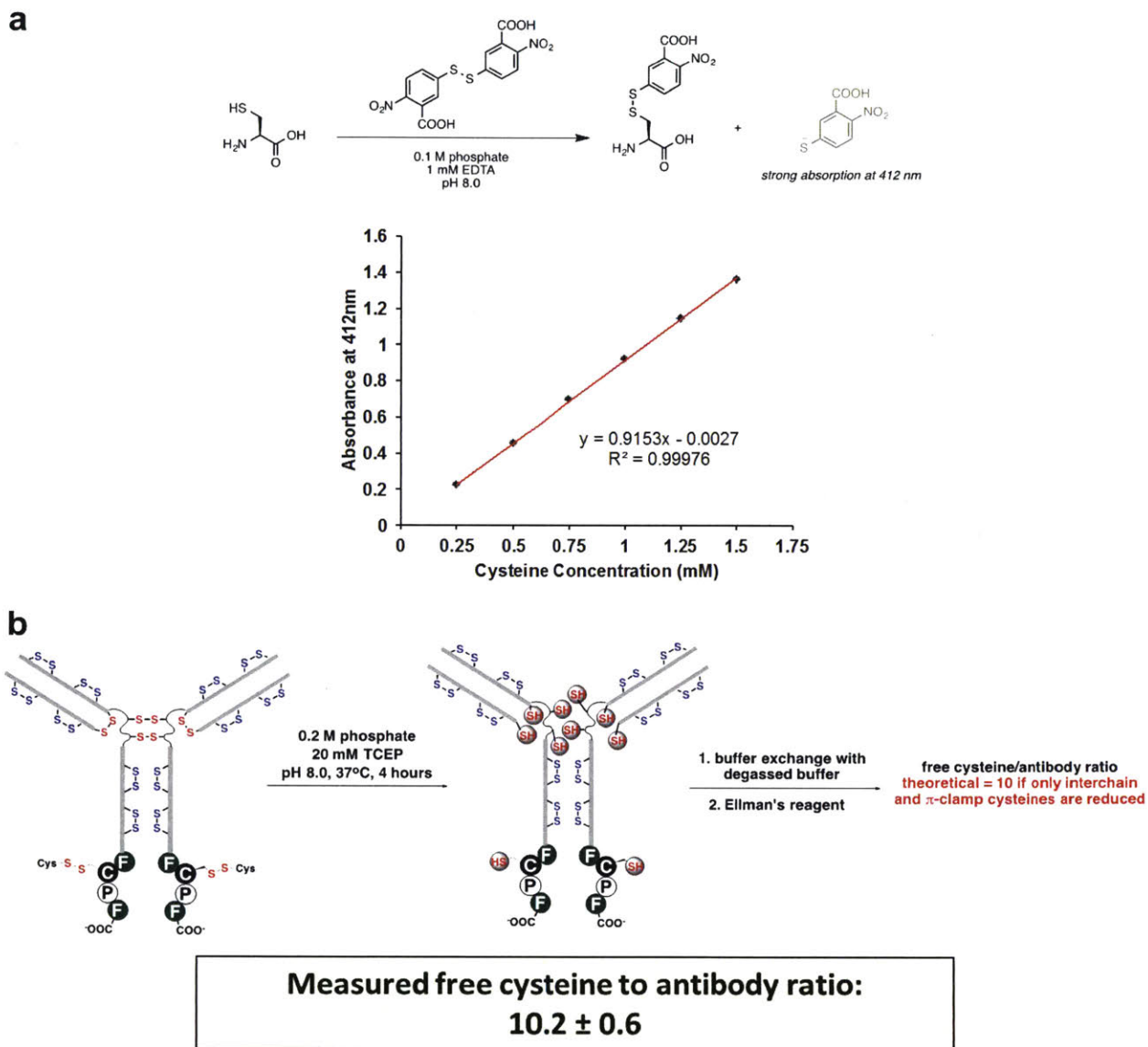
Figure 3.29. Using Ellman's reagent to determine the free cysteine-to-antibody ratio under the developed reaction conditions for the π-clamp-mediated conjugation.

(**a**) A standard curve is generated using various concentrations of L-cysteine. (**b**) Under the developed reaction conditions, only around 10 free thiols were generated and we hypothesized them to be from the reduction of the inter-chain disulfide and the π-clamp cysteine residues. Protein concentration is determined by measuring UV absorption at 280 nm; and the free cysteine concentration is determined by reacting with Ellman's reagent and then measure the absorption at 412 nm. The measured free cysteine-to-antibody ratio is the average of three independent experiments.
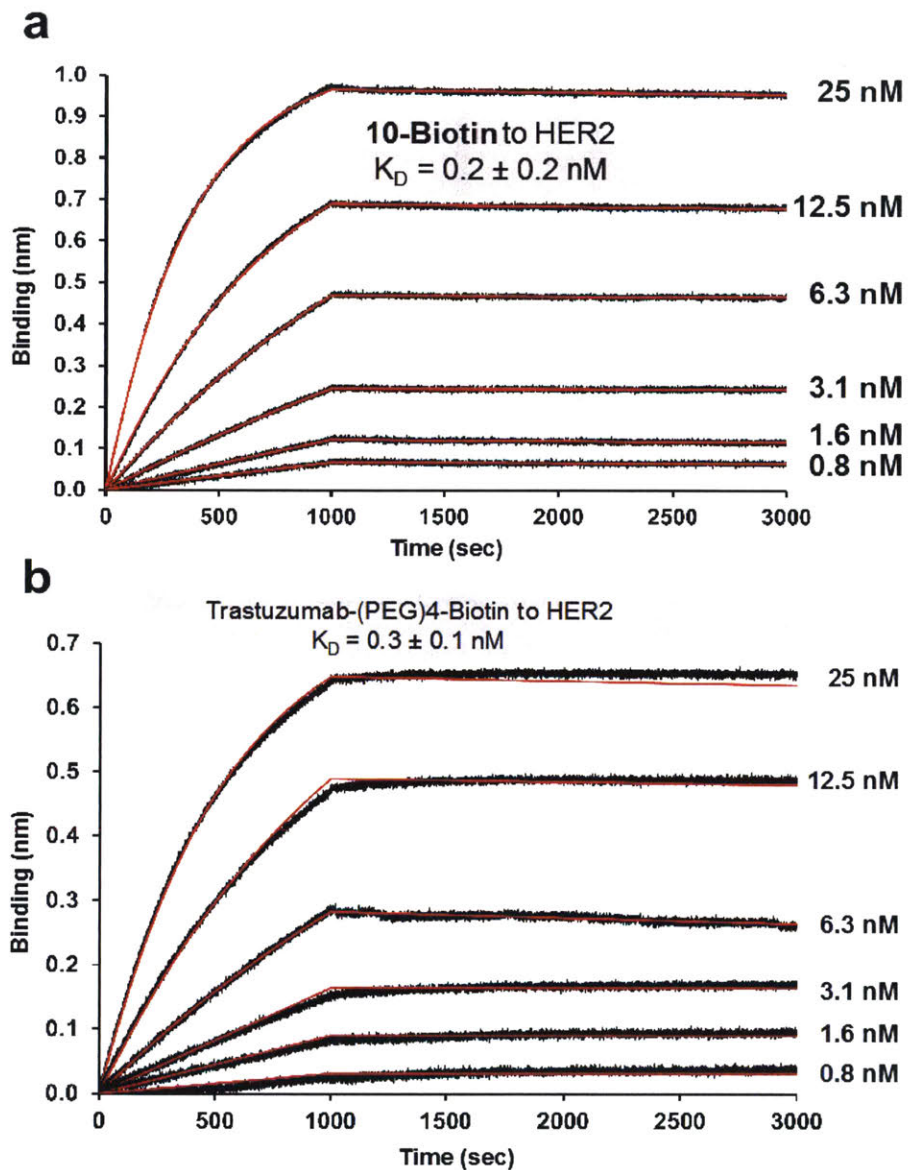
Figure 3.30. BioLayer interferometry assays for measuring the binding of biotin-linked antibodies to HER2.

The biotin-conjugated π-clamp trastuzumab (**10-Biotin**) binds to HER2 in the BioLayer interferometry assay ($K_D = 0.2 \pm 0.2$ nM). The concentration of recombinant HER2 in each experiment is shown next to the curve (see Experimental section 3.4.13 for detailed protocol).
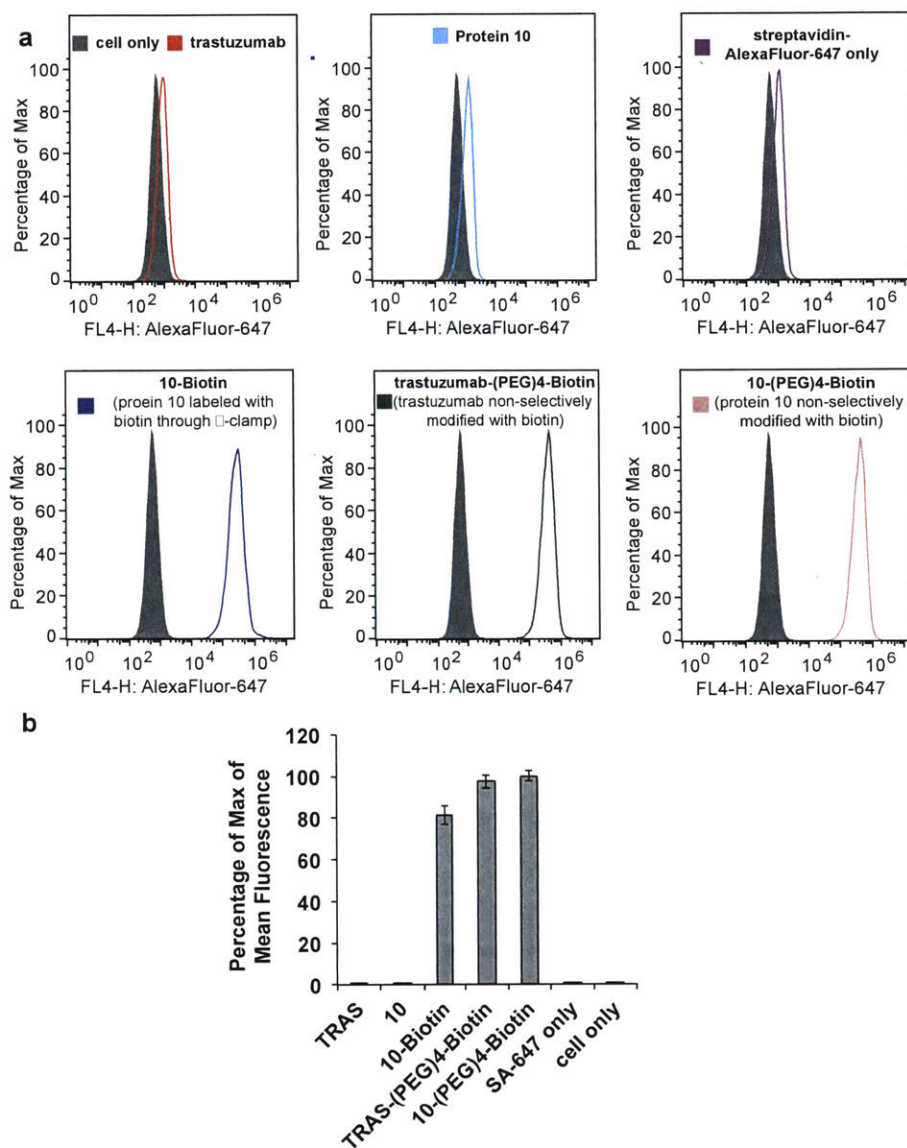
Figure 3.31. Flow cytometry assays for measuring the binding of modified trastuzumab variants to BT474 cells (HER2 positive) through secondary staining of biotin.

Streptavidin-647 was used as the secondary staining reagent to bind to biotin. (**a**) Histograms of AlexaFluor-647 fluorescence. (**b**) Bar graph showing the mean AlexaFluor-647 fluorescence in each sample. **10-Biotin** retained binding to BT474 cells (HER2-positive) compared to the controls. Cells were treated with **10-Biotin** or controls, washed with phosphate buffer saline (PBS) with 0.1% BSA, and then treated with streptavidin-AlexaFluor-647 before analyzed by flow cytometer. **10**-(PEG)4-Biotin and trastuzumab-(PEG)4-Biotin were prepared from reacting Biotin-(PEG)4-NHS with protein **10** or trastuzumab, respectively (see Experimental section 3.4.14 for detailed protocol).
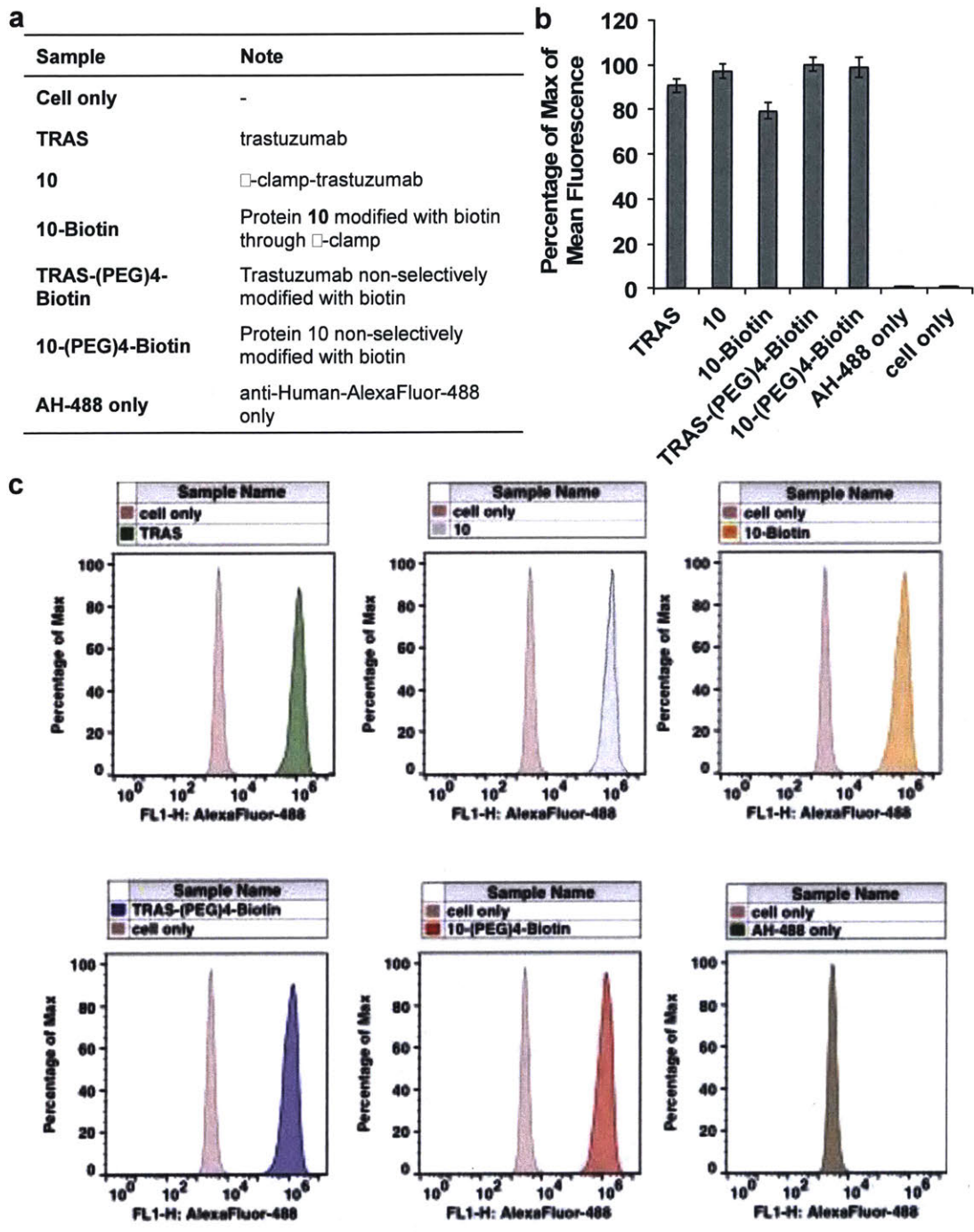
Figure 3.32. Flow cytometry assays for measuring the binding of modified trastuzumab variants to BT474 cells (HER2 positive) through secondary staining of antibody Fc region. AntiHuman Fc-AlexaFluor-488 as the secondary staining reagent. (**a**) Sample description. (**b**) Bar graph showing the mean AlexaFluor-488 fluorescence in each sample. (**c**) Histograms of AlexaFluor-488 fluorescence.
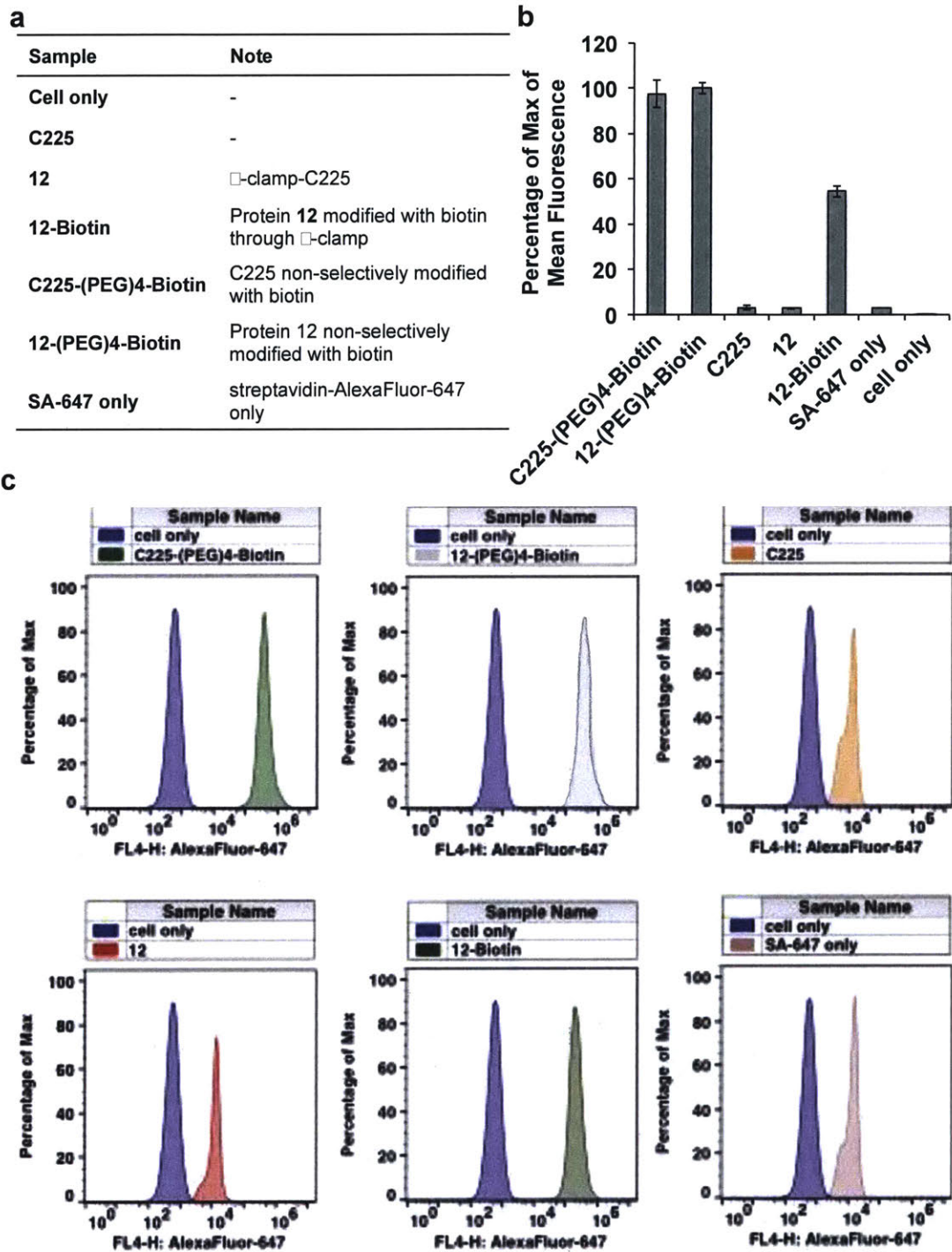
Figure 3.33. Flow cytometry assays for measuring the binding of modified C225 variants to A431 cells (EGFR positive) through secondary staining of biotin.

Streptavidin-AlexaFluor-647 as the secondary staining reagent. (**a**) Sample description. (**b**) Bar graph showing the mean AlexaFluor-647 fluorescence in each sample. (**c**) Histograms of AlexaFluor-647 fluorescence.

210
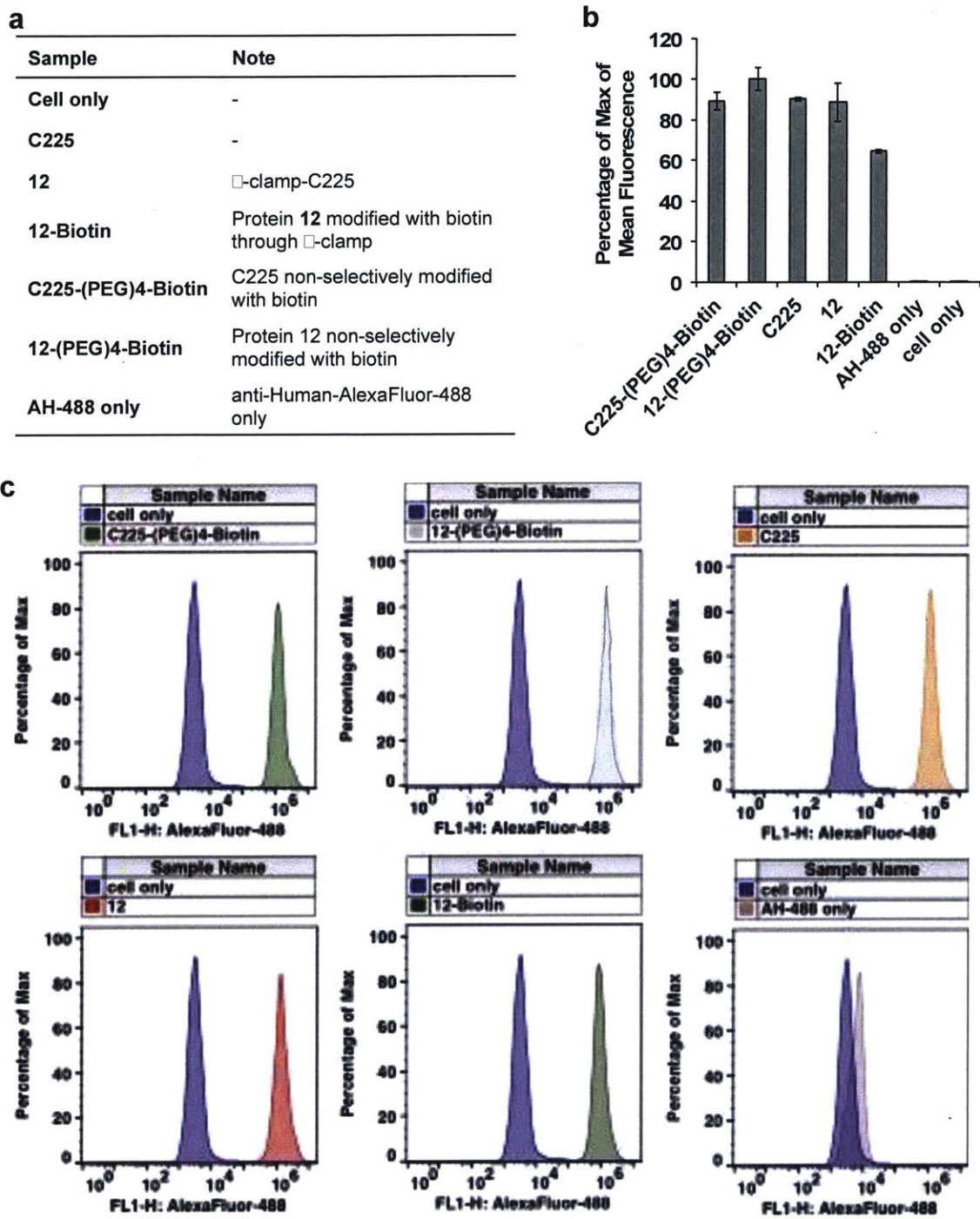
Figure 3.34. Flow cytometry assays for measuring the binding of modified C225 variants to A431 cells (EGFR positive) through secondary staining of antibody Fc region.

Antihuman Fc-AlexaFluor-488 as the secondary staining reagent. (**a**) Histograms of AlexaFluor-488 fluorescence. (**b**) Bar graph showing the mean AlexaFluor-488 fluorescence in each sample.
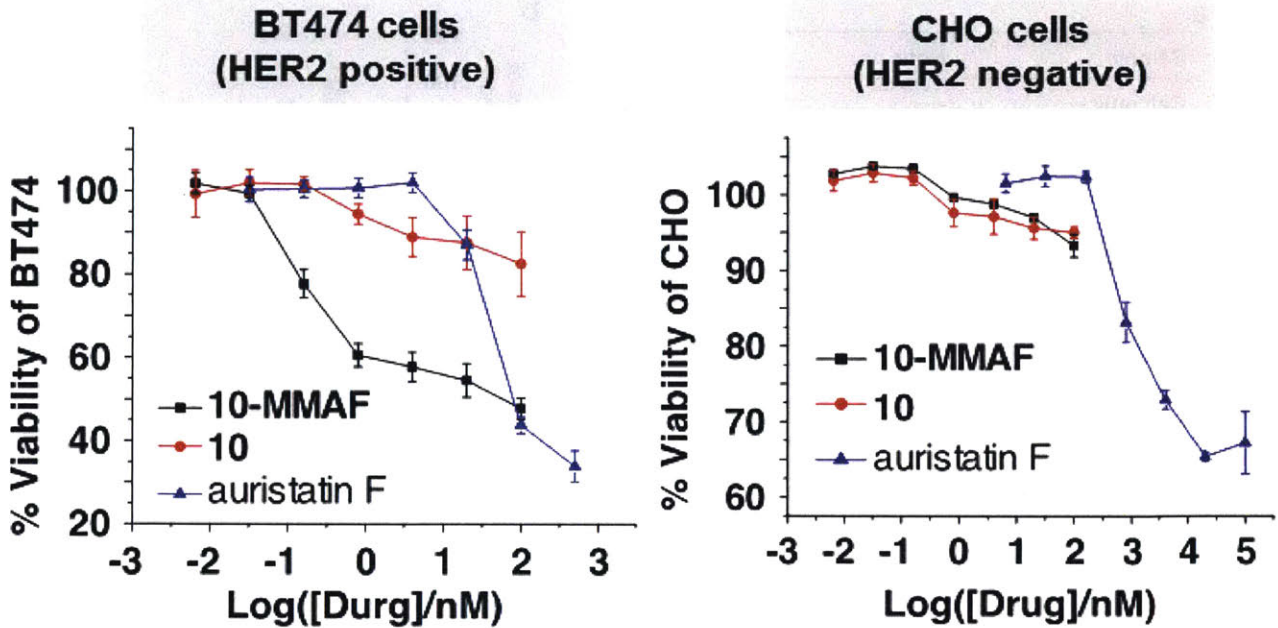
Figure 3.35. Cell viability assays showed selective killing of HER2 positive cells by π-clamp antibody-drug conjugates.

10-MMAF killed BT474 cells (HER2 positive) but was not effective for CHO cells (HER2 negative). EC_{50} values for BT474 cells were 0.19 nM for **10-MMAF** and 41 nM for auristatin F. EC_{50} value of auristatin F for CHO cells is 1.3 μM. Cell viability was quantified using CellTiter Glo assay and was normalized to cell only (see Experimental section 3.4.15 for detailed protocol). Experiments were done in triplicate for each dose; error bar shown indicates the standard deviation from the average of three experiments.

3.1.6. Cysteine pK_a is not the determining factor for π-clamp reactivity

To determine whether the increased arylation rate of π-clamp results from the decreased pK_a of cysteine inside the π-clamp sequence, we measured the pK_a of cysteines in the π-clamp peptide **1E** and in the double glycine mutant **1A** (Figure 3.36) according to previously reported method.[56] The π-clamp cysteine in **1E** has a pK_a of 7.69 $\pm$ 0.09 while the cysteine in **1A** has a pK_a of 8.30 $\pm$ 0.05. This decreased pK_a of cysteine in the π-clamp might result from the stabilization of thiolate by sulfur-π interaction with the phenyl rings.[57]

However, the decreased pK_a is not the major reason for the enhanced reactivity of π-clamp for perfluoroaryl probes. We performed competing arylation reactions for peptide **1A** and **1E** at pH 6.0 to 10.0. All reactions showed selective arylation of π-clamp peptide **1A** in the presence of competing cysteine species **1A** (Figure 3.37). This is in sharp contrast to the competing alkylation reactions with iodoacetamide, where all reactions showed non-selective alkylation of peptides **1A** and **1E** at pH 6.0 to 10.0 (Figure 3.38).
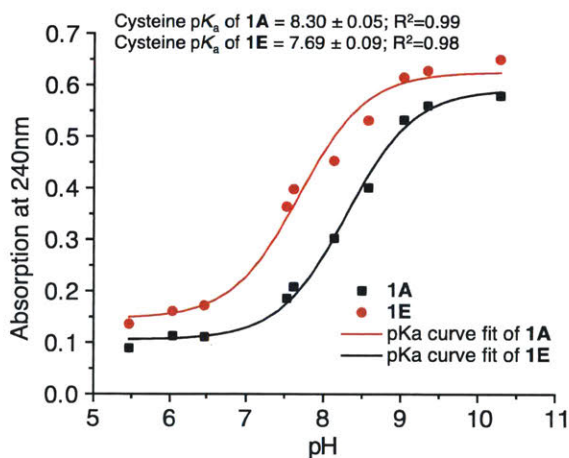
Figure 3.36. Determination of Cysteine pK_a of the π-clamp and the double glycine control by measuring absorbance at 240 nm across a range of pH values.

Conditions: 0.2 mM peptide, 0.1 M buffer. Following buffers and pH values are used: MES buffers at pH 5.47, 6.04, and 6.47; phosphate buffers at pH 7.53 and 7.62; Tris buffers at pH 8.00 and 8.50; borate buffers at pH 9.04 and 9.35; and CAPS buffer at pH 10.29. The UV absorption at 240 nm at each pH value was measured and the absorption of buffer-only was subtracted as background. The UV absorption values were then plotted against pH and the pK_a was determined by previously reported method.[56]
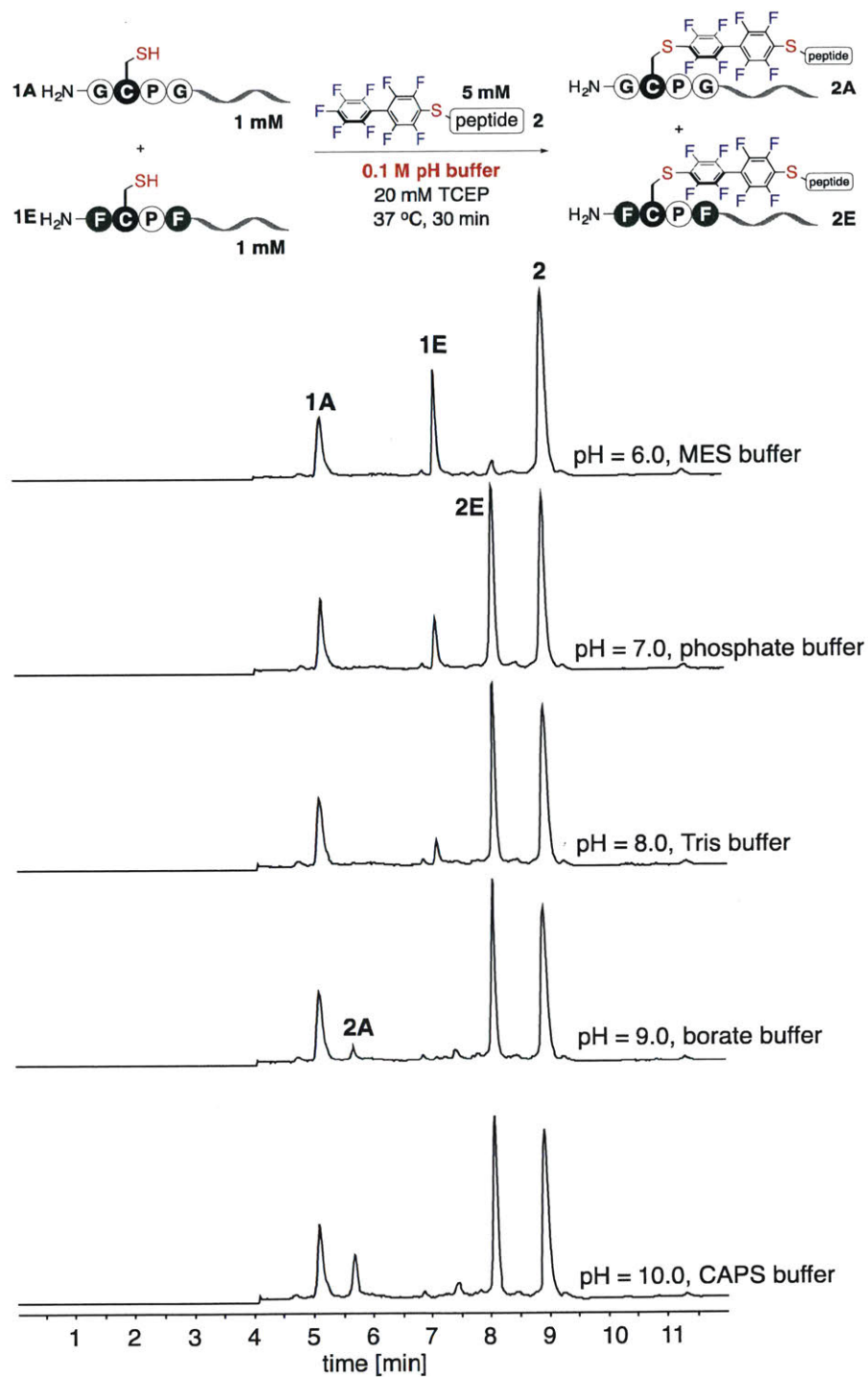
Figure 3.37. Selective arylation of the π-clamp cysteine in the presence of another competing cysteine across a range of pH values.

Reaction conditions: 1 mM **1A**, 1 mM **1E**, 5 mM **2**, 0.1 M buffer, 20 mM TCEP, at 37 °C for 30 minutes. 10 μL of each reaction was quenched by addition of 100 μL of 50% A: 50% B and analyzed by LC-MS.

214
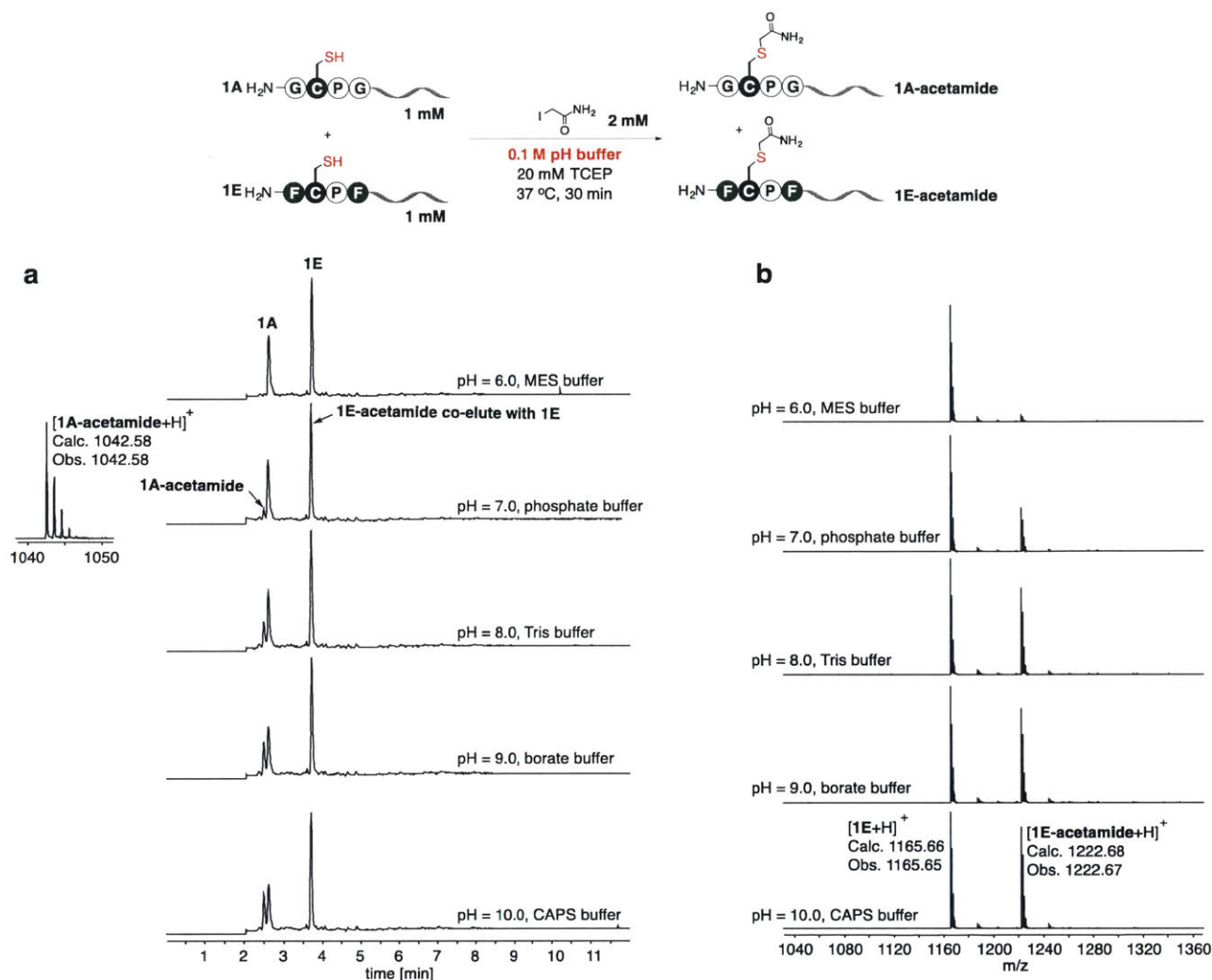
Figure 3.38. Reactions of the π-clamp cysteine and another competing cysteine across a range of pH values using iodoacetamide, showing that alkylation reaction is not selective. (**a**) LC-MS chromatograms for reactions at various pH values, the π-clamp alkylated product (**1E-acetamide**) co-eluted with peptide **1E**, mass spectrums of integration of the whole TIC peaks are shown in (**b**). Reaction conditions: 1 mM **1A**, 1 mM **1E**, 2 mM iodoacetamide, 0.1 M buffer, 20 mM TCEP, at room temperature for 10 minutes. 10 μL of each reaction mixture was quenched by addition of 100 μL of 50% A: 50% B and analyzed by LC-MS.

3.1.7. Molecular dynamics (MD) simulation of π-clamp structure

To further investigate the mechanism of the π-clamp mediated reaction, we used molecular dynamics (MD) to sample the conformational arrangements of the π-clamp peptide (**1E**) (Figure 3.39). Simulations indicated that **1E** adopts four primary conformations when a *cis*-Pro is present: a "π-clamp" (**S1**) with the phenyl rings of Phe-1 and Phe-4 interacting face-on with the Cys-2 thiol; a "half-clamp" (**S2**) where only the Phe-4 side chain interacts with the Cys-2 thiol; **S3** in which the Phe-1 and Phe-4 side chains are stacked together, leaving the Cys-2 thiol exposed; and an open configuration (**S4**) where all side chains are too far apart to interact. MD simulation for π-clamp peptide (**1E**) with a *trans*-Pro indicated two "open" structures with the cysteine thiol not interacting a Phe residue and one structure with Phe-4 side chain interacting with Cys-2 thiol (Figure 3.40).
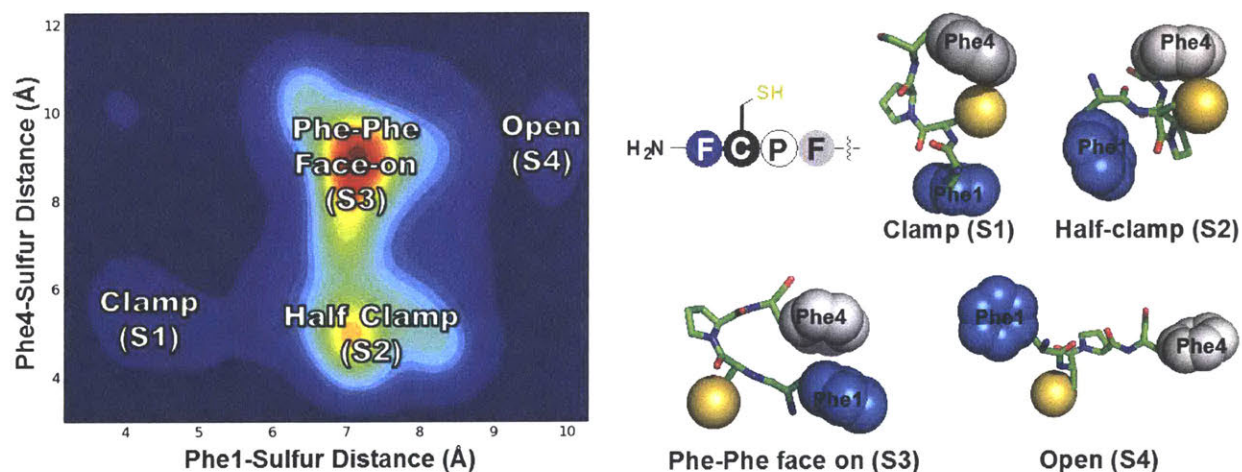
Figure 3.39. Molecular dynamics simulation reveals the structural landscape of π-clamp with a *cis*-proline configuration.

Four primary structures **S1 – S4** were identified from molecular dynamics simulation of π-clamp peptide **1E**. The phenyl rings and cysteine thiol are shown as spheres; the rest of the peptide is drawn as sticks. The heat map was plotted by summarizing all the obtained structures according to the distances between the sulfur atom and the centers of the phenyl rings.
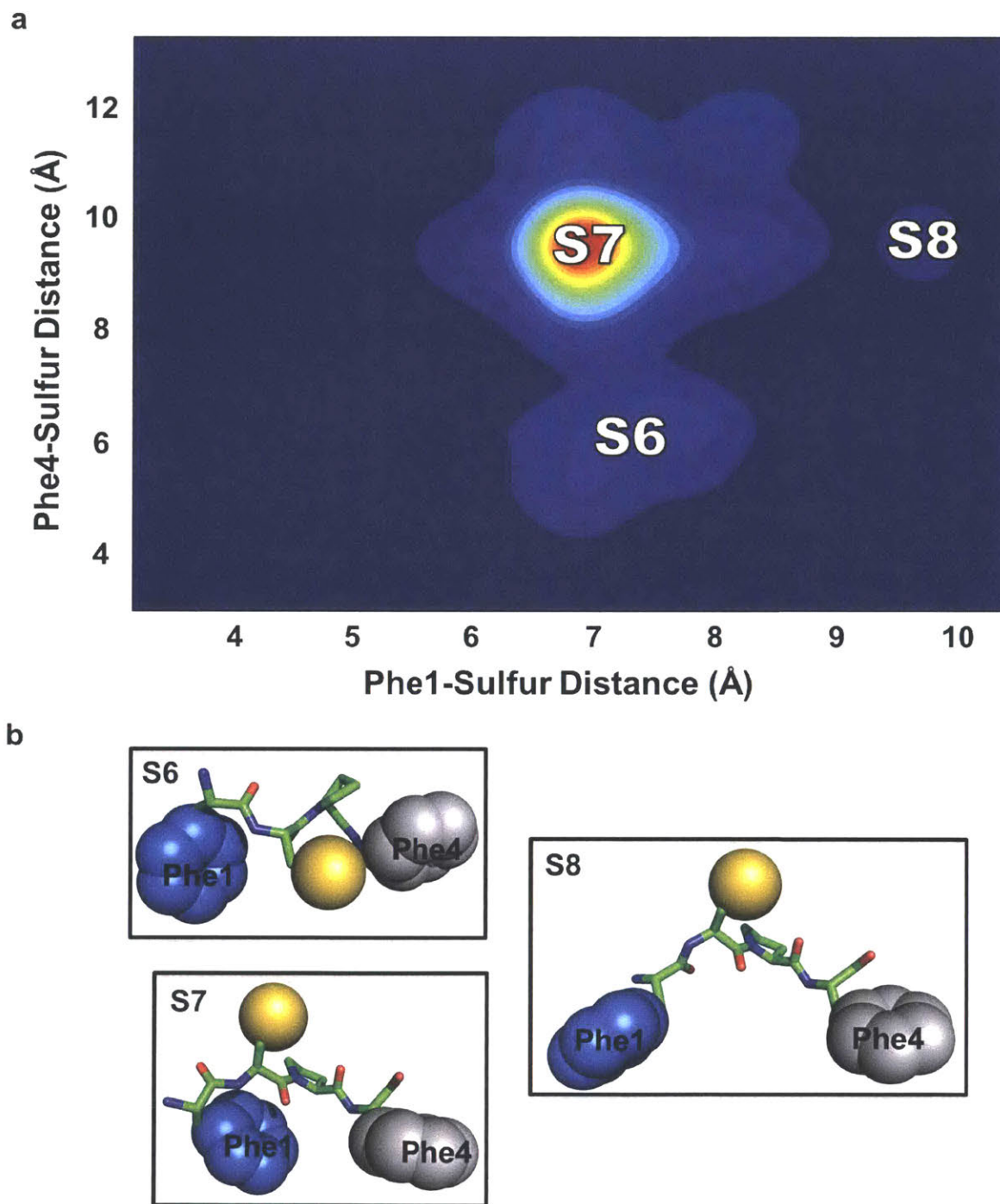
Figure 3.40. Molecular dynamics simulation reveals the structural landscape of π-clamp with a *trans*-proline configuration.

Three primary structures **S5 - S7** (**b**) were identified from the heat map sampling MD structures of peptide **1E** with a trans-proline (**a**). **S5** has Phe4 residue interacts with the sulfur, while **S6** and **S7** show no interactions between the sulfur and the phenylalanine rings.

217

3.1.8. Density functional theory (DFT) calculation of π-clamp mechanism

With these MD structures in hand, we used density functional theory (DFT) to investigate the nucleophilic aromatic substitution energy pathway for structures with a *cis*-Pro (Figure 3.41). We found that the half-clamp structure **S2** stabilized the arylation product by approximately 5 kcal/mol compared to the double glycine mutant, indicating the important role of Phe-4 in promoting the arylation reaction. This is consistent with our mutation studies showing that Phe-4 alone can partially mediate the arylation reaction (Figure 3.4, Entry 3). The product generated from the open structure (**S4**) has similar free energy compared to that of the double glycine mutant, further substantiating the hypothesis that the two phenylalanine side chains are important for the arylation reaction with the perfluoroaryl groups.

The most stable product was observed with the "π-clamp" structure (**S1**) of which the free energy was approximately 7 kcal/mol lower than that of the double glycine mutant. We further found that the activation energy for the formation of the transition state[58] (**III** in Figure 3.41) was decreased by approximately 3 kcal/mol when the π-clamp (**S1**) was present, presumably because of the phenyl rings recognizing the perfluoroaryl group and activating cysteine sulfur before conjugation. Collectively, these DFT calculations indicated that the π-clamp offers both a kinetic advantage (lower activation energy) and a thermodynamic advantage (lower free energy) over the double glycine mutant for the selective reaction with the perfluoroaryl reagents.

218
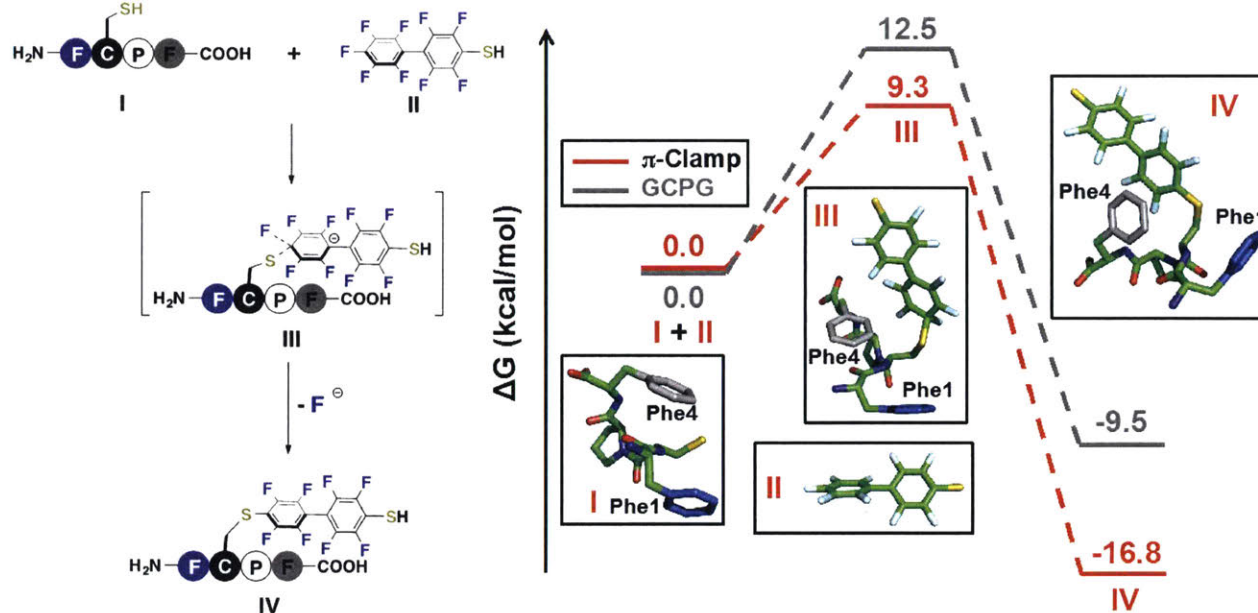
Figure 3.41. Density functional theory calculation of the reaction pathways of π-clamp and control.

Conjugation to the π-clamp is energetically favored compared to the double glycine mutant. Left, proposed nucleophilic aromatic substitution pathway for arylation at the π-clamp. Right, computed geometries and free energy surface of the nucleophilic aromatic substitution at the π-clamp (red). The free energy surface of the double glycine control is also shown (grey).

3.3. Discussion and Conclusion

Here we describe the discovery of a π-clamp to mediate site-selective cysteine conjugation. The π-clamp is composed of natural amino acids and shares some essential features of large enzymes, yet it mediates a purely abiotic cysteine perfluoroarylation reaction. The π-clamp tunes the reactivity of a cysteine thiol in its "active-site", recognizes the perfluoroaromatic reaction partner, and decreases the activation energy for the reaction. In addition, the π-clamp has practical applications in protein labeling.[4] The reported reaction is site-specific, operational under physiologically relevant conditions, enzyme-free, and as efficient as the commonly used azide-alkyne click chemistry[59,60] (π-clamp rate constant: $0.73~M^{-1}S^{-1}$).

Compared to existing bioconjugation techniques[38], the advantages of the π-clamp include its (1) small size that offers minimal structural perturbation to the target protein; (2) genetic encodability for straightforward incorporation; (3) ability to perform protecting-group-free dual cysteine modification; (4) and reaction mode that tunes the kinetic parameters to favor the cysteine perfluoroarylation reaction. This mode of reaction is distinct when compared to other advanced cysteine bioconjugations that use entropy to favor conjugation.[40–42]

The unexpected mode of site-specificity provided by the π-clamp requires further mention. In all existing conjugation methods[38], selectivity results from the judicious choice of certain functional groups so that each reaction pair undergoes conjugation in the presence of many other potentially reactive groups. For example, the unnatural handles used for click reactions are orthogonal to other functional groups on the target of interest.[12] In contrast, selectivity in the π-clamp mediated conjugation is achieved by fine-tuning the local chemical environment and reactivity as proteins do. This provides a complementary strategy to non-natural amino acid-mediated bioconjugation.[61] By fine-tuning the peptide microenvironment to allow for selective modification, the π-clamp significantly expands the chemistry available for selectively tailoring biomolecules.

3.4. Experimental

3.4.1. Chemicals

Decafluorobiphenyl was purchased from Oakwood Chemicals (West Columbia, SC). *Tris*(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) was purchased from Hampton Research (Aliso Viejo, CA). 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), Fmoc-Rink amide linker, D-Biotin, fluorescein isothiocyanate isomer I, Fmoc-18-amino-4,7,10,13,16-pentaoxaoctadecanoic acid (Fmoc-(PEG)$_5$-COOH), Fmoc-L-Propargylglycine-OH, Fmoc-L-Gly-OH, Fmoc-L-Leu-OH, Fmoc-L-Ile-OH, Fmoc-L-Val-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Ala-OH, Fmoc-L-Cys(Trt)-OH, Fmoc-L-Cys(*S*-tBu)-OH, Fmoc-L-Gln(Trt)-OH, Fmoc-L-Asn(Trt)-OH, Fmoc-L-Glu(OtBu)-OH, Fmoc-L-Asp(OtBu)-OH, Fmoc-L-Arg(Pbf)-OH, Fmoc-L-Phe-OH, Fmoc-L-Trp(Boc)-OH, Fmoc-His(Boc)-OH, Fmoc-L-Ser(tBu)-OH, Fmoc-L-Thr(tBu)-OH, Fmoc-L-Tyr(tBu)-OH, Fmoc-L-Pro-OH, and Fmoc-D-Pro-OH were purchased from Chem-Impex International (Wood Dale, IL). Aminomethyl polystyrene resin was prepared in house. Peptide synthesis-grade *N*, *N*-dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were obtained from VWR International (Philadelphia, PA). All other reagents were purchased from Sigma-Aldrich and used as received unless otherwise noted.

3.4.2. Peptide synthesis

All peptides were synthesized on a 0.2 mmol scale using manual Fmoc-SPPS chemistry under flow using a 3 minute cycle for each amino acid.[62] Specifically, all reagents and solvents are delivered to a stainless steel reactor containing resins at a constant flow rate using HPLC pump; temperature of the reactor was maintained at 60 °C during the synthesis using water bath. Procedure for each amino acid coupling cycle included a 30 second coupling with 1 mmol Fmoc-protected amino acid, 1 mmol HBTU, and 500 μL of diisopropyl ethyl amine (DIEA) in 2.5 mL of DMF at a flow rate of 6 mL/min (note that for coupling of cysteine and tryptophan, 190 μL of DIEA was used to prevent racemization); 1 minute wash with DMF at a flow rate of 20 mL/min; 20 second deprotection with 50% (v/v) piperidine in DMF at a flow rate of 20 mL/min; and 1 minute wash with DMF at a flow rate was 20 mL/min. After completion of the stepwise SPPS, the resin was washed thoroughly with DCM and dried under vacuum. The peptide is simultaneously

cleaved from the resin and side-chain deprotected by treatment with 2.5% (v/v) water, 2.5% (v/v) 1,2-ethanedithiol (EDT), and 1% (v/v) triisoproprylsilane in neat trifluoroacetic acid (TFA) for 2 hours at room temperature. The resulting solution containing peptide was evaporated by blowing a stream of nitrogen gas over its surface for 15 minutes, then triturated and washed with cold diethyl ether three times. The obtained gummy-like solid was dissolved in 50% H_2O: 50% acetonitrile containing 0.1% TFA and lyophilized. *These same solvent compositions were used in majority of experiments and will be referred to as A: 0.1% TFA in H_2O and B: 0.1% TFA in acetonitrile.*

3.4.3. Peptide purification

The crude peptide was dissolved in 95% A: 5% B with 6 M guanidinium hydrochloride and purified by semi-preparative RP-HPLC (Agilent Zorbax SB C_{18} column: 21.2 x 250 mm, 7 µm, linear gradient: 5-50% B over 90 min, flow rate: 5 mL/min). 1 µL of each HPLC fraction was mixed with 1 µL of alpha-cyano-4-hydroxycinnamic acid (CHCA) matrix in 75% A: 25% B, spotted with MALDI, and checked for fractions with desired molecular mass. The purity of fractions was confirmed by analytical RP-HPLC (Agilent Zorbax SB C_3 column: 2.1 x 150 mm, 5 µm, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min). HPLC fractions containing only product materials were confirmed by LC-MS analysis, combined, and then lyophilized. Peptides synthesized using fast flow-based SPPS and purified by RP-HPLC are listed in Table 3.1.

Table 3.1. Sequences and masses for peptides synthesized by fast-flow peptide synthesizer.

Peptide	Sequence	Calculated mass	Observed mass
1A	H_2N-GCPGGLLKNK-$CONH_2$	984.55	984.55
1B	H_2N-FCPGGLLKNK-$CONH_2$	1074.60	1074.60
1C	H_2N-GCPFGLLKNK-$CONH_2$	1074.60	1074.60
1D*	H_2N-FC(^DP)FGLLKNK-$CONH_2$	1164.65	1164.65
1E	H_2N-FCPFGLLKNK-$CONH_2$	1164.65	1164.65
1F	H_2N-FCPYGLLKNK-$CONH_2$	1180.64	1180.64
1G	H_2N-FCPWGLLKNK-$CONH_2$	1203.66	1203.66
1H	H_2N-YCPFGLLKNK-$CONH_2$	1180.64	1180.64
1I	H_2N-YCPYGLLKNK-$CONH_2$	1196.64	1196.64
1J	H_2N-YCPWGLLKNK-$CONH_2$	1219.65	1219.65
1K	H_2N-WCPFGLLKNK-$CONH_2$	1203.66	1203.66
1L	H_2N-WCPYGLLKNK-$CONH_2$	1219.65	1219.65
1M	H_2N-WCPWGLLKNK-$CONH_2$	1242.67	1242.67
1N	H_2N-KNKLLGFCPF-$CONH_2$	1164.65	1164.65
1O	H_2N-KNKLLGFCPFGLLKNK-$CONH_2$	1818.07	1818.07
2-Cys	H_2N-VTLPSTCGAS-$CONH_2$	933.46	933.46
3-Cys[#]	Biotin-RRC-$CONH_2$	658.32	658.32
3'-Cys[#]	Biotin-ENLYFQGCKKK-$CONH_2$	1581.78	1581.78
4-Cys[†]	FITC-(β-A)-GLRLKNKC-$CONH_2$	1389.63	1389.63
5-Cys[‡]	H_2N-Pra-TLPSTCGAS-$CONH_2$	929.43	929.43
7-Pep	H_2N-GGGGGNKRENLYFQGFCPF-$CONH_2$	2045.95	2045.95
8-Pep	H_2N-FCPWGLPSTGG-$CONH_2$	1119.52	1119.52

*^DP represents D-proline.

[#]Biotin was installed as the last amino acid under flow.

[†]FITC was installed under batch conditions; β-A represents β-alanine.

[‡]Pra represents L-propargylglycine.

3.4.4. LC-MS analysis

LC-MS chromatograms and associated mass spectra were acquired using Agilent 6520 ESI-Q-TOF mass spectrometer. Following LC-MS methods were used:

Method A LC conditions: Zorbax 300SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300 – 3000 *m/z*, temperature of drying gas = 350 °C, flow rate of drying gas = 11 L/min, pressure of nebulizer gas = 60 psi, the capillary, fragmentor, and octupole rf voltages were set at 4000, 175, and 750, respectively.

Method B LC conditions: Zorbax 300SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 minutes 5% B, 2-21 minutes 5-65% B, 21-22 minutes 65% B, flow rate: 0.8 mL/min. MS conditions are same as *Method A*.

Method C LC conditions: Zorbax 300SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 minutes 5% B, 2-21 minutes 5-95% B, 21-22 minutes 95% B, flow rate: 0.8 mL/min. MS conditions are same as *Method A*.

Method D LC conditions: Zorbax 300SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 75 °C, gradient: 0-2 minutes 5% B, 2-11 minutes 5-65% B, 11-12 minutes 65% B, flow rate: 0.8 mL/min. MS conditions: positive electrospray ionization (ESI) high mass mode in mass range 1000 – 7000 *m/z*, temperature of drying gas = 350 °C, flow rate of drying gas = 10 L/min, pressure of nebulizer gas = 20 psi, the capillary, fragmentor, and octupole rf voltages were set at 5000, 350, and 750, respectively.

Data were processed using Agilent MassHunter software package. Deconvoluted masses of proteins were obtained using maximum entropy algorithm.

LC-MS data shown were acquired using Method A unless noted. Y-axis in all chromatograms shown in supplementary figures represents total ion current (TIC) unless noted; mass spectrum insets correspond to the integration of the TIC peak unless noted.

3.4.5. Determination of reaction yields

All yields reported were determined by integrating TIC spectra. First, using Agilent MassHunter software package, the peak area for all relevant peptidic species on the chromatogram were integrated. Because no side product was generated in all experiments, the conversion of the limiting reagent equals to the yield of the product. Then the yield was calculated as following: %yield = %conversion = 1 - S_t/S_0 where S_t is the peak area of the limiting reagent at time t, and S_0 is the peak area of the limiting reagent at time 0.

3.4.6. Preparation of S-perfluoroarylated electrophiles

Probes **2** and **3** are prepared as previously reported.[47]

Synthesis of Probes 4–6

To 20 μmoles of solid sample of thiol-containing probe (**4-Cys**, **5-Cys**, or **PEG2000-thiol**) dissolved in 100 mM NEt3 in 1mL of DMF in a plastic Eppendorf tube was added 800 μmoles of decafluorobiphenyl. The tube was vortexed and sonicated to ensure complete reagent mixing and dissolution, the reaction mixture was left at room temperature for 30 minutes. 1 μL of each reaction mixture was quenched by addition of 20 μL of 50% A: 50% B and was then analyzed by LC-MS. Resulting reaction mixtures were quenched by addition of 20 mL of 95% A: 5% B, excess decafluorobiphenyl was precipitated as white solid. The resulting sample was centrifuged at 4,000 rpm for 10 minutes. The supernatant was filtered through 0.22 μm nylon syringe filter, and purified by RP-HPLC.

LC-MS data for all starting materials and HPLC-purified products of probes **4 – 6** were shown in Figure 3.42.
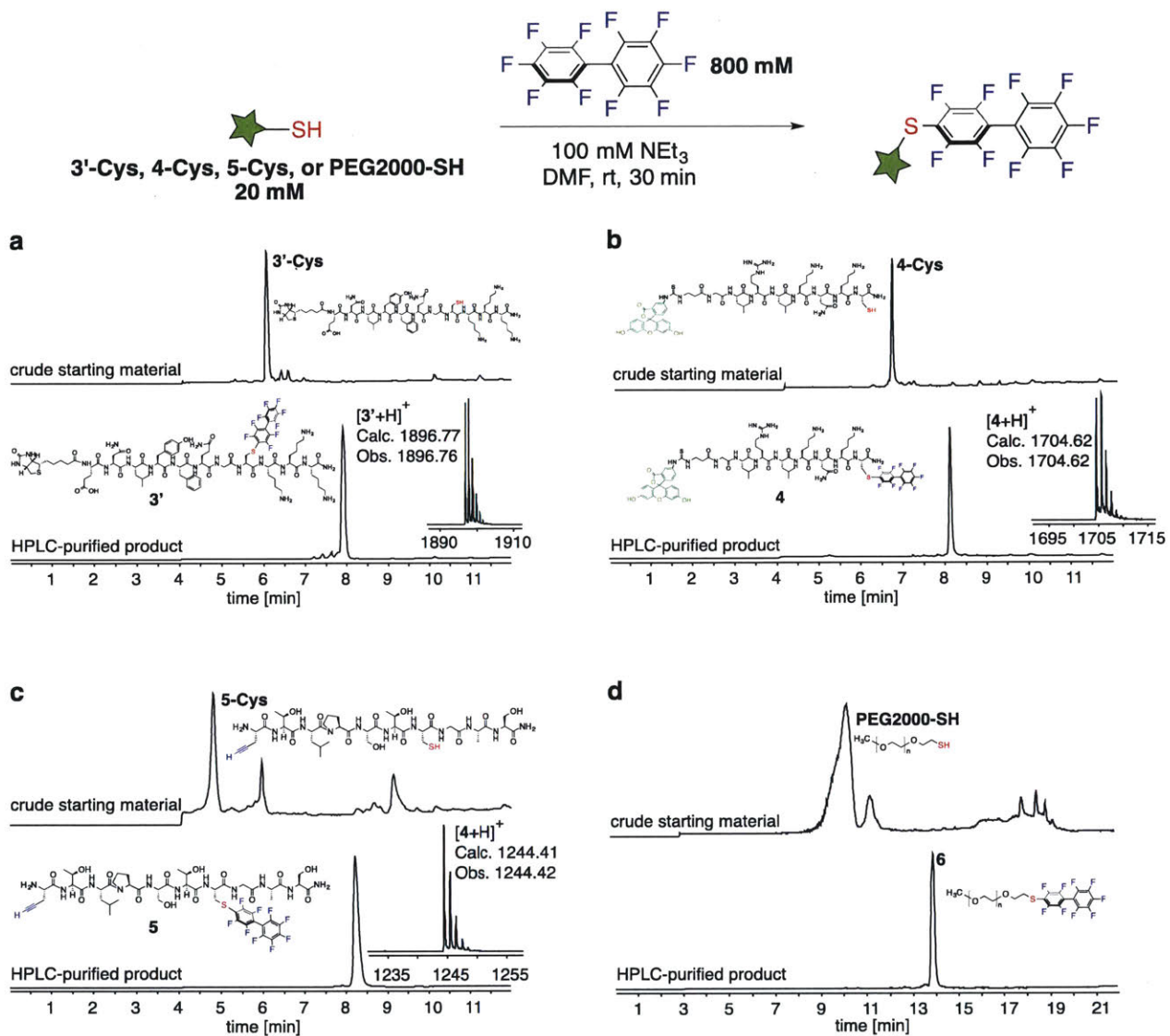
Figure 3.42. Synthesis of perfluoroaryl probes.

Shown are LC-MS chromatograms for crude starting material and HPLC-purified products of probes (**a**) **3'**, (**b**) **4**, (**c**) **5**, and (**d**) **6**. Data in (**d**) were acquired using LC-MS *Method C*.

Synthesis of 11-Biotin

The peptidic part (VTLPSTCGAS) of **11-Biotin** was synthesized under flow using a Fmoc-Cys(S-tBu)-OH amino acid. The (PEG)$_5$ linker and the biotin moiety were incorporated by coupling Fmoc-(PEG)$_5$-COOH and D-biotin to the resin under batch SPPS conditions (Figure 2.4a).

The perfluoroaryl group was installed using an on-resin perfluoroarylation procedure (Fig. S29a). The S-tBu protecting group was removed on-resin by incubation of 300 mg of resin in 5 mL of DMF solution containing 280 mg of dithiothreitol (DTT) and 190 μL of triethylamine at 60 °C for 5 minutes. The resin was washed 4 times with DMF. Then 5 mL of DMF solution containing 450 mg of decafluorobiphenyl and 190 μL of triethylamine was added, and the reaction mixture was rotated at room temperature for 1 hour. The resin was washed 4 times with DMF, 4 times with DCM, and then dried *in vacuo*. The resulting product **11-Biotin** was cleaved from the resin and purified by RP-HPLC (Figure 2.4b) following standard procedures.
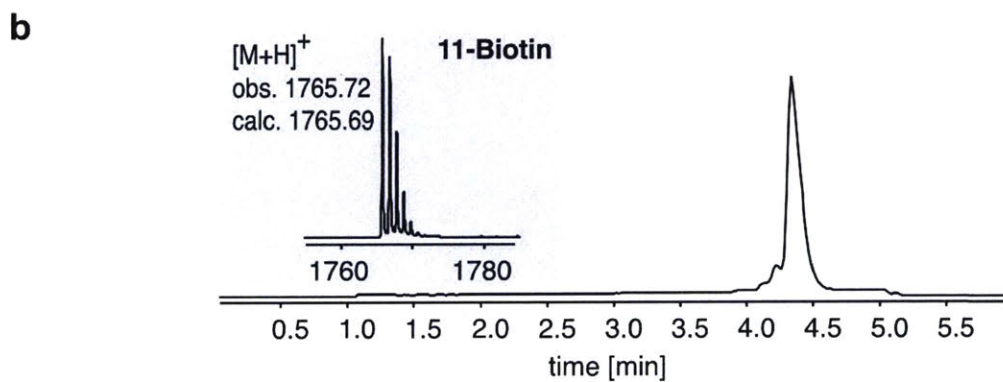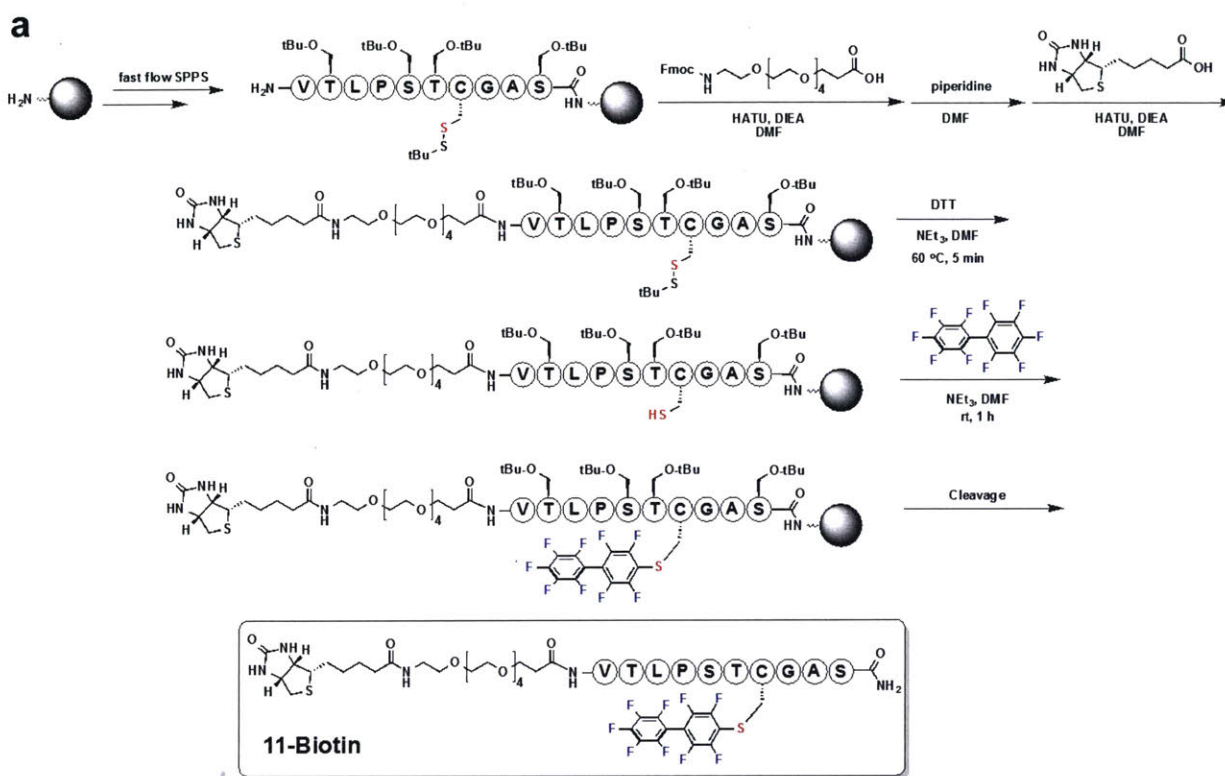
Figure 3.43. Synthesis of 11-Biotin.

(a) Synthetic route. (b) LC-MS analysis of the purified product. Total ion current (TIC) chromatogram was shown and the mass spectrum of the highest point of the TIC peak was shown as insets.

Synthesis of 11-MMAF

The peptidic part (VTLPSTCGAS) of **11-MMAF** was synthesized under flow using a Fmoc-Cys(*S*tBu)-OH amino acid. The (PEG)$_5$ linker and the Arg-Cys moiety were incorporated by coupling Fmoc-(PEG)$_5$-COOH, Fmoc-Arg(Pbf)-OH, and Fmoc-Cys(Trt)-OH under batch SPPS conditions (Figure 3.44a).

The perfluoroaryl group was installed using an on-resin perfluoroarylation procedure (Figure 3.44a). The *S*-tBu protecting group was removed on-resin by incubation of 300 mg resin in 5 mL of DMF solution containing 280 mg of dithiothreitol (DTT) and 190 μL of triethylamine at 60 °C for 5 minutes. The resin was washed 4 times with DMF. Then 5 mL of DMF solution containing 450 mg of decafluorobiphenyl and 190 μL of triethylamine was added, and the reaction mixture was rotated at room temperature for 1 hour. The resin was washed 4 times with DMF, 4 times with DCM, and then dried *in vacuo*. The resulting product **11-Cys** was cleaved from the resin and purified by RP-HPLC following standard procedures.

MMAF was linked to the perfluoroaryl moiety by reacting **11-Cys** with maleimide-MMAF (Concortis Biosystems). To a solution of 3.6 mg **11-Cys** in 500 μL Tris buffer (0.1 M, pH 8.0) was added 20 μL of maleimide-MMAF (20 mM in DMSO) and 500 μL of acetonitrile. The reaction mixture was vortexed for 20 seconds to allow proper mixing and was left at room temperature for 10 minutes. 20 μL of glutathione (250 mM in water) was added to quench the excess maleimide-MMAF. The resulting mixture was directly purified by RP-HPLC to afford **11-MMAF** (Figure 3.44b).
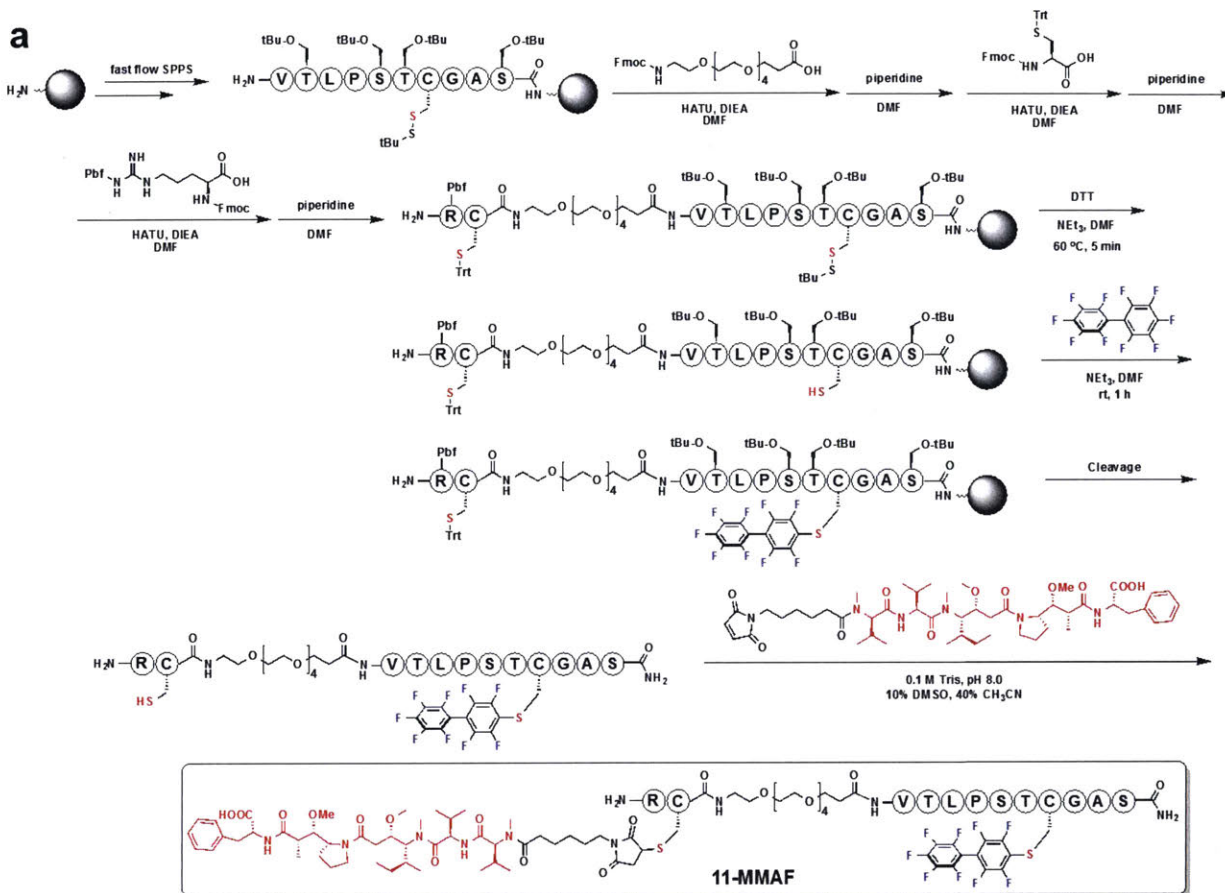
Figure 3.44. Synthesis of 11-MMAF.

(a) Synthetic route. (b) LC-MS analysis of the purified product. Total ion current (TIC) chromatogram was shown and the mass spectrum of the highest point of the TIC peak was shown as insets.

230

3.4.7. Identification of the Phe-Cys-Pro-Trp sequence through library selection

Split-Pool Synthesis

All split-pool syntheses were performed using batch SPPS conditions. Peptides were synthesized on aminomethyl polystyrene resins. 1 gram of resin was used for the library synthesis. The *C*-terminal fixed region (Gly-Leu-Leu-Lys-Rink) was synthesized using the following protocol for each amino acid. Amino acids and the rink amide linker were incorporated into peptide sequence through a cycle of coupling, wash, deprotection, and wash steps. Procedure for the coupling of amino acid included a 20 minutes coupling with 5 mmol Fmoc-L-amino acid, 5 mmol HATU, and 2.5 □L of DIEA in 12.5 mL of DMF. After coupling, the resin was washed with DMF (4 times). 20 mL of piperidine (20% (v/v) in DMF) was added to the resin for 10 minutes (twice: 5 min each). The resin was washed with DMF (4 times) and then subjected to coupling of the next amino acid. The DIEA amount for coupling of the fixed cysteine was decreased to 950 µL to prevent racemization.

For coupling of each amino acid in the variable region, the resins were first split into 19 portions into 10 mL fritted syringes. 0.2 mmol Fmoc-L-amino acid, 0.2 mmol HATU, and 100 µL of DIEA in 500 µL of DMF were added to the resins for coupling of each amino acid. The DIEA amount for coupling of tryptophan and histidine was decreased to 38 µL to prevent racemization. After coupling, the resins were washed with DMF (4 times) and were combined. 20 mL of piperidine (20% (v/v) in DMF) was added to the resin for 10 minutes (twice: 5 min each). The resin was washed with DMF (4 times) and then subjected to coupling of the next amino acid.

After the spilt-pool synthesis, the resins were washed 4 times with DCM and were dried under vacuum. The crude peptide was obtained using previously mentioned cleavage protocol (see the Peptide Synthesis section). The lyophilized crude peptide library was used directly in reactions with biotin-perfluoroaryl probe without any further purification.

Library Reaction and Selection

200 µL of the peptide library (1 mg/mL in 0.5 M phosphate buffer, 50 mM TCEP, pH 8.0) was incubated with the probe **3'** (1 mM) at room temperature for 30 minutes. After the reaction, streptavidin magnetic beads (Pierce Biotechnology, 100 µL in PBS) were added to the resulting solution to capture the reacted peptide hits. After incubation at room temperature for 30 minutes,

the beads were washed 3 times with phosphate buffer (0.1 M, pH 8.0), 3 times with TEV protease buffer (50 mM Tris, 0.5 mM EDTA, 1 mM DTT, pH 8.0). The beads were incubated with 500 µL of TEV protease (0.1 mg/mL in TEV protease buffer) overnight at room temperature. The supernatant was collected and was subjected to LC-MS/MS analysis for peptide sequencing. The compound mass list was generated using Agilent MassHunter software package. The identified compounds were manually matched to the mass of possible peptide sequences generated using Microsoft Excel.

First round of selection generated three possible hits with sequences Phe-Cys-Pro-Trp, Phe-Cys-Ile/Leu-Ile/Leu, and Phe-Cys-Ile/Leu-Met. We resynthesized these three hits and only Phe-Cys-Pro-Trp was identified to be the positive hit with the improved reactivity with the perfluoroaryl electrophile (Figure 3.3). This sequence was investigated further as described in Results section 3.1.2.

3.4.8. Preparation of protein 7

Protein **7** was prepared via sortagging reaction between 1C-LF$_N$-DTA-LPSTGGHis$_5$ (**7-pro**) and G5-TEVsite-π-clamp peptide (**7-pep**). The synthetic scheme and results are shown in Figure 3.45. Experimental protocols are described below.

Expression and Purification of 1C-LF$_N$-DTA-LPSTGG- His$_5$ (7-pro) in E. Coli

pET-SUMO-LF$_N$-DT$_A$-LPSTGG-His$_5$ plasmid was constructed as reported previously.[63] The *N*-terminal cysteine was introduced by site-directed mutagenesis using QuickChange Lightning Single Site-directed Mutagenesis Kit (Agilent) following the manufacturer's instructions. The generated pET-SUMO-1C-LF$_N$-DT$_A$-LPSTGG-His$_5$ construct encodes for the following protein sequence (LF$_N$ is highlighted in green, DT$_A$ is highlighted in blue, and *N*-terminal cysteine is highlighted in red):

1C-LF$_N$-DTA-LPSTGG-His$_5$

CGGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLKEIMKHIVKIEVKGEEAVKKEAAEKLLEKVPSDV
LEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDIYGKDALLHEHYVYAKEGYEPVLVIQSSEDYVEN
TEKALNVYYEIGKILSRDILSKINQPYQKFLDVLNTIKNASDSDGQDLLFTNQLKEHPTDFSVEFLEQNS
NEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKFNEQEINLSLEELKDQRSGRELERGADDVVDSSK
SFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGV

VKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYI
NNWEQAKALSVELEINFETRGKRGQDAMYEYMAQASAGNRLPSTGGHHHHH

E. coli BL21(DE3) cells transformed with pET-SUMO-1C-LF$_N$-DT$_A$-LPSTGG-His$_5$ plasmid were grown in 1 L of LB medium containing kanamycin (30 µg/mL) at 37 °C until OD$_{600}$ = 0.6. Then, expression was induced by addition of 0.5 mM IPTG overnight at 30 °C. After harvesting the cells by centrifugation (6,000 rpm for 10 min), the cell pellet was lysed by sonication in 25 mL of 50 mM Tris and 150 mM NaCl (pH 7.5) buffer containing 15 mg lysozyme (Calbiochem), 1 mg DNase I (Sigma-Aldrich), and 0.5 tablet of protease inhibitor cocktail (Roche Diagnostics, Germany). The suspension was centrifuged at 17,000 rpm for 30 min to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK), first washed with 40 mL of 20 mM Tris and 150 mM NaCl (pH 8.5), and then washed with 40 mL of 40 mM imidazole in 20 mM Tris and 150 mM NaCl (pH 8.5). The protein was eluted from the column with buffer containing 500 mM imidazole in 20 mM Tris and 150 mM NaCl (pH 8.5). Imidazole was removed from protein using a HiPrep 26/10 Desalting column (GE Healthcare, UK), the protein was eluted into 20 mM Tris and 150 mM NaCl (pH 7.5) buffer. The protein was analyzed by LC-MS to confirm its purity and molecular weight.

SUMO group on SUMO-1C-LF$_N$-DTA-LPSTGG-His$_5$ was cleaved by incubating 1 µg of SUMO protease per mg of protein at room temperature for 60 minutes. The crude reaction mixture was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK) and the flow through containing 1C-LF$_N$-DTA-LPSTGG-His$_5$ was collected. The protein was analyzed by LC-MS confirming sample purity and molecular weight.

Sortagging Reaction with G5-TEVsite-π-clamp peptide (7-pep)

Sortagging reaction was performed on a 250 µL scale using triple mutant sortase (SrtA*) evolved by *Chen, et al*[64]. Reaction conditions are: 100 µM **7-pro**, 1 mM **7-pep**, 5 µM SrtA*, SrtA* buffer (10 mM CaCl$_2$ in 50 mM Tris and 150 mM NaCl), 5 mM TCEP (pH 7.5). The reaction mixture was incubated at room temperature for 30 minutes, 1 µL of the reaction mixture was quenched by the addition of 20 µL of 50% A: 50% B and was analyzed by LC-MS to confirm the completion of the reaction. Then 100 µL of Ni-NTA was added, and the mixture was further incubated at room temperature for 15 minutes to remove SrtA* and G-His$_5$ fragment. The mixture

was centrifuged for 1 min at 17,000 rpm to remove Ni-NTA beads. The supernatant was buffer exchanged for three times with 20 mM Tris and 150 mM NaCl (pH 7.5) using Amicon Ultra concentrator (EMD-Millipore) to remove excess **7-pep**. The concentrated protein sample was analyzed by LC-MS to confirm its purity and molecular weight. Resulting protein **7** has the following sequence (LF$_N$ is highlighted in green, DT$_A$ is highlighted in blue, *N*-terminal cysteine is highlighted in red, and the TEV site is highlighted in orange)

Protein 7

CGGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLKEIMKHIVKIEVKGEEAVKKEAAEKLLEKVPSDV
LEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDIYGKDALLHEHYVYAKEGYEPVLVIQSSEDYVEN
TEKALNVYYEIGKILSRDILSKINQPYQKFLDVLNTIKNASDSDGQDLLFTNQLKEHPTDFSVEFLEQNS
NEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKFNEQEINLSLEELKDQRSGRELERGADDVVDSSK
SFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGV
VKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYI
NNWEQAKALSVELEINFETRGKRGQDAMYEYMAQASAGNRLPSTGGGGGNKRENLYFQGFCPF-CONH$_2$
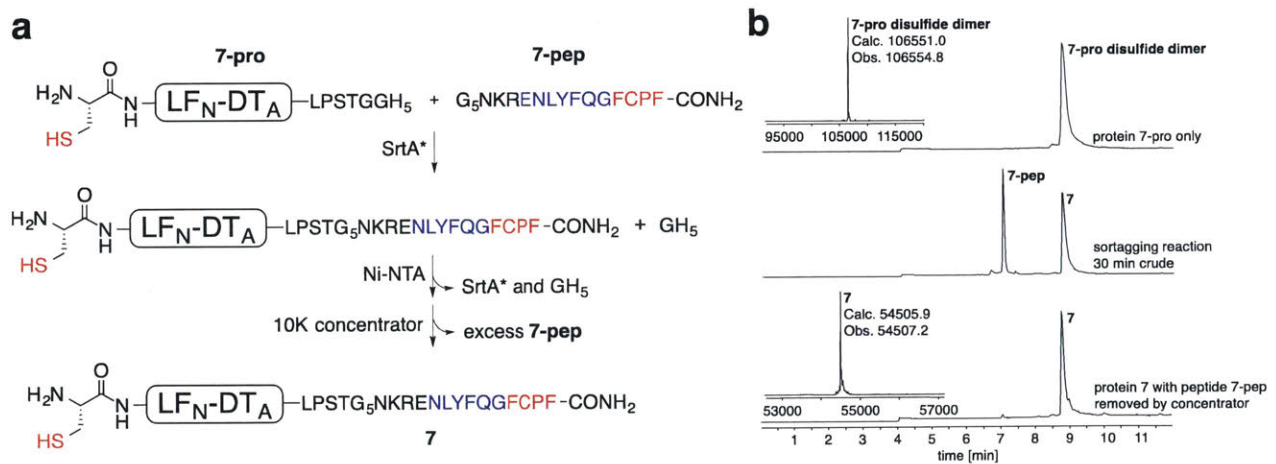
Figure 3.45. Preparation of protein 7.

(**a**) Scheme for synthesis of protein **7** via sortagging reaction. (**b**) LC-MS chromatograms and deconvoluted protein masses for starting material (top), crude sortagging reaction before purification (center), and purified protein **7** (bottom).

3.4.9. Expression and purification of Sortase A variants

pET-21b-SrtA-His$_6$ plasmid was constructed as reported previously.[65] pET-21b-FCPF-SrtA-His$_6$ plasmid with the π-clamp inserted between ^{5}Gly and ^{6}Gly was constructed by site-directed mutagenesis using QuickChange Lightning Single Site-directed Mutagenesis Kit (Agilent) following the manufacturer's instructions. Sequences of SrtA (**9**) π-Clamp SrtA (**8**) are (π-clamp is highlighted in blue and active-site cysteine of SrtA is highlighted in red):

SrtA (9)

ASMTGGQQMGRDPNSQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATSEQLNRGVSFAEENESLDDQN
ISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTDVEVLDEQKGKDKQLTLIT**C**
DDYNEKTGVWETRKIFVATEVKLEHHHHHH

π-Clamp SrtA (8)

ASMTG**FCPF**GQQMGRDPNSQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATSEQLNRGVSFAEENESL
DDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRNVKPTDVEVLDEQKGKDKQLT
LIT**C**DDYNEKTGVWETRKIFVATEVKLEHHHHHH

E. coli BL21(DE3) cells transformed with sortase plasmids were grown in 1 L of LB medium containing kanamycin (30 μg/mL) at 37 °C until OD$_{600}$ = 0.6. Then, expression was induced by addition of 0.5 mM IPTG overnight at 30 °C. After harvesting the cells by centrifugation (6,000 rpm, 10 min), the cell pellet was lysed by sonication in 25 mL of 50 mM Tris and 150 mM NaCl (pH 7.5) containing 15 mg lysozyme (Calbiochem), 1 mg DNase I (Sigma-Aldrich), and 0.5 tablet of protease inhibitor cocktail (Roche Diagnostics, Germany). The suspension was centrifuged at 17,000 rpm for 30 min to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK), first washed with 40 mL of 20 mM Tris and 150 mM NaCl (pH 8.5), and then washed with 40 mL of 40 mM imidazole in 20 mM Tris and 150 mM NaCl (pH 8.5). The protein was eluted from the column with buffer containing 500 mM imidazole in 20 mM Tris and 150 mM NaCl (pH 8.5). Imidazole was removed from the protein using a HiPrep 26/10 Desalting column (GE Healthcare, UK), protein was eluted into 20 mM Tris and 150 mM NaCl (pH 7.5). The proteins were analyzed by LC-MS to confirm their purity and molecular weight.

3.4.10. Expression and purification of antibodies

The gWiz-trastuzumab-HC-GFCPF and the gWiz-C225-HC-GFCPF plasmids were constructed by inserting the π-clamp at the *C*-terminus of the trastuzumab and C225 respectively following a glycine linker, using the QuickChange Lightning Single Site-directed Mutagenesis Kit (Agilent) per manufacturer's protocol. The light chain and heavy chain sequences for the trastuzumab, π-clamp-trastuzumab (**10**), C225, π-clamp-C225 (**12**) are listed below (the π-clamp is highlighted in blue and the trastuzumab cysteines are highlighted in red):

Trastuzumab-Light Chain

DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSGTD
FTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTRTVAAPSVFIFPPSDEQLKSGTASVVCLLNN
FYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKS
FNRGEC

Trastuzumab-Heavy Chain

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTI
SADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSG
GTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPS
NTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNW
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ
VYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW
QQGNVFSCSVMHEALHNHYTQKSLSLSPGK

π-Clamp-Trastuzumab (10)-Light Chain

DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSGTD
FTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTRTVAAPSVFIFPPSDEQLKSGTASVVCLLNN
FYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKS
FNRGEC

π-Clamp-Trastuzumab (10)-Heavy Chain

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTI
SADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSG

GTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPS
NTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNW
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ
VYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW
QGNVFSCSVMHEALHNHYTQKSLSLSPGKGFCPF

C225-Light Chain

DILLTQSPVILSVSPGERVSFSCRASQSIGTNIHWYQQRTNGSPRLLIKYASESISGIPSRFSGSGSGTD
FTLSINSVESEDIADYYCQQNNNWPTTFGAGTKLELKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFY
PREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFN
RGEC

C225-Heavy Chain

QVQLKQSGPGLVQPSQSLSITCTVSGFSLTNYGVHWVRQSPGKGLEWLGVIWSGGNTDYNTPFTSRLSIN
KDNSKSQVFFKMNSLQSNDTAIYYCARALTYYDYEFAYWGQGTLVTVSAASTKGPSVFPLAPSSKSTSGG
TAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSN
TKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNW
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ
VYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ
QGNVFSCSVMHEALHNHYTQKSLSLSPGK

π-clamp-C225 (12)-Light Chain

DILLTQSPVILSVSPGERVSFSCRASQSIGTNIHWYQQRTNGSPRLLIKYASESISGIPSRFSGSGSGTD
FTLSINSVESEDIADYYCQQNNNWPTTFGAGTKLELKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFY
PREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFN
RGEC

π-clamp-C225 (12)-Heavy Chain

QVQLKQSGPGLVQPSQSLSITCTVSGFSLTNYGVHWVRQSPGKGLEWLGVIWSGGNTDYNTPFTSRLSIN
KDNSKSQVFFKMNSLQSNDTAIYYCARALTYYDYEFAYWGQGTLVTVSAASTKGPSVFPLAPSSKSTSGG
TAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSN
TKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNW

VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYK**C**KVSNKALPAPIEKTISKAKGQPREPQV

YTLPPSRDELTKNQVSLT**C**LVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ

QGNVFS**C**SVMHEALHNHYTQKSLSLSPGKG**FCPF**

The antibodies were expressed via transient transfections of HEK293F cells (Invitrogen) as previously described,[54] and purified using Protein A affinity chromatography (Genscript) following manufacturer's instructions. The purified IgGs were analyzed by LC-MS to confirm their molecular weight and purity, and stored in phosphate buffer saline (PBS) at -80 °C. Mass spectrometry analysis of the intact antibody indicates that the π-clamp-C225 (**12**) is expressed with the π-clamp cysteine residues capped by a disulfide bond with free cysteines (Figure 3.46).
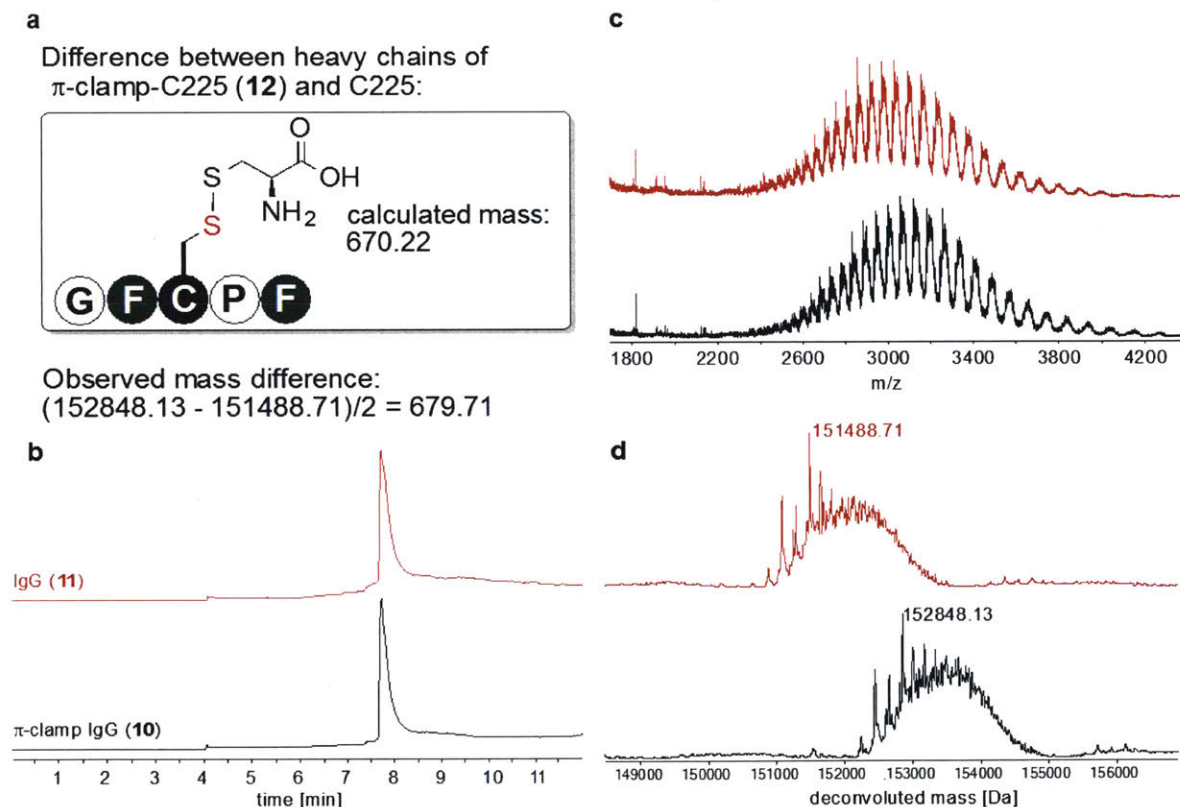
Figure 3.46. LC-MS analysis of intact C225 antibodies.

(**a**) Observed mass difference between intact IgG (**11**) and π-clamp IgG (**10**) matches the calculated mass difference assuming that the π-clamp cysteines are capped by disulfide bond with free cysteines. (**b**) TIC spectrum for IgGs analyzed. (**c**) Mass spectra obtained by integration of the whole TIC peaks. (**d**) deconvoluted mass spectra for IgGs. Data were acquired using LC-MS *Method D*.

238

3.4.11. Kinetic studies on π-clamp and controls

All reactions are performed on a 10-μL scale. For each peptide, reactions are quenched by addition of 100 μL of 50% A: 50% B at time point 0 min, 2 min, 5 min, 10 min, 20 min, and 30 min and then analyzed by LC-MS. TIC chromatograms for all reactions were shown in Figure 3.6 to Figure 3.12, Figure 3.15, and Figure 3.16.

The reaction yield at each time point was calculated as described in section 1e. The second-order rate constants (k_2) were determined by fitting the data to the following equation:

$$y = \frac{\left(ln \frac{[peptide]_0[probe]_t}{[peptide]_t[probe]_0}\right)}{([probe]_0 - [peptide]_0)} = k_2 t$$

where $[peptide]_0$ and $[probe]_0$ are the initial concentrations of the peptide and the probe; and $[peptide]_t$ and $[probe]_t$ are the concentrations of the peptide and the probe at time t.

All kinetic curves generated using OriginPro 8.0 software package are summarized in Figure 3.5.

3.4.12. Non-selective biotinylation of antibodies

Trastuzumab, C225, π-clamp-trastuzumab (**10**), and π-clamp-C225 (**12**) were linked to biotin through reactions with Biotin-(PEG)4-NHS (Life Technologies). A solution of EZ-Link NHS-(PEG)4-Biotin (10 μL, 200 μM in PBS) was added to the corresponding protein (10 μL, 20 μM in PBS), after which the reaction mixture was pipetted up and down for 20 times to allow proper mixing and was left at room temperature for 30 minutes. The crude reaction mixture was buffer exchanged with PBS for 5 times to remove the excess of NHS-(PEG)4-Biotin. LC-MS analysis of the biotinylated π-clamp-traszutumab (10-(PEG)4-Biotin) showed that the biotin-to-antibody ratio was around 1.6 (Figure 3.47).

239
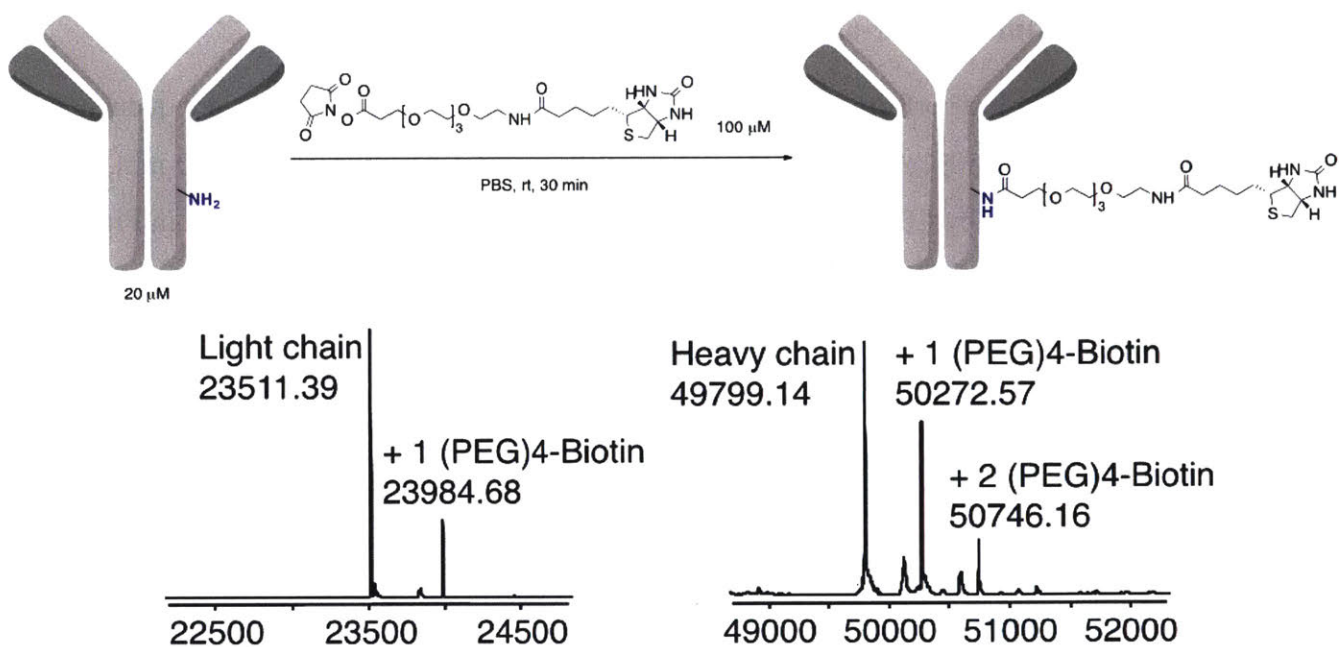
Figure 3.47. Non-selective biotinylation of antibody.

3.4.13. BioLayer interferometry binding assay

In vitro binding assays were performed using Fortebio Octet BioLayer Interferometry system at 37 °C. Briefly, streptavidin tips were dipped into 200 µL of biotin-linked antibody solution (20 nM of **10-Biotin** or trastuzumab-(PEG)4-Biotin in PBS with 0.1% BSA and 0.02% tween) for the loading of the antibodies. The tips loaded with antibody were sampled with recombinant HER2 (R&D Systems) at various HER2 concentrations in PBS with 0.1% BSA and 0.02% tween to obtain the association curve, buffer only serves as the reference. After association, the tips were dipped into PBS with 0.1% BSA and 0.02% tween to obtain the dissociation curve. Following the protocols provided by Fortebio Biosystems, the association and dissociation curves of each sample were manually fitted using Solver in excel to obtain the K_D. The final K_D was reported as the average of the K_D obtained from experiments with serially diluted HER2 (Figure 3.30).

3.4.14. Flow cytometry

Cells were seeded in 96-well V-bottom plates at a density of 20,000/well. Cells were spun down and washed with 150 µL of PBSA (phosphate buffer saline with 0.1% BSA). Cells were

then treated with 100 μL of biotin-linked antibodies or controls (20 nM in PBSA) for 1 hour on ice. Cells were spun down, washed with PBSA, and then treated with either streptavidin-AlexaFluro-647 or antiHuman Fc-AlexaFluro-488 for 30 minutes on ice. Cells were spun down, washed with PBSA, and then suspended in 100 μL of PBSA and analyzed by the Accuri C6 (BD Biosciences) flow cytometer in the CEHS Instrument Facility of MIT. All experiments were done in triplicate.

The flow cytometry results were summarized in Figure 3.31 to Figure 3.34.

3.4.15. Cell viability assay

Cells were seeded in 96-well white opaque plate at a density of 5 $\times10^3$/well (CHO) or 10 $\times10^3$/well (BT474). Cells were allowed to attach for 24 hours at 37 °C and 5% CO_2 in humidified atmosphere. Cells were then treated with serial dilutions of auristatin F, **10-MMAF**, or **10** for 96 hours (BT474) or 72 hours (CHO, treatment time was shortened to prevent overgrowth). The viability of cells was measured using CellTiter Glo reagents following the manufacturer's protocol and was normalized to the viability of cells without any treatment. The data were plotted using OriginLab software and the EC_{50} values were obtained by fitting the viability curves with a sigmoidal Boltzmann fit.

3.4.16. Molecular dynamics simulation

Molecular dynamics simulations were performed on the π-clamp peptide **1E**. The double glycine mutant peptide **1A** was also studied as a control.

Calculations were performed using the GROMACS 4.6.5[66] molecular dynamics package. The peptide was described by the AMBER 2003[67] force field. This force field was chosen because it best reproduced the Ramachandran plot proline, a key structural feature in the π-clamp sequence, when compared to CHARMM 2.7[68], OPLS-AA[69,70], and GROMOS96 45A3[71]. In addition, it has been shown to be accurate for dispersive interactions between phenylalanine and sulfur[72,73], presumably the main interaction in the π-clamp. We sampled the peptide structure with the proline adopts both a cis conformation (Figure 3.39) and a trans conformation (Figure 3.40) The peptide was solvated in 3382 explicit TIP3P[74] waters. Periodic boundary conditions were employed in a $(4.7 \text{ nm})^3$ simulation box. Simulations were performed in the NVT ensemble with temperature set to 300 K and enforced by the Nose-Hoover thermostat[75,76]. Simulations were run for 500 ns with

a time step of 2 fs. The linear peptide ($\phi = \psi = 180°$) was used as the initial configuration and equilibrated for 50 ns. For heat map shown in Figure 3.39 and Figure 3.40, sampling was performed at 10 ps intervals.

We pick the MD structures from the *cis*-proline to further investigate their arylation reaction energy pathway using DFT calculations. Four clamp structures (clamp_A – D), three half-clamp structures (Phe-4 interacting with perfluoroaryl group, Phe-1 interacting with sulfur, half-clamp_A – C), two open structures (open_A and B), and one double glycine mutant structure were optimized with DFT. These MD structures are shown in Figure 3.48.
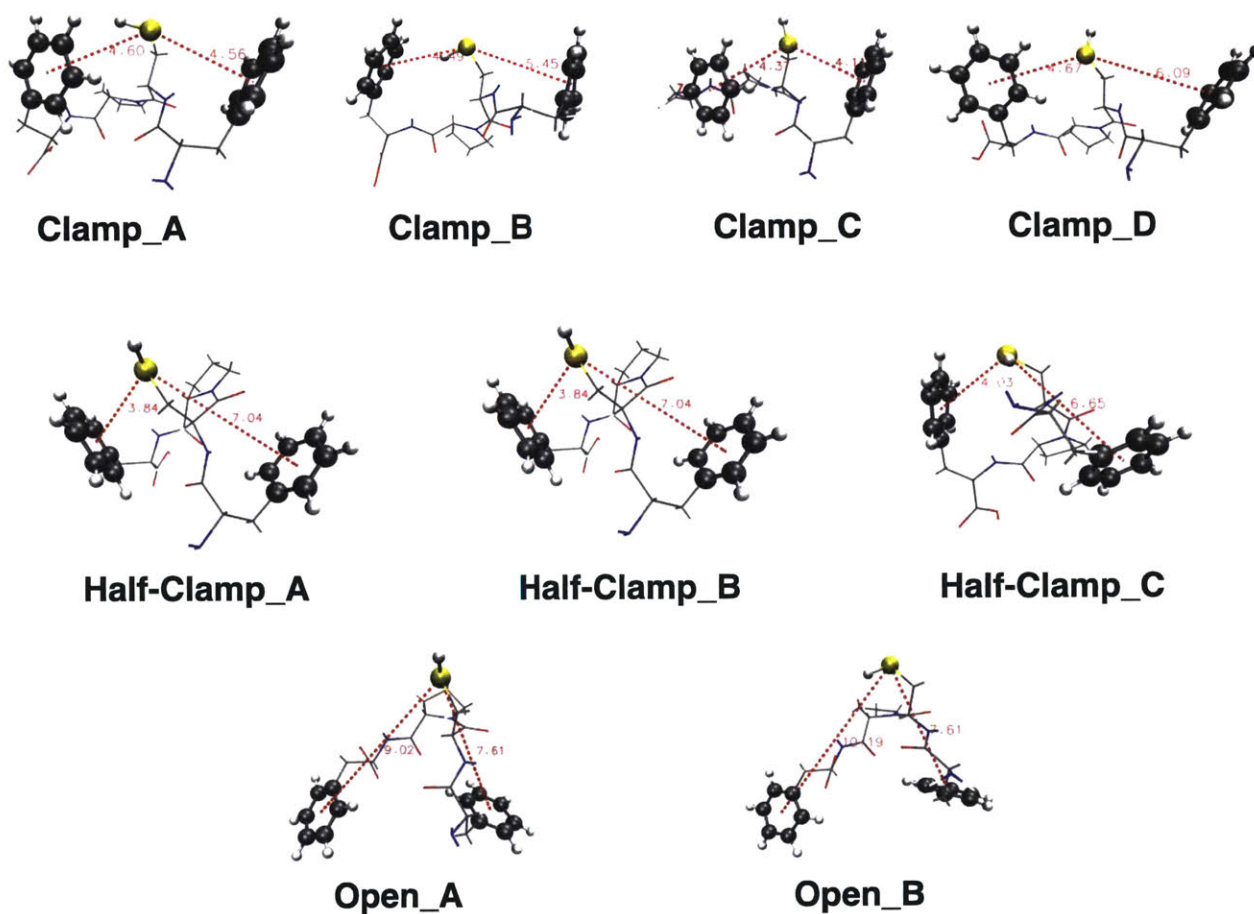
Figure 3.48. MD Snapshots used as starting tetra-peptide structures in DFT calculations. Dotted red lines show the distance between the center of each phenyl ring and the sulfur atom of cysteine. Note only the Phe-Cys-Pro-Phe part of the peptide is shown for clarity.

3.4.17. Density functional theory calculation

All DFT computations were carried out using the Q-Chem 4.1[77] software package. To reduce the computational cost, we deleted the redundant sequence and only kept the π-clamp or Gly-Cys-Pro-Gly as the cysteine peptide part, and kept perfluorobiphenyl thiol (**II**) as the perfluoroaryl probe for DFT calculation. The initial geometries used were were obtained from snapshots in MD simulation. The free energy (ΔG) was calculated as: $\Delta G = E_{Product} + E_{HF} - E_{Peptide} - E_{Perfluoroaromatics}$, where $E_{product}$ is the energy of the arylated product, E_{HF} is the energy of the solvated hydrogen fluoride, $E_{peptide}$ is the energy of the cysteine peptide (π-clamp or double glycine control), and $E_{perfluoroaromatics}$ is the energy of perfluorobiphenyl thiol. We extracted 4 snapshots from MD simulations for different starting structures of peptides in DFT calculations. For the product's starting structure, we manually connected the perfluoroaryl group to the cysteine of peptide.

In each case, four gas-phase geometry optimizations were performed on structures sampled from the MD trajectory, using the B3LYP exchange-correlation functional[78] in the 6-31G* basis set[79]. To account for π-π interactions, we also include Grimme's DFT-D3 empirical dispersion correction[80] for the optimization. Once a potential energy minimum was located, we refined the energy by preforming a single point energy calculation with the more accurate combination of the rPW86 exchange functional[81], the PBE local correlation functional[82], and the VV10 non-local correlation functional[23] to accurately handle the long-range dispersions critical to the π-π interactions. For these calculations, we also employed the larger 6-31G** basis set[79] and a large non-local integration grid (Lebedev, 75 radial points, 302 angular points)[24]. We then calculated the binding energies in both the gas phase and in water. We approximate the latter by the polarizable continuum model (PCM)[83], using 302 PCM grid points and a dielectric constant of 78.39. The optimized product geometries obtained from DFT calculations are shown in Figure 3.49, and the calculated free energy results are summarized in Table 3.2.

.
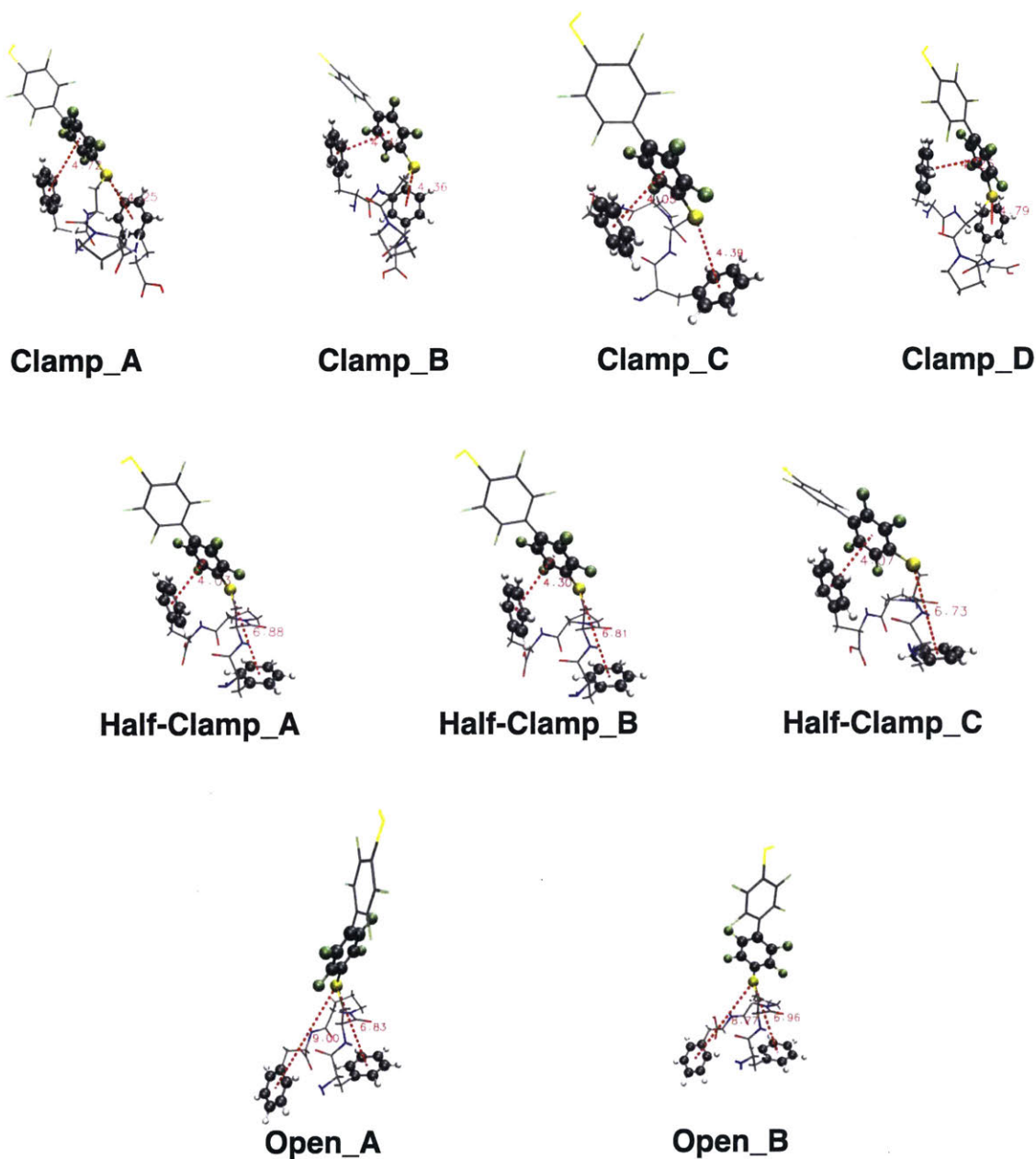
Figure 3.49. Optimized product geometries obtained from DFT calculations.

For clamp and half-clamp structures, dotted lines show the distance from the perfluoroaryl ring to the phenyl ring of Phe-4 and the distance from cysteine sulfur to the phenyl ring of Phe-1. For other structures, dotted red lines show the distance between the center of each phenyl ring and the sulfur atom of cysteine.

244

Table 3.2. Calculated free energies of reactions with structures obtained from MD simulation.

Structure	Gas Phase Δ (kcal/mol)	PCM Solvated Δ (kcal/mol)
Clamp_A	-12.97	-15.22
Clamp_B	-12.60	-15.61
Clamp_C	-17.22	-19.61
Clamp_D	-13.20	-16.62
Clamp Average	**-14.00**	**-16.76**
Half-Clamp_A	-11.89	-14.57
Half-Clamp_B	-11.80	-14.81
Half-Clamp_C	-11.32	-15.80
Half-Clamp Average	**-11.67**	**-15.06**
Open_A	-7.16	-9.36
Open_B	-5.69	-9.78
Open Average	**-6.43**	**-9.57**
GCPG	**-6.83**	**-9.48**

We've also computed the activation energy for the formation of the transition state for the π-clamp (Figure 3.39) and the double glycine control (Figure 3.50). Transition state (TS) searches were performed at the B3LYP/DFT-D3/6-31G* level of theory, using a Hessian eigenvector following method[84]. Following the TS search, we carried out a vibration frequency calculation at the same level of theory to confirm the structure was a first-order saddle point. The activation energy (ΔE) was calculated as: $\Delta E = E_{TS} - E_{peptide} - E_{perfluoroaromatics}$, where E_{TS} is the energy of the transition state. All single point calculations were performed using the long-range corrected version (LC-VV10) of the method we used previously for binding energy single point calculations in both gas phase and PCM water. The calculated activation energies are summarized in Table 3.3.

GCPG (I')   **GCPG_TS (III')**   **GCPG_product (IV')**
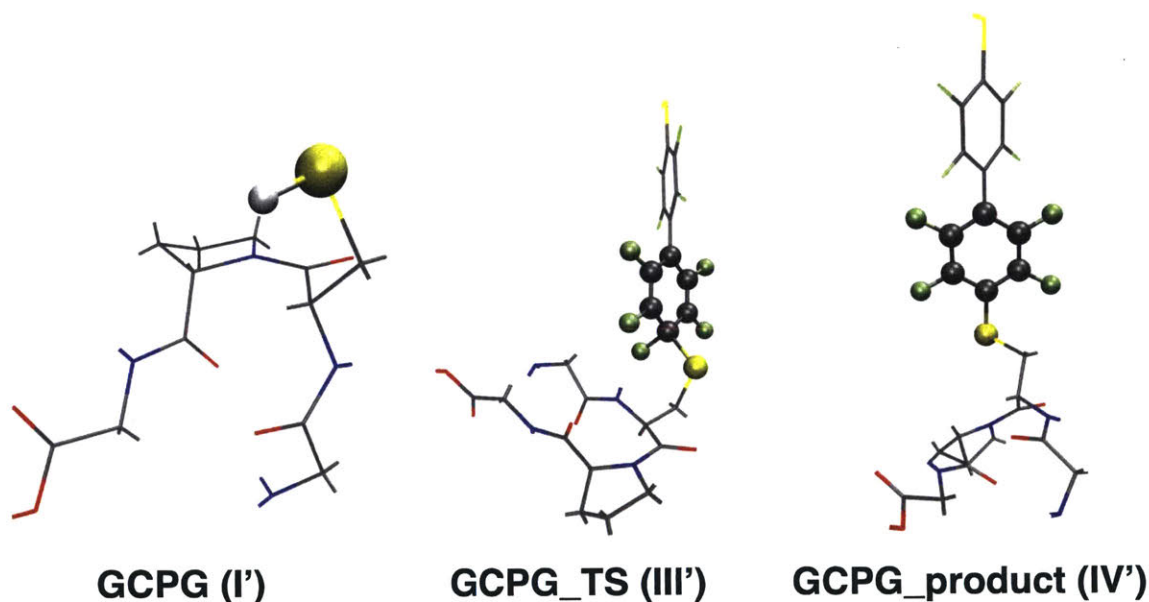
Figure 3.50. Optimized geometries for the starting peptide, the transition state, and the product for the double glycine mutant tetra-peptide (Gly-Cys-Pro-Gly).
The free energy of reaction pathway is shown in Figure 3.41.

Table 3.3. Calculated activation energies for the formation of transition state from the π-clamp and the double glycine control.

Structure	Gas Phase ΔE (kcal/mol)	PCM Solvated ΔE (kcal/mol)
π-clamp	0.56	9.29
GCPG	5.84	12.50
Difference	5.28	3.21

3.5. Acknowledgements

3.6. References

(1)   Carrico, I. S. Chemoselective Modification of Proteins: Hitting the Target. *Chem. Soc. Rev.* **2008**, *37*, 1423–1431.

(2)   Hackenberger, C. P. R.; Schwarzer, D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chemie Int. Ed.* **2008**, *47*, 10030–10074.

(3)   Rabuka, D. Chemoenzymatic Methods for Site-Specific Protein Modification. *Curr. Opin. Chem. Biol.* **2010**, *14*, 790–796.

(4)   Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*.

(5)   Take Aim. *Nat. Chem.* **2012**, *4*, 955.

(6)   Panowski, S.; Bhakta, S.; Raab, H.; Polakis, P.; Junutula, J. R. Site-Specific Antibody Drug Conjugates for Cancer Therapy. *MAbs* **2014**, *6*, 34–45.

(7)   Kochendoerfer, G. G. Site-Specific Polymer Modification of Therapeutic Proteins. *Curr. Opin. Chem. Biol.* **2005**, *9*, 555–560.

(8)   Uttamapinant, C.; White, K. A.; Baruah, H.; Thompson, S.; Fernández-Suárez, M.; Puthenveetil, S.; Ting, A. Y. A Fluorophore Ligase for Site-Specific Protein Labeling inside Living Cells. *Proc. Natl. Acad. Sci.* **2010**, *107*, 10914–10919.

(9)   Krishna, O. D.; Kiick, K. L. Protein- and Peptide-Modified Synthetic Polymeric Biomaterials. *Biopolymers* **2010**, *94*, 32–48.

(10)   Lichtor, P. A.; Miller, S. J. Combinatorial Evolution of Site- and Enantioselective Catalysts for Polyene Epoxidation. *Nat. Chem.* **2012**, *4*, 990–995.

(11)   Wilcock, B. C.; Uno, B. E.; Bromann, G. L.; Clark, M. J.; Anderson, T. M.; Burke, M. D. Electronic Tuning of Site-Selectivity. *Nat. Chem.* **2012**, *4*, 996–1003.

(12)   Sletten, E. M.; Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chemie - Int. Ed.* **2009**, *48*, 6974–6998.

(13)   Afagh, N. A.; Yudin, A. K. Chemoselectivity and the Curious Reactivity Preferences of Functional Groups. *Angew. Chemie Int. Ed.* **2010**, *49*, 262–310.

(14) Lewis, C. A.; Miller, S. J. Site-Selective Derivatization and Remodeling of Erythromycin A by Using Simple Peptide-Based Chiral Catalysts. *Angew. Chemie Int. Ed.* **2006**, *45*, 5616–5619.

(15) Chen, M. S.; White, M. C. A Predictably Selective Aliphatic C–H Oxidation Reaction for Complex Molecule Synthesis. *Science.* **2007**, *318*, 783–787.

(16) Snyder, S. A.; Gollner, A.; Chiriac, M. I. Regioselective Reactions for Programmable Resveratrol Oligomer Synthesis. *Nature* **2011**, *474*, 461–466.

(17) Pathak, T. P.; Miller, S. J. Site-Selective Bromination of Vancomycin. *J. Am. Chem. Soc.* **2012**, *134*, 6120–6123.

(18) Wender, P. A.; Hilinski, M. K.; Mayweg, A. V. W. Late-Stage Intermolecular CH Activation for Lead Diversification: A Highly Chemoselective Oxyfunctionalization of the C-9 Position of Potent Bryostatin Analogues. *Org. Lett.* **2004**, *7*, 79–82.

(19) Peddibhotla, S.; Dang, Y.; Liu, J. O.; Romo, D. Simultaneous Arming and Structure/Activity Studies of Natural Products Employing O–H Insertions: An Expedient and Versatile Strategy for Natural Products-Based Chemical Genetics. *J. Am. Chem. Soc.* **2007**, *129*, 12222–12231.

(20) Saxon, E.; Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science (80-.).* **2000**, *287*, 2007–2010.

(21) Rostovtsev, V. V; Green, L. G.; Fokin, V. V; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chemie Int. Ed.* **2002**, *41*, 2596–2599.

(22) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide–Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

(23) Song, W.; Wang, Y.; Qu, J.; Lin, Q. Selective Functionalization of a Genetically Encoded Alkene-Containing Protein via "Photoclick Chemistry" in Bacterial Cells. *J. Am. Chem. Soc.* **2008**, *130*, 9654–9655.

(24) Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. A "Traceless" Staudinger Ligation for the

Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2*, 2141–2143.

(25) Blackman, M. L.; Royzen, M.; Fox, J. M. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels−Alder Reactivity. *J. Am. Chem. Soc.* **2008**, *130*, 13518–13519.

(26) Lin, C.-W.; Ting, A. Y. Transglutaminase-Catalyzed Site-Specific Conjugation of Small-Molecule Probes to Proteins in Vitro and on the Surface of Living Cells. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.

(27) Wollack, J. W.; Silverman, J. M.; Petzold, C. J.; Mougous, J. D.; Distefano, M. D. A Minimalist Substrate for Enzymatic Peptide and Protein Conjugation. *Chembiochem* **2009**, *10*, 2934–2943.

(28) Yin, J.; Straight, P. D.; McLoughlin, S. M.; Zhou, Z.; Lin, A. J.; Golan, D. E.; Kelleher, N. L.; Kolter, R.; Walsh, C. T. Genetically Encoded Short Peptide Tag for Versatile Protein Labeling by Sfp Phosphopantetheinyl Transferase. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15815–15820.

(29) Cull, M. G.; Schatz, P. J. Biotinylation of Proteins in Vivo and in Vitro Using Small Peptide Tags. In *Methods in Enzymology*; Jeremy Thorner, S. D. E. J. N. A., Ed.; Academic Press, 2000; Vol. Volume 326, pp. 430–440.

(30) Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Xie, K. T.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. Redirecting Lipoic Acid Ligase for Cell Surface Protein Labeling with Small-Molecule Probes. *Nat. Biotechnol.* **2007**, *25*, 1483–1487.

(31) Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. Sortagging: A Versatile Method for Protein Labeling. *Nat. Chem. Biol.* **2007**, *3*, 707–708.

(32) Whitford, D. *Proteins: Structure and Function*; J. Wiley & Sons: Hoboken, NJ, 2005.

(33) Walsh, C. Enabling the Chemistry of Life. *Nature* **2001**, *409*, 226–231.

(34) Giles, N. M.; Giles, G. I.; Jacob, C. Multiple Roles of Cysteine in Biocatalysis. *Biochem. Biophys. Res. Commun.* **2003**, *300*, 1–4.

(35) Weerapana, E.; Wang, C.; Simon, G. M.; Richter, F.; Khare, S.; Dillon, M. B. D.; Bachovchin, D. a; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative Reactivity Profiling

Predicts Functional Cysteines in Proteomes. *Nature* **2010**, *468*, 790–795.

(36) Nathani, R. I.; Moody, P.; Chudasama, V.; Smith, M. E. B.; Fitzmaurice, R. J.; Caddick, S. A Novel Approach to the Site-Selective Dual Labelling of a Protein via Chemoselective Cysteine modification†Electronic Supplementary Information (ESI) Available: LC-MS, ES-MS, Deconvoluted Spectra and Fluorescence Emission Spectra for All Reactions with P. *Chem. Sci.* **2013**, *4*, 3455–3458.

(37) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. – An Asian J.* **2009**, *4*, 630–640.

(38) *Bioconjugate Techniques (Third edition)*; Hermanson, G. T., Academic Press: Boston, **2015**.

(39) Sun, M. M. C.; Beam, K. S.; Cerveny, C. G.; Hamblett, K. J.; Blackmore, R. S.; Torgov, M. Y.; Handley, F. G. M.; Ihle, N. C.; Senter, P. D.; Alley, S. C. Reduction−Alkylation Strategies for the Modification of Specific Monoclonal Antibody Disulfides. *Bioconjug. Chem.* **2005**, *16*, 1282–1290.

(40) Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific Covalent Labeling of\nRecombinant Protein Molecules\nInside Live Cells. *Science (80-.)*. **1998**, *281*, 269.

(41) Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124*, 6063–6076.

(42) Wilson, P.; Anastasaki, A.; Owen, M. R.; Kempe, K.; Haddleton, D. M.; Mann, S. K.; Johnston, A. P. R.; Quinn, J. F.; Whittaker, M. R.; Hogg, P. J.; *et al.* Organic Arsenicals as Efficient and Highly Specific Linkers for Protein/peptide-Polymer Conjugation. *J. Am. Chem. Soc.* **2015**, *137*, 4215–4222.

(43) Stroffekova, K.; Proenza, C.; Beam, K. G. The Protein-Labeling Reagent FLASH-EDT2 Binds Not Only to CCXXCC Motifs but Also Non-Specifically to Endogenous Cysteine-Rich Proteins. *Pflugers Arch. J. Physiol.* **2001**, *442*, 859–866.

(44) Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa Jr, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem.*

Biol. **2008**, *15*, 128–136.

(45) Rush, J. S.; Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **2008**, *130*, 12240–12241.

(46) Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.

(47) Zhang, C.; Spokoyny, A. M.; Zou, Y.; Simon, M. D.; Pentelute, B. L. Enzymatic "Click" Ligation: Selective Cysteine Modification in Polypeptides Enabled by Promiscuous Glutathione S-Transferase. *Angew. Chem., Int. Ed.* **2013**, *52*, 14001–14005.

(48) Zhang, C.; Dai, P.; Spokoyny, A. M.; Pentelute, B. L. Enzyme-Catalyzed Macrocyclization of Long Unprotected Peptides. *Org. Lett.* **2014**, *16*, 3652–3655.

(49) Zou, Y.; Spokoyny, A. M.; Zhang, C.; Simon, M. D.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. Convergent Diversity-Oriented Side-Chain Macrocyclization Scan for Unprotected Polypeptides. *Org. Biomol. Chem.* **2014**, *12*, 566–573.

(50) Sievers, E. L.; Senter, P. D. Antibody-Drug Conjugates in Cancer Therapy. *Annu. Rev. Med.* **2013**, *64*, 15–29.

(51) Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; *et al.* Site-Specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index. *Nat Biotech* **2008**, *26*, 925–932.

(52) Junutula, J. R.; Bhakta, S.; Raab, H.; Ervin, K. E.; Eigenbrot, C.; Vandlen, R.; Scheller, R. H.; Lowman, H. B. Rapid Identification of Reactive Cysteine Residues for Site-Specific Labeling of Antibody-Fabs. *J. Immunol. Methods* **2008**, *332*, 41–52.

(53) Piccart-Gebhart, M. J.; Procter, M.; Leyland-Jones, B.; Goldhirsch, A.; Untch, M.; Smith, I.; Gianni, L.; Baselga, J.; Bell, R.; Jackisch, C.; *et al.* Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer. *N. Engl. J. Med.* **2005**, *353*, 1659–1672.

(54) Spangler, J. B.; Manzari, M. T.; Rosalia, E. K.; Chen, T. F.; Wittrup, K. D. Triepitopic Antibody Fusions Inhibit Cetuximab-Resistant BRAF and KRAS Mutant Tumors via EGFR

Signal Repression. *J. Mol. Biol.* **2012**, *422*, 532–544.

(55)   Perrotte, P.; Matsumoto, T.; Inoue, K.; Kuniyasu, H.; Eve, B. Y.; Hicklin, D. J.; Radinsky, R.; Dinney, C. P. N. Anti-Epidermal Growth Factor Receptor Antibody C225 Inhibits Angiogenesis in Human Transitional Cell Carcinoma Growing Orthotopically in Nude Mice. *Clin. Cancer Res.* **1999**, *5*, 257–264.

(56)   Nelson, K. J.; Parsonage, D.; Hall, A.; Karplus, P. A.; Poole, L. B. Cysteine pKa Values for the Bacterial Peroxiredoxin AhpC. *Biochemistry* **2008**, *47*, 12860–12868.

(57)   Salonen, L. M.; Ellermann, M.; Diederich, F. Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew. Chemie Int. Ed.* **2011**, *50*, 4808–4842.

(58)   Artamkina, G. A.; Egorov, M. P.; Beletskaya, I. P. Some Aspects of Anionic .sigma.-Complexes. *Chem. Rev.* **1982**, *82*, 427–459.

(59)   Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004–2021.

(60)   Jewett, J. C.; Bertozzi, C. R. Cu-Free Click Cycloaddition Reactions in Chemical Biology. *Chem. Soc. Rev.* **2010**, *39*, 1272–1279.

(61)   Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **2014**, *114*, 4764–4806.

(62)   Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; *et al.* Rapid Flow-Based Peptide Synthesis. *Chembiochem* **2014**, *15*, 713–720.

(63)   Liao, X.; Rabideau, A. E.; Pentelute, B. L. Delivery of Antibody Mimics into Mammalian Cells via Anthrax Toxin Protective Antigen. *Chembiochem* **2014**, *15*, 2458–2466.

(64)   Chen, I.; Dorr, B. M.; Liu, D. R. A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display. *Proc. Natl. Acad. Sci.* **2011**, *108*, 11399–11404.

(65)   Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L. Protein Thioester Synthesis Enabled by Sortase. *J. Am. Chem. Soc.* **2012**, *134*, 10749–10752.

(66)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*

2008, *4*, 435–447.

(67) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; *et al.* A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(68) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; *et al.* CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(69) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(70) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides†. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(71) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. An Improved GROMOS96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase. *J. Comput. Chem.* **2001**, *22*, 1205–1218.

(72) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. Performance of Empirical Potentials (AMBER, CFF95, CVFF, CHARMM, OPLS, POLTEV), Semiempirical Quantum Chemical Methods (AM1, MNDO/M, PM3), and Ab Initio Hartree–Fock Method for Interaction of DNA Bases: Comparison with Nonempirical beyond Hartree–Fock Res. *J. Comput. Chem.* **1997**, *18*, 1136–1150.

(73) Kolář, M.; Berka, K.; Jurečka, P.; Hobza, P. On the Reliability of the AMBER Force Field and Its Empirical Dispersion Contribution for the Description of Noncovalent Complexes. *ChemPhysChem* **2010**, *11*, 2399–2408.

(74) Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(75) Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

(76)   Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(77)   Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V; Levchenko, S. V; O'Neill, D. P.; *et al.* Advances in Methods and Algorithms in a Modern Quantum Chemistry Program Package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(78)   Becke, A. D. A New Mixing of Hartree–Fock and Local Density-functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

(79)   Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(80)   Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*.

(81)   Murray, É. D.; Lee, K.; Langreth, D. C. Investigation of Exchange Energy Density Functional Accuracy for Interacting Molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2754–2762.

(82)   Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(83)   Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24*, 669–681.

(84)   Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, *7*, 385–395.

Chapter 4. Organometallic palladium reagents for cysteine arylation

The work presented in this chapter is a part of the following manuscripts and is reproduced here with permission from the Nature Publication Group and Macmillan Publishers Limited.

Vinogradova, E. V.*; Zhang, C.*; Spokoyny, A. M.; Pentelute, B. L.[#]; Buchwald, S. L.[#]

Organometallic Palladium Reagents for Cysteine Bioconjugation

Nature **2015**, *526*, 687–691. DOI: 10.1038/nature15739

*These authors contributed equally

[#]Co-corresponding authors

4.1. Introduction

Post-translational modifications greatly expand the function of proteins.[1] Chemists aim to mimic nature's success through the development of chemo- and regioselective reactions of proteins. The diversity of potentially reactive functional groups presents in biomolecules (e.g., amides, acids, alcohols, amines) combined with the requirement for fast kinetics and mild reaction conditions (e.g., aqueous solvent, pH 6–8, T<37 °C) set a high bar for the development of new techniques to functionalize proteins. Nevertheless, methods have emerged for bioconjugation with natural and unnatural amino acids in protein molecules.[2,3] Cysteine is a key residue for the chemical modification of proteins owing to the unique reactivity of the thiol functional group and the low abundance of cysteine residues in naturally occurring proteins.[4,5] Michael addition to maleimides and S_N2 reaction with alkyl halides are commonly used for cysteine modification. The resulting conjugates tend to decompose in the presence of external bases or thiol nucleophiles,[6] which prompted the recent development of advanced cysteine bioconjugations for the improved stability of the conjugates.[7]

The ability to achieve high levels of chemo- and regioselectivity through the judicious choice of metal and ligand design suggest metal-mediated processes could be very attractive for the development of new bioconjugations. Existing metal based transformations often rely on the use of functional handles[8] or unnatural amino acids such as 4-iodophenylalanine, aldehyde- or alkyne-containing amino acids,[9–11] and require high concentrations (mM) of derivatizing agents, which can cause off-target reactivity or purification problems. We hypothesized that palladium complexes resulting from the oxidative addition of aryl halides or trifluoromethanesulfonates[12] could be used for the transfer of aryl groups to cysteine residues in proteins (Fig. 1a).[13] The efficiency and selectivity of the proposed reaction with the highly active palladium species may be hampered by the presence of a variety of functional groups within complex biopolymers.[14,15] However, we envisioned that careful choice of ligand would provide stable, yet highly reactive reagents for the desired transformations (Figure 4.1b), while the interaction between the soft nucleophile cysteine thiol and the aryl palladium(II) species would guide its selectivity.

4.2. Results

Here we report that palladium(II) complexes can be used as storable reagents for efficient and highly selective cysteine conjugation (Figure 4.1). The bioconjugation is fast and robust across a range of biocompatible reaction conditions. The straightforward synthesis of the palladium reagents from diverse and easily accessible aryl halide and trifluoromethanesulfonate precursors makes the method highly practical, providing access to a large structural space for protein modification. The resulting aryl bioconjugates are stable towards acids, bases, oxidants, and external thiol nucleophiles. The broad utility of the new bioconjugation platform was further corroborated by the synthesis of novel classes of stapled peptides and antibody-drug conjugates. Thus, these palladium complexes (Figure 4.1b) show potential as a new set of benchtop reagents for diverse bioconjugation applications.

4.1.1. Reactions of organometallic palladium reagents with model peptides

We began our study with a palladium-tolyl complex (**1A-OTf**) using 2-dicyclohexylphosphino-2',6'-diisopropoxybiphenyl (RuPhos) as the ligand and trifluoromethanesulfonate as the counterion (Figure 4.1b). A model peptide (**P1**) was used for the optimization of the reaction conditions and for exploration of the substrate scope. Full conversion of the starting peptide to the corresponding aryl product was observed in less than 5 minutes at low micromolar concentrations of reagents (Figure 2.1a). Further, the reaction was selective for cysteine. No reaction was observed using a control peptide wtih cysteine mutated to serine (Figure 2.1b), in contrast to the palladium-mediated protein allylation, which is selective for tyrosine (O-allylation) over lysine and cysteine (N- and S-allylation).[16] These results highlight the importance of the ligand choice to facilitate C–S reductive elimination and the electrophilicity of the palladium center to tune the selectivity of the transformation.

a

b

1A-OTf
(X-ray)[a]

(RuPhos)Pd(Ar)X[b]

X = I, Br, Cl, OTf

solid storable reagents

- ✓ chemoselective
- ✓ benign conditions
- ✓ broad substrate scope
- ✓ stable products
- ✓ low μM concentrations
- ✓ fast kinetics
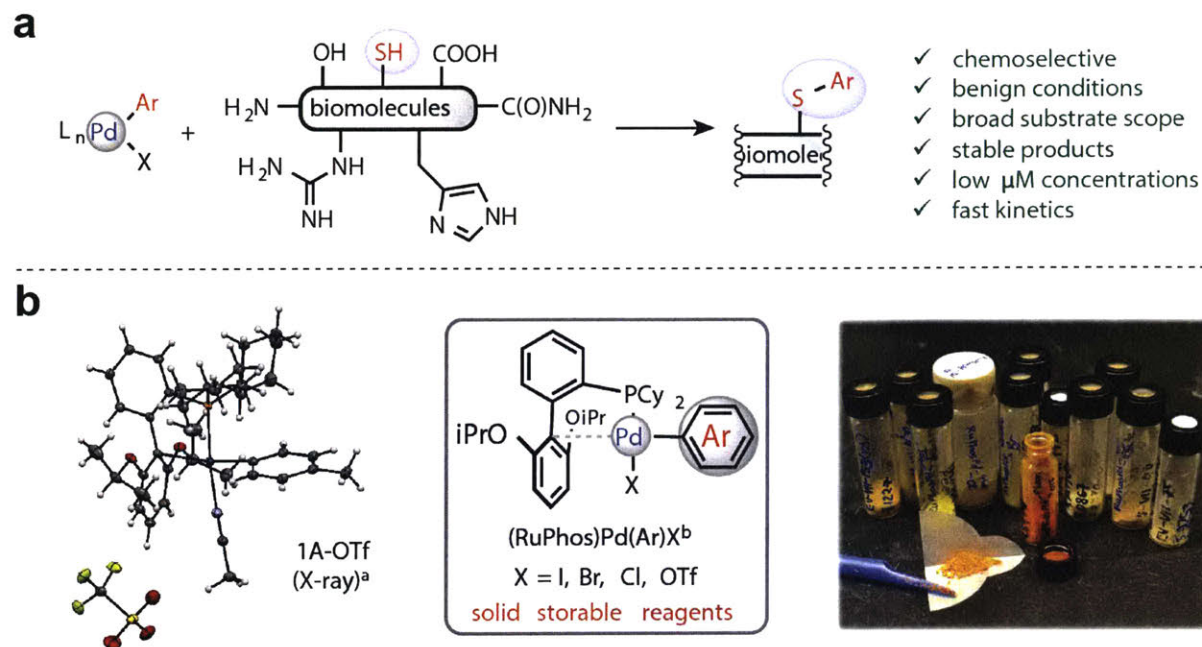
Figure 4.1. Organometallic palladium reagents for cysteine bioconjugation.
(**a**) Proposed cysteine bioconjugation using palladium reagents; (**b**) Palladium reagents for cysteine bioconjugation: X-ray structure, storage and operational simplicity. The crystal structure shows an example of palladium complexes made. [a]Crystals were obtained by vapor diffusion of an Et$_2$O/CH$_3$CN solution of **1A-OTf** with pentane. [b]RuPhos = 2-dicyclohexylphosphino-2',6'-diisopropoxybiphenyl.
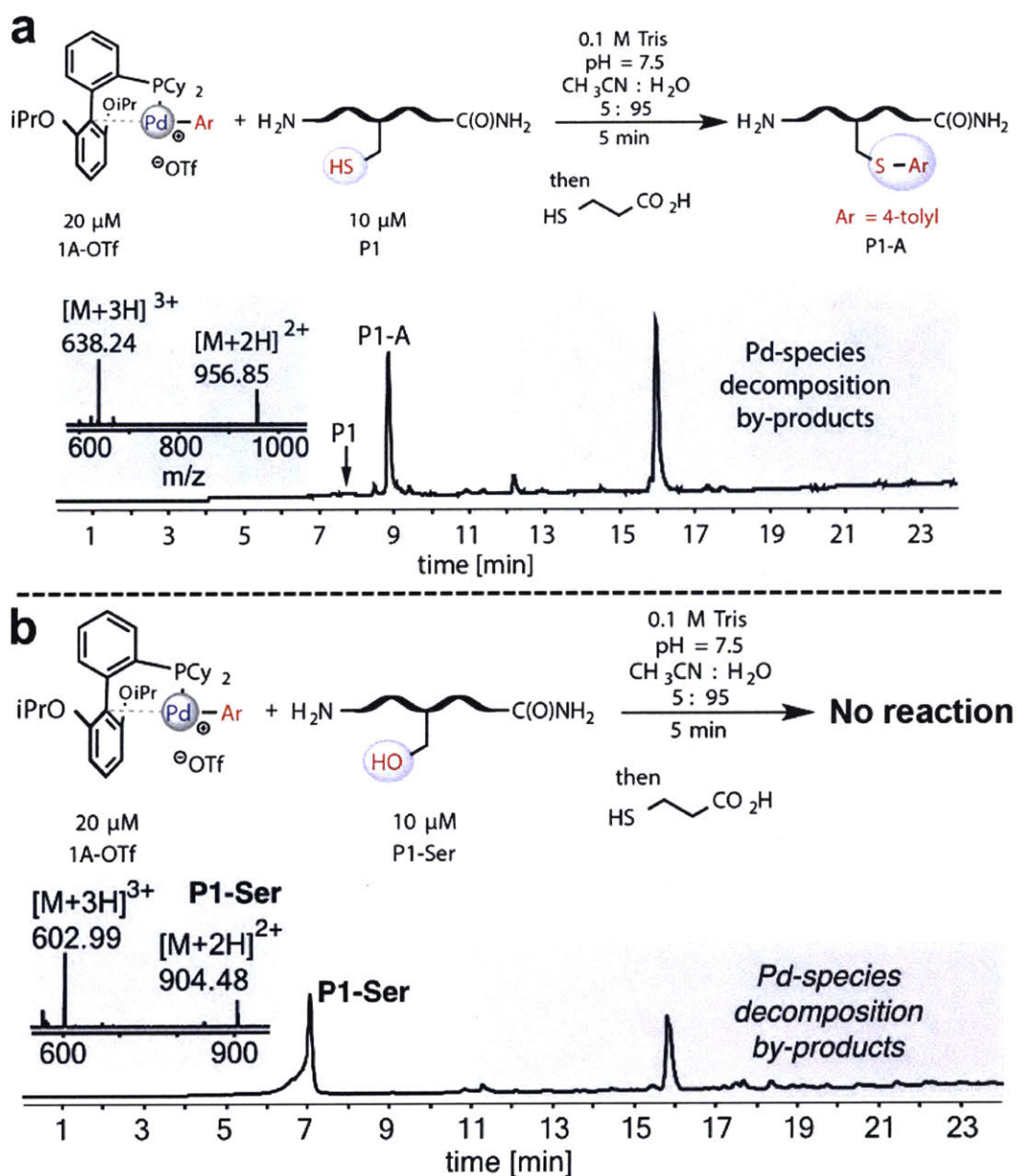
Figure 4.2. Organometallic palladium reagents for reactions with a cysteine peptide and a serine control.

(a) Model reaction with a peptide substrate and the LC-MS trace of the crude reaction mixture after 5 min. The mass spectrum of the arylated product is shown in the inset. Peptide **P1** sequence: NH_2-RSNFYLGCAGLAHDKAT-C(O)NH_2. Reaction conditions: **P1** (10 µM), complex **1A-OTf** (20 µM), Tris buffer (100 mM; pH 7.5), $CH_3CN : H_2O = 5 : 95$, rt. The reaction was quenched by the addition of 3-mercaptopropionic acid (3 equivalents to **1A-OTf**) before LC-MS analysis. (b) Reaction with the serine control peptide.

Most cysteine conjugation reactions operate at nearly neutral to slightly basic pH values. Further evaluation of the reaction conditions using palladium reagents revealed quantitative conversion of the starting peptide to the corresponding S-aryl cysteine conjugate within a broad pH range (5.5 – 8.5) using common organic cosolvents (5% of DMF, DMSO, CH_3CN) in various buffers (Table 4.1, entries 3, 4, 5, 13, and 14). Remarkably, even in 0.1% TFA solution (pH 2.0) the reaction yielded 59% of the S-arylated product after 7 hours (Table 4.1, entry 17). The process was also compatible with the protein disulfide reducing agent $tris$(2-carboxyethyl)phosphine (TCEP, Table 4.1, entry 20) that has been shown to hamper bioconjugations by reacting with maleimide and α-haloacyl groups.[17]

The palladium mediated conjugation is fast and complete product formation occurs within 15 seconds at 4 °C. The reaction rate was estimated by competition experiments against the commonly used N-methyl maleimide cysteine ligation (Figure 4.3).[18] At pH 7.5 the rate of the palladium-mediated reaction was comparable to that of the maleimide ligation, where 70% of the products resulted from the reaction with palladium-tolyl complex (**1A-OTf**). Notably, the palladium-mediated conjugation outperformed the maleimide ligation at pH 5.5 where only the arylated product was formed.

The optimized conditions (0.1 M Tris buffer, 5% CH_3CN, pH 7.5, room temperature) were used for further evaluation of the substrate scope. Palladium complexes containing chloride, bromide and iodide counterions were all found to produce the desired product (Figure 4.4, **1A-Cl**, **1A-Br**, and **1A-I**). This method can be used to functionalize unprotected peptides with a variety of important groups including fluorescent tags (**1C, 1D**), affinity labels (**1E**), bioconjugation handles (aldehyde **1F**, ketone **1G**, and alkyne **1H**), photochemical crosslinkers (**1I**), complex drug molecules (**1J**), and vinyl groups (**1- vinyl**, see Experimental section Figure 4.26).

Importantly, the palladium(II) complexes are stable under ambient conditions, and can be stored in closed vials under air at 4 °C for over four months. Long-term stability of **1A-I**, **1A-Br**, **1A-Cl**, and **1A-OTf** was evaluated using NMR, where only the complex bearing the trifluoromethanesulfonate counterion (**1A-OTf**) showed some degradation (≤15%) after 20 weeks (see Experimental section 4.4.12 for details). Nevertheless, the "aged" reagents still exhibited reactivity comparable to the freshly made complexes (Figure 4.5).

Table 4.1. Evaluation of reaction conditions.[a]

Entry	Buffer	Peptide Conc.	pH	Solvent	Yield
1	100 mM Tris	1 mM	8.5	$H_2O:CH_3CN$ (2:1)	93 %
2	100 mM Tris	**100 μM**	8.5	$H_2O:CH_3CN$ (**95:5**)	85 %
3	100 mM Tris	**10 μM**	**8.5**	$H_2O:CH_3CN$ (95:5)	>99 %
4	100 mM Tris	10 μM	**8**	$H_2O:CH_3CN$ (95:5)	>99 %
5	100 mM Tris	10 μM	**7.5**	$H_2O:CH_3CN$ (95:5)	>99 %
6	100 mM **HEPES**	10 μM	7.5	$H_2O:CH_3CN$ (95:5)	>99 %
7	100 mM **MOPS**	10 μM	7.5	$H_2O:CH_3CN$ (95:5)	>99 %
8	100 mM **Na_2HPO_4/NaH_2PO_4**	10 μM	7.5	$H_2O:CH_3CN$ (95:5)	>99 %
9	**25 mM Tris**	10 μM	7.5	$H_2O:CH_3CN$ (95:5)	93 %
10	100 mM Tris	10 μM	**7**	$H_2O:CH_3CN$ (95:5)	84 %
11[b]	**200 mM Tris**	10 μM	**7**	$H_2O:CH_3CN$ (95:5)	91 %
12	**100 mM MOPS**	10 μM	**7**	$H_2O:CH_3CN$ (95:5)	>99 %
13	100 mM MOPS	10 μM	**6.5**	$H_2O:CH_3CN$ (95:5)	>99 %
14	100 mM MES	10 μM	**5.5**	$H_2O:CH_3CN$ (95:5)	95 %
15[c]	100 mM MES	10 μM	**5.5**	$H_2O:CH_3CN$ (95:5)	>99 %
16	0.1 % TFA	10 μM	**2.0**	$H_2O:CH_3CN$ (95:5)	18 %
17[d]	0.1 % TFA	10 μM	**2.0**	$H_2O:CH_3CN$ (95:5)	59 %
18	100 mM Tris	10 μM	7.5	$H_2O:$**DMF** (95:5)	>99 %
19	100 mM Tris	10 μM	7.5	$H_2O:$**DMSO** (95:5)	>99 %
20[e]	100 mM Tris	10 μM	7.5	$H_2O:CH_3CN$ (95:5)	>99 %

[a]Optimal conditions used for further substrate scope evaluation are highlighted in grey; [b]Further increase in the molarity of the Tris buffer (400 mM) did not have an effect on the product yield. [c]Reaction time: 10 min; [d]Reaction time: 7 h 20 min; [e]Reaction performed in the presence of TCEP (20 μM). Method for calculating yields is described in Experimental section 4.4.4. See Experimental section 4.4.9 for the LC-MS chromatograms of these reactions.
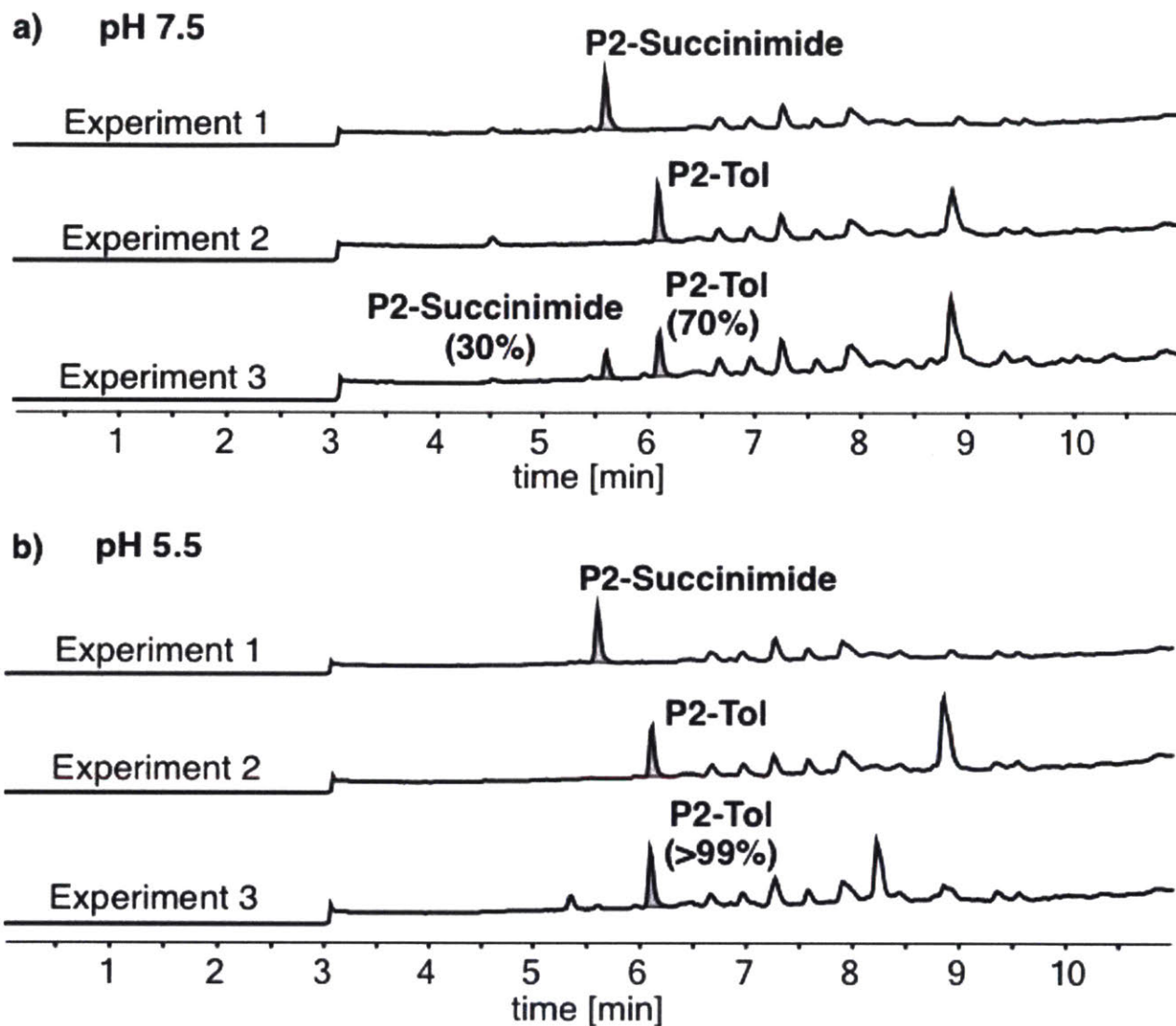
Figure 4.3. Competition experiments comparing the rates of palladium reagents and maleimides at different pH values.

LC-MS TIC curves for the experiments evaluating relative rates of cysteine bioconjugation with *N*-ethyl maleimide and palladium reagent **1A-OTf** at (**a**) pH 7.5 and (**b**) pH 5.5.
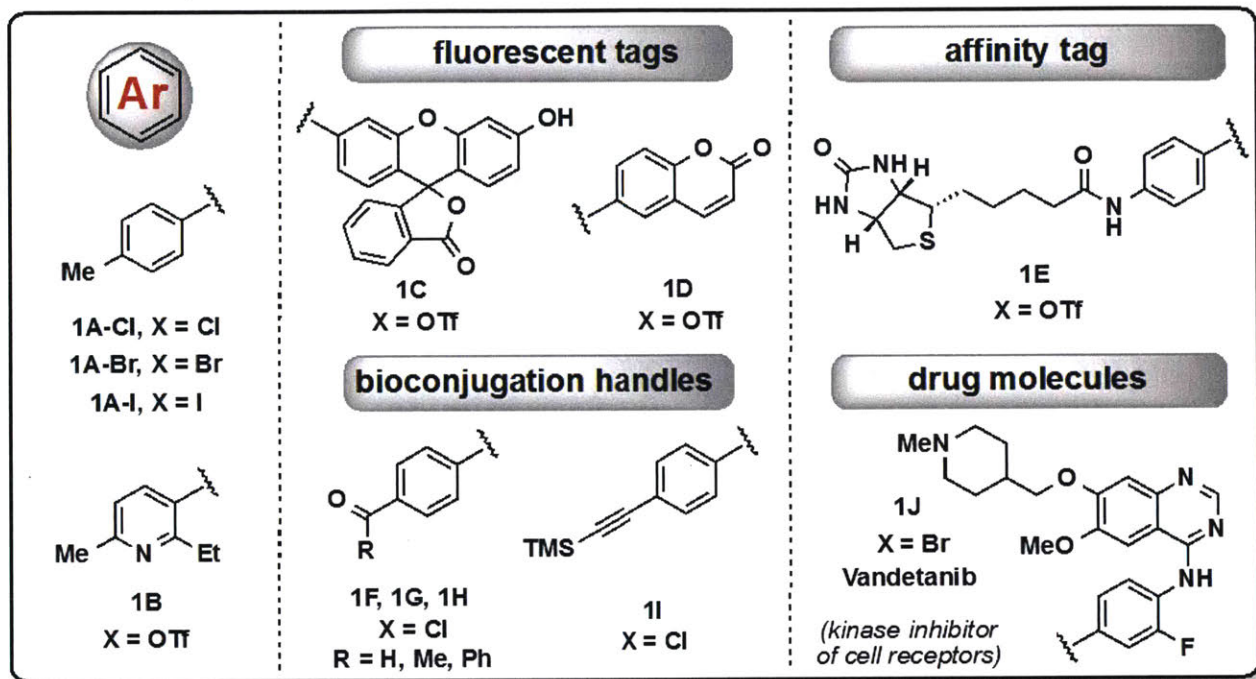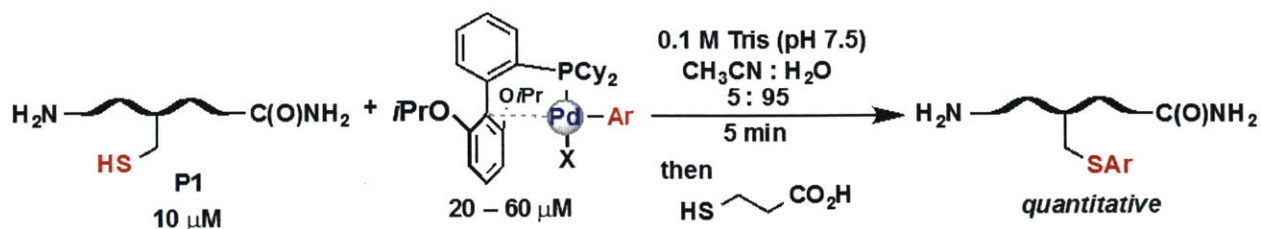
Figure 4.4. The substrate scope of cysteine arylation using organometallic palladium reagents.

Full conversion of starting peptide **P1** into the corresponding arylated products was observed in all the cases as confirmed by LC-MS. For exact reaction procedures and conditions and LC-MS chromatograms, see Experimental section 4.4.10.
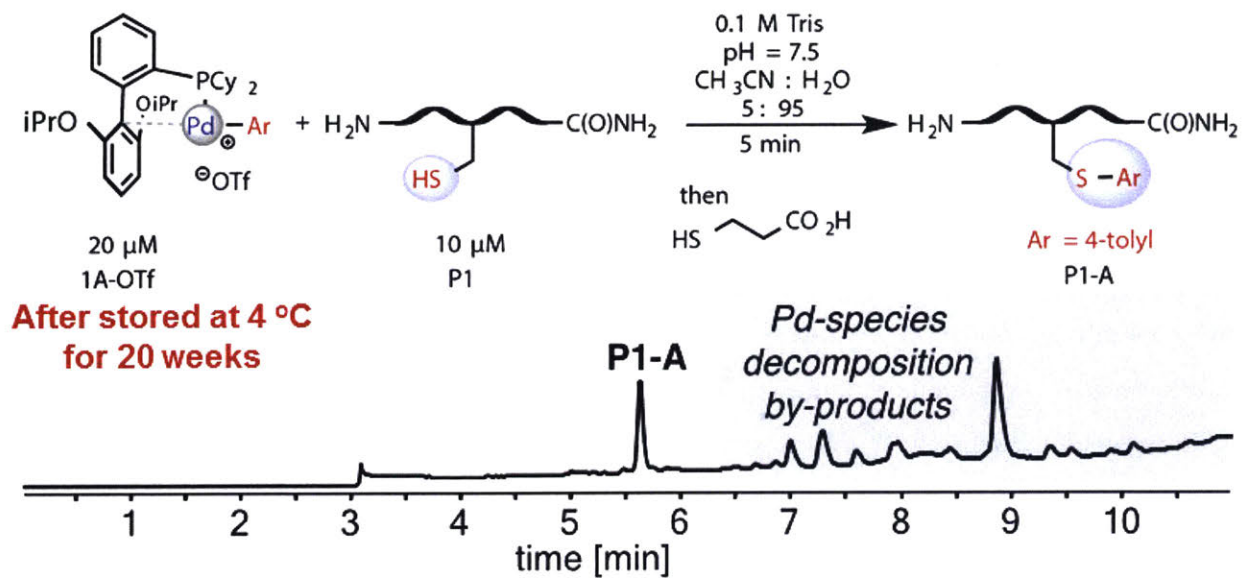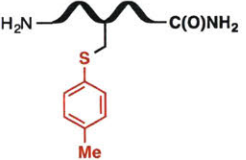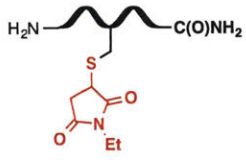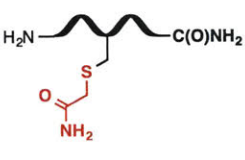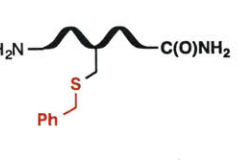
Figure 4.5. LC-MS analysis of a crude reaction using "aged" 1A-OTf reagent.

4.1.2. Stability of aryl-thiol linkages versus alkyl-thiol linkages in peptides

The stability of our arylated peptides was compared to that of conjugates formed from reactions with reagents including *N*-ethyl maleimide, 2-bromoacetamide, and benzyl bromide. The *S*-arylated peptide was shown to be stable toward acids, bases, and external thiol nucleophiles (Table 4.2). In contrast, the corresponding acetamide derivative was unstable under acidic and basic conditions and the maleimide conjugate decomposed in the presence of base and exogenous thiol. Finally, comparable stability of both aryl and benzyl conjugates to treatment with the periodic acid oxidant at 37 °C was observed. However, additional tuning of the electronic properties of the aromatic ring of the arylated peptide by installing a *para*-electron withdrawing cyano-group could be achieved. This modification significantly decreased the amount of oxidation producing the most stable peptides across all the evaluated conjugates. Notably, installing the para cyano-group in the benzyl conjugates did not have any effect toward oxidation (Figure 4.6).

Table 4.2. Stability of the cysteine conjugates in the presence of base, acid, external thiol nucleophiles, and under oxidative conditions.

	A	B	C	D
			% remaining peptide	
base	83 %	0 %	66 %	84 %
acid	83 %	84 %	63 %	85 %
GSH	90 %	37 %	85 %	88 %
oxidation[a]	73 %	82 %	73 %	39 %

[a]Reaction time – 30 min. See Experimental section 4.4.13 for detailed reaction conditions and LC-MS chromatograms.
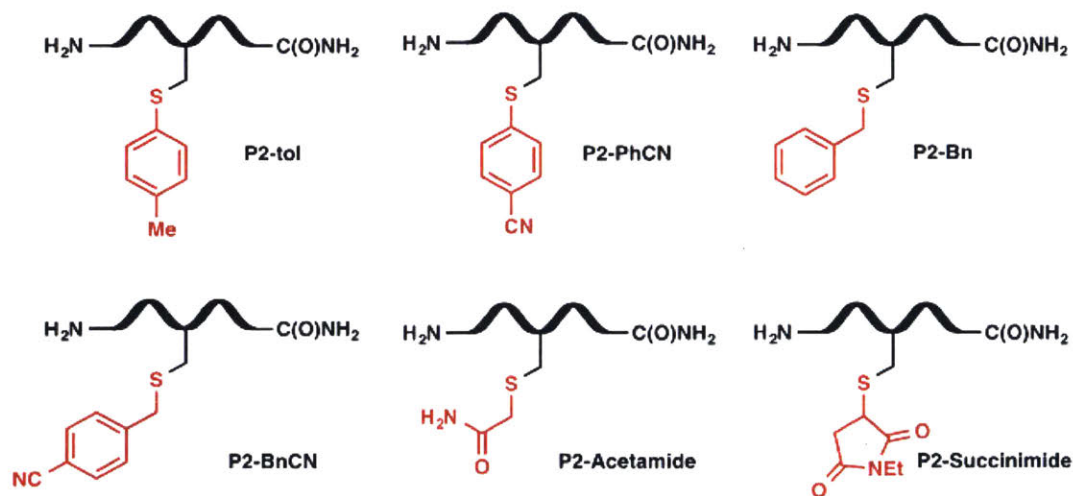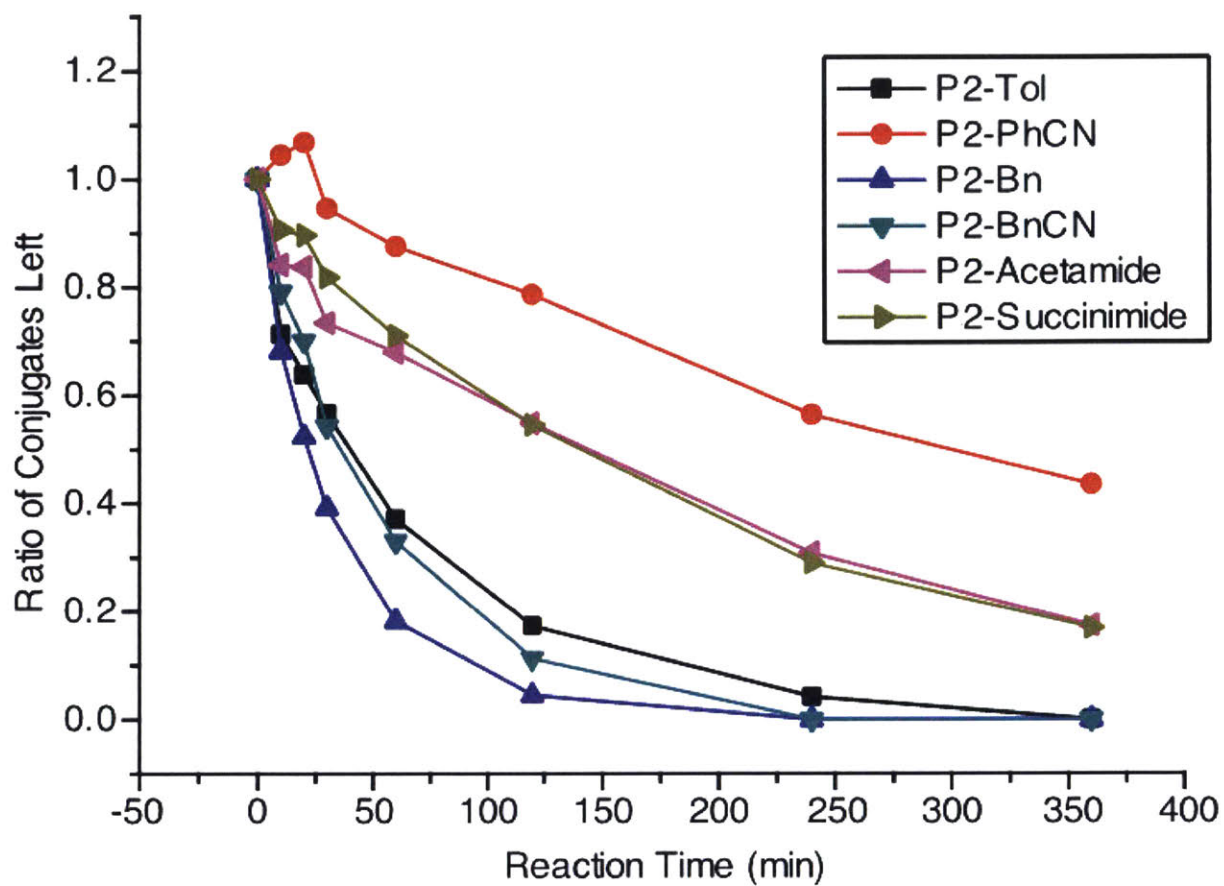
Figure 4.6. Stability of cysteine conjugates under oxidative conditions.

See Experimental section 4.4.13 for detailed reaction conditions and LC-MS chromatograms.

4.1.3. Selective labeling of antibody mimetic proteins

We further explored this reaction with proteins. Three antibody mimetic proteins[19] (**P4-P6**) were expressed that contained a cysteine at structurally distinct positions including the *N*-terminus, *C*-terminus, and a loop. The same proteins without cysteine were used as controls to confirm the selectivity of the reaction (**P7-P9**). All three proteins (**P4-P6**) were quantitatively tagged with either coumarin (Figure 4.7) or a drug molecule (Figure 4.9) within 30 minutes at 1 μM protein concentration. No arylated product was generated for proteins lacking a cysteine (Figure 4.8 and Figure 4.10). The fast kinetics and high efficiency of the reactions at low micromolar and low nanomolar (Figure 4.11) protein concentrations are in contrast to reported bioconjugation methods using organometallic reagents, where longer reaction times were needed and generally lower conversions were observed.[9,20] 5% of organic co-solvent is required for the reaction, as protein labeling with palladium reagents in an all-aqueous buffer resulted in lower conversions (Figure 4.12).
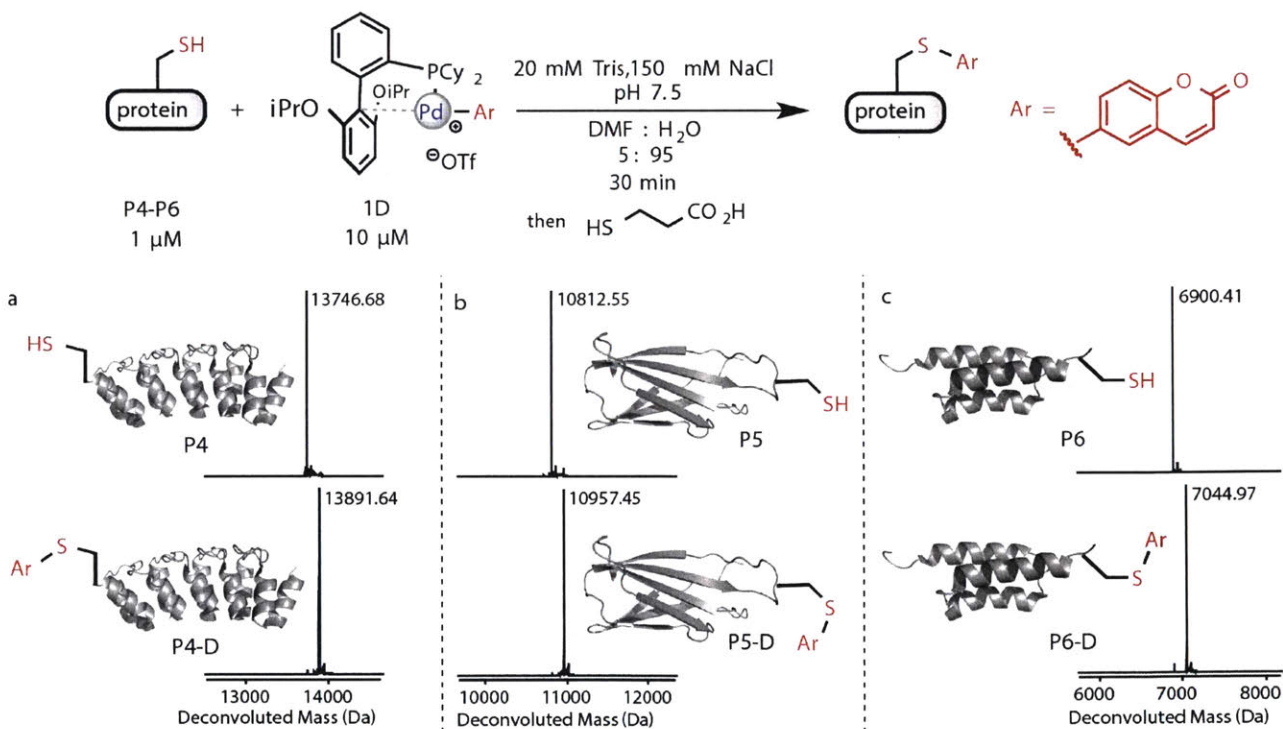
Figure 4.7. Modification of antibody mimetic proteins using the developed palladium reagents.

Cysteine residues at (**a**) the *N*-terminus (**P4**), (**b**) a loop (**P5**), and (**c**) the *C*-terminus (**P6**) of proteins are quantitatively modified with coumarin after the reaction with palladium complex **1D**.
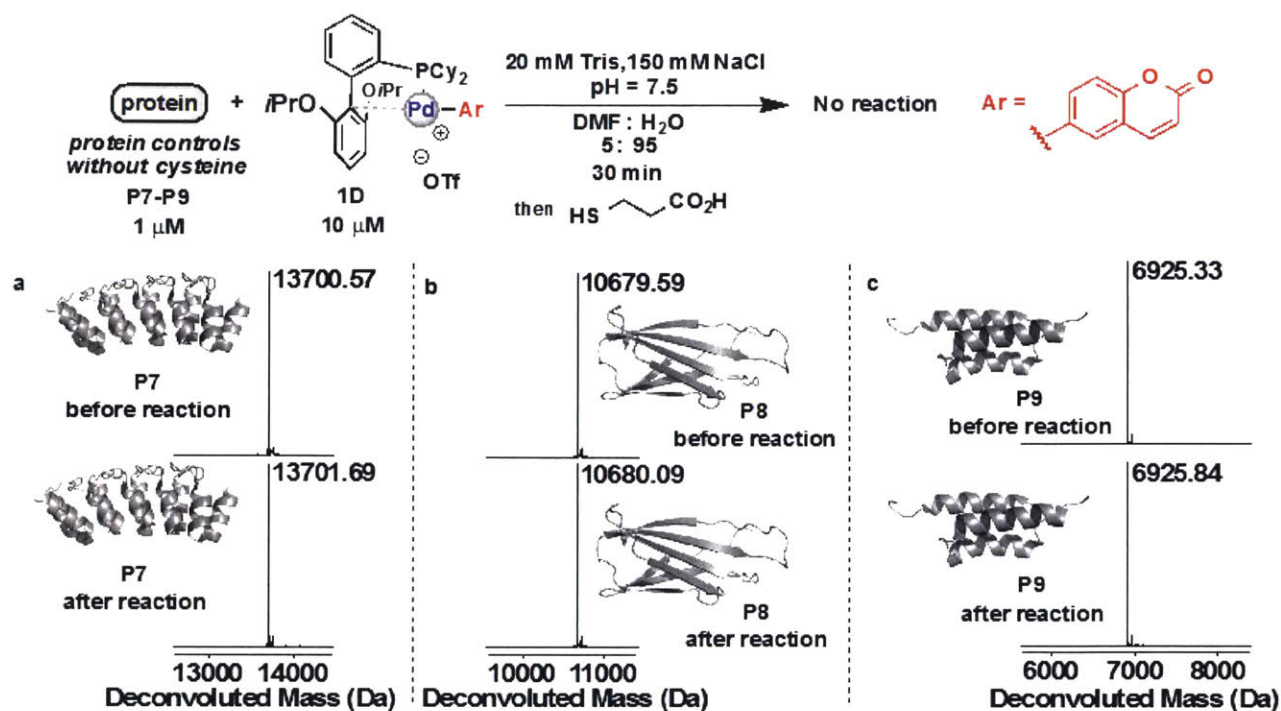
Figure 4.8. Control reactions for labeling of proteins without cysteine residues using palladium complex 1D.
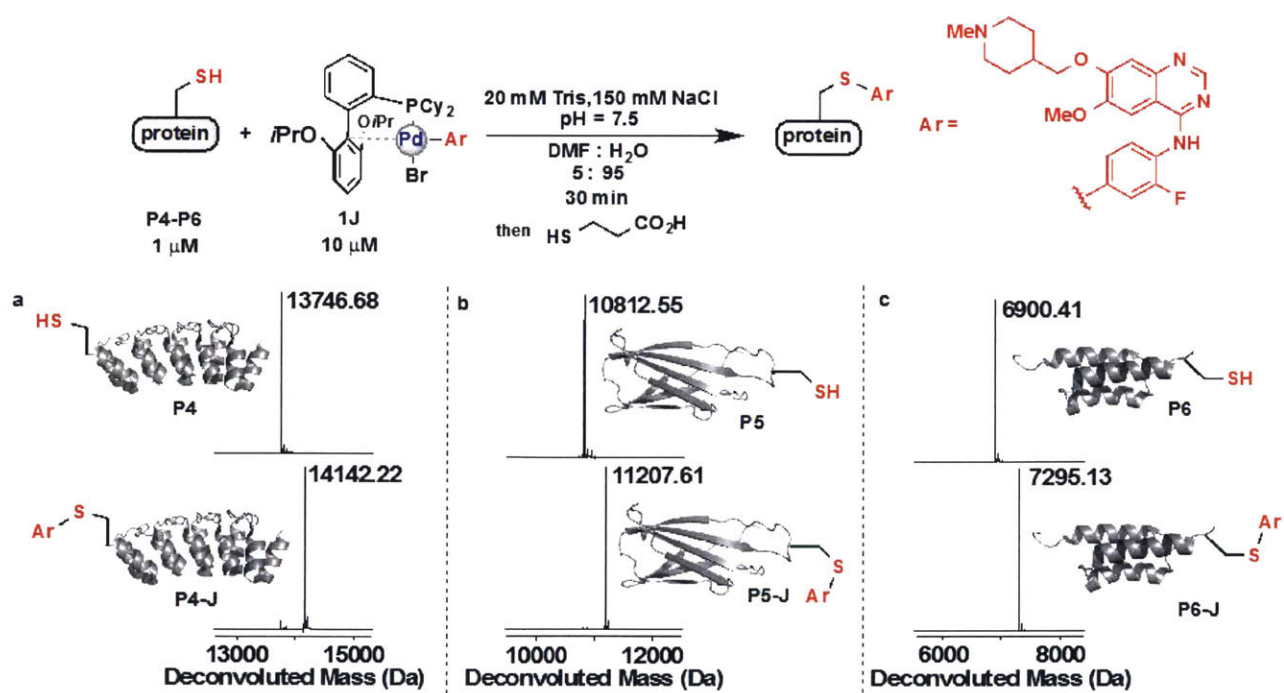
Figure 4.9. Modification of antibody mimetic proteins using palladium complex 1J.
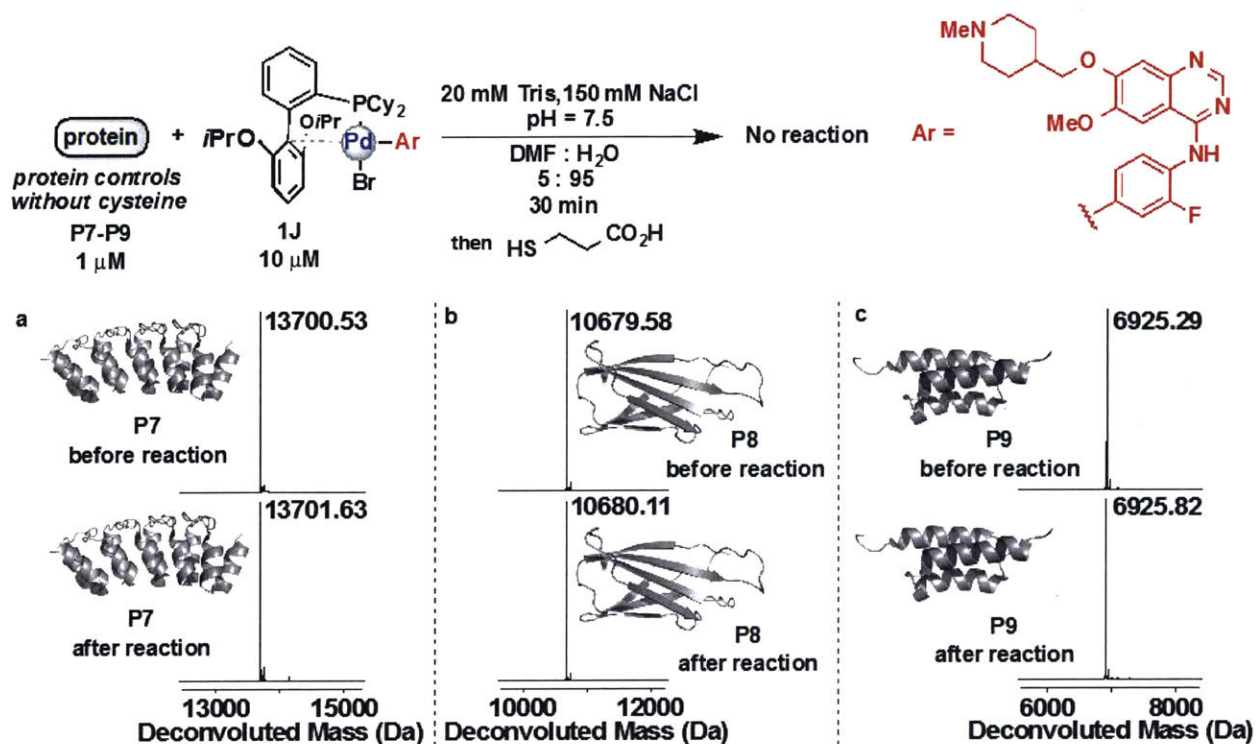
Figure 4.10. Control reactions for labeling of proteins without cysteine residues using palladium complex 1J.
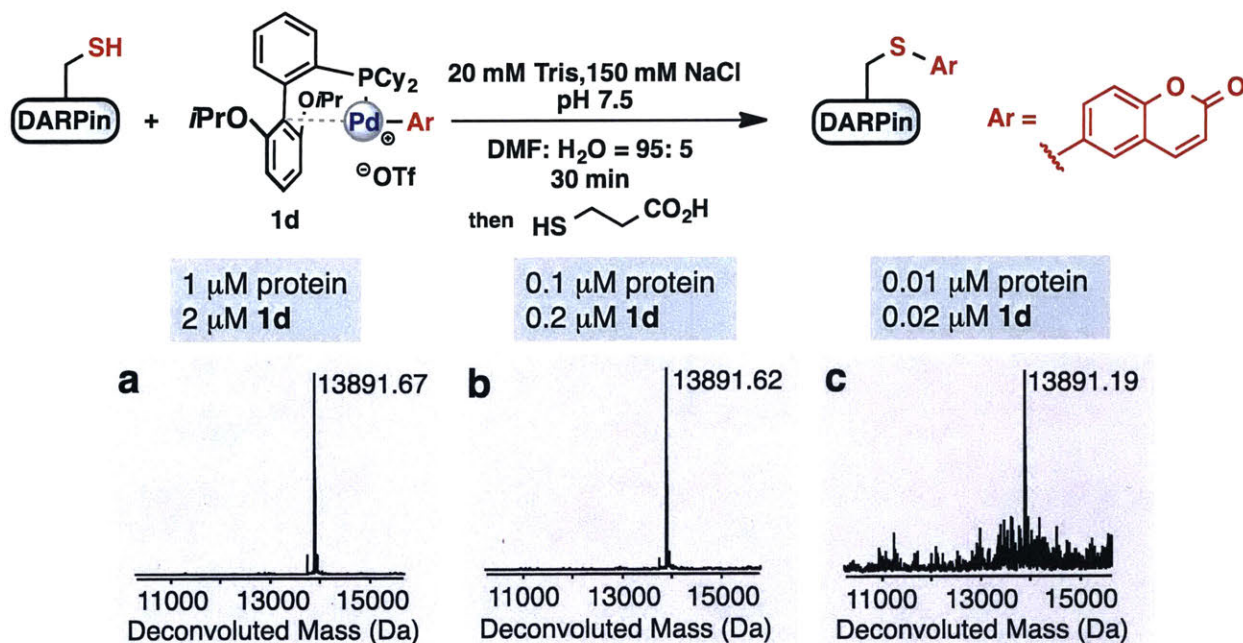
Figure 4.11. Protein labeling experiments at low concentrations.
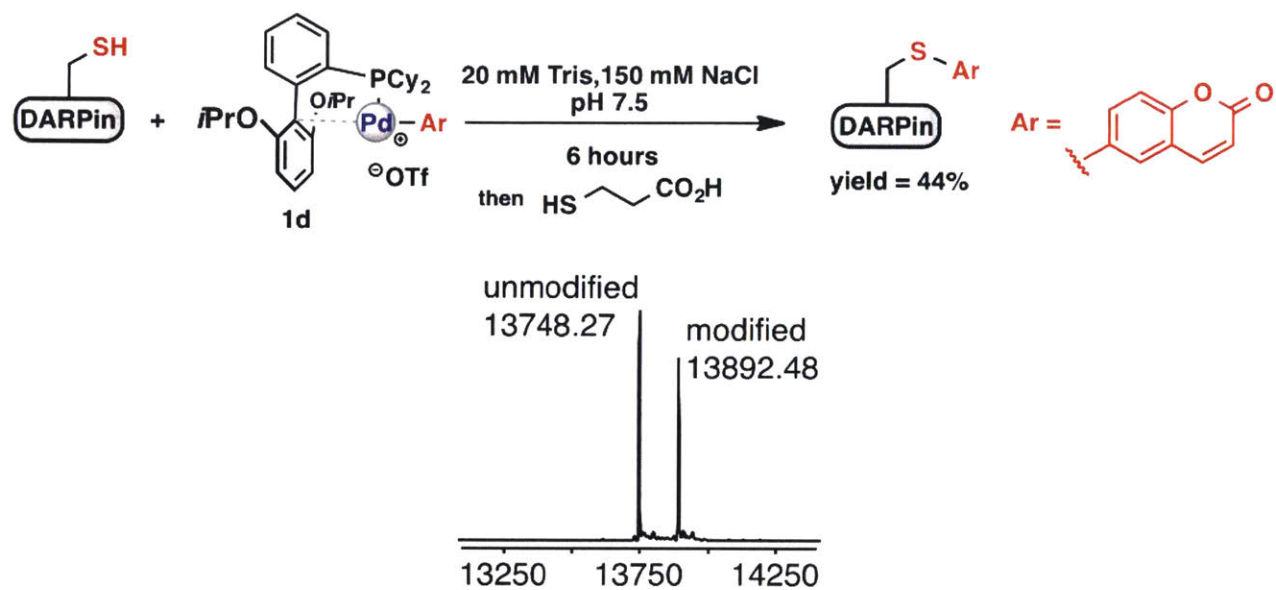
Figure 4.12. Protein labeling experiments without organic solvent.

4.1.4. Selective labeling of bacterial toxin protein variants

The developed protocol was used to arylate an engineered cysteine residue in the C-terminal region of Diphtheria toxin A-chain (DTA) fused to the lethal factor N-terminal domain (LF$_N$-DTA-Cys, see).[21] The modified LF$_N$-DTA-Cys variant was readily separated from the remaining palladium species, ligands, and other small molecules using commercially available size-exclusion chromatography (SEC) columns and displayed similar activity (EC$_{50}$ = 0.40 ± 0.09 nM) in a cell-based protein synthesis inhibition assay compared to the control (a serine mutant LFN-DTA-Ser showed EC$_{50}$ = 0.25 ± 0.05 nM, a cysteine-alkylated variant showed EC$_{50}$ = 0.20 ± 0.01 nM) (Figure 4.14).
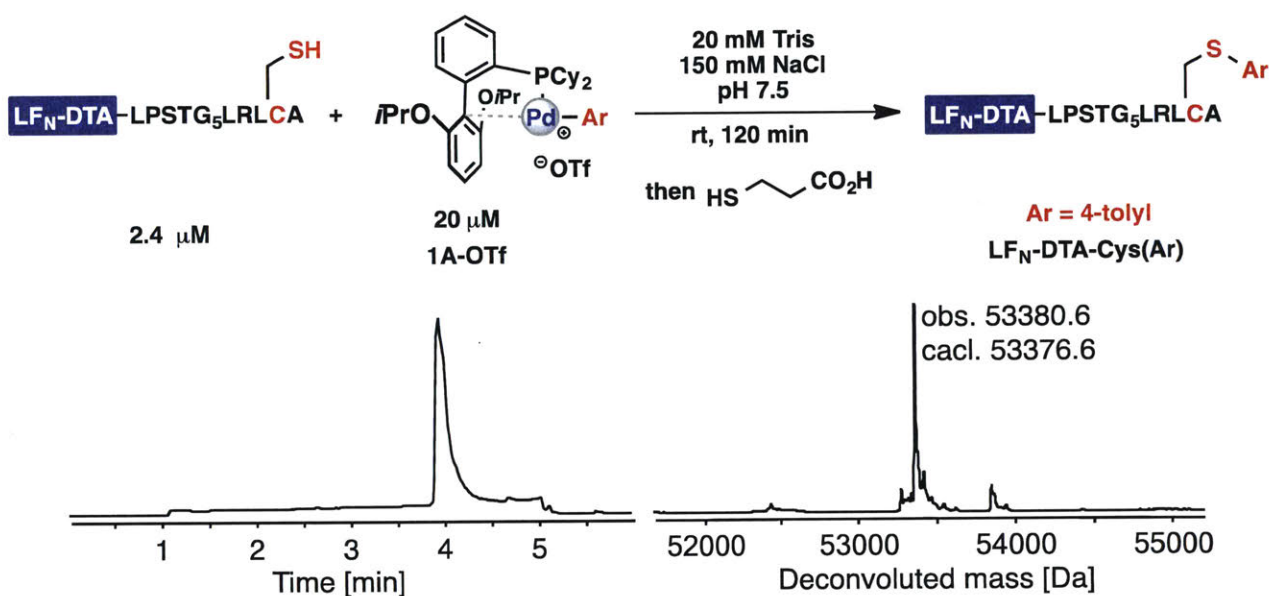
Figure 4.13. LC-MS analysis of the SEC-purified LF$_N$-DTA-Cys(Ar).

Total ion current chromatogram (right) and the deconvoluted mass spectrum of the full protein peak (left) are shown. See Experimental section 4.4.17 for detailed protocol for protein labeling and purification.
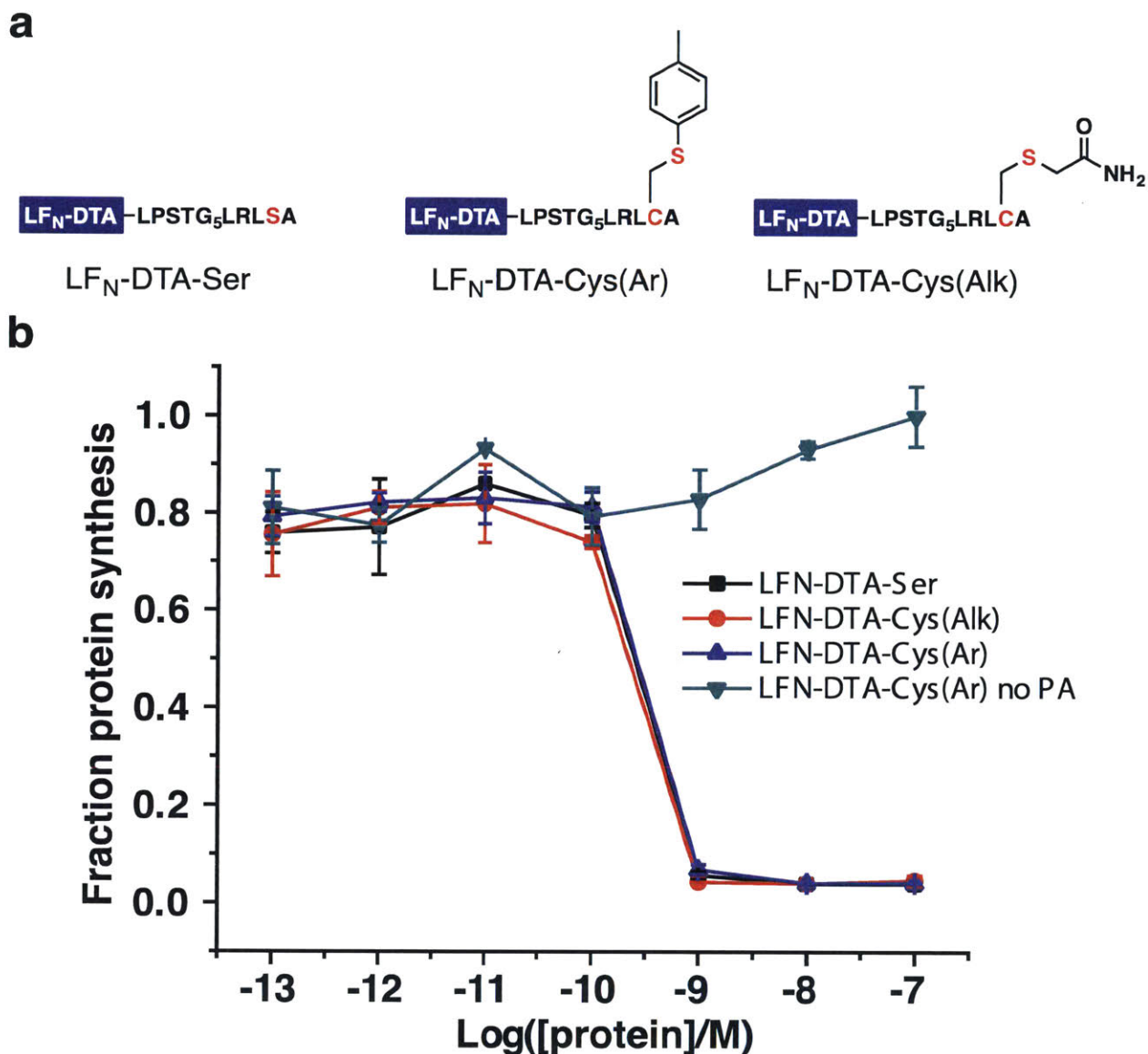
Figure 4.14. Translocation efficiencies of the LF_N-DTA variants were analyzed by protein synthesis inhibition assay.

LF_N-DTA-Cys(Ar) showed similar protein synthesis inhibition profile as the controls (LF_N-DTA-Cys(Alk) and LF_N-DTA-Ser), indicating that the palladium-mediated bioconjugation and the following purification procedure did not significantly alter the activity of the protein. EC_{50} values for each protein are: 0.40 ± 0.09 nM for LF_N-DTA-Cys(Ar), 0.20 ± 0.01 nM for LF_N-DTA-Cys(Alk), and 0.25 ± 0.05 nM for LF_N-DTA-Ser. EC_{50} values were obtained by fitting the curves to sigmoidal Boltzmann equation using OriginLab 8.0 software (Northhamptown, MA). See experimental section 4.4.17 for detailed protocol for the cell assay.

4.1.5. Synthesis of linker-free antibody-drug conjugates

Antibody-drug conjugates (ADCs) are a promising class of biotherapeutics, which combine the potency of cytotoxic drugs with the target specificity of monoclonal antibodies.[22] We aimed to attach drug molecules directly to cysteine residues in antibodies through the developed palladium conjugation chemistry. The drug payload Vandetanib was used to form palladium complex (**1J**) by making use of the aryl bromide present in its structure. Treating partially reduced Trastuzumab antibody[23] with **1J** readily produced ADCs with a 4.4 drug to antibody ratio (DAR) (Figure 4.15). The purified arylated ADCs retained binding affinity ($K_D = 0.3 \pm 0.2$ nM) to recombinant HER2 compared to the unmodified trastuzumab antibody (Figure 4.16). While traditional ADCs use various linkers to attach drug molecules to antibodies, our method significantly expands the structural space of ADCs by providing the capability to directly attach drug molecules containing native or pre-installed aryl halide or phenol functional groups. The therapeutic potential of this class of "linker-free" ADCs will be investigated in the future.
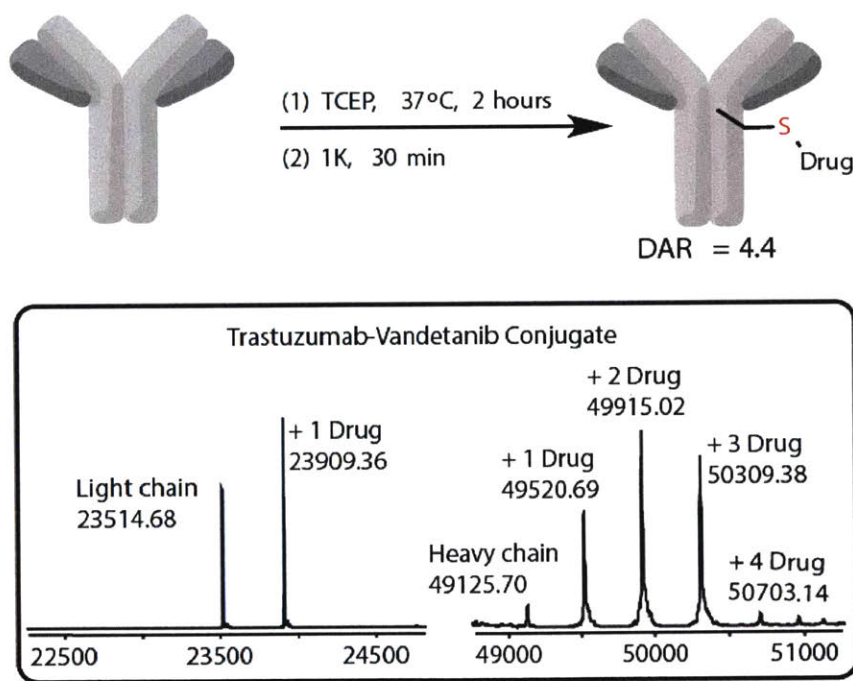
Figure 4.15. Synthesis of linker-free antibody-drug conjugates using a palladium-drug complex.

Mass spectra of the fully reduced and deglycosylated ADCs were shown. Drug-to-Antibody Ratio (DAR) represents an average number of drugs per antibody. See Experimental section for detailed protocol for antibody reduction and labeling.
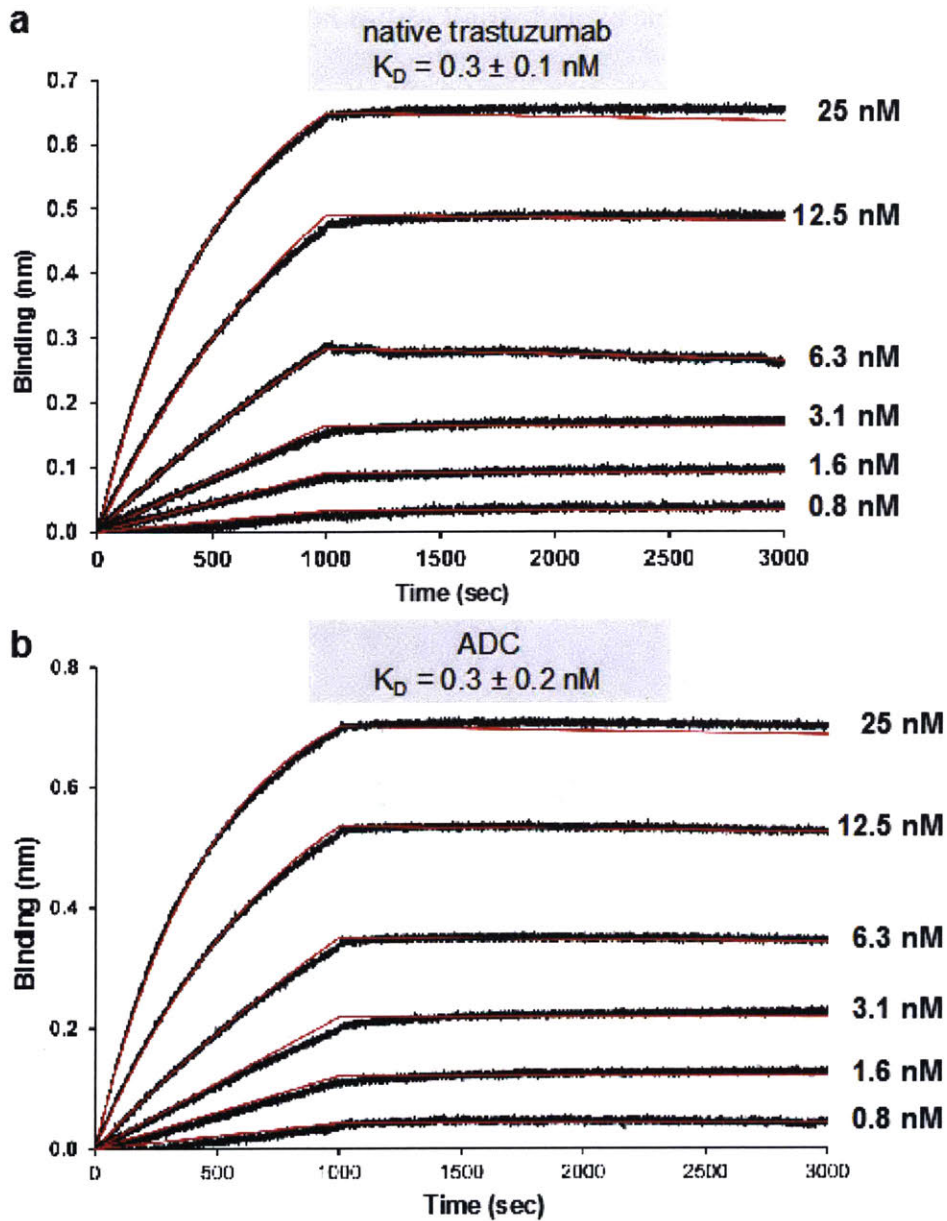
Figure 4.16. The ADC and the native trastuzumab showed similar binding affinity to HER2 in the BioLayer Interferometry assay.

Black curves are the experimental results, and red curves are the fitting results. The concentration of the recombinant HER2 used in each experiment is listed next to the corresponding curve. See Experimental section 4.4.18 for the detailed protocol of the BioLayer Interferometry assay.

4.1.6. Peptide macrocyclization using bispalladium reagents

Stapled peptides have shown significant promise as next generation therapeutics.[24,25] However, there are limited methods for the synthesis of these bioconjugates with structurally diverse linkers,[26] which hinders the systematic investigation of the effect of the linker on the properties of stapled peptides.[27] We envisioned that palladium reagents containing two electrophilic metal centers could be efficiently used to crosslink two cysteine residues on a peptide chain, thereby providing access to stapled peptides with various aryl linkers. Indeed, running the reaction at 10 μM concentration of peptide in a 1:1 (v/v) acetonitrile/water solvent mixture at pH 7.5 using a two-fold excess of the bis-palladium complex (**2A**) resulted in quantitative formation of the target stapled peptide within 10 minutes (Figure 4.17). Considering the availability of commercially or otherwise easily accessible diarylhalide reagents, this approach provides facile access to a diverse aryl-linker space for stapled peptides.[28]
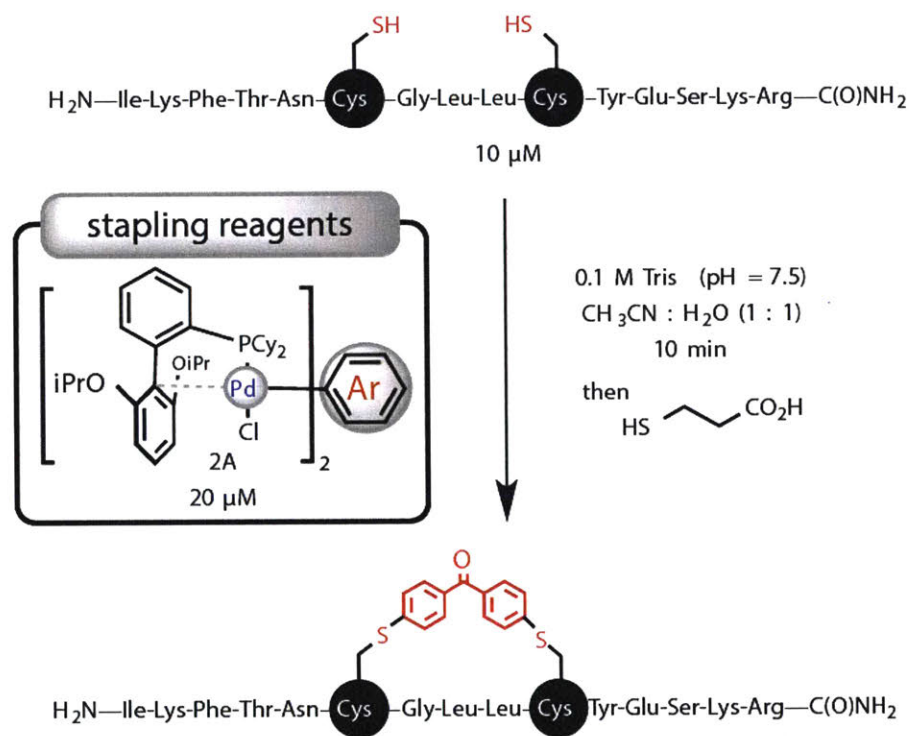
Figure 4.17. Stapling of a model peptide using a bis-palladium reagent 2A.
Reaction conditions: **P3** (10 μM), complex **2A** (20 μM), Tris buffer (100 mM; pH 7.5), CH₃CN : H₂O = 1 : 1, rt, 10 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6 equivalents to **2A**) before the LC-MS analysis.

4.3. Discussion and Conclusion

In conclusion, we have introduced a new approach for cysteine bioconjugation and have shown for the first time that palladium(II) complexes can be used for fast and chemoselective arylation of cysteine residues in complex biomolecules. The versatility of the method is particularly notable, since a large variety of aryl halides or trifluoromethanesulfonates are commercially available or readily accessible from simple precursors. Finally, the ease of preparation, storage, and use of the palladium reagents make them particularly attractive for routine application in chemistry, biology, medicine, and materials science. Further evolution of the metals and ligands employed will likely provide an extended set of organometallic bioconjugation reagents with altered selectivity and efficiency, allowing for functionalization of other amino acid residues.

4.4. Experimental

4.4.1. Chemicals

Tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) was purchased from Hampton Research (Aliso Viejo, CA). 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU), D-Biotin, Fmoc-Rink amide linker, Fmoc-L-Gly-OH, Fmoc-L-Leu-OH, Fmoc-L-Lys(Boc)-OH, Fmoc-L-Ala-OH, Fmoc-L-Cys(Trt)-OH, Fmoc-L-Gln(Trt)-OH, Fmoc-L-Asn(Trt)-OH, Fmoc-L-Glu(O*t*Bu)-OH, Fmoc-L-Arg(Pbf)-OH, Fmoc-L-Phe-OH, Fmoc-L-Ser(*t*Bu)-OH, Fmoc-L-Thr(*t*Bu)-OH, Fmoc-L-Tyr(*t*Bu)-OH, and Fmoc-L-His(Trt)-OH were purchased from Chem-Impex International (Wood Dale, IL). Aminomethyl polystyrene resin was prepared according to an in-house protocol.[29] Peptide synthesis-grade *N,N*-dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, HPLC-grade acetonitrile, and guanidine hydrochloride were obtained from VWR International (Philadelphia, PA). Aryl halides and aryl trifluoromethanesulfonates were purchased from Aldrich Chemical Co., Alfa Aesar, or Matrix Scientific and were used without additional purification. All deuterated solvents were purchased from Cambridge Isotopes and used without further purification. All other reagents were purchased from Sigma-Aldrich and used as received.

All reactions with peptides, proteins, and antibodies were set up on the bench top and carried out under ambient conditions. For procedures carried out in the nitrogen-filled glovebox, the dry degassed THF was obtained by passage through activated alumina columns followed by purging with argon. Anhydrous pentane, cyclohexane, and acetonitrile were purchased from Aldrich Chemical Company in Sureseal® bottles and were purged with argon before use.

4.4.2. NMR, FTIR, and elemental analysis

All small-molecule organic and organometallic compounds were characterized by ^{1}H, ^{13}C NMR, and IR spectroscopy, as well as elemental analysis (unless otherwise noted). ^{19}F NMR spectroscopy was used for organometallic complexes containing a trifluoromethanesulfonate counterion. ^{31}P NMR spectroscopy was used for characterization of palladium complexes. Copies of the ^{1}H, ^{13}C, ^{31}P, and ^{19}F NMR spectra can be found at the end of the Supporting Information. Nuclear Magnetic Resonance spectra were recorded on a Bruker 400 MHz instrument and a Varian 300 MHz instrument. Unless otherwise stated, all ^{1}H NMR experiments are reported in δ units, parts per million (ppm), and were measured relative to the signals of the residual proton resonances CH_2Cl_2 (5.32 ppm) or CH_3CN (1.94 ppm) in the deuterated solvents. All ^{13}C NMR spectra are measured decoupled from ^{1}H nuclei and are reported in δ units (ppm) relative to CD_2Cl_2 (54.00 ppm) or CD_3CN (118.69 ppm), unless otherwise stated. All ^{31}P NMR spectra are measured decoupled from ^{1}H nuclei and are reported relative to H_3PO_4 (0.00 ppm). ^{19}F NMR spectra are measured decoupled from ^{1}H nuclei and are reported in ppm relative to $CFCl_3$ (0.00 ppm) or α,α,α-trifluorotoluene (-63.72 ppm). All FT-IR spectra were recorded on a Thermo Scientific – Nicolet iS5 spectrometer (iD5 ATR – diamond). Elemental analyses were performed by Atlantic Microlabs Inc., Norcross, GA.

4.4.3. LC-MS analysis

LC-MS chromatograms and associated mass spectra were acquired using Agilent 6520 ESI-Q-TOF mass spectrometer. Solvent compositions used in the majority of experiments are 0.1% TFA in H_2O (solvent A) and 0.1% TFA in acetonitrile (solvent B). The following LC-MS methods were used:

__Method A__ LC conditions: Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-3 min 5% B, 3-22 min 5-95% B, 22-24 min 95% B, flow rate: 0.8 mL/min. MS conditions: positive electrospray ionization (ESI) extended dynamic mode in mass range 300 – 3000 *m/z*, temperature of drying gas = 350 °C, flow rate of drying gas = 11 L/min, pressure of nebulizer gas = 60 psi, the capillary, fragmentor, and octupole rf voltages were set at 4000, 175, and 750, respectively.

__Method B__ LC conditions: Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 40 °C, gradient: 0-2 min 5% B, 2-11 min 5-65% B, 11-12 min 65% B, flow rate: 0.8 mL/min. MS conditions are same as *Method A*.

__Method C__ LC conditions: Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, column temperature: 40 °C, gradient: gradient: 0-2 min 5% B, 2-10 min 5-95% B, 10-11 min 95% B, flow rate: 0.8 mL/min. MS conditions are same as *Method A*.

Data were processed using Agilent MassHunter software package. Deconvoluted masses of proteins were obtained using maximum entropy algorithm.

LC-MS data shown were acquired using Method A, unless otherwise noted; Y-axis in all chromatograms shown in supplementary figures represents total ion current (TIC); mass spectrum insets correspond to the integration of the TIC peak unless otherwise noted.

4.4.4. Determination of reaction yields

All reported yields were determined by integrating TIC spectra. First, the peak areas for all relevant peptide-containing species on the chromatogram were integrated using Agilent MassHunter software package. Since no peptide-based side products were generated in the experiments, the yields shown in Table 2 were determined as follows: %yield = S_{pr}/S_{total} where S_{pr} is the peak area of the product and S_{total} is the peak area of combined peptide-containing species (product and starting material). The yield of the stapled peptide **P3-A** was calculated as follows: %yield = $k \cdot S_{pr}/S_{st}$ where S_{pr} is the peak area of the reaction product, S_{st} is the peak area of a known amount of purified product, and k equals to the ratio of the known amount of standard divided by the initial amount of starting material. For peptide stability experiments the conversion was calculated as following: %remaining peptide = S_t/S_0 where S_t is the peak area of the corresponding cysteine conjugate at time t, and S_0 is the peak area of the conjugate at time 0.

4.4.5. X-ray structure determination of 1A-OTf

Low-temperature diffraction data (φ- and ω-scans) was collected on a Bruker-AXS X8 Kappa Duo diffractometer coupled to a Smart Apex2 CCD detector with Mo $K\alpha$ radiation (λ = 0.71073 Å) from an $I\mu S$ micro-source for the structure of compound **1A-OTf**. The structure was solved by direct methods using SHELXS[30] and refined against F^2 on all data by full-matrix least squares with SHELXL-97[31] using established refinement techniques.[32] All non-hydrogen atoms were refined anisotropically. All hydrogen atoms were included into the model at geometrically calculated positions and refined using a riding model. The isotropic displacement parameters of all hydrogen atoms were fixed to 1.2 times the U value of the atoms they are linked to (1.5 times for methyl groups). All disordered atoms were refined with the help of similarity restraints on the 1,2- and 1,3-distances and displacement parameters as well as rigid bond restraints for anisotropic displacement parameters unless otherwise noted below.

X-ray quality crystals of **1A-OTf•CH₃CN** were obtained by vapor diffusion of an Et₂O/CH₃CN solution of **1A-OTf** with pentane. Under these conditions, a molecule of CH₃CN coordinates to palladium in **1A-OTf**, displacing the trifluoromethanesulfonate anion. The resulting **1A-OTf•CH₃CN** crystallizes in the monoclinic space group $P2_1$ with one molecule in the asymmetric unit.

Table 4.3. Crystal data and structure refinement for 1A-OTf•CH₃CN.

Empirical formula	$C_{40}H_{53}F_3NO_5PPdS$
Formula weight	854.26
Temperature	100(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	$P2_1$
Unit cell dimensions	a = 10.5958(3) Å $\quad$ α= 90°.
	b = 17.8048(6) Å $\quad$ β= 98.6068(10)°.
	c = 10.9802(4) Å $\quad$ γ = 90°.
Volume	2048.15(12) $Å^3$
Z	2
Density (calculated)	1.385 Mg/m^3
Absorption coefficient	0.599 mm^{-1}
F(000)	888
Crystal size	0.300 x 0.250 x 0.070 mm^3
Theta range for data collection	1.876 to 38.568°.
Index ranges	-18<=h<=18, -31<=k<=31, -19<=l<=19
Reflections collected	177765
Independent reflections	23140 [R(int) = 0.0297]
Completeness to theta = 25.242°	100.0 %
Absorption correction	Semi-empirical from equivalents
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	23140 / 1 / 475
Goodness-of-fit on F^2	1.023
Final R indices [I>2sigma(I)]	R1 = 0.0230, wR2 = 0.0518
R indices (all data)	R1 = 0.0265, wR2 = 0.0529
Absolute structure parameter	-0.009(2)
Extinction coefficient	n/a
Largest diff. peak and hole	0.588 and -0.463 e.$Å^{-3}$

4.4.6. Synthesis of palladium complexes

Aryl halide and trifluoromethanesulfonate precursors

4-Aminophenyl trifluoromethanesulfonate,[33] fluorescein monotriflate,[34] and 2-ethyl-6-methylpyridin-3-yl trifluoromethanesulfonate[35] were prepared according to literature procedures.

Aryl trifluoromethanesulfonate S1
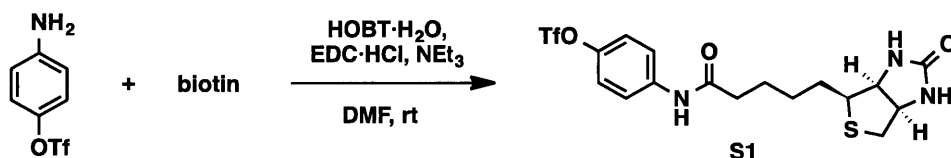
In an oven-dried round-bottom flask (250 mL) biotin (810.3 mg, 3.32 mmol) was suspended in anhydrous DMF (20 mL). Subsequently, 1-hydroxybenzotriazole hydrate (609.4 mg, 3.98 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (763.0 mg, 3.98 mmol), and NEt$_3$ (0.88 mL, 3.32 mmol) were added to the reaction and the mixture was stirred at room temperature for 15 min. After this time, 4-aminophenyl trifluoromethanesulfonate (800 mg, 3.32 mmol) was added and the reaction was stirred for an additional 3 h. Upon completion of the reaction, DMF was removed under vacuum. Addition of CH$_2$Cl$_2$ resulted in the formation of a white precipitate, which was filtered, washed with additional CH$_2$Cl$_2$ and pentane, and dried under reduced pressure to afford the final product as a white solid (829.6 mg, 54%).

^{1}H NMR (400 MHz, DMSO-d_6) δ⎕ 12.01 (s, br, NH), 10.18 (s, br, NH), 7.74 (d, J = 8.8 Hz, 1H), 7.41 (d, J = 8.8 Hz, 1H), 6.43 (s, 1H), 6.36 (s, 1H), 4.31 (m, 1H), 4.13 (m, 1H), 3.10 (m, 1H), 2.82 (dd, J = 12.5, 4.8 Hz, 2H), 2.58 (d, J = 12.4 Hz, 1H), 2.33 (t, J = 7.5 Hz, 1H), 2.20 (t, J = 7.4 Hz, 1H), 1.62 (m, 2H), 1.50 (m, 2H), 1.35 (m, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 174.48, 171.62, 162.77, 162.76, 144.01, 139.67, 121.86, 120.46, 118.28 (q, J = 321 Hz, CF$_3$), 61.10, 61.09, 59.25, 59.23, 55.43, 55.41, 36.23, 33.52, 28.22, 28.15, 28.11, 28.08, 25.02, 24.57 (observed complexity is due to $C–F$ coupling). ^{19}F NMR (282 MHz, DMSO-d_6) δ -73.16. FT-IR (neat, cm^{-1}): 3262.53, 2923.74, 1698.21, 1662.2, 1529.16, 1501.48, 1464.35, 1405.69, 1323.64, 1250.64, 1210.37, 1137.17, 883.25, 839.93, 694.8, 667.57, 633.6, 603.6. HRMS electrospray (*m/z*): [M + H]$^+$ calcd for C$_{17}$H$_{21}$F$_3$N$_3$O$_5$S$_2$: 467.0869, found 468.0885; [2M + H]$^+$ calcd for C$_{34}$H$_{41}$F$_6$N$_6$O$_{10}$S$_4$: 935.1671, found 935.1766.

Synthesis of Oxidative Addition Complexes

[(1,5-COD)PdCl$_2$] was prepared according to literature procedure.[36] [(1,5-COD)Pd(CH$_2$TMS)$_2$] was prepared according to a modified literature procedure.[37] COD is 1,5-cyclooctadiene.

Synthesis of [(1,5-COD)Pd(CH$_2$TMS)$_2$]

A flame-dried Schlenk flask (100 mL), equipped with a magnetic stir bar, was filled with argon and charged with (1,5-COD)PdCl$_2$ (3.15g, 11.05 mmol). The flask was put under vacuum and filled back with argon. The procedure was repeated 3 consecutive times. Diethyl ether (49.3 mL) was introduced via syringe, the reaction was cooled to –40 °C (acetonitrile/dry ice bath) and TMSCH$_2$MgCl (23.4 mL, 1.0 M) was added dropwise over 10–20 min. The reaction was stirred at –40 °C for 1 h and then at 0 °C (ice/water bath) for an additional 20 min. Acetone (1.3 mL) was added at 0 °C, the reaction mixture was stirred for 5 min, after which the solvent was removed under vacuum using an external trap (the flask was kept at 0 °C). The flask was then opened to air, pentane (100 mL) was added and the crude material was filtered through a pad of Celite into a new round-bottom flask (500 mL) at 0 °C. The filter cake was washed with pentane (50 mL × 2). Pentane from the combined washes was removed with the aid of a rotary evaporator at 0 °C (ice/water bath). The resulting white solid was dried under vacuum for 2 h at 0 °C, and transferred into a 20 mL scintillation vial in the glovebox (3.00 g, 70%). The ^{1}H and ^{13}C NMR spectra of the obtained material are identical to those reported in the literature.[10] The title compound was stored in the glovebox at –20 °C.

Representative Procedure for the Synthesis of Oxidative Addition Complexes.

In a nitrogen-filled glovebox, an oven-dried scintillation vial (10 mL), which was equipped with a magnetic stir bar, was charged with RuPhos (1.1 equiv), Ar–X (1.1 equiv), and cyclohexane. Solid (COD)Pd(CH$_2$SiMe$_3$)$_2$ (1 equiv) was added rapidly in one portion and the resulting solution was stirred for 16 h at rt. After this time, pentane (3 mL) was added and the resulting mixture was placed into a –20 °C freezer for 3 h. The vial was then taken outside of the glovebox, and the resulting precipitate was filtered, washed with pentane (3 × 3 mL), and dried under reduced pressure to afford the oxidative addition complex.
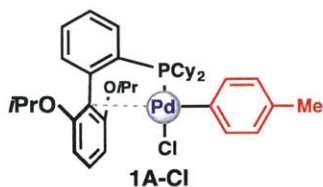
1A-Cl

Following the general procedure, a mixture containing 4-chlorotoluene (17 μL, 0.14 mmol), RuPhos (66 mg, 0.14 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (50 mg, 0.13 mmol) was stirred at rt in cyclohexane (1.5 mL) for 16 h. General work up afforded **1A-Cl** as a white solid (68.7 mg, 77%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.62 (t, J = 8.4 Hz, 2H), 7.43 (tt, J = 7.6, 1.5 Hz, 2H), 7.38 (m, 2H), 6.93 (dd, J = 8.2, 2.1 Hz, 2H), 6.86 (ddd, J = 7.6, 3.0, 1.3 Hz, 1H), 6.77 (d, J = 8.1 Hz, 2H), 6.63 (d, J = 8.5 Hz, 2H), 4.62 (hept, J = 6.1 Hz, 2H), 2.21 (s, 3H), 2.12 (m, 2H), 1.75 (d, J = 12.1 Hz, 6H), 1.63 (m, 6H), 1.38 (d, J = 6.0 Hz, 6H), 1.16 (m, 6H), 1.01 (d, J = 6.1 Hz, 6H), 0.78 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 158.94, 145.05, 144.88, 136.63, 136.60, 134.55, 133.94, 133.75, 133.74, 133.59, 132.58, 132.48, 132.37, 131.03, 130.50, 130.48, 127.88, 127.86, 126.32, 126.26, 111.65, 111.61, 107.08, 70.83, 33.86, 33.59, 28.26, 27.77, 27.75, 27.29, 27.16, 26.99, 26.95, 26.84, 26.12, 21.95, 21.40, 20.33 (observed complexity is due to C–P coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 31.40. FT-IR (neat, cm^{-1}): 2973.28, 2927.26, 2850.11, 1595.76, 1585.67, 1482.98, 1452.13, 1382.47, 1367.3, 1293.75, 1271.69, 1238.07, 1136.18, 1114.6, 1052.64, 1014.05, 1002.69, 948.47, 941.24, 915.67, 894.84, 888.54, 847.55, 797.78, 759.42, 745.72, 733.6, 720.03, 668.01, 610.21. Anal. Calcd. for C$_{37}$H$_{50}$ClO$_2$PPd: C, 63.52; H, 7.20. Found: C, 63.47; H, 7.24.
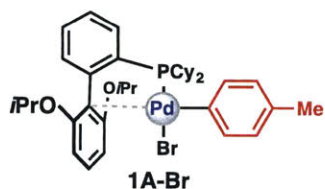
1A-Br

Following the general procedure, a mixture containing 4-bromotoluene (24.2 mg, 0.14 mmol), RuPhos (66.0 mg, 0.14 mmol), and (COD)Pd(CH₂SiMe₃)₂ (50.0 mg, 0.13 mmol) was stirred at rt in cyclohexane (1 mL) for 16 h. General work up afforded **1A-Br** as an off-white solid (78.4 mg, 82%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.61 (m, 2H), 7.43 (tt, J = 7.5, 1.6 Hz, 1H), 7.37 (m, 1H), 6.91 (dd, J = 8.2, 2.3 Hz, 2H), 6.86 (ddd, J = 7.8, 3.1, 1.5 Hz, 1H), 6.76 (d, J = 8.0 Hz, 2H), 6.64 (d, J = 8.4 Hz, 2H), 4.60 (hept, J = 6.1 Hz, 2H), 2.22 (s, 3H), 2.14 (m, 2H), 1.77 (m, 6H), 1.60 (m, 6H), 1.38 (d, J = 6.0 Hz, 6H), 1.17 (m, 6H), 1.01 (d, J = 6.0 Hz, 6H), 0.78 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 159.42, 145.38, 145.20, 137.74, 137.70, 134.88, 134.18, 133.84, 133.11, 133.01, 132.94, 131.62, 131.56, 130.99, 130.97, 128.20, 126.81, 126.76, 112.44, 112.41, 107.88, 71.44, 34.40, 34.14, 28.73, 28.17, 28.15, 27.82, 27.69, 27.49, 27.46, 27.35, 26.60, 22.46, 21.93, 20.79 (observed complexity is due to C–P coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 29.89. FT-IR (neat, cm^{-1}): 2980.24, 2917.56, 2844.37, 1587.1, 1478.11, 1464.73, 1451.26, 1442.93, 1384.38, 1372.21, 1349.62, 1333.2, 1289.54, 1255.37, 1207.89, 1112.89, 1069.63, 1054.03, 1010.41, 902.77, 892.23, 851.92, 837.45, 791.2, 785.02, 767.28, 760.24, 747.31, 736.44, 700.12, 675.71, 612.24. HRMS electrospray (m/z): [M – Br]$^+$ calcd for C$_{37}$H$_{50}$O$_2$PPd: 663.2596, found 663.2603.

1A-I
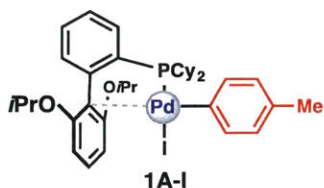
Following the general procedure, a mixture containing 4-iodotoluene (61.7 mg, 0.28 mmol), RuPhos (131.9 mg, 0.28 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (100.0 mg, 0.26 mmol) was stirred at rt in cyclohexane (1.5 mL) for 16 h. General work up afforded **1A-I** as a bright yellow solid (180.0 mg, 89%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.60 (m, 2H), 7.42 (m, 1H), 7.36 (m, 1H), 6.89 (dd, J = 8.2, 2.0 Hz, 2H), 6.84 (ddd, J = 7.5, 2.9, 1.2 Hz, 2H), 6.74 (d, J = 8.0 Hz, 2H), 6.66 (d, J = 8.4 Hz, 2H), 4.59 (hept, J = 6.2 Hz, 2H), 2.22 (s, 3H), 2.14 (m, 2H), 1.77 (m, 6H), 1.59 (m, 6H), 1.38 (d, J = 6.0 Hz, 6H), 1.18 (m, 6H), 1.01 (d, J = 6.1 Hz, 6H), 0.77 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 159.56, 145.15, 144.97, 138.88, 138.85, 134.81, 133.96, 133.64, 133.19, 133.08, 132.67, 131.63, 130.96, 130.94, 127.85, 127.84, 126.76, 126.71, 126.62, 112.67, 112.63, 108.32, 71.59, 34.38, 34.13, 28.63, 28.03, 28.00, 27.87, 27.74, 27.48, 27.37, 26.57, 26.55, 22.48, 21.97, 20.71 (observed complexity is due to *C–P* coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 25.64. FT-IR (neat, cm^{-1}): 2975.8, 2922.64, 2844.86, 1585.12, 1569.19, 1477.72, 1454.19, 1386.62, 1374.01, 1326.62, 1282, 1251.88, 1207.49, 1171.57, 1135.22, 1112.08, 1065.38, 1055.2, 1010.27, 997.53, 846.52, 787.24, 772.21, 767.54, 759.04, 747.91, 734.94, 727.13, 696.29, 676.75, 637.81, 612.42, 602.4, 579.03. Anal. Calcd. for C$_{37}$H$_{50}$IO$_2$PPd: C, 56.18; H, 6.37. Found: C, 55.69; H, 6.42.

1A-OTf
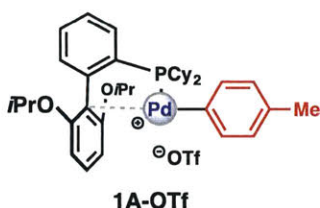
Following the general procedure, a mixture containing 4-tolyl trifluoromethanesulfonate (100.0 mg, 0.42 mmol), RuPhos (194.0 mg, 0.42 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (147.0 mg, 0.38 mmol) was stirred at rt in cyclohexane (1.5 mL) for 16 h. General work up afforded **1A-OTf** as an off-white solid (270.0 mg, 88%).

^{1}H NMR (400 MHz, CD$_3$CN) δ 7.73 (t, J = 7.7 Hz, 1H), 7.64 (t, J = 8.5 Hz, 1H), 7.52 (m, 1H), 7.47 (m, 1H), 6.93 (m, 4H), 6.77 (m, 3H), 4.68 (hept, J = 6.2 Hz, 2H), 2.22 (m, 5H), 1.79 (m, 6H), 1.60 (m, 6H), 1.41 (d, J = 6.0 Hz, 6H), 1.20 (m, 6H), 1.03 (d, J = 6.0 Hz, 6H), 0.69 (m, 2H). ^{13}C NMR (101 MHz, CD$_3$CN) δ 163.57, 145.97, 145.80, 138.39, 137.14, 137.10, 136.03, 134.17, 133.30, 133.28, 133.06, 133.01, 132.89, 130.30, 130.27, 128.37, 128.30, 107.26, 72.63, 35.36, 35.05, 29.55, 29.44, 28.02, 27.95, 27.82, 27.75, 27.63, 27.09, 27.07, 22.58, 22.30, 21.04 (observed complexity is due to *C–P* coupling). ^{31}P NMR (121 MHz, CD$_3$CN) δ 41.32. ^{19}F NMR (376 MHz, CD$_3$CN) δ -82.72. FT-IR (neat, cm^{-1}): 2973.6, 2921.73, 2850.32, 1589.06, 1480.22, 1451.38, 1389, 1378.25, 1308.92, 1255.2, 1229.22, 1208.11, 1177.28, 1141.2, 1110.23, 1066.14, 1055.85, 1016.65, 900.16, 849.85, 791.27, 767.96, 736.56, 628.06, 613.09. Anal. Calcd. for C$_{38}$H$_{50}$F$_3$O$_5$PPdS: C, 56.12; H, 6.20. Found: C, 56.39; H, 6.46.

1B
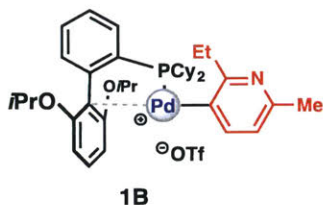
Following the general procedure, a mixture containing 2-ethyl-6-methylpyridin-3-yl trifluoromethanesulfonate (76.0 mg, 0.28 mmol, *Note:* 2.2 equiv was used), RuPhos (66.0 mg, 0.141 mmol), and (COD)Pd(CH₂SiMe₃)₂ (50.0 mg, 0.129 mmol) was stirred at rt in cyclohexane (0.75 mL) for 16 h. General work up afforded **1B** as a light yellow solid (95.0 mg, 88%).

^{1}H NMR (400 MHz, CD$_3$CN) δ 7.70 (m, 2H), 7.53 (tt, J = 7.6, 1.6 Hz, 1H), 7.47 (tt, J = 7.4, 1.6 Hz, 1H), 6.99 (dd, J = 7.9, 2.8 Hz, 1H), 6.79 (m, 4H), 4.71 (m, 1H), 4.63 (m, 1H), 3.17 (m, 2H), 2.40 (m, 5H), 2.12 (m, 2H), 1.76 (m, 6H), 1.49 (m, 5H), 1.40 (d, J = 6.0 Hz, 3H), 1.34 (d, J = 6.0 Hz, 3H), 1.25 (m, 5H), 1.13 (m, 9H), 0.92 (d, J = 6.0 Hz, 3H), 0.17 (m, 1H). ^{13}C NMR (101 MHz, CD$_3$CN) δ 164.70, 164.45, 164.09, 154.73, 145.65, 145.48, 143.59, 143.54, 138.99, 135.58, 135.14, 133.08, 133.06, 132.54, 132.42, 132.31, 130.75, 128.13, 128.07, 122.77, 106.93, 106.75, 73.27, 72.21, 36.84, 35.72, 35.42, 33.66, 33.37, 31.96, 31.93, 29.44, 27.82, 27.75, 27.66, 27.60, 27.57, 27.45, 27.11, 26.99, 26.88, 26.83, 26.67, 26.61, 23.36, 22.37, 22.24, 22.09, 21.80, 13.29 (observed complexity is due to *C–P* coupling). ^{31}P NMR (121 MHz, CD$_3$CN) δ 41.30. ^{19}F NMR (376 MHz, CD$_3$CN) δ -78.65. FT-IR (neat, cm^{-1}): 2926.73, 2852.81, 1589.56, 1558.05, 1447.32, 1385.47, 1307.23, 1274.47, 1252.38, 1225.28, 1200.75, 1173.43, 1143.04, 1104.51, 1061.88, 1030.39, 1010.98, 892.25, 849.62, 820.09, 806.38, 769.71, 748.12, 737.28, 627.7, 609.26. Anal. Calcd. for C$_{39}$H$_{53}$F$_3$NO$_5$PPdS: C, 55.61; H, 6.34. Found: C, 55.15; H, 6.19.

1C
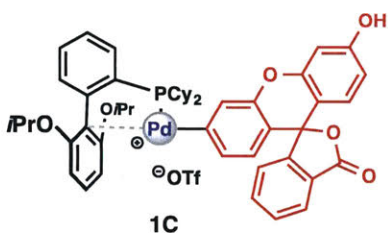
Following the general procedure, a mixture containing fluorescein monotrifluoromethanesulfonate (52.5 mg, 0.11 mmol, *Note:* used as the limiting reagent), RuPhos (66.0 mg, 0.14 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (50.0 mg, 0.13 mmol) was stirred in THF (0.75 mL) at rt for 16 h using aluminum foil for light exclusion. General work up afforded **1C** as a bright orange precipitate (107.5 mg, 92%).

^{1}H NMR (400 MHz, DMSO) *Complex spectrum observed (see Appendix)* ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 44.41. ^{19}F NMR (376 MHz, CD$_2$Cl$_2$) δ -79.08. FT-IR (neat, cm^{-1}): 2929.22, 2851.87, 1761.86, 1625.02, 1587.77, 1564.66, 1509.79, 1443.16, 1384.65, 1346.61, 1274.48, 1252.52, 1222.15, 1154.31, 1106.89, 1083.41, 1064.08, 1029.68, 981.25, 949.65, 934.76, 891.29, 849.46, 820.67, 763.46, 695.31, 660.91, 636.48, 626.65. Anal. Calcd. for C$_{51}$H$_{54}$F$_3$O$_9$PPdS: C, 59.05; H, 5.25. Found: C, 58.87; H, 5.42.

1D
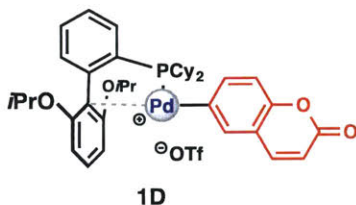
Following the general procedure, a mixture containing 2-oxo-2*H*-chromen-6-yl trifluoromethanesulfonate (38.2 mg, 0.13 mmol, *Note:* 1.01 equiv was used), RuPhos (66.0 mg, 0.14 mmol), and (COD)Pd(CH₂SiMe₃)₂ (50.0 mg, 0.13 mmol) was stirred at rt in THF (0.75 mL) for 16 h. General work up afforded **1D** as a light yellow solid (103.3 mg, 93%).

^{1}H NMR (400 MHz, CD$_3$CN) δ 7.73 (d, J = 8.0 Hz, 1H), 7.58 (m, 3H), 7.45 (m, 1H), 7.29 (d, J = 8.5 Hz, 1H), 7.14 (s, 1H), 7.08 (d, J = 8.6 Hz, 1H), 6.88 (m, 1H), 6.78 (d, J = 8.5 Hz, 2H), 6.33 (d, J = 9.5 Hz, 1H), 4.69 (hept, J = 6.1 Hz, 2H), 2.10 (s, br, 2H), 1.81 (m, 6H), 1.64 (m, 6H), 1.39 (d, J = 6.0 Hz, 6H), 1.28 (m, 2H), 1.13 (m, 4H), 1.05 (d, J = 6.1 Hz, 6H), 0.84 (m, 2H). ^{13}C NMR (101 MHz, CD$_3$CN) δ 161.41, 153.57, 144.67, 139.93, 139.90, 135.33, 135.29, 132.85, 128.01, 127.95, 117.07, 116.41, 107.42, 107.38, 72.69, 35.73 (br), 29.66, 27.66, 27.61, 27.53, 27.41, 27.24, 26.73, 26.72, 22.14 (observed complexity is due to *C–P* coupling). ^{31}P NMR (121 MHz, CD$_3$CN) δ 42.43. ^{19}F NMR (376 MHz, CD$_3$CN) δ -78.89. FT-IR (neat, cm^{-1}): 2979.13, 2924.12, 2848.66, 1721.51, 1584.47, 1543.58, 1453.67, 1413.89, 1384.87, 1315.94, 1254.76, 1227.39, 1213.36, 1195.09, 1159.85, 1139.1, 1112.08, 1071.78, 1030.22, 1004.07, 891.57, 877.55, 851.16, 826.12, 772.82, 758.61, 747.4, 737.17, 632.64. Anal. Calcd. for C$_{40}$H$_{48}$F$_3$O$_7$PPdS: C, 55.40; H, 5.58. Found: C, 55.77; H, 5.49.

1E
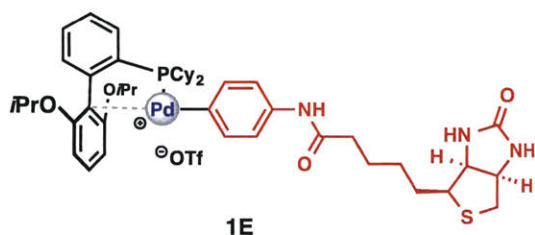
Following the general procedure, a mixture containing aryl trifluoromethanesulfonate **S1** (100.0 mg, 0.21 mmol, *Note:* 1 equiv was used), RuPhos (109.8 mg, 0.24 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (83.2 mg, 0.21 mmol) was stirred in THF (1.5 mL) at rt for 16 h. General work up afforded **1E** as a light orange solid (179.0 mg, 80%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) *Complex spectrum observed (see Appendix).* ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 42.30. ^{19}F NMR (376 MHz, CD$_2$Cl$_2$) δ -78.99. FT-IR (neat, cm^{-1}): 3301.3 (br), 2929.22, 2853.24, 1699.37, 1579.49, 1519.47, 1480.55, 1449.2, 1383.79, 1330.61, 1250.31, 1154.46, 1106.11, 1063.64, 1029.55, 1004.78, 898.47, 849.61, 819.71, 763.79, 736.07, 667.69, 636.99. HRMS electrospray (*m/z*): [M – OSO$_2$CF$_3$]$^+$ calcd for C$_{46}$H$_{47}$N$_3$O$_4$PPdS$_2$: 890.3329, found 890.3325.

1F
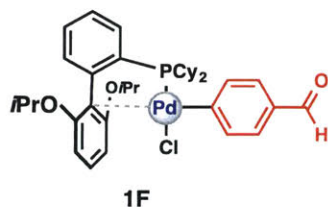
Following the general procedure, a mixture containing 4-chlorobenzaldehyde (39.7 mg, 0.28 mmol), RuPhos (131.9 mg, 0.28 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (100.0 mg, 0.26 mmol) was stirred in cyclohexane (1.5 mL) at rt for 16 h. General work up afforded **1F** as a white solid (166.0 mg, 91%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 9.83 (s, 1H), 7.64 (m, 2H), 7.46 (tt, J = 7.7, 1.6 Hz, 1H), 7.40 (m, 5H), 6.89 (ddd, J = 7.7, 3.2, 1.4 Hz, 1H), 6.66 (d, J = 8.5 Hz, 2H), 4.64 (hept, J = 6.1 Hz, 2H), 2.11 (m, 2H), 1.68 (m, 12H), 1.39 (d, J = 6.0 Hz, 6H), 1.18 (m, 6H), 1.03 (d, J = 6.1 Hz, 6H), 0.77 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 192.86, 159.84, 154.40, 145.37, 145.19, 138.25, 138.21, 135.60, 133.61, 133.26, 133.10, 132.99, 131.50, 131.32, 131.30, 127.36, 127.06, 127.00, 111.36, 111.32, 107.76, 71.62, 34.46, 34.19, 28.82, 28.37, 28.35, 27.68, 27.54, 27.48, 27.35, 27.23, 26.59, 26.58, 22.44, 21.86 (observed complexity is due to $C–P$ coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 33.19. FT-IR (neat, cm^{-1}): 2928.23, 2849.21, 2812.38, 2721.21, 1693.19, 1652.71, 1591.79, 1572.27, 1551.15, 1506.8, 1456.37, 1382.73, 1371.34, 1329.84, 1295.43, 1268.02, 1243.71, 1216.23, 1166.08, 1113.54, 1074.07, 1054.93, 1011.02, 1004.34, 939.3, 917.45, 896.81, 888.89, 848.95, 837.59, 810.58, 778.57, 763.38, 744.18, 737.98, 724, 674.14, 612.18. Anal. Calcd. for C$_{37}$H$_{48}$ClO$_3$PPd: C, 62.27; H, 6.78. Found: C, 62.45; H, 6.93.

1G
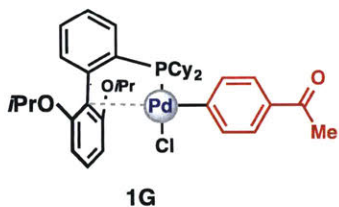
Following the general procedure, a mixture containing 4-chloroacetophenone (36.7 μL, 0.28 mmol), RuPhos (131.9 mg, 0.28 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (100.0 mg, 0.26 mmol) was stirred in cyclohexane (1.5 mL) at rt for 16 h. General work up afforded **1G** as a white solid (187.1 mg, 80%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.63 (m, 2H), 7.47 (m, 3H), 7.40 (m, 1H), 7.27 (m, 2H), 6.88 (ddd, J = 7.7, 3.2, 1.4 Hz, 1H), 6.65 (d, J = 8.5 Hz, 2H), 4.64 (hept, J = 6.1 Hz, 2H), 2.49 (s, 3H), 2.12 (m, 2H), 1.68 (m, 12H), 1.39 (d, J = 6.0 Hz, 6H), 1.18 (m, 6H), 1.02 (d, J = 6.1 Hz, 6H), 0.79 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 198.60, 159.72, 151.23, 151.22, 145.37, 145.20, 137.70, 137.67, 135.45, 133.70, 133.58, 133.33, 133.09, 132.99, 131.51, 131.25, 131.23, 127.01, 126.95, 126.31, 126.30, 111.61, 111.58, 107.76, 71.59, 34.45, 34.18, 28.83, 28.38, 28.36, 27.68, 27.55, 27.49, 27.36, 27.24, 26.74, 26.61, 26.60, 22.45, 21.86 (observed complexity is due to C–P coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 33.19. FT-IR (neat, cm^{-1}): 2937.05, 2849.21, 1672.81, 1590.71, 1570.43, 1541.83, 1453.46, 1383.73, 1370.46, 1350.76, 1296.23, 1270.06, 1246.23, 1177.32, 1131.31, 1112.59, 1075.99, 1057.28, 1008.91, 952, 919.43, 900.05, 847.69, 820.15, 807.81, 777.15, 758.83, 739.52, 724.42, 667.86, 614.63, 596.58, 578.79, 572.06, 557.85, 555.93, 553.94. Anal. Calcd. for C$_{38}$H$_{50}$ClO$_3$PPd: C, 62.72; H, 6.93. Found: C, 62.68; H, 6.98.
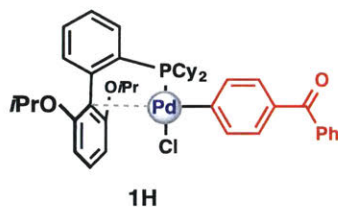
1H

Following the general procedure, a mixture containing 4-chlorobenzophenone (61.2 mg, 0.28 mmol), RuPhos (131.9 mg, 0.28 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (100.0 mg, 0.26 mmol) was stirred in cyclohexane (1.5 mL) at rt for 16 h. General work up afforded **1H** as a white solid (170.3 mg, 84%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.72 (m, 2H), 7.65 (m, 2H), 7.56 (m, 1H), 7.46 (m, 3H), 7.39 (m, 3H), 7.31 (dd, J = 8.5, 2.0 Hz, 2H), 6.89 (ddd, J = 7.6, 3.1, 1.3 Hz, 1H), 6.66 (d, J = 8.4 Hz, 2H), 4.64 (hept, J = 6.2 Hz, 2H), 2.15 (m, 2H), 1.68 (m, 12H), 1.39 (d, J = 6.0 Hz, 6H), 1.18 (m, 6H), 1.03 (d, J = 6.0 Hz, 6H), 0.79 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 197.01, 159.80, 150.68, 145.42, 145.25, 139.09, 137.53, 137.49, 135.53, 133.75, 133.38, 133.30, 133.09, 132.98, 132.21, 131.53, 131.28, 131.26, 130.25, 128.58, 128.33, 128.31, 127.02, 126.96, 111.48, 111.44, 107.73, 71.57, 34.40, 34.13, 28.79, 28.35, 28.33, 27.72, 27.59, 27.36, 27.25, 26.58, 22.44, 21.88 (observed complexity is due to $C–P$ coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 33.29. FT-IR (neat, cm^{-1}): 2974.7, 2926.53, 2849.69, 1652.22, 1567.46, 1538.81, 1456.24, 1382.54, 1371.38, 1314.3, 1300.72, 1278.58, 1244.96, 1174.95, 1109.61, 1052.35, 1011.48, 1000.34, 935.09, 919.51, 849.46, 832.32, 785.17, 758.91, 735.49, 725.02, 699.26, 668.36, 651.79, 612.29. Anal. Calcd. for C$_{43}$H$_{52}$ClO$_3$PPd: C, 65.40; H, 6.64. Found: C, 65.62; H, 6.38.

1I
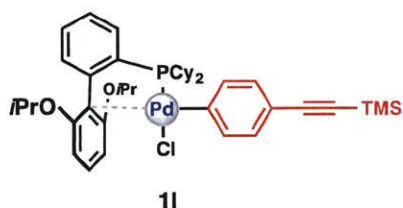
Following the general procedure, a mixture containing (4-chlorophenylethynyl)trimethylsilane (71.6 mg, 0.34 mmol), RuPhos (131.9 mg, 0.28 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (100.0 mg, 0.26 mmol) was stirred in cyclohexane (1.5 mL) at rt for 16 h. General work up afforded **1I** as a white solid (157.7 mg, 78%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.62 (m, 2H), 7.45 (m, 1H), 7.39 (m, 1H), 7.08 (dd, J = 8.4, 2.0 Hz, 2H), 7.01 (d, J = 8.1 Hz, 2H), 6.88 (ddd, J = 7.6, 3.1, 1.4 Hz, 1H), 6.65 (d, J = 8.5 Hz, 2H), 4.63 (hept, J = 6.1 Hz, 2H), 2.21 (m, 2H), 1.67 (m, 13H), 1.38 (d, J = 6.0 Hz, 6H), 1.17 (m, 6H), 1.01 (d, J = 6.1 Hz, 6H), 0.82 (m, 2H), 0.22 (s, 9H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 159.52, 145.34, 145.17, 142.81, 137.56, 137.52, 135.24, 133.74, 133.37, 133.11, 133.00, 131.50, 131.14, 130.13, 126.97, 126.91, 118.26, 111.99, 111.96, 107.81, 106.45, 92.74, 71.56, 34.42, 34.15, 28.87, 28.34, 27.72, 27.59, 27.49, 27.39, 27.28, 26.64, 22.46, 21.84, 0.30 (observed complexity is due to *C–P* coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 32.69. FT-IR (neat, cm^{-1}): 2920.64, 2850.24, 2153.99, 1573.62, 1475.48, 1452.03, 1383.03, 1367.97, 1294.31, 1273.73, 1238.97, 1175.48, 1110.6, 1054.15, 1012.36, 1003.56, 941.96, 916.46, 866.35, 839.95, 815.21, 787.82, 754.48, 745.71, 730.56, 720.42, 698.18, 659.3. Anal. Calcd. for C$_{41}$H$_{56}$ClO$_2$PPdSi: C, 62.99; H, 7.22. Found: C, 63.25; H, 7.09.
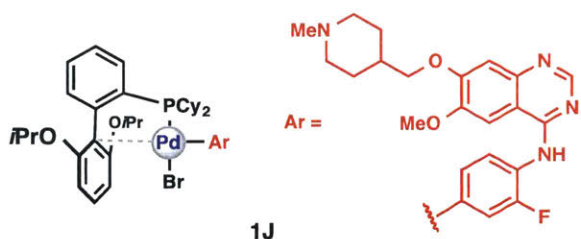
Following the general procedure, a mixture containing Vandetanib (61.7 mg, 0.13 mmol, *Note:* 1.01 equiv was used), RuPhos (66.0 mg, 0.14 mmol), and (COD)Pd(CH₂SiMe₃)₂ (50.0 mg, 0.13 mmol) was stirred in THF (1.5 mL) at rt for 16 h. General work up afforded **1J** as an off-white solid (119.0 mg, 88%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 8.54 (s, 1H), 7.85 (t, J = 8.6 Hz, 1H), 7.65 (m, 2H), 7.46 (t, J = 7.5 Hz, 1H), 7.40 (t, J = 7.4 Hz, 1H), 7.20 (s, 1H), 7.11 (m, 1H), 7.04 (s, 1H), 6.93 (m, 2H), 6.88 (dd, J = 7.6, 3.2 Hz, 1H), 6.67 (d, J = 8.5 Hz, 2H), 4.63 (hept, J = 6.0 Hz, 2H), 3.98 (s, 3H), 2.85 (m, 2H), 2.23 (m, 5H), 1.85 (m, 17H), 1.40 (m, 8H), 1.22 (m, 8H), 1.03 (d, J = 6.1 Hz, 6H), 0.90 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 159.92, 156.93, 154.90, 154.48, 153.96, 150.35, 147.98, 145.33, 145.16, 135.49, 133.80, 133.52, 133.45, 133.14, 133.03, 132.04, 131.99, 131.57, 131.27, 131.25, 126.97, 126.92, 124.08, 123.88, 122.96, 122.85, 122.49, 111.38, 111.34, 109.50, 109.05, 107.96, 100.30, 74.14, 71.64, 56.91, 55.89, 46.82, 35.58, 34.50, 34.23, 29.68, 28.77, 28.34, 27.80, 27.67, 27.39, 27.28, 26.58, 22.43, 22.39, 21.91 (observed complexity is due to *C–P* coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 32.24. FT-IR (neat, cm^{-1}): 2926.73, 2849, 2780.97, 1619.76, 1584.47, 1495.57, 1452.88, 1416.66, 1381.79, 1333.79, 1248.7, 1209.65, 1138.06, 1110.84, 1061.64, 1000.91, 940.37, 884.58, 849.88, 817.62, 762.84, 737.25, 667.56, 630.8. Anal. Calcd. for C$_{52}$H$_{67}$BrFN$_4$O$_4$PPd: C, 59.57; H, 6.44. Found: C, 59.24; H, 6.41.

2A
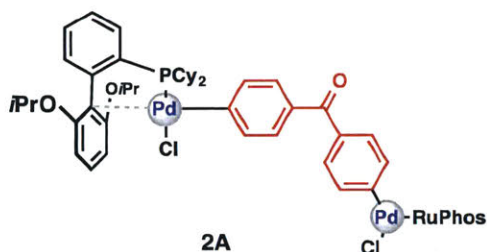
Following a slightly modified general procedure, a mixture of 4,4'-dichlorobenzophenone (30.0 mg, 0.12 mmol, 1 equiv), RuPhos (139.4 mg, 0.30 mmol, 2.5 equiv), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (116.2 mg, 0.30 mmol, 2.5 equiv) was stirred in cyclohexane (1.2 mL) at rt for 16 h. General work up afforded **2A** as a beige solid (146.8 mg, 88%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.64 (m, 4H), 7.45 (m, 2H), 7.39 (m, 2H), 7.32 (d, J = 8.0 Hz, 4H), 7.25 (dd, J = 8.4, 2.1 Hz, 4H), 6.88 (ddd, J = 7.7, 3.1, 1.3 Hz, 2H), 6.65 (d, J = 8.5 Hz, 4H), 4.64 (hept, J = 6.1 Hz, 4H), 2.14 (m, 4H), 1.70 (m, 24H), 1.39 (d, J = 6.0 Hz, 12H), 1.20 (m, 12H), 1.02 (d, J = 6.0 Hz, 12H), 0.75 (m, 4H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 197.01, 159.78, 149.09, 145.47, 145.30, 137.25, 137.21, 135.49, 134.06, 133.94, 133.58, 133.06, 132.95, 131.55, 131.23, 131.21, 128.34, 126.98, 126.92, 111.50, 107.69, 71.53, 34.39, 34.12, 28.78, 28.32, 27.73, 27.59, 27.38, 27.27, 26.59, 22.44, 21.89 (observed complexity is due to C–P coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 33.27. FT-IR (neat, cm^{-1}): 2924.96, 2849.75, 1652.54, 1645.7, 1569.12, 1538.83, 1456.32, 1382.56, 1371.3, 1278.13, 1244.17, 1173.52, 1110.74, 1051.09, 1010.83, 921.17, 848.83, 814.88, 784.1, 751.87, 667.12, 623.81, 612.26, 579.08. Anal. Calcd. for C$_{73}$H$_{94}$Cl$_2$O$_5$P$_2$Pd$_2$: C, 62.75; H, 6.78. Found: C, 62.36; H, 6.83.
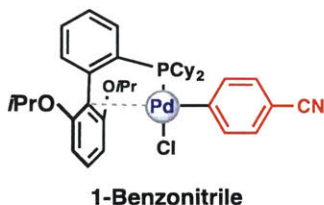
1-Benzonitrile

Following the general procedure, a mixture of 4-chlorobenzonitrile (42.4 mg, 0.31 mmol), RuPhos (144.0 mg, 0.31 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (100.0 mg, 0.26 mmol) was stirred in cyclohexane (1.5 mL) at rt for 16 h. General work up afforded **1-Benzonitrile** as a white solid (186.4 mg, 99%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.62 (m, 2H), 7.51 (tt, J = 7.5, 1.6 Hz, 1H), 7.44 (m, 1H), 7.27 (dd, J = 8.0, 2.0 Hz, 2H), 7.20 (d, J = 8.0 Hz, 2H), 6.93 (dd, J = 7.8, 3.1 Hz, 1H), 6.70 (d, J = 8.4 Hz, 2H), 4.65 (hept, J = 6.1 Hz, 2H), 2.08 (m, 2H), 1.89 (m, 2H), 1.78 (m, 4H), 1.67 (m, 4H), 1.56 (m, 2H), 1.37 (d, J = 6.0 Hz, 6H), 1.20 (m, 6H), 1.03 (d, J = 6.0 Hz, 6H), 0.85 (m, 2H). ^{13}C NMR (101 MHz, CD$_2$Cl$_2$) δ 160.15, 139.17, 139.13, 133.70, 133.60, 132.71, 131.75, 129.68, 127.56, 127.50, 120.54, 108.36, 107.04, 72.10, 35.11, 34.84, 29.69, 29.16, 27.81, 27.69, 27.61, 27.47, 26.85, 26.84, 22.49, 21.94. (observed complexity is due to C–P coupling). ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 33.46. FT-IR (neat, cm^{-1}): 2927.23, 2849.55, 2223.84, 1591.35, 1575.29, 1473.98, 1450.97, 1383.19, 1371.34, 1274.04, 1243.98, 1174.33, 1130.21, 1114.05, 1055.96, 1014.33, 1003.94, 849.24, 834.6, 809.39, 778.82, 761.92, 743.75, 737.28, 723.9, 714.69, 612.21. HRMS electrospray (m/z): [M – Cl]$^+$ calcd for C$_{37}$H$_{47}$NO$_2$PPd: 674.2392, found 674.2411.

1-Vinyl
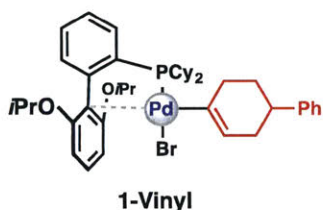
Following the general procedure, a mixture containing 4-bromo-1,2,3,6-tetrahydro-1,1'-biphenyl (30.0 mg, 0.127 mmol), RuPhos (59.0 mg, 0.127 mmol), and (COD)Pd(CH$_2$SiMe$_3$)$_2$ (44.7 mg, 0.115 mmol) was stirred in cyclohexane (0.75 mL) at rt for 16 h. General work up afforded **1-Vinyl** as a yellow solid (80.0 mg, 86%).

^{1}H NMR (400 MHz, CD$_2$Cl$_2$) δ 7.66 (td, J = 7.0, 1.9 Hz, 1H), 7.56 (t, J = 8.4 Hz, 1H), 7.40 (m, 2H), 7.23 (m, 4H), 7.14 (m, 1H), 6.82 (m, 1H), 6.60 (m, 2H), 4.94 (s, 1H), 4.56 (m, 2H), 2.90 (m, 3H), 1.90 (m, 18H), 1.25 (m, 22H). ^{13}C NMR: *could not be obtained due to complex decomposition in solution (CD$_2$Cl$_2$; > 4 h).* ^{31}P NMR (121 MHz, CD$_2$Cl$_2$) δ 29.17, 28.89. FT-IR (neat, cm^{-1}): 2970.4, 2923.13, 2844.48, 1576.18, 1455.04, 1387.01, 1375.4, 1283.43, 1252.95, 1201.89, 1172.85, 1141.69, 1110.95, 1065.66, 1026.88, 997.97, 968.22, 902.49, 891.58, 847.07, 775.05, 767.23, 759.11, 748.35, 741.77, 726.46, 694.67, 676.18, 612.58. Anal. Calcd. for C$_{42}$H$_{56}$BrO$_2$PPd: C, 62.26; H, 6.97. Found: C, 62.03; H, 6.85.

4.4.7. Peptide synthesis

All peptides were synthesized on a 0.2 mmol scale using manual Fmoc-SPPS chemistry under flow using a 3 min cycle for each amino acid.[38] Specifically, all reagents and solvents were delivered to a stainless steel reactor containing resins at a constant flow rate using HPLC pump; the temperature of the reactor was maintained at 60 °C during the synthesis using a water bath. The procedure for each amino acid coupling cycle included: 1) a 30 s coupling with 1 mmol of the corresponding Fmoc-protected amino acid, 1 mmol HBTU, and 500 μL of diisopropyl ethyl amine (DIPEA) in 2.5 mL of DMF at a flow rate of 6 mL/min (note that for the coupling of cysteine and tryptophan, 190 μL of DIPEA was used to prevent racemization); 2) 1 min wash with DMF at a flow rate of 20 mL/min; 3) 20 s deprotection with 50% (v/v) piperidine in DMF at a flow rate of 20 mL/min; and 4) 1 min wash with DMF at a flow rate of 20 mL/min. After completion of the stepwise SPPS, the resin was washed thoroughly with DCM and dried under vacuum. The peptide was simultaneously cleaved from the resin and deprotected on the side-chains by treatment with 2.5% (v/v) water, 2.5% (v/v) 1,2-ethanedithiol (EDT), and 1% (v/v) triisoproprylsilane in neat trifluoroacetic acid (TFA) for 2 h at room temperature. The solvent from the resulting solution containing the target peptide was evaporated by purging with nitrogen gas for 15 min. The residue was then triturated and washed with cold diethyl ether three times. The obtained gummy-like solid was dissolved in 50% H_2O: 50% acetonitrile containing 0.1% TFA and lyophilized.

4.4.8. Peptide purification

The crude peptide was dissolved in 95% A: 5% B with 6 M guanidinium hydrochloride and purified by semi-preparative RP-HPLC (Agilent Zorbax SB C_{18} column: 21.2 x 250 mm, 7 μm, linear gradient: 5-50% B over 90 min, flow rate: 5 mL/min). Each HPLC fraction was analyzed by MALDI-TOF mass-spectrometry. Specifically, 1 μL of each HPLC fraction was mixed with 1 μL of α-cyano-4-hydroxycinnamic acid (CHCA) matrix in 75% A: 25% B and the resulting mixture was analyzed for the desired molecular mass using MALDI-TOF. The purity of the fractions containing the desired peptide was further analyzed by analytical RP-HPLC (Agilent Zorbax SB C_3 column: 2.1 x 150 mm, 5 μm, gradient: 0-2 min 5% B, 2-11 min 5-65% B, 11-12 min 65% B, flow rate: 0.8 mL/min). HPLC fractions containing pure product were further confirmed by LC-MS, combined, and lyophilized. Peptides synthesized using fast flow-based SPPS and purified by RP-HPLC are listed in Table 4.4.

Table 4.4. Sequences and masses of peptides synthesized by fast-flow peptide synthesizer.

Peptide	Sequence[a]	Calculated mass	Observed mass
P1	NH$_2$-RSNFYLG**C**AGLAHDKAT-CONH$_2$	1821.89	1821.89
P1-Ser	NH$_2$-RSNFYLG**S**AGLAHDKAT-CONH$_2$	1805.92	1805.92
P2	NH$_2$-RSNFFLG**C**AGA-CONH$_2$	1140.55	1140.55
P3	NH$_2$-IKFTN**C**GLL**C**YESKR-CONH$_2$	1772.91	1772.91

[a]Cysteine residues and the corresponding serine residue in P1-Ser are highlighted in red.

4.4.9. Procedures and LC-MS analysis of palladium-mediated conjugation under different reaction conditions

General procedure A (Table 4.1, entries 3–7, 10, 13–15, 18, 19). Peptide **P1** (4 μL, 150 μM in water), H_2O (47 μL), organic solvent (1 μL), and the buffer (6 μL, 1 M) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of the palladium complex (2 μL, 600 μM) in organic solvent was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 3 equiv to the palladium complex). After an additional 5 min, a solution of 50% A: 50% B (v/v, 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS. Final concentrations of the reaction before quenching: peptide **P1** – 10 μM, **Pd-complex** – 20 μM, Buffer – 100 mM; organic solvent: H_2O = 5: 95.

Effect of concentrations of peptides and palladium complexes (Table 4.1, entries 1–3).

Procedure for entries 1 and 21. Peptide **P1** or **P1-Ser** (4 μL, 15 mM in water), H_2O (32 μL), and Tris base buffer (4 μL, 1 M, pH 8.5) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of complex **1A-OTf** (20 μL, 6 mM) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 5 μL/mL solution in water, 3 equiv to **1A-OTf**). After an additional 5 min a solution of 50% A: 50% B (v/v, 550 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS. Final concentration of the reaction before quenching: peptide – 1 mM, **1A-OTf** – 2 mM, Tris buffer – 100 mM; CH_3CN: H_2O = 1: 2.

Procedure for entry 2. Peptide **P1** (4 μL, 1.5 mM in water), H_2O (47 μL), CH_3CN (1 μL), and Tris buffer (6 μL, 1 M, pH 8.5) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of the palladium complex (2 μL, 6 mM) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.5 μL/mL solution in water, 3 equiv to the palladium

303

complex). After an additional 5 min, a solution of 50% A: 50% B (v/v, 60 µL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS. Final concentrations of the reaction before quenching: peptide – 100 µM, **1A-OTf** – 200 µM, Tris buffer – 100 mM; CH₃CN : H₂O = 5 : 95.
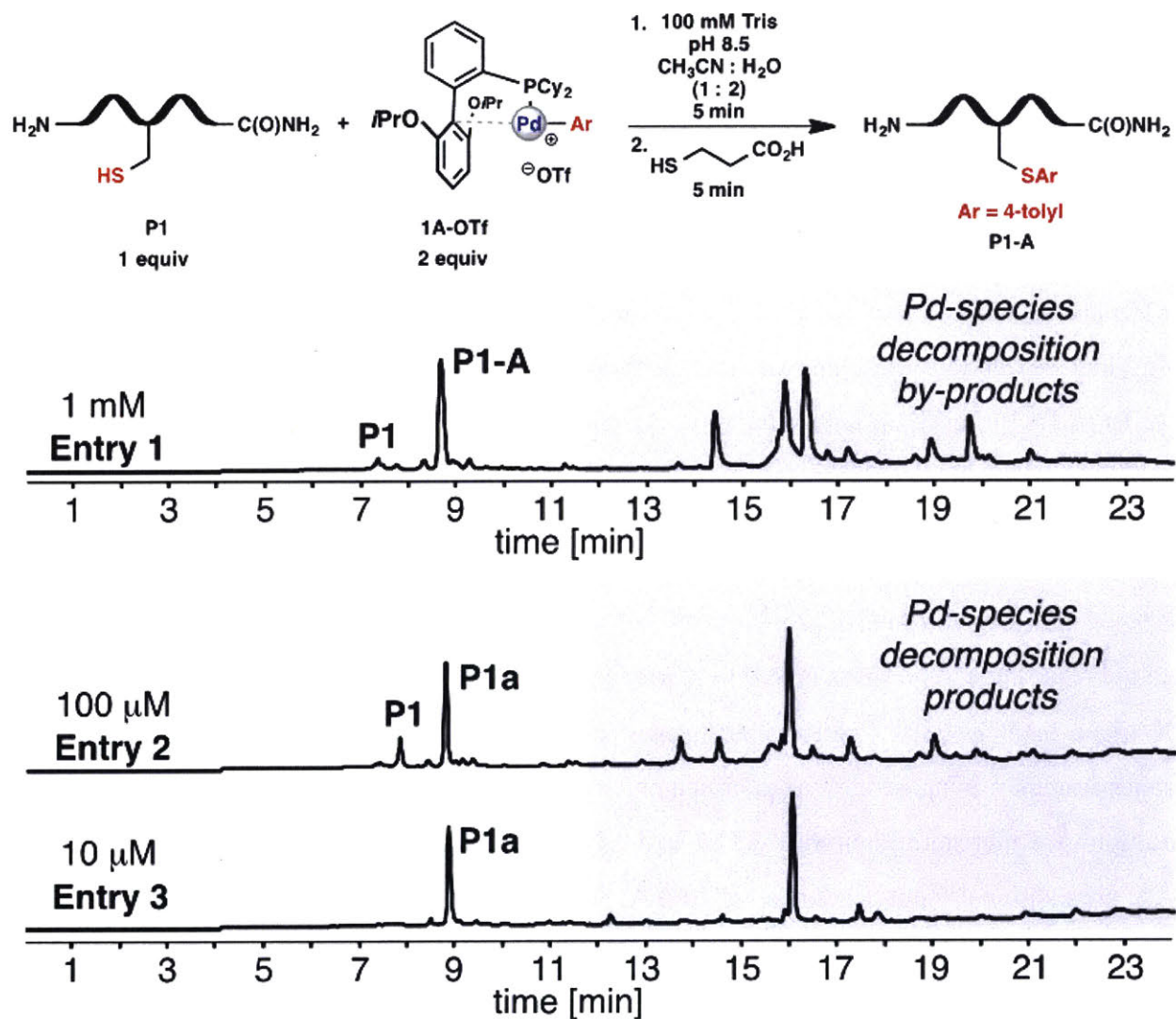

Figure 4.18. LC-MS chromatograms of reactions with different concentrations of peptide and palladium reagents.

Effect of pH on the bioconjugation reaction (entries 3–5, 10–17)

Experimental procedure for entries 3-5 followed general procedure A.

Experimental procedure for entries 11 and 12. Peptide **P1** (6 μL, 100 μM in water), H_2O (39 μL), and the buffer (12 μL, 1 M Tris or 0.5 M MOPS) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of the palladium complex (3 μL, 400 μM) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 3 equiv to the palladium complex). After an additional 5 min, a solution of 50% A: 50% B (v/v, 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS (*LC-MS data shown was acquired using Method C)*. Final concentrations of the reaction before quenching: peptide **P1** – 10 μM, **Pd-complex** – 20 μM, Buffer – 200 mM Tris or 100 mM MOPS; organic solvent: H_2O = 5: 95.

Experimental procedure for entries 16 and 17. Peptide **P1** (4 μL, 150 μM in water), 0.1% TFA solution in H_2O (53 μL), and CH_3CN (1 μL) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of **1A-OTf** (2 μL, 600 μM) in organic solvent was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min or 7 h 20 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 3 equiv to **1A-OTf**). After an additional 5 min, a solution of 50% A: 50% B (v/v, 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS. Final concentrations of the reaction before quenching: peptide – 10 μM, **1A-OTf** – 20 μM, pH = 2.0, CH_3CN: H_2O = 5: 95.

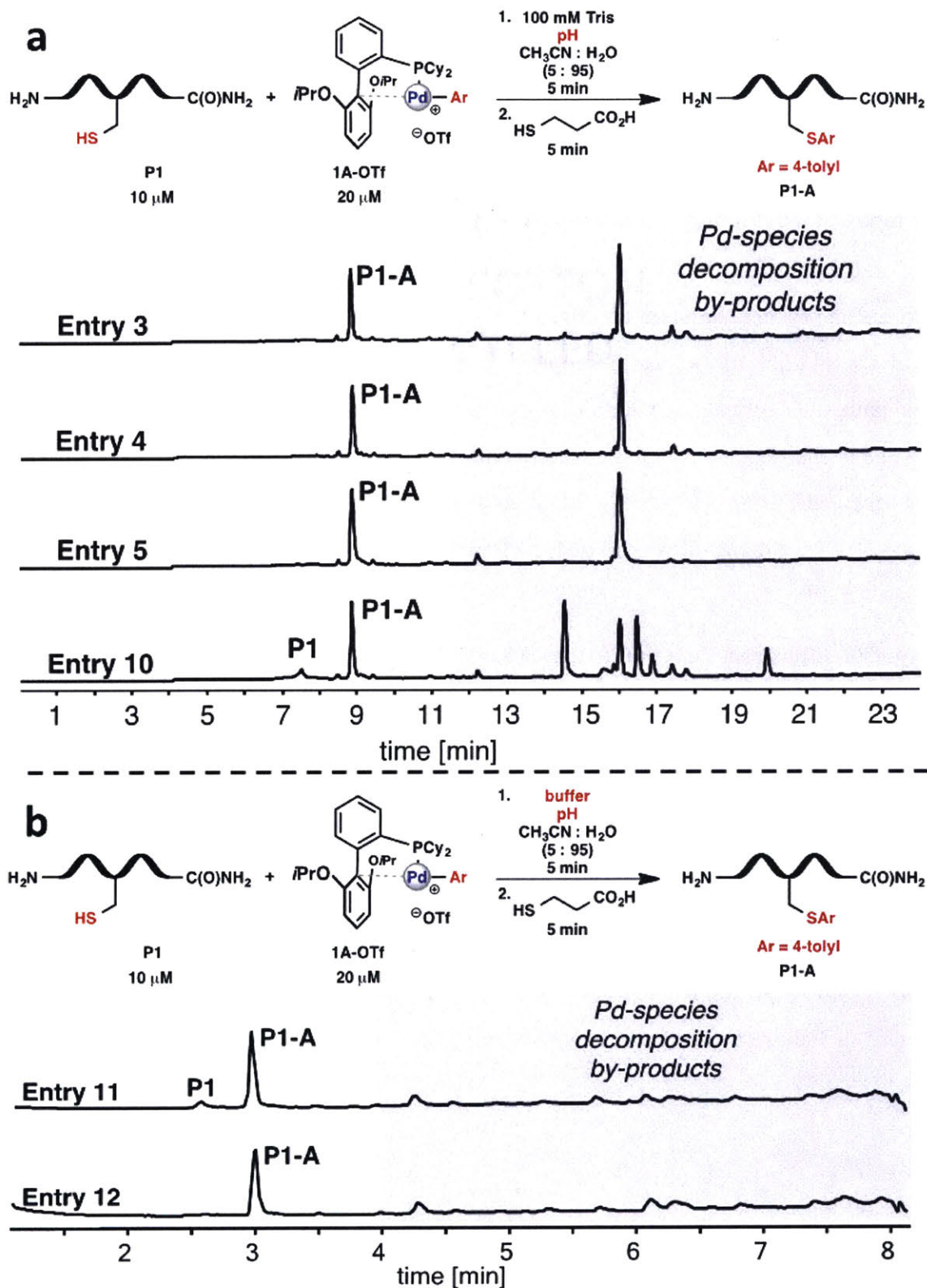**Figure 4.19. LC-MS chromatograms of reactions at different pH (pH >= 7).**

(**a**) Reactions using Tris buffers with pH > 7. (**b**) Reactions using buffers with pH = 7.
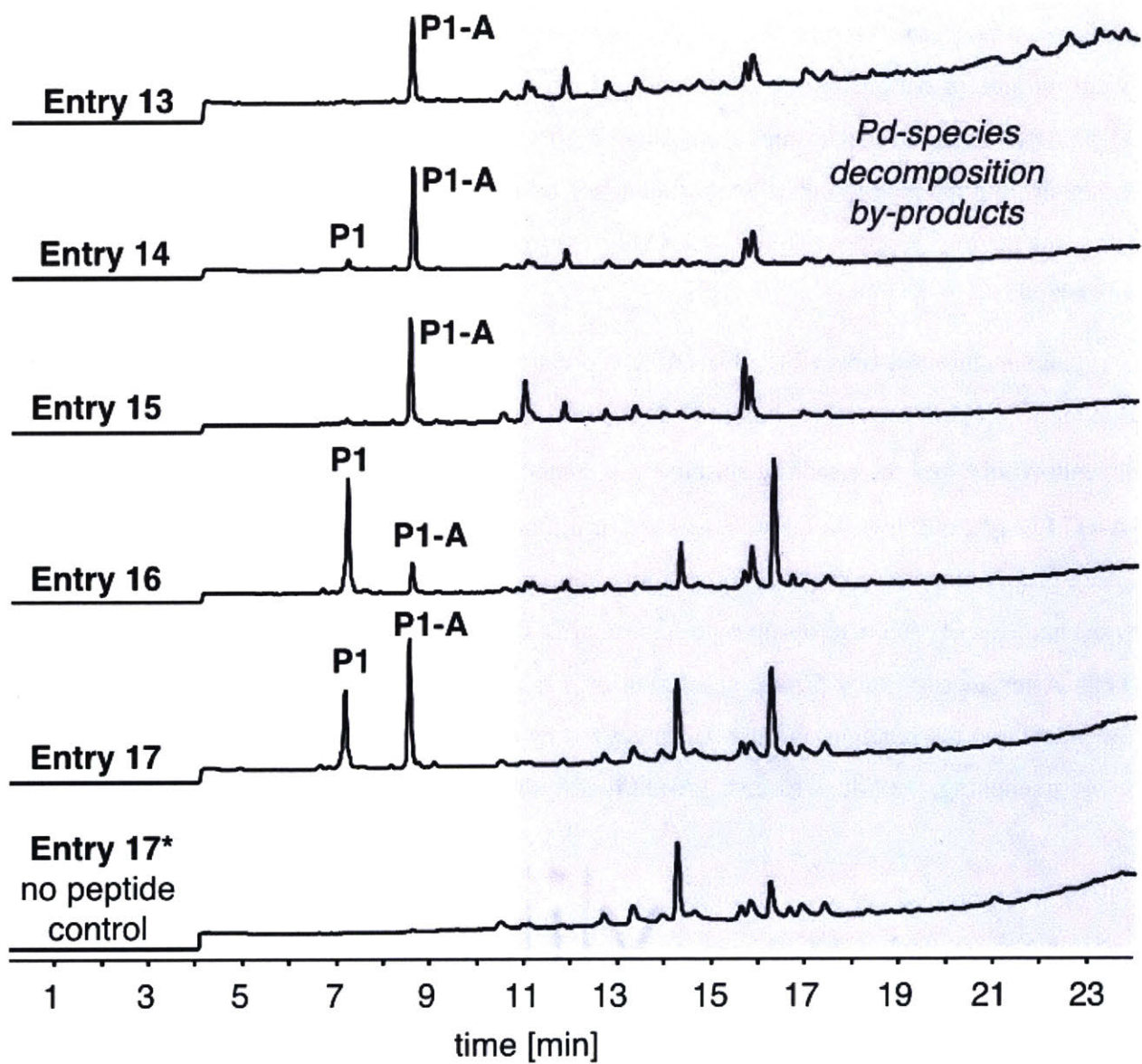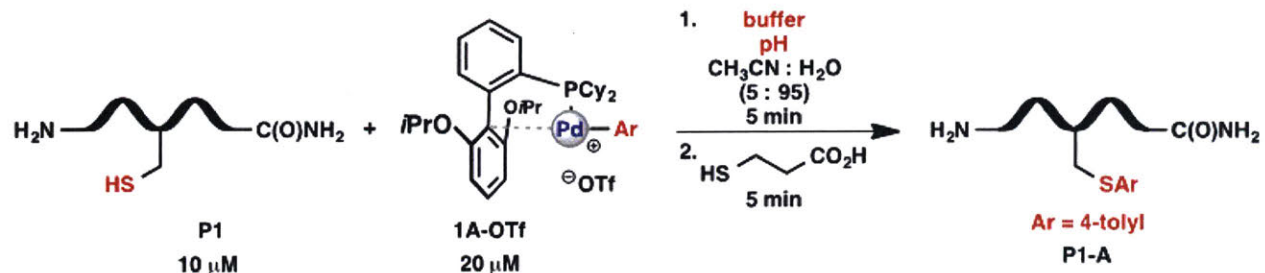
306

Figure 4.20. LC-MS chromatograms of reactions at different pH (pH < 7).

Effect of the buffer on the bioconjugation reaction (Table 4.1, entries 5–9)

Experimental procedure for entry 8. Peptide **P1** (4 μL, 150 μM in water), H_2O (23 μL), CH_3CN (1 μL), and Na_2HPO_4/NaH_2PO_4 buffer (30 μL, 0.2 M, pH 7.5) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of **1A-OTf** (2 μL, 600 μM) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 3 equiv to **1A-OTf**). After an additional 5 min, a solution of 50% A: 50% B (v/v, 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS. Final concentration of the reaction before quenching: peptide – 10 μM, **1A-OTf** – 20 μM, Na_2HPO_4/NaH_2PO_4 – 100 mM; CH_3CN: H_2O = 5: 95.

Experimental procedure for entry 9. Peptide **P1** (4 μL, 150 μM in water), H_2O (51.5 μL), CH_3CN (1 μL), and Tris buffer (1.5 μL, 1 M, pH 7.5) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of **1A-OTf** (2 μL, 600 μM) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 3 equiv to **1A-OTf**). After an additional 5 min, a solution of 50% A: 50% B (v/v, 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS. Final concentration of the reaction before quenching: peptide – 10 μM, **1A-OTf** – 20 μM, Tris Buffer – 25 mM; CH_3CN: H_2O = 5: 95.

Figure 4.21. LC-MS chromatograms of reactions with different buffers.

Effect of organic co-solvent on the bioconjugation reaction (Table 4.1, entries 5, 18, 19).
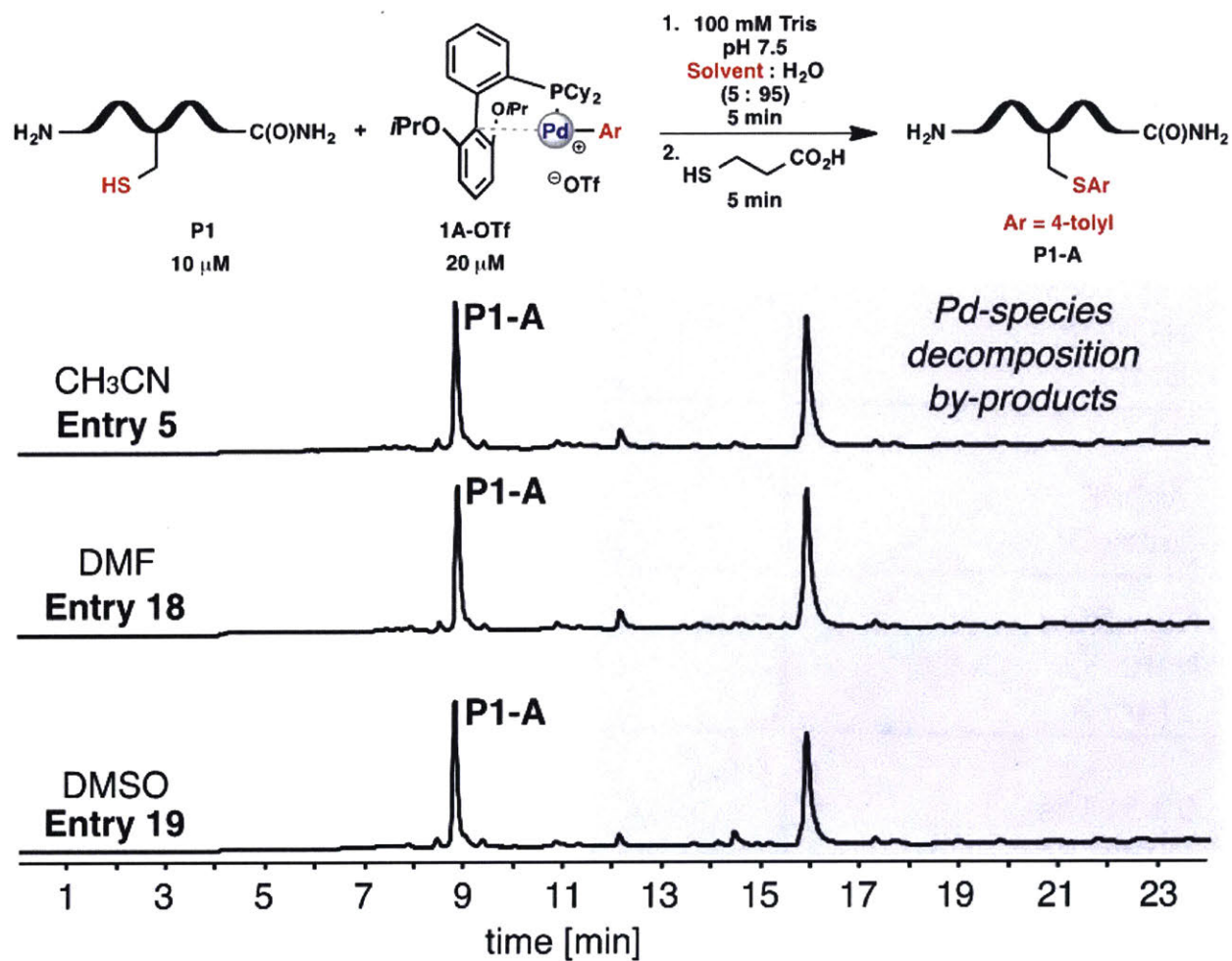
Figure 4.22. LC-MS chromatograms of reactions with different organic co-solvents.

Reaction in the presence of TCEP (entry 20).

Reaction procedure. Peptide **P1** (4 μL, 150 μM in water), TCEP (2 μL, 600 μM), H₂O (45 μL), CH₃CN (1 μL), and Tris buffer (6 μL, 1 M, pH 7.5) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of **1A-OTf** (2 μL, 600 μM) in CH₃CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.1 μL/mL solution in water, 6 equiv to **1A-OTf**). After an additional 5 min, a solution of 50% A: 50% B (v/v; 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS (*LC-MS data shown was acquired using Method C*). Final concentration of the reaction before quenching: peptide – 10 μM, **1A-OTf** – 20 μM, TCEP – 20 μM; Tris buffer – 100 mM; CH₃CN: H₂O = 5: 95.
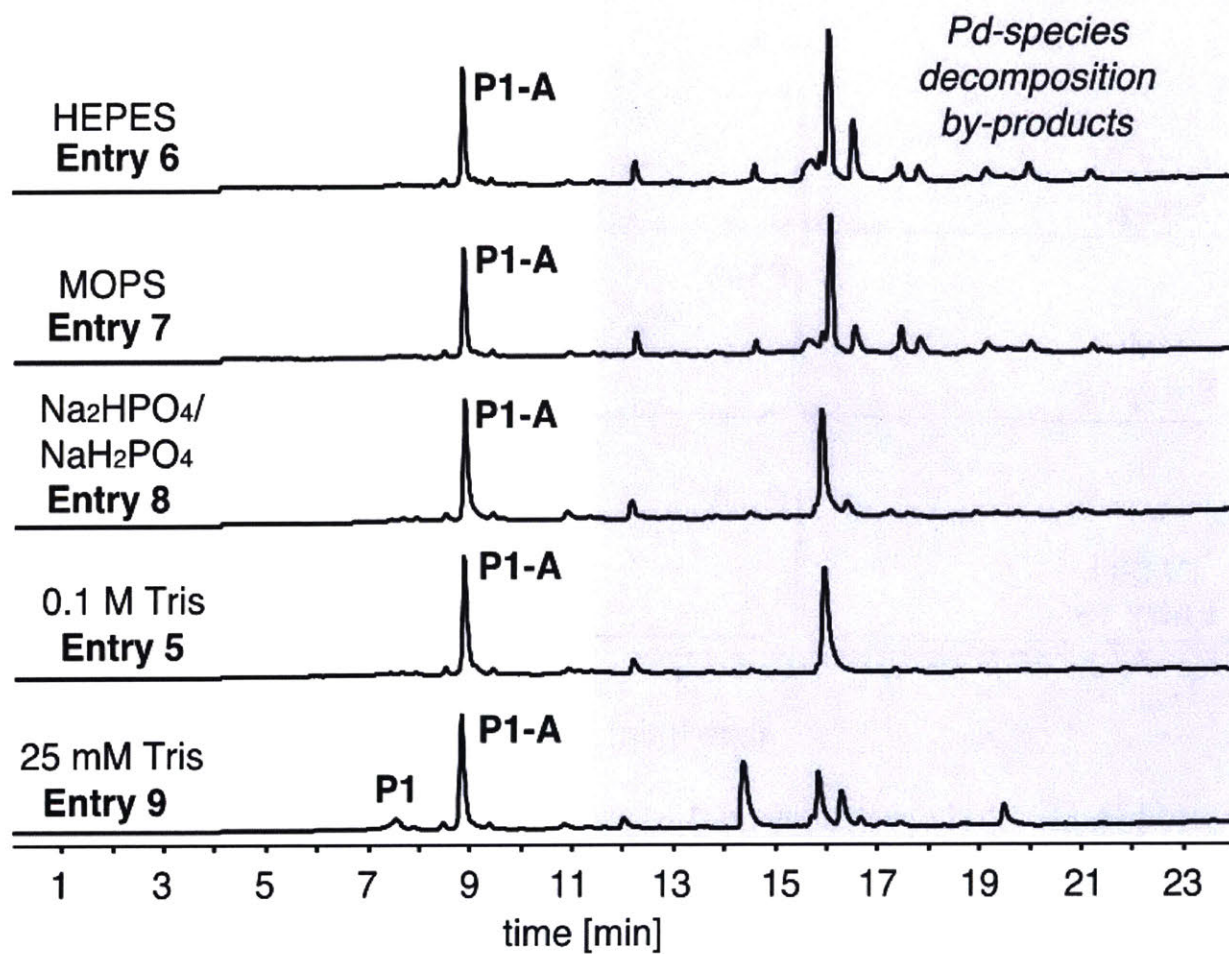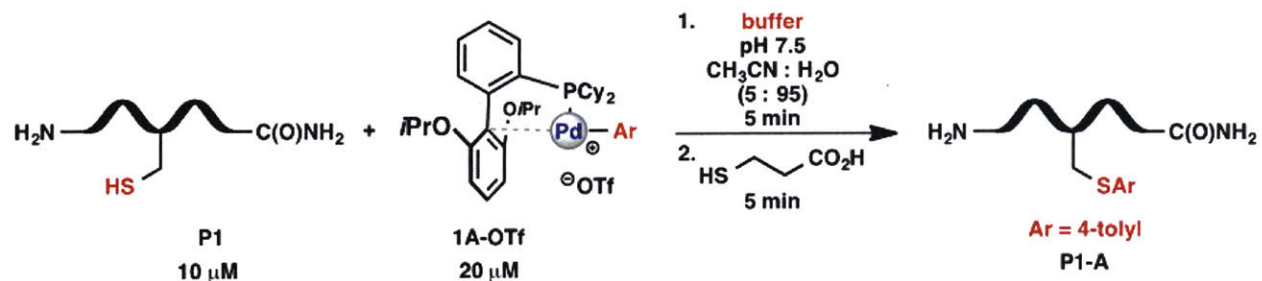

Figure 4.23. LC-MS chromatograms of the reaction in the presence of TCEP.

311

4.4.10. Procedures and LC-MS analysis of palladium-mediated conjugation with different palladium complexes

The modified peptide **P1-A** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1A-OTf – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

The modified peptide **P1-B** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1B – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

The modified peptide **P1-C** was synthesized according to standard procedure A. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 2 equiv to 1C). Final conditions before quenching: peptide – 10 μM, 1C – 30 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

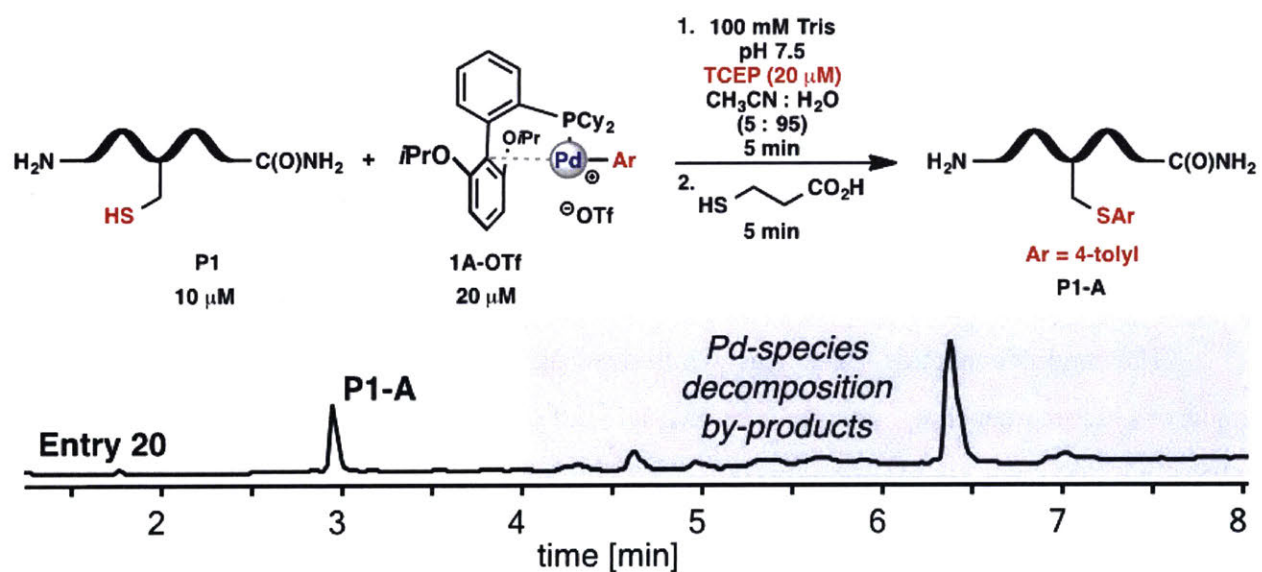The modified peptide **P1-D** was synthesized according to standard procedure A. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 2 equiv to 1D). Final conditions before quenching: peptide – 10 μM, 1D – 30 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

The modified peptide **P1-E** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1E – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

4.4.10. Procedures and LC-MS analysis of palladium-mediated conjugation with different palladium complexes

The modified peptide **P1-A** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1A-OTf – 20 μM, 0.1 M Tris (pH 7.5), CH_3CN: H_2O = 5: 95.

The modified peptide **P1-B** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1B – 20 μM, 0.1 M Tris (pH 7.5), CH_3CN: H_2O = 5: 95.

The modified peptide **P1-C** was synthesized according to standard procedure A. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 2 equiv to 1C). Final conditions before quenching: peptide – 10 μM, 1C – 30 μM, 0.1 M Tris (pH 7.5), CH_3CN: H_2O = 5: 95.

The modified peptide **P1-D** was synthesized according to standard procedure A. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 2 equiv to 1D). Final conditions before quenching: peptide – 10 μM, 1D – 30 μM, 0.1 M Tris (pH 7.5), CH_3CN: H_2O = 5: 95.

The modified peptide **P1-E** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1E – 20 μM, 0.1 M Tris (pH 7.5), CH_3CN: H_2O = 5: 95.

Figure 4.24. LC-MS analysis of peptide conjugation reactions with palladium complexes **1A-OTf (a), 1B (b), 1C (c), 1D (d),** and **1E (e).**

The modified peptide **P1-A** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1A-X (X = Cl, Br, I) – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

The modified peptide **P1-F** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1F – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95. Note: The product peak in the TIC contained a shoulder with the observed masses: $[M+3H]^{3+}$ = 636.99; $[M+2H]^{2+}$ = 954.98. This data corresponds to a product with $M_{shoulder}$ = $[M_{P1g} - H_2O]$, which can form as a result of reversible imine formation in the reaction between the aldehyde and the lysine residue within the peptide.

The modified peptide **P1-G** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1G – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.
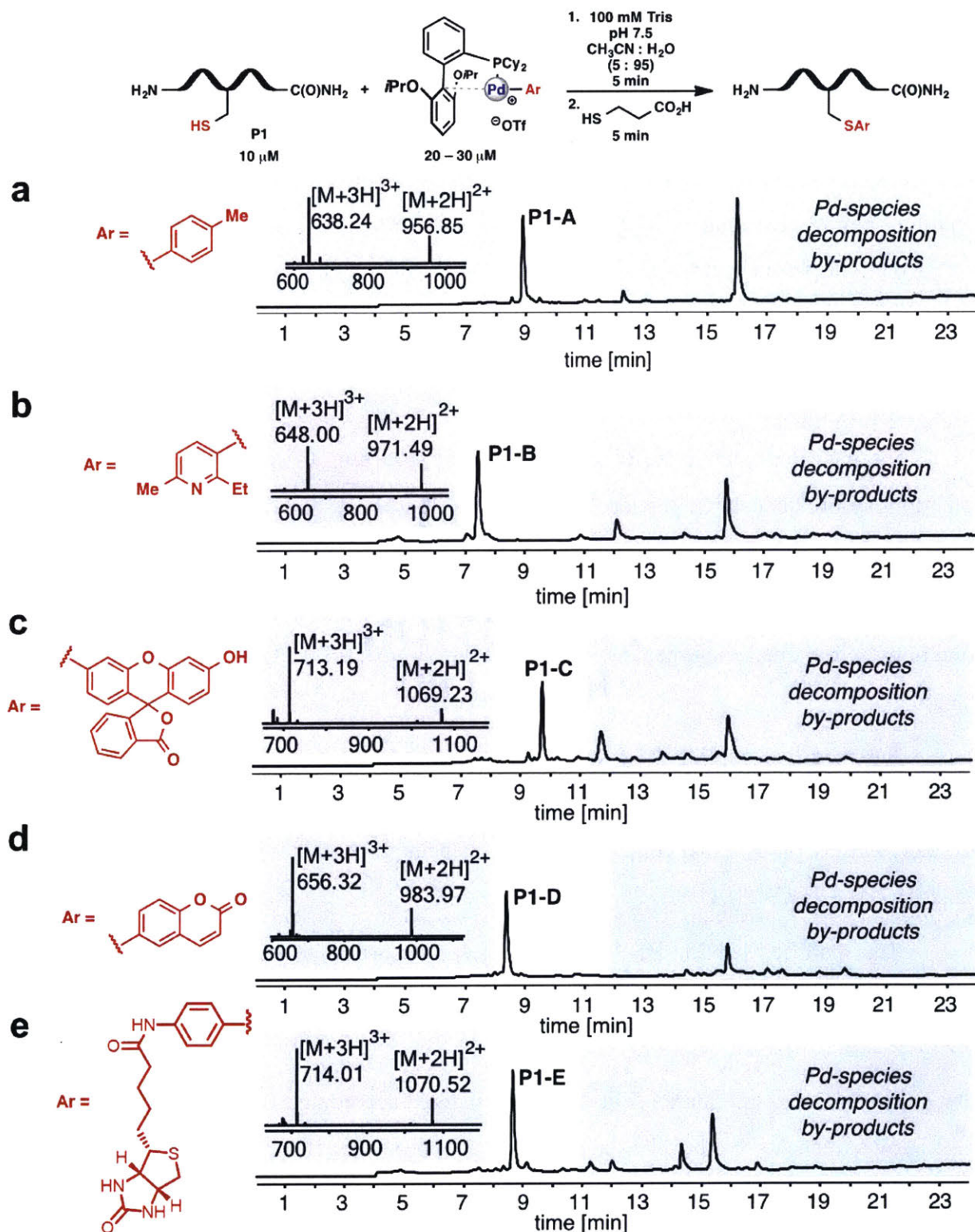
The modified peptide **P1-H** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1H – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN : H₂O = 5 : 95.

The modified peptide **P1-I** was synthesized according to standard procedure A. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 1 equiv to 1I). Final conditions before quenching: peptide – 10 μM, 1I – 60 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

The modified peptide **P1-J** was synthesized according to standard procedure A. Final conditions before quenching: peptide – 10 μM, 1J – 20 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

The modified peptide **P1-Vinyl** was synthesized according to standard procedure A. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution in water, 1.5 equiv to 1-Vinyl). Final conditions before quenching: peptide – 10 μM, 1-Vinyl – 40 μM, 0.1 M Tris (pH 7.5), CH₃CN: H₂O = 5: 95.

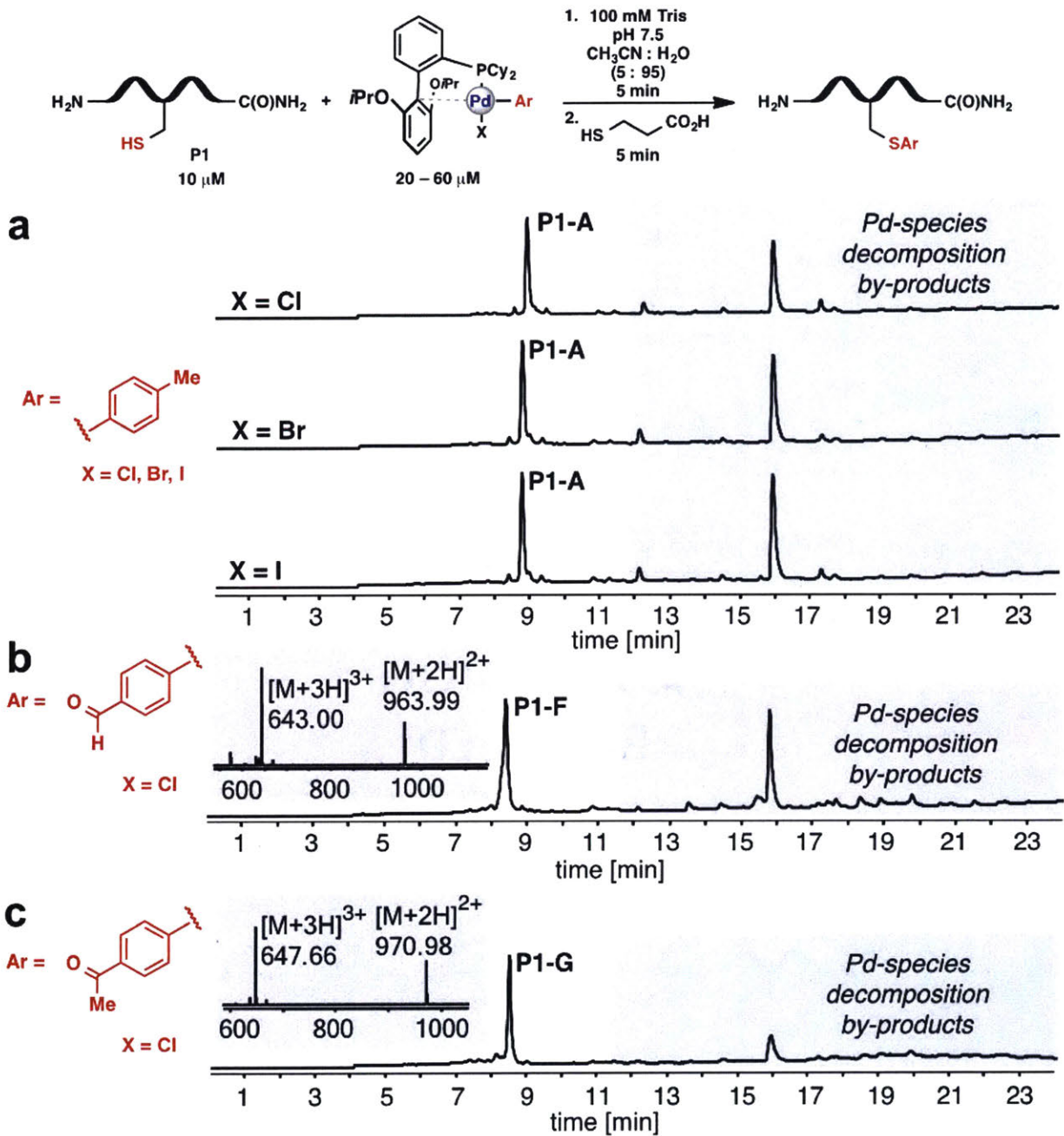**Figure 4.25. LC-MS analysis of peptide conjugation reactions with palladium complexes 1A-halide (halide = Cl, Br, or I) (a), 1F (b), and 1G (c).**
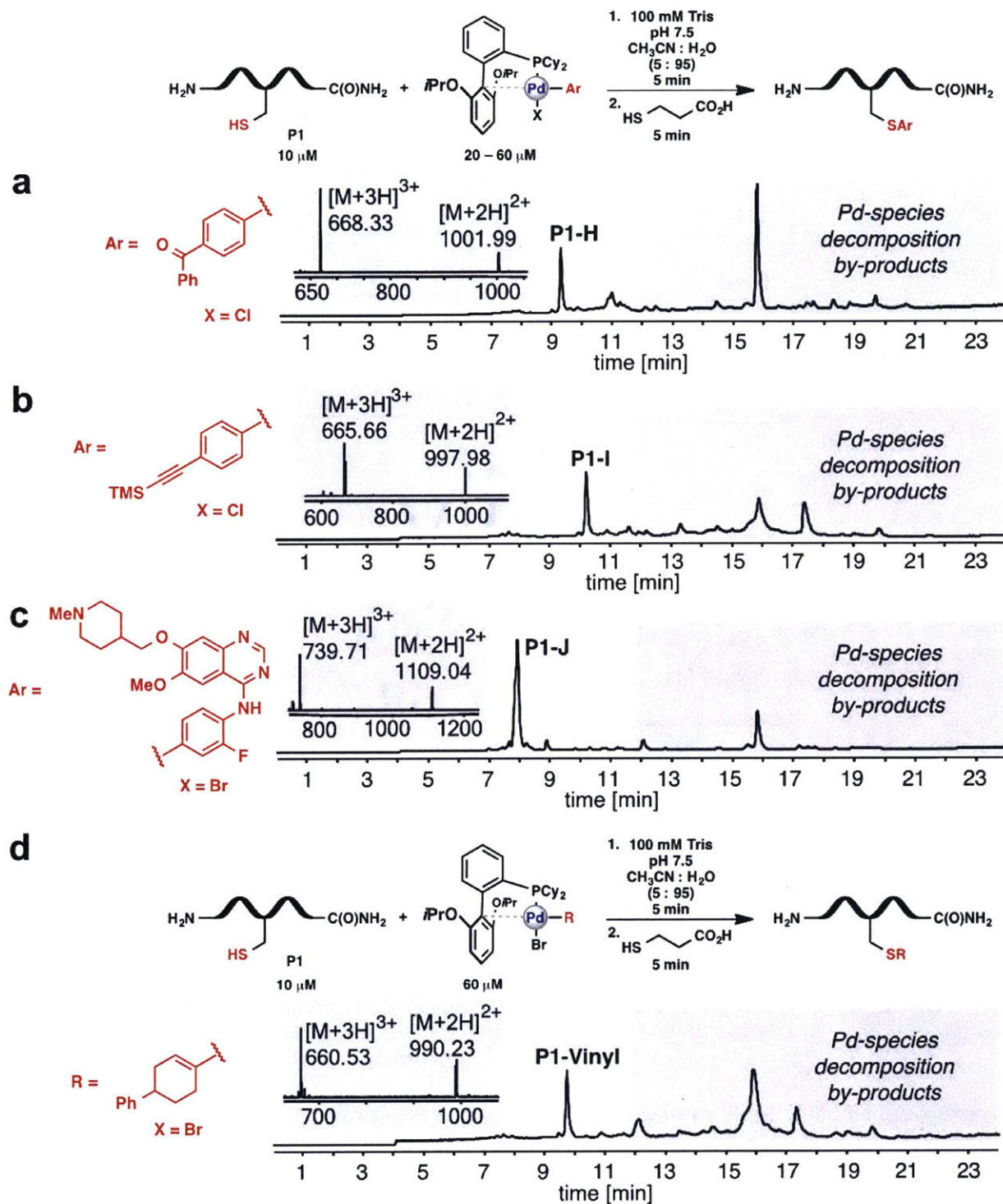
Figure 4.26. LC-MS analysis of peptide conjugation reactions with palladium complexes 1H (a), 1I (b), 1J (c), and 1-Vinyl (d).

4.4.11. Procedures for peptide stapling reactions

Peptide **P3** (4 μL, 150 μM in water), H_2O (23 μL), and Tris buffer (3 μL, 1 M, pH 7.5) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed by vortexing for 10 **s**. A stock solution of palladium complex **2A** (30 μL, 40 μM) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 10 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.1 μL/mL solution in water, 6 equiv. to **2A**). After an additional 5 min a solution of 50% A: 50% B (v/v, 60 μL) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS (injection: 3 μL). Final concentration of the reaction before quenching: peptide **P3** – 10 μM, **2A** – 20 μM, Tris buffer – 100 mM; pH 7.5; CH_3CN: H_2O = 1: 1.
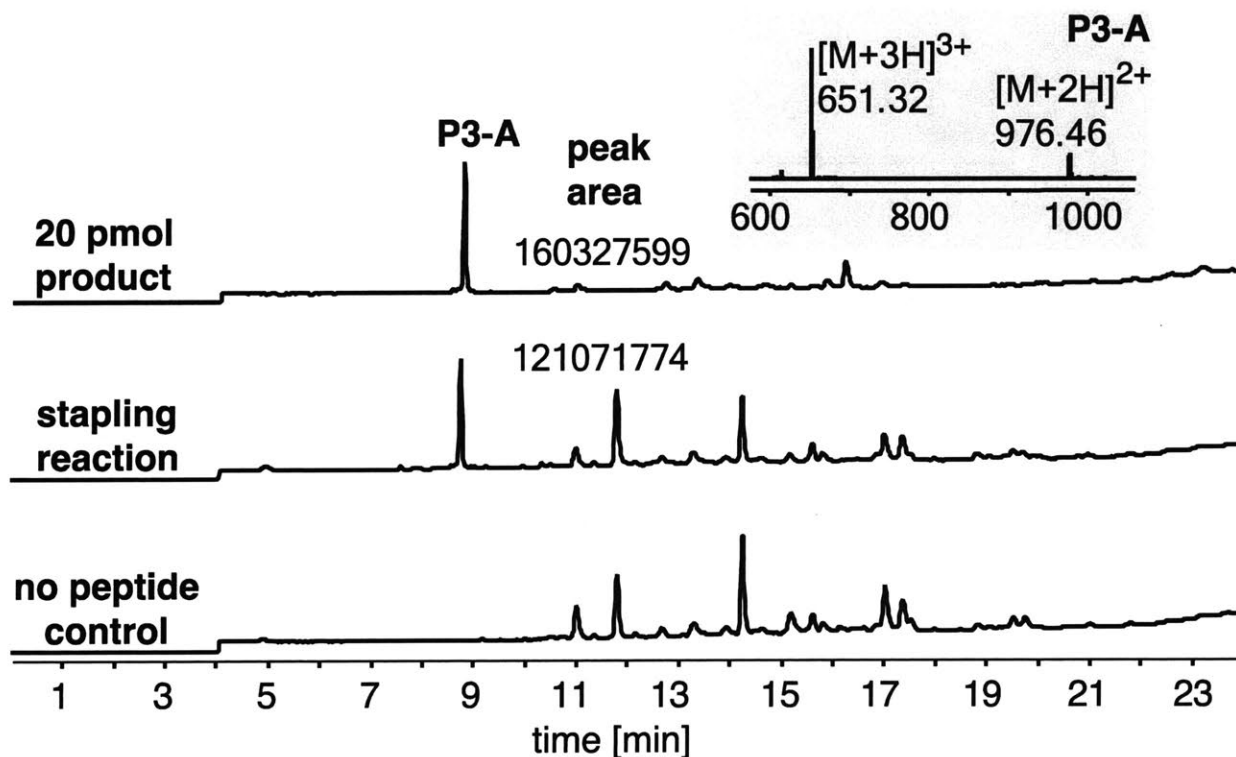
Figure 4.27. LC-MS chromatograms for peptide stapling reactions and controls.

a

100 pmol   area: 1538970279

50 pmol   area: 805676042

25 pmol   area: 407313558

10 pmol   area: 149260697

5 pmol   area: 66652067

time [min]

b

area [10^8]

$R^2 = 0.999$

P3-A [pmol]
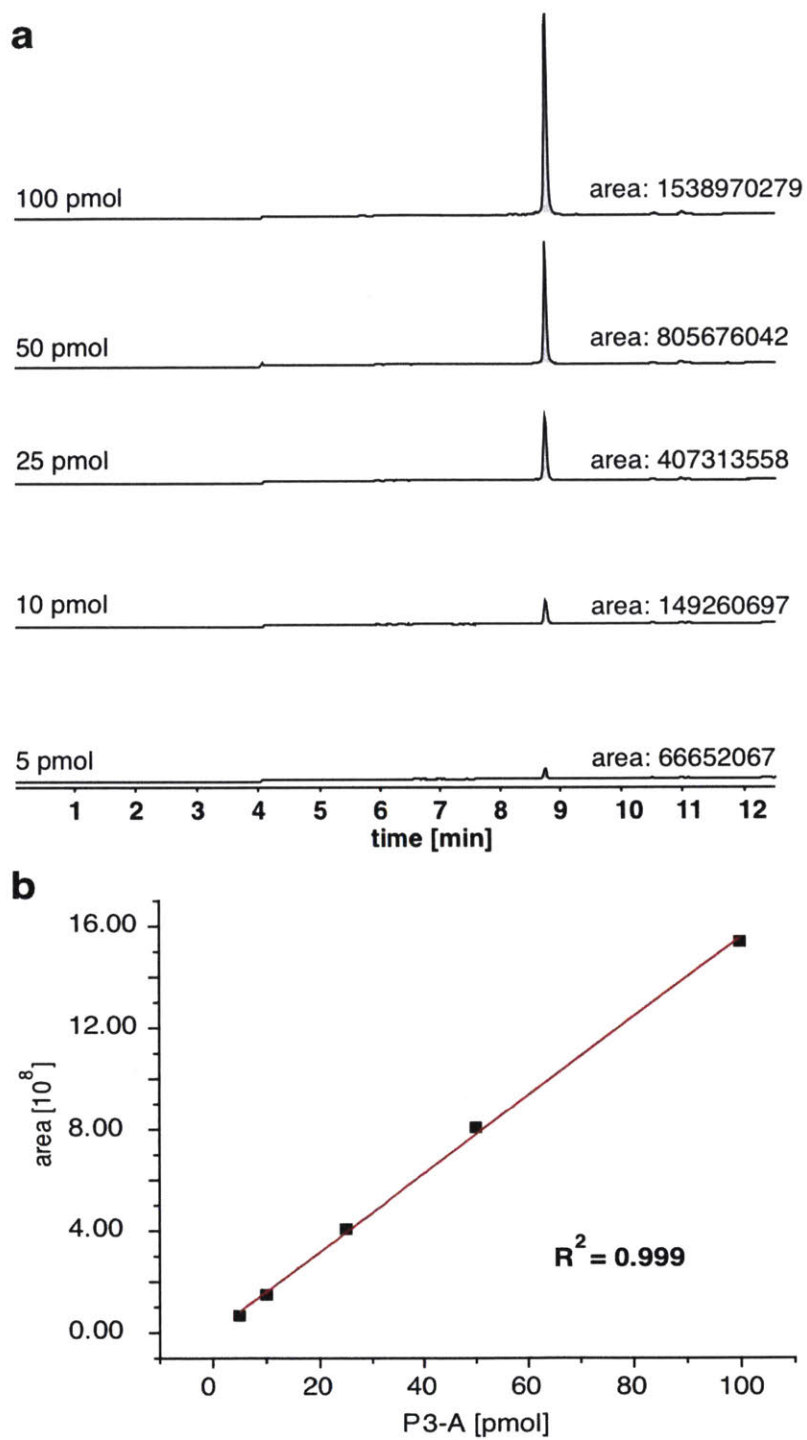
Figure 4.28. LC-MS standard curve for the stapled peptide product P3-A.
(**a**) LC-MS chromatograms for analysis of various amount of **P3-A**. (**b**) Linear fitting of the amounts of **P3-A** to the area under the LC-MS TIC curves.

318

4.4.12. Stability of palladium complexes

All synthesized palladium(II) complexes were stored in closed 5 mL scintillation vials under air at 4 °C. The long-term stability (4.5 – 10.5 months) of the 4-tolyl complexes (**1A-I**, **1A-Br**, **1A-Cl**, and **1A-OTf**) was evaluated by ^{1}H NMR spectroscopy in CD$_2$Cl$_2$. Only **1A-OTf** showed some decomposition after 4.5 months of storage in air (15% of a new RuPhos-containing by-product was detected). The reactivity of this "aged" complex was tested under standard conditions for cysteine arylation of peptide **P1** (Procedure A), and proved to be similar to the reactivity of the freshly made **1A-OTf** (Figure 4.5).

Figure 4.29. ^{1}H-NMR spectra of palladium complexes 1A-I (a) and 1A-Br (b) after months of storage.

Figure 4.30. ^{1}H-NMR spectra of palladium complexes 1A-Cl (a) and 1A-OTf (b) after months of storage.
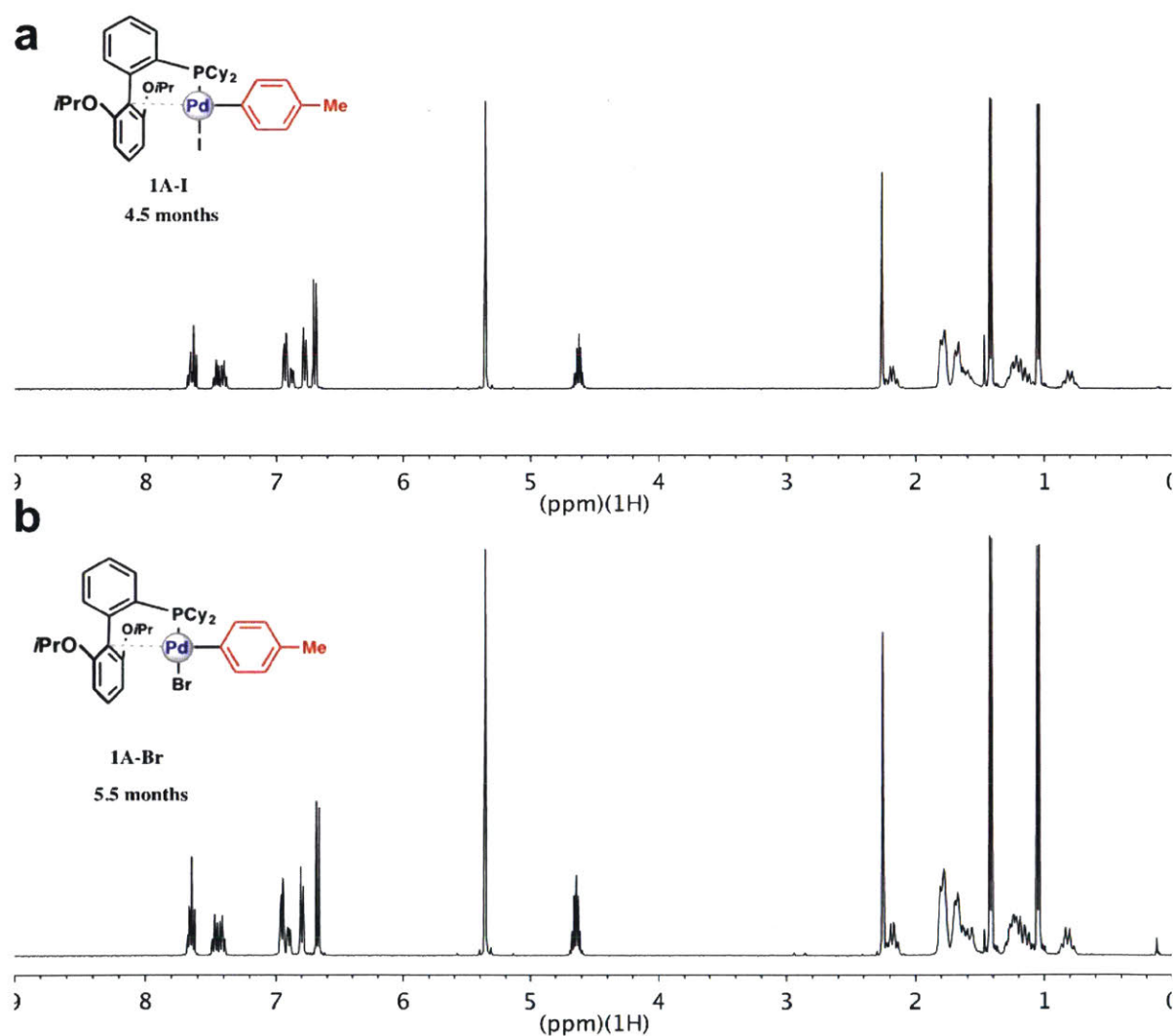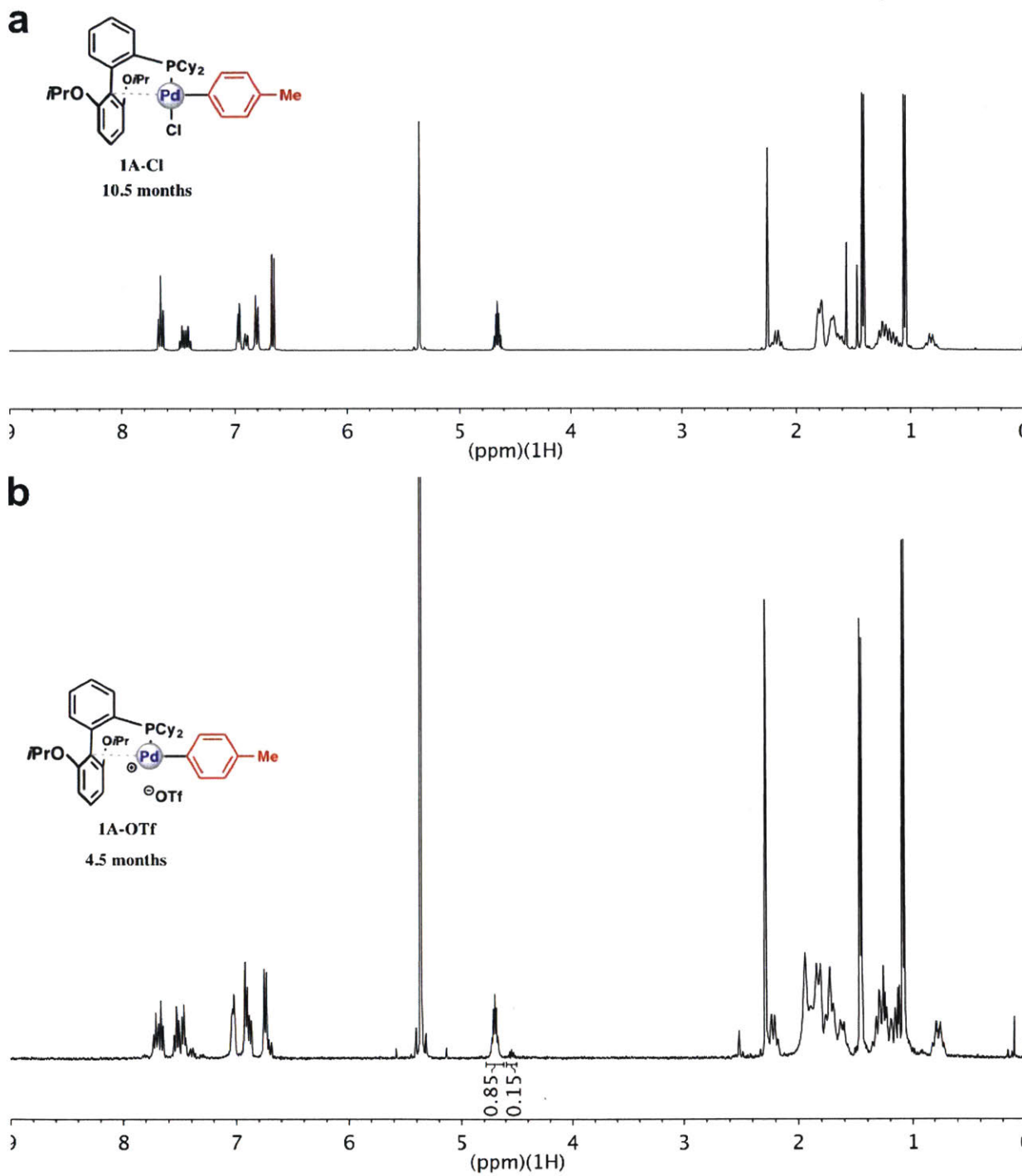
4.4.13. Stability evaluation of cysteine peptide conjugates

Synthesis and purification of modified model peptides
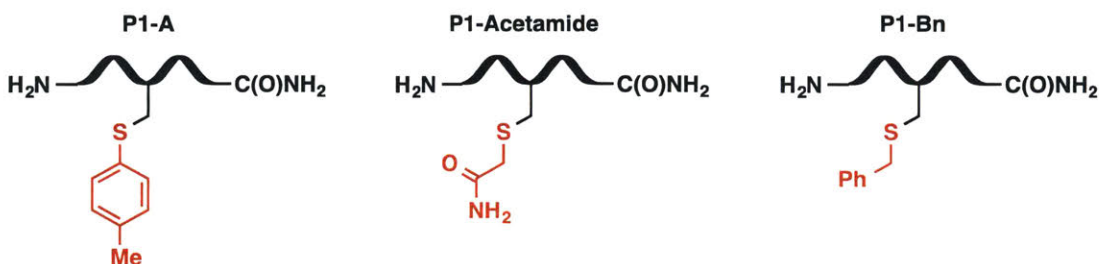
To a solution of peptide P1 (3 μmoles) in Tris buffer (2 mL, 0.1 M, pH 8.0) in a 5-mL Eppendorf tube was added the corresponding labeling reagent (15 μmoles; 1A for P1-A, iodoacetamide for P1-Acetamide, and benzylbromide for P1-Bn) dissolved in CH₃CN (1 mL). The reaction was vortexed for 20 s to ensure proper reagent mixing and left at room temperature for 30 min. After this time, 20 mL of solvent A (0.1% TFA in water) was added to quench the reaction. The resulting solution was filtered and subjected to HPLC purification.
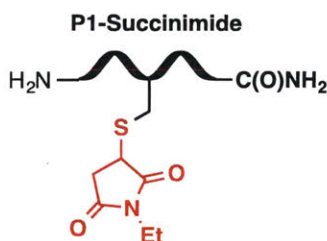
To a solution of peptide P1 (3 μmoles) in MES buffer (2 mL, 0.1 M, pH 6.0) in a 5-mL Eppendorf tube was added *N*-ethyl maleimide (15 μmoles) dissolved in CH₃CN (1mL). The reaction was vortexed for 20 s to ensure proper reagent mixing and left at room temperature for 30 min. After this time, 20 mL of solvent A (0.1% TFA in water) was added to quench the reaction. The resulting solution was filtered and subjected to HPLC purification.
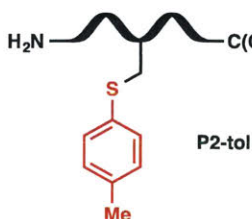
To a solution of peptide P2 (9.5 μmoles) in Tris buffer (1.1 mL, 0.1 M, pH 8.0) in a 5-mL Eppendorf tube was added 1A (19 μmoles) dissolved in CH₃CN (1.1 mL). The reaction was

vortexed for 20 s to ensure proper reagent mixing and left at room temperature for 30 min. After this time, thiopropionic acid (30.6 μL in 2 mL H$_2$O) was added to quench the reaction. The resulting precipitate was spinned down, separated from the solution and washed with an additional 1 mL of 0.1% TFA in water. The resulting solutions were combined and subjected to HPLC purification. *Note:* ICP-MS analysis of the pure isolated peptide showed 2.9 ppm remaining palladium content.

To a solution of peptide P2 (10 μmoles) in Tris buffer (1 mL, 0.1 M, pH 8.0) in a 5-mL Eppendorf tube was added the corresponding labeling reagent (11 μmoles; 1-Benzonitrile for P2-PhCN, benzylbromide for P2-Bn, *p*-cyanobenzylbromide for P2-BnCN, and iodoacetamide for P2-Acetamide) dissolved in CH$_3$CN (1mL). The reaction was vortexed for 20 s to ensure proper reagent mixing and left at room temperature for 30 min. After this time, 10 mL of solvent A (0.1% TFA in water) was added to quench the reaction. The resulting solution was filtered and subjected to HPLC purification.
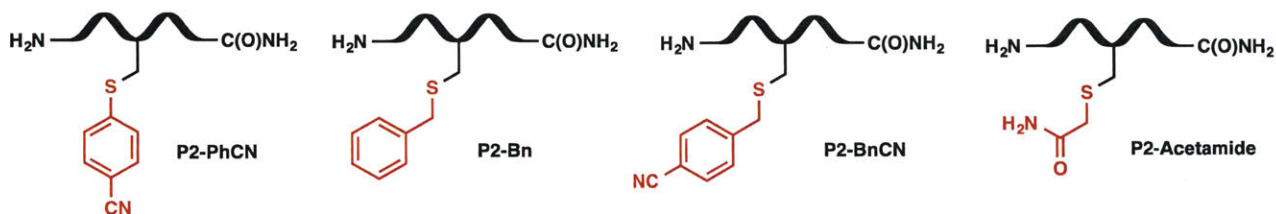

To a solution of peptide P2 (10 μmoles) in MES buffer (1 mL, 0.1 M, pH 6.0) in a 5-mL Eppendorf tube was added *N*-ethyl maleimide (11 μmoles) dissolved in CH$_3$CN (1 mL). The reaction was vortexed for 20 s to ensure proper reagent mixing and left at room temperature for 30 min. 10 mL of solvent A (0.1% TFA in water) was added to quench the reaction. The resulting solution was filtered and then subjected to HPLC purification.
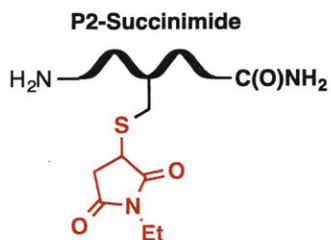
Stability evaluation in the presence of base, acid or an external thiol nucleophile

Peptide **P1** conjugates were pre-dissolved in water in plastic Eppendorfs to afford the 1.11 mM stock solutions used in the stability evaluation experiments. For each experiment, the corresponding cysteine conjugate (1.11 mM; 18 μL) and stability test reagent (2 μL, 50 mM in H$_2$O or 50 mM in 1M Tris, pH 7.4) were combined in a plastic Eppendorf and left at rt for 2 days, followed by 4 days at 37 °C. After this time, individual reactions were quenched with a solution of 50% A: 50% B (v/v, 200 μL) and the resulting samples were analyzed by LC-MS (Figure 4.31).

Basic conditions

Stability test reagent: K$_2$CO$_3$ (2 μL, 50 mM in H$_2$O);

Final conditions before quenching: 1mM peptide, 5 mM K$_2$CO$_3$; 2 d at rt, then 4 d at 37 °C.

Acidic Conditions

Stability test reagent: HCl (2 μL, 1 M in H$_2$O);

Final conditions before quenching: 1 mM peptide, 0.1 M HCl; 2 d at rt, then 4 d at 37 °C.

Presence of External Thiol Nucleophiles: GSH

Stability test reagent: Glutathione (2 μL, 50 mM in 1 M Tris; pH 7.4);

Final conditions before quenching: 1 mM peptide, 5 mM GSH, 0.1M Tris, pH 7.4; 2 d at rt, then 4 d at 37 °C.

Stability of cysteine conjugates toward oxidation

Peptide **P2** conjugates were pre-dissolved in water in plastic Eppendorfs to afford the 111.1 μM stock solutions used in the oxidation stability evaluation experiments. The corresponding cysteine conjugates (18 μL, 111.1 μM in H$_2$O) and H$_5$IO$_6$ (2 μL, 4 mM in H$_2$O) were then combined in a plastic Eppendorf, mixed using a vortexer and transferred into a pre-heated water bath at 37 °C. Individual reactions were quenched with Na$_2$SO$_3$ (20 μL, 4 mM in H$_2$O) after 10 min, 30 min, 1 h, 2 h, 4 h, and 6 h, and the resulting mixtures were kept at rt for an additional 10 min. Subsequently, a solution of 50% A: 50% B (v/v, 160 μL) was added and the resulting samples were analyzed by LC-MS (Figure 4.32). Final conditions before quenching: 100 μM peptide, 400 μM H$_5$IO$_6$, 37 °C.
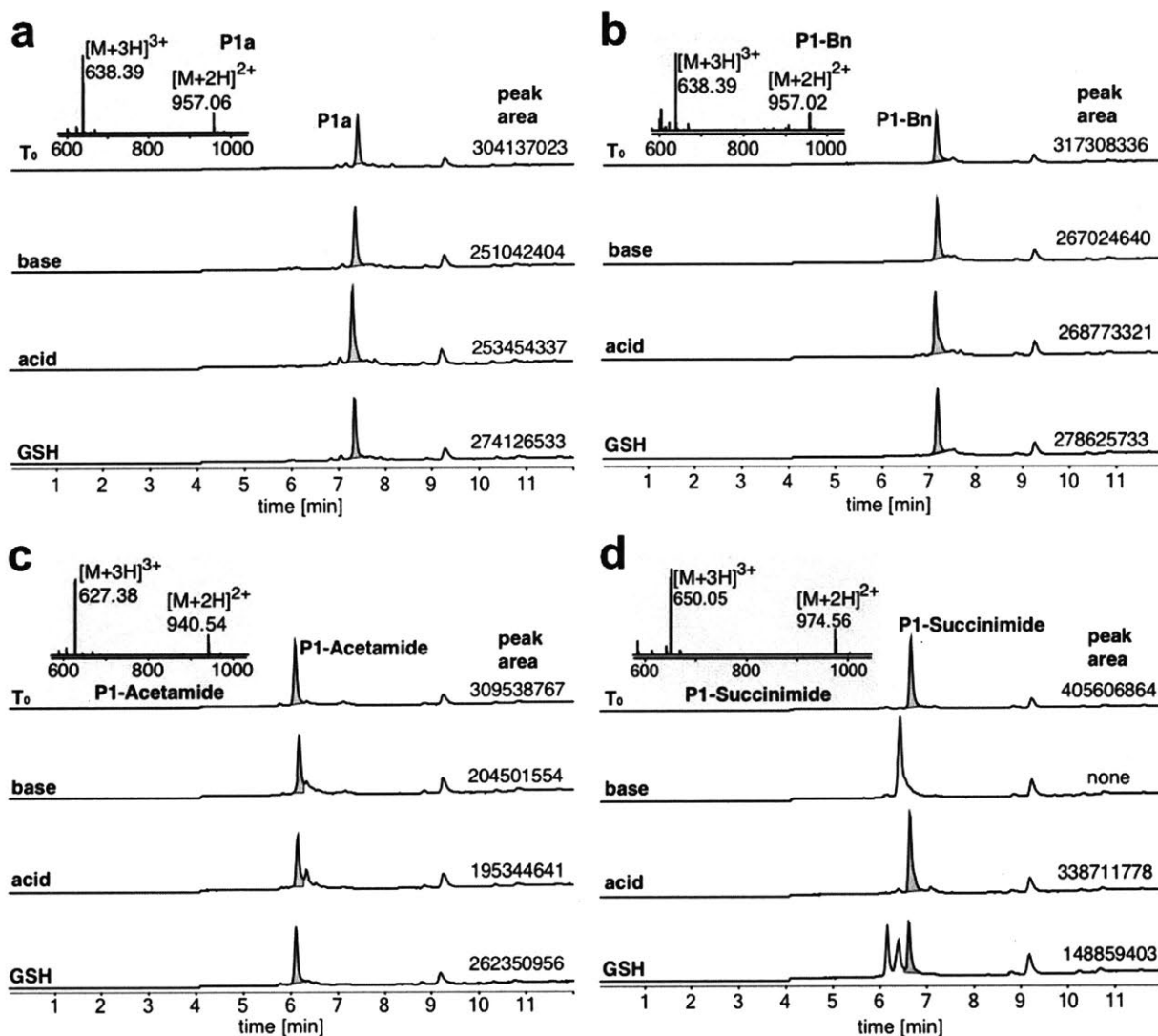
Figure 4.31. LC-MS chromatograms for the experiments evaluating the stability of cysteine conjugates in the presence of base, acid, and external thiol nucleophiles (GSH).

(**a**) Analysis of the stability of **P1a**. (**b**) Analysis of the stability of **P1-Bn**. (**c**) Analysis of the stability of **P1-Acetamide**. (**d**) Analysis of the stability of **P1-Succinimide**.
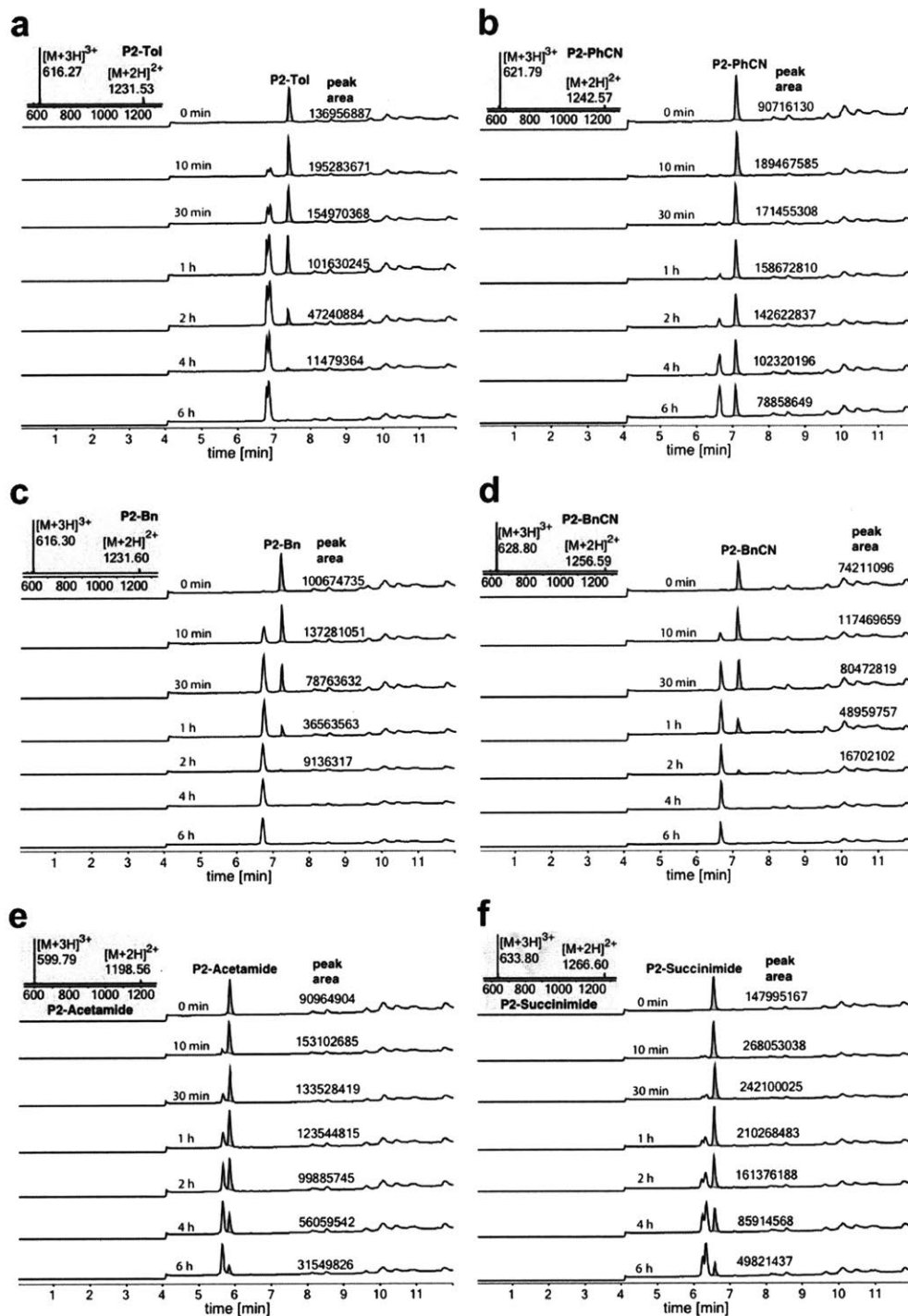
Figure 4.32. LC-MS chromatograms for the experiments evaluating the stability of cysteine conjugates against oxidation.

(**a**) Analysis of the stability of **P2-Tol**. (**b**) Analysis of the stability of **P2-PhCN**. (**c**) Analysis of the stability of **P2-Bn**. (**d**) Analysis of the stability of **P2-BnCN**. (**e**) Analysis of the stability of **P2-Acetamide**. (**f**) Analysis of the stability of **P2-Succinimide**.

4.4.14. Evaluation of the reaction rate of palladium-mediated cysteine conjugation

Kinetic parameters of the maleimide conjugation at biologically relevant pH are typically determined by extrapolation of the results obtained at lower pH, at which the reaction is significantly slower.[18] This approach could not be applied to our system due to potential variation of the reactive species structure at different pH. Both of the bioconjugation reactions are extremely fast and reach completion at room temperature at pH 7.5 in less than 30 s. Therefore, the rate of the process was estimated through a competition experiment between palladium reagent **1A-OTf** and *N*-ethyl maleimide at pH 7.5 and pH 5.5 in the reaction with peptide **P2**. First, the orthogonality of the palladium complex and the maleimide reagent was established by dissolving equimolar quantities of *N*-ethyl maleimide and complex **1A-OTf** in CD_3CN and subsequent analysis of the resulting solution by ^1H, ^{31}P and ^{19}F NMR spectroscopies (Figure 4.33).

Procedure for Competition Experiments.

Peptide **P2** (4 μL, 150 μM), H_2O (47 μL), Tris buffer (6 μL, 1 M, pH 7.5 or pH 5.5), and CH_3CN (1 μL) were combined in a 0.6 mL plastic Eppendorf tube and the resulting solution was mixed using a vortexer. A stock solution of the bioconjugation reagent(s) (2 μL, 600 μM *N*-ethyl maleimide, 600 μM **1A-OTf**, or [600 μM *N*-ethyl maleimide and 600 μM **1A-OTf**]) in CH_3CN was added in one portion, the reaction tube was vortexed to ensure proper reagent mixing and left at room temperature for 5 min. The reaction was quenched by the addition of 3-mercaptopropionic acid (6.3 μL, 0.05 μL/mL solution, 3 equiv to **1A-OTf**). After an additional 5 min a solution of 50% A: 50% B (v/v, 60 □L) was added to the Eppendorf and the reaction mixture was analyzed by LC-MS (Fig. S6).

Final conditions before quenching:

Experiment 1. **P2** – 10 μM, *N*-ethyl maleimide – 20 μM, 0.1 M Tris (pH 7.5 or pH 5.5), CH_3CN: H_2O = 5: 95.

Experiment 2. **P2** – 10 μM, **1A-OTf** – 20 μM, 0.1 M Tris (pH 7.5 or pH 5.5), CH_3CN: H_2O = 5: 95.

Experiment 3. **P2** – 10 μM, **1A-OTf** – 20 μM, *N*-ethyl maleimide – 20 μM, 0.1 M Tris (pH 7.5 or pH 5.5), CH_3CN: H_2O = 5: 95. *Note: N-ethyl maleimide and 1A-OTf were added together as a solution in CH₃CN.*
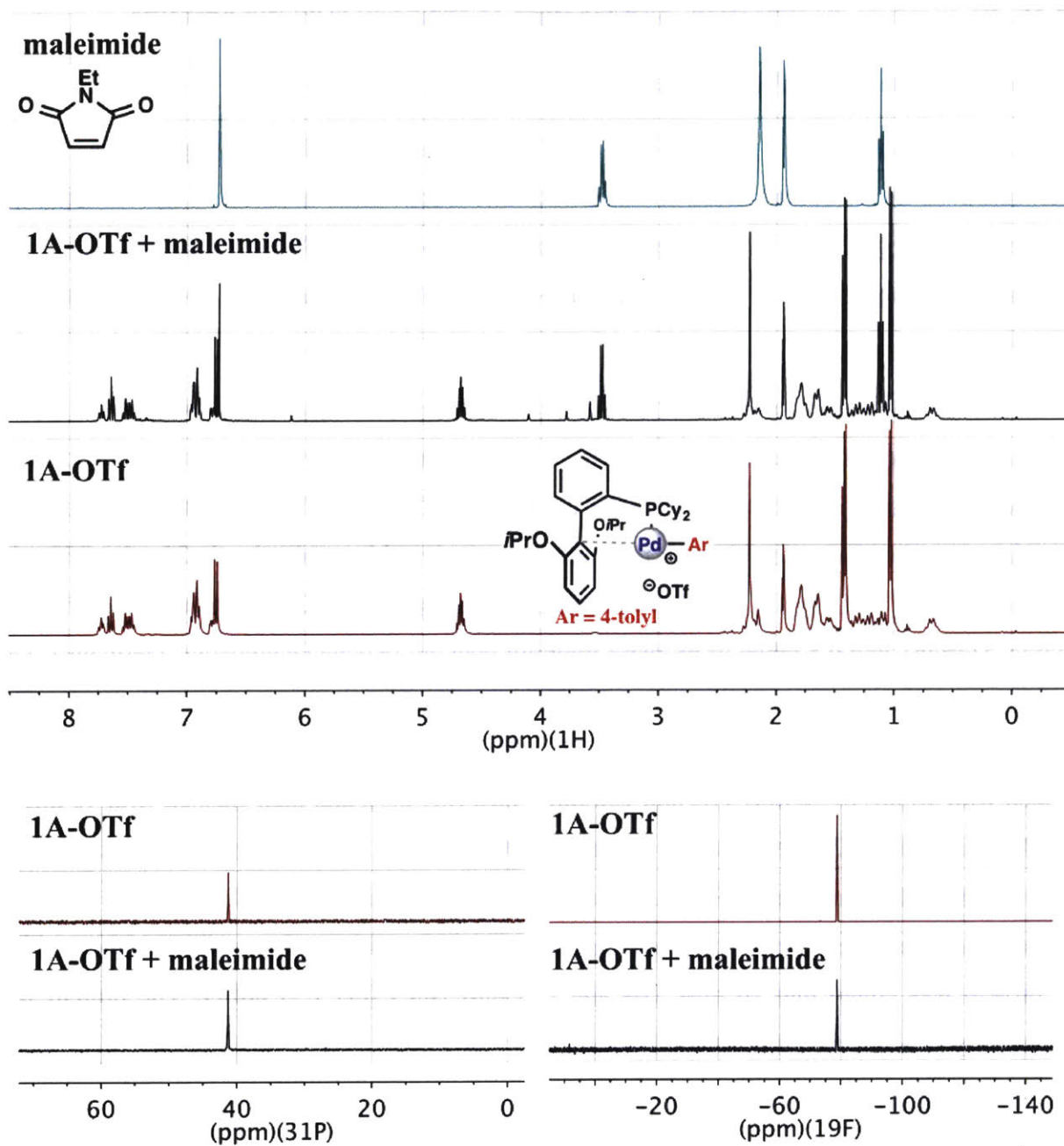
Figure 4.33. Comparison of the ^{1}H, ^{31}P and ^{19}F NMR spectra of *N*-ethyl maleimide, palladium reagent 1A-OTf and their equimolar mixture.

4.4.15. Expression, purification, and labeling of antibody mimetic proteins

Protein Expression and Purification

pET-SUMO-DARPin, pET-SUMO-10FN3, pET-SUMO-Affibody plasmids were constructed as reported previously.[39] Cysteine mutations were introduced by site-directed mutagenesis using QuickChange Lightning Single Site-directed Mutagenesis Kit (Agilent) following manufacturer's instructions. Sequences of generated protein constructs are summarized in Table S6.

E. coli BL21(DE3) cells transformed with pET-SUMO-Protein plasmid were grown in 1 L of LB medium containing kanamycin (30 μg/mL) at 37 °C until $OD_{600} = 0.6$. Then, expression was induced by the addition of 0.5 mM IPTG overnight at 30 °C. After harvesting the cells by centrifugation (6,000 rpm for 10 min), the cell pellet was lysed by sonication in 25 mL of 50 mM Tris and 150 mM NaCl (pH 7.5) buffer containing 15 mg lysozyme (Calbiochem), 1 mg DNase I (Sigma-Aldrich), and 0.5 tablet of protease inhibitor cocktail (Roche Diagnostics, Germany). The resulting suspension was centrifuged at 17,000 rpm for 30 min to remove cell debris. The supernatant was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK), first washed with 40 mL of 20 mM Tris and 150 mM NaCl (pH 8.5), and then washed with 40 mL of 40 mM imidazole in 20 mM Tris and 150 mM NaCl (pH 8.5). The protein was eluted from the column with buffer containing 500 mM imidazole in 20 mM Tris and 150 mM NaCl (pH 8.5). Imidazole was removed from protein using a HiPrep 26/10 Desalting column (GE Healthcare, UK), the protein was eluted into 20 mM Tris and 150 mM NaCl (pH 7.5) buffer. The protein was analyzed by LC-MS to confirm its purity and molecular weight.

SUMO group on SUMO-Protein was cleaved by incubating 1 μg of SUMO protease per mg of protein at room temperature for 60 min. The crude reaction mixture was loaded onto a 5 mL HisTrap FF crude Ni-NTA column (GE Healthcare, UK) and the flow through containing the desried protein was collected. The protein was analyzed by LC-MS confirming sample purity and molecular weight. Purified proteins were concentrated using Amicon 3K concentrator (50 mL, EMD Millipore); protein aliquots were flash frozen and stored in −80 °C freezer. Sequences and calculated masses of proteins are summarized in Table 4.5.

Protein Labeling Experiments

To a solution of protein (500 pmoles) in 475 µL of 20 mM Tris and 150 mM NaCl buffer (pH 7.5) was added palladium-coumarin complex **1D** or palladium-drug complex **1J** (25 µL, 200 µM) in DMF. The solution was pipetted up and down 20 times to ensure proper reagent mixing. The reaction mixture was left at room temperature for 30 min. After this time, the reaction was quenched by the addition of 3-mercaptopropionic acid (25 µL, 2 mM) dissolved in 20 mM Tris and 150 mM NaCl buffer (pH 7.5). After an additional 5 min at rt, 500 µL of 1: 1 CH_3CN/H_2O (v/v) containing 0.2% TFA was added and the resulting mixture was analyzed by LC-MS.

Table 4.5. Protein sequences and masses.

Protein **P4**: DARPin-Cys	**Calculated Mass: 13747.3 Da**

Sequence[a]:

GGCGGSDLGKKLLEAARAGQDDEVRILMANGADVNAYDDNGVTPLHLAAFLGHLEIVEVLLKYGADVNA
ADSWGTTPLHLAATWGHLEIVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQKLN

Protein **P7**: DARPin	**Calculated Mass: 13701.3 Da**

Sequence:

GGGGGSDLGKKLLEAARAGQDDEVRILMANGADVNAYDDNGVTPLHLAAFLGHLEIVEVLLKYGADVNA
ADSWGTTPLHLAATWGHLEIVEVLLKHGADVNAQDKFGKTAFDISIDNGNEDLAEILQKLN

Protein **P5**: 10FN3-Cys	**Calculated Mass: 10813.1 Da**

Sequence:

SVSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKSTATISGLKPGVDY
TITVYAVTLPSTCGASSKPISINYRTEIDKPSQ

Protein **P8**: 10FN3	**Calculated Mass: 10679.9 Da**

Sequence:

VSDVPRDLEVVAATPTSLLISWDAPAVTVRYYRITYGETGGNSPVQEFTVPGSKSTATISGLKPGVDYT
ITVYAVTLPSTGGASSKPISINYRTEIDKPSQ

Protein **P6**: Affibody-Cys	**Calculated Mass: 6900.6**

Sequence:

GGGGGVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDACAPK

Protein **P9**: Affibody	**Calculated Mass: 6925.6 Da**

Sequence:

GGGGGVDNKFNKEQQNAFYEILHLPNLNEEQRNAFIQSLKDDPSQSANLLAEAKKLNDAQAPK

[a]Cysteine residues are highlighted in red.

4.4.16. Antibody expression and purification

Expression and Purification of Trastuzumab

Trastuzumab was expressed from the gWiz-trastuzumab plasmids. The light chain and heavy chain sequences for trastuzumab used in this study are:

Trastuzumab-Light Chain

DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSGTD
FTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTRTVAAPSVFIFPPSDEQLKSGTASVVCLLNN
FYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKS
FNRGEC

Trastuzumab-Heavy Chain

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTI
SADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSG
GTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPS
NTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNW
YVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQ
VYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW
QQGNVFSCSVMHEALHNHYTQKSLSLSPGK

The IgGs were expressed via transient transfections of HEK293F cells (Invitrogen) as previously described,[40] and purified using Protein A affinity chromatography (Genscript) following manufacturer's instructions. The purified IgGs were analyzed by LC-MS to confirm their molecular weight and purity, and stored in PBS at -80 °C.

Synthesis and purification of an Antibody-Drug Conjugate (ADC).

TCEP (1 μL, 25 mM in water) and Tris buffer (5 μL, 1 M, pH 8.0) were added to a solution of trastuzumab (57 μM, 44 μL) in PBS. The reaction mixture was pipetted up and down 20 times and incubated in a water bath at 37 °C for 2 h. The final reaction conditions for the reduction are: 50 μM antibody, 500 μM TCEP, 0.1 M Tris, pH 8.0.

Tris buffer (187.5 μL, 0.1 M, pH 8.0) and palladium-drug complex **1J** (12.5 μL, 2 mM in DMF) were added to the partially reduced antibody, the reaction mixture was pipetted up and down

20 times and was left at room temperature for 30 minutes. The final reaction conditions are: 10 μM antibody, 100 μM **1J**, 0.1 M Tris, pH 8.0, 5% DMF. The reaction mixture was purified using three protocols shown below to assess the protein recovery and palladium removal efficiency in each protocol.

Purification protocol 1

The crude reaction mixture was directly loaded on a size-exclusion chromatography (SEC) column (GE Healthcare) for purification using buffer P (20 mM Tris, 150 mM NaCl, pH 7.5) as the elution buffer. The fractions were collected and analyzed by gel electrophoresis. The fractions containing antibodies were then combined, concentrated, and buffer exchanged into phosphate buffer saline (PBS) using 10K spin concentrator (EMD Millipore). The antibody was recovered in 47% yield as measured by UV280. This purification protocol allowed for the removal of 81% of the palladium species from the crude reaction mixture as determined by ICP-MS (see details in the ICP-MS section). This protein sample was used in the Octet BioLayer Interferometry binding assay.

Purification protocol 2

The crude reaction mixture was quenched with a solution of thiopropionic acid (0.5 μL, 50 mM in buffer P, 10 equivalents relative to the amount of the palladium reagent used), and the resulting solution was incubated at room temperature for 5 minutes. The solution was filtered through a 0.2 μm nylon spin filter (PALL Life Sciences). The filtered solution was loaded on a PD-10 desalting column (GE Healthcare) pre-equilibrated with TPA buffer (20 mM Tris, 150 mM NaCl, 1 mM thiopropionic acid). The antibody was eluted into 3.5 mL of TPA buffer following the manufacturer's protocol. The resulting solution was buffer exchanged into phosphate buffer saline (PBS) using 10K spin concentrator (EMD Millipore). The antibody was recovered in 86% yield as measured by UV280. This purification procedure allowed for removal of 73% of the palladium species from the crude reaction mixture as determined by ICP-MS (see details in the ICP-MS section).

Purification protocol 3

The crude reaction mixture was quenched with a solution of thiopropionic acid (0.5 μL, 50 mM in buffer P, 10 equivalents relative to the amount of the palladium reagent used), and the

resulting solution was incubated at room temperature for 5 minutes. The solution was filtered through a 0.2 μm nylon spin filter (PALL Life Sciences). The filtered solution was loaded on a size-exclusion chromatography (SEC) column (GE Healthcare) pre-equilibrated with TPA buffer (20 mM Tris, 150 mM NaCl, 1 mM thiopropionic acid). The antibody was purified using TPA buffer as the elution buffer. The fractions were collected and analyzed by gel electrophoresis, and the fractions containing antibodies were combined, concentrated, and buffer exchanged into phosphate buffer saline (PBS) using 10K spin concentrator (EMD Millipore). The antibody was recovered in 39% yield as measured by UV280. The purification procedure allowed for the removal of 94% of the palladium species from the crude reaction mixture as determined by ICP-MS (see details in the ICP-MS section).

LC-MS analysis of the purified ADCs

N-linked glycans were removed by the addition of 1 μL of PNGase F (New England BioLabs) to the antibody (100 μg in 100 μL 0.1 M Tris, pH 8.0) and incubation of the mixture for 1 h at 45 °C. The resulting solution was reduced by the addition of 1/10 volume of 200 mM TCEP solution (pH 8.0) and incubation for 30 min at 37 °C. The resulting mixture was quenched by the addition of 10/1 volume of 50% A: 50% B (v/v) and analyzed by LC-MS. The ADC purified using *protocol 1* showed the drug-to-antibody ratio of 4.4.

4.4.17. Modification of LF$_N$-DTA and protein synthesis inhibition assay

Synthesis and purification of modified LF$_N$-DTA variants

Lethal factor N-terminal domain fused to diphtheria toxin A-chain (LF$_N$-DTA) was expressed and purified as described before.[41] A C-terminal LPSTGGH$_6$ tag was included for the Sortase-mediated peptide conjugation. Cysteine-containing peptide NH$_2$-GGGGGLRLCA-C(O)NH$_2$ and its serine variant NH$_2$-GGGGGLRLSA-C(O)NH$_2$ were synthesized, purified, and conjugated to the LF$_N$-DTA-LPSTGGH$_6$ tag following the previously described protocol (Fig. S12).[41] Sequences for the resulting proteins **LF$_N$-DTA-Cys** and **LF$_N$-DTA-Ser** are shown below.

LF$_N$-DTA-Cys

AGGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLKEIMKHIVKIEVKGEEAVKKEAAEKLLEKVPSDV
LEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDIYGKDALLHEHYVYAKEGYEPVLVIQSSEDYVEN
TEKALNVYYEIGKILSRDILSKINQPYQKFLDVLNTIKNASDSDGQDLLFTNQLKEHPTDFSVEFLEQNS

NEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKFNEQEINLSLEELKDQRSGRELERGADDVVDSSK
SFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGV
VKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYI
NNWEQAKALSVELEINFETRGKRGQDAMYEYMAQASAGNRLPSTGGGGGLRL**C**A-C(O)NH₂

LF_N-DTA-Ser

AGGHGDVGMHVKEKEKNKDENKRKDEERNKTQEEHLKEIMKHIVKIEVKGEEAVKKEAAEKLLEKVPSDV
LEMYKAIGGKIYIVDGDITKHISLEALSEDKKKIKDIYGKDALLHEHYVYAKEGYEPVLVIQSSEDYVEN
TEKALNVYYEIGKILSRDILSKINQPYQKFLDVLNTIKNASDSDGQDLLFTNQLKEHPTDFSVEFLEQNS
NEVQEVFAKAFAYYIEPQHRDVLQLYAPEAFNYMDKFNEQEINLSLEELKDQRSGRELERGADDVVDSSK
SFVMENFSSYHGTKPGYVDSIQKGIQKPKSGTQGNYDDDWKGFYSTDNKYDAAGYSVDNENPLSGKAGGV
VKVTYPGLTKVLALKVDNAETIKKELGLSLTEPLMEQVGTEEFIKRFGDGASRVVLSLPFAEGSSSVEYI
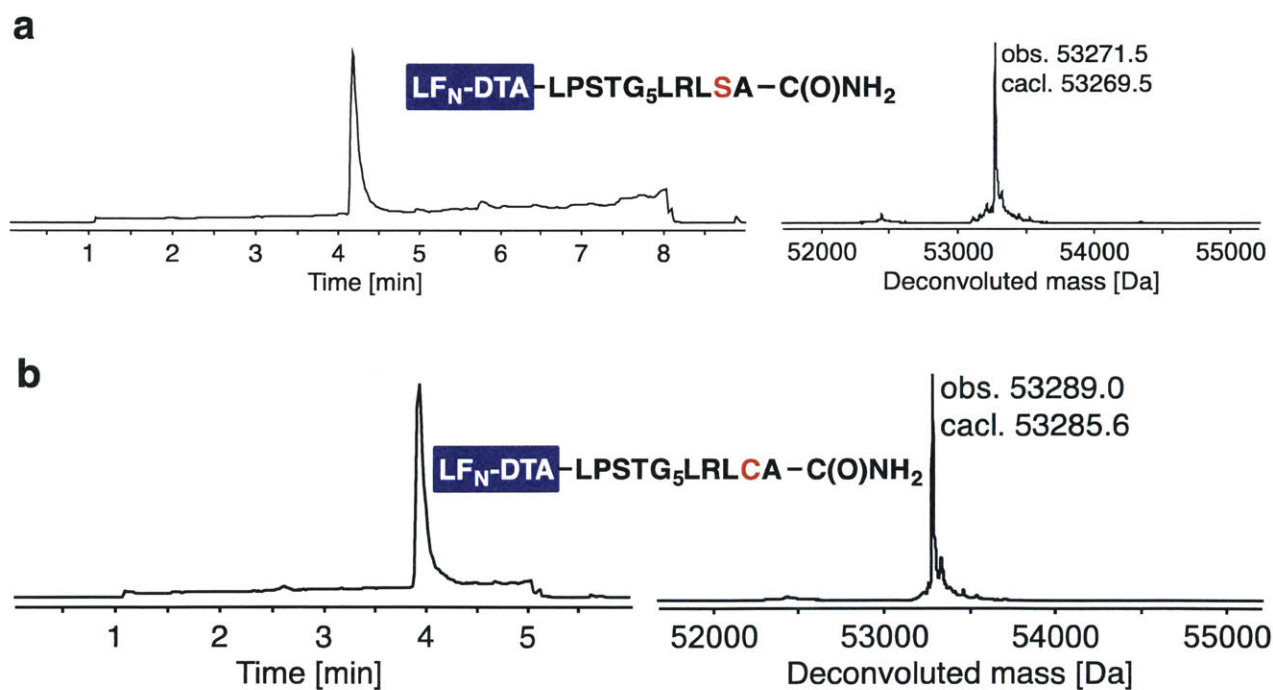NNWEQAKALSVELEINFETRGKRGQDAMYEYMAQASAGNRLPSTGGGGGLRL**S**A-C(O)NH₂

Figure 4.34. LC-MS analysis of the purified LF_N-DTA-Ser (a) and LF_N-DTA-Cys (b).
Total ion current chromatograms (left) and the deconvoluted mass spectra (right) of the full protein
peaks are shown.

Arylation of LF_N-DTA-Cys using 1A-OTf

Complex **1A-OTf** (100 μL; 400 μM in DMF) was added to a solution of **LF_N-DTA-Cys** (2.5 μM, 1.9 mL) in buffer P (20 mM Tris, 150 mM NaCl, pH 7.5). The final reaction conditions are: 2.4 μM of **LF_N-DTA-Cys**, 20 μM of **1A-OTf**, 150 mM NaCl, 20 mM Tris, pH 7.5, 5% DMF. The reaction mixture was pipetted up and down 20 times to afford proper mixing and was rotated at room temperature for 2 h, after which 20 μL of 20 mM thiopropionic acid in buffer P (10 equivalents of thiopropionic acid relative to the palladium reagent) was added to quench the reaction. The reaction mixture was incubated at room temperature for additional 5 min and was then subjected to purification by size-exclusion chromatography (SEC) using buffer P (20 mM Tris, 150 mM NaCl, pH 7.5) as the elution buffer. The fractions were analyzed by gel electrophoresis and the fractions containing the desired protein product were combined, concentrated, and buffer exchanged into phosphate buffer saline (PBS) using 10K spin concentrator (EMD Millipore). The resulting modified protein **LF_N-DTA-Cys(Ar)** was recovered in 56% yield as measured by UV280. This procedure allowed for the removal of 91% of the palladium species in the crude reaction mixture as measured by ICP-MS (see the ICP-MS section for details).

Alkylation of LF_N-DTA-Cys

Tris buffer (60 μL, 1M, pH 8.0) and bromoacetamide (60 μL, 10 mM in water) were added to a solution of **LF_N-DTA-Cys** (11 μM, 480 μL) in buffer P. The final reaction conditions are: 9 μM LF_N-DTA-Cys, 1 mM bromoacetamide, 0.1 M Tris, pH 8.0. The reaction mixture was pipetted up and down 20 times to afford proper mixing and was left at room temperature for 30 minutes. Upon completion, the reaction mixture was buffer exchanged 5 times with PBS buffer using 10K spin concentrator (EMD Millipore). The purified protein is denoted as **LF_N-DTA-Cys(Alk)**.
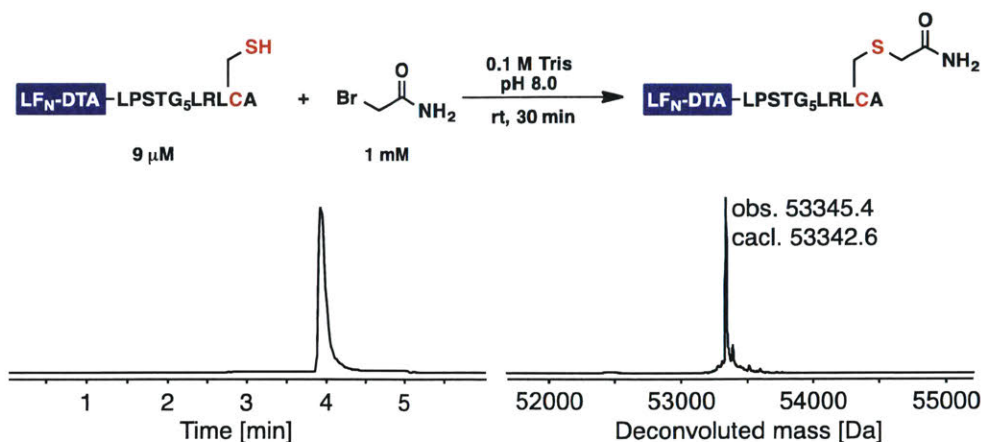
Figure 4.35. Alkylation of the LF$_N$-DTA-Cys and LC-MS analysis of the purified LF$_N$-DTA-Cys(Alk).

Total ion current chromatogram (left) and the deconvoluted mass spectrum of the full protein peak (right) are shown.

Protein synthesis inhibition assay

The translocation assay was performed as described previously.[14] The CHO-K1 cells were maintained in the F-12K media supplemented with 10% (v/v) fetal bovine serum at 37 °C and 5% CO_2 in humidified atmosphere. The cells were plated at 3.0×10^4/well in a 96-well white opaque plate one day prior to the assay. Modified LF$_N$-DTA variants (Figure S14a) were prepared in ten-fold serial dilutions in 50 μL of F-12K media each, after which additional 50 μL of F-12K media containing 20 nM of protective antigen (PA) were added. The resulting 100 μL samples were added to the CHO-K1 cells and incubated for 2 hours at 37 °C and 5% CO_2. The cells were then washed three times with PBS, after which 100 μL of leucine-free F-12K medium supplemented with 1 μCi/mL ^{3}H-leucine (Perkin Elmer, MA) was added. The resulting mixture was incubated for an additional 1 h at 37 °C with 5% CO_2. The cells were washed three times with PBS and suspended in 150 μL of scintillation fluid. ^{3}H-Leu incorporation into cellular proteins was measured to determine the inhibition of protein synthesis by the LF$_N$-DTA variants. Scintillation counts from cells treated with only PA were used as control values for normalization. Each experiment was done in triplicate.

4.4.18. BioLayer interferometry binding assay

The purified ADC and the native trastuzumab were biotinylated using EZ-Link NHS-(PEG)$_4$-Biotin (Life Technologies). A solution of EZ-Link NHS-(PEG)$_4$-Biotin (10 µL, 200 µM in PBS) was added to the corresponding protein (10 µL, 20 µM in PBS), after which the reaction mixture was pipetted up and down 20 times to allow proper mixing and was left at room temperature for 30 min. The crude reaction mixture was buffer exchanged with PBS for 5 times to remove the excess of NHS-(PEG)$_4$-Biotin.

The biotinylated proteins were immobilized on streptavidin tips and were sampled with the serially diluted concentrations of recombinant HER2 (R&D Biosystems) in the Octet BioLayer Interferometry system. The obtained results were fitted according to the protocol provided by the manufacturer of the Octet BioLayer Interferometry system.

4.4.19. Evaluation of the palladium removal efficiency using ICP-MS

ICP-MS was used to measure the remaining palladium content in the purified LF$_N$-DTA-Cys(Ar) protein and the purified antibody-drug conjugate (ADC). Protein samples were diluted in buffer solutions and lyophilized. The lyophilized powders were sent to Merck for ICP-MS analysis. The amount of remaining palladium in the purified sample was obtained from the calculations based on the measured palladium content in the ICP-MS samples. The palladium removal efficiency was reported as the percentage of palladium being removed from the sample compared to the palladium used in the crude reaction mixture. The obtained results are summarized in Table 4.6.

Table 4.6. Sample information for ICP-MS analysis of purified protein samples.

Protein Sample	Volume	Buffer	Concentration	Lyophilized Mass	Pd concentration
LF$_N$-DTA-Cys(Ar)	0.1 mL	Buffer P	0.285 mg/mL	2.93 mg	26 ppm
ADC (Purification protocol 1)	0.5 mL	PBS	0.1 mg/mL	5.26 mg	28 ppm
ADC (Purification protocol 2)	0.5 mL	PBS	0.1 mg/mL	4.48 mg	25 ppm
ADC (Purification protocol 3)	0.5 mL	PBS	0.1 mg/mL	4.74 mg	12 ppm

Calculation of the palladium removal efficiency.

LF$_N$-DTA-Cys(Ar)

100 μL of 400 μM of palladium complex **1A-OTf** was used in the crude reaction; this corresponds to the total palladium amount:

[Total palladium] = (100 μL)*(400 μM)*(106.42 g/mol) = 4.26 μg

56% of the protein was recovered, the total amount of protein recovered is:

[Total protein recovered] = (2.5 μM)*(1.9 mL)*(53286 g/mol)*0.56 = 0.14 mg

Palladium concentration in the protein sample is:

[Palladium concentration in protein] = (26 ppm)*(2.93 mg)/((0.1 mL)*(0.285 mg/mL)) = 2673 ppm

Palladium remaining in the purified protein is:

[Palladium remaining] = (2673 ppm)*(0.14 mg) = 0.37 μg

Palladium removal efficiency is:

[Palladium removal] = 1-[Palladium remaining]/[Total palladium] = 1-(0.37 μg)/(4.26 μg) = 91%

ADC (Protocol 1)

12.5 μL of 2 mM of palladium complex **1J** was used in the crude reaction; this gives the total palladium amount in the crude reaction:

[Total palladium] = (12.5 μL)*(2 mM)*(106.42 g/mol) = 2.66 μg

86% of the protein was recovered, the total amount of protein recovered is:

[Total protein recovered] = (57 μM)*(44 μL)*(149*10^3 g/mol)*0.86 = 0.321 mg

Palladium concentration in the protein sample is:

[Palladium concentration in protein] = (25 ppm)*(4.48 mg)/((0.5 mL)*(0.1 mg/mL)) = 2240 ppm

Palladium remaining in the purified protein is:

[Palladium remaining] = (2240 ppm)*(0.321 mg) = 0.719 μg

Palladium removal efficiency is:

[Palladium removal] = 1-[Palladium remaining]/[Total palladium] = 1-(0.719 μg)/(2.66 μg) = 73%

ADC (Protocol 2)

12.5 μL of 2 mM of palladium complex **1J** was used in the crude reaction; this gives the total palladium amount in the crude reaction:

[Total palladium] = (12.5 μL)*(2 mM)*(106.42 g/mol) = 2.66 μg

47% of the protein was recovered, the total amount of protein recovered is:

[Total protein recovered] = (57 μM)*(44 μL)*(149*10³ g/mol)*0.47 = 0.175 mg

Palladium concentration in the protein sample is:

[Palladium concentration in protein] = (28 ppm)*(5.26 mg)/((0.5 mL)*(0.1 mg/mL)) = 2946 ppm

Palladium remaining in the purified protein is:

[Palladium remaining] = (2946 ppm)*(0.175 mg) = 0.516 μg

Palladium removal efficiency is:

[Palladium removal] = 1-[Palladium remaining]/[Total palladium] = 1-(0.516 μg)/(2.66 μg) = 81%

ADC (Protocol 3)

12.5 μL of 2 mM of palladium complex **1J** was used in the crude reaction; this gives the total palladium amount in the crude reaction:

[Total palladium] = (12.5 μL)*(2 mM)*(106.42 g/mol) = 2.66 μg

39% of the protein was recovered, the total amount of protein recovered is:

[Total protein recovered] = (57 μM)*(44 μL)*(149*10³ g/mol)*0.39 = 0.146 mg

Palladium concentration in the protein sample is:

[Palladium concentration in protein] = (12 ppm)*(4.74 mg)/((0.5 mL)*(0.1 mg/mL)) = 1138 ppm

Palladium remaining in the purified protein is:

[Palladium remaining] = (1138 ppm)*(0.146 mg) = 0.166 μg

Palladium removal efficiency is:

[Palladium removal] = 1-[Palladium remaining]/[Total palladium] = 1-(0.166 μg)/(2.66 μg) = 94%

4.5. Acknowledgements

4.6. References

(1)   Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angew. Chemie Int. Ed.* **2005**, *44*, 7342–7372.

(2)   Rabuka, D. Chemoenzymatic Methods for Site-Specific Protein Modification. *Curr. Opin. Chem. Biol.* **2010**, *14*, 790–796.

(3)   Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*.

(4)   Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. a.; Davis, B. G. Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. - An Asian J.* **2009**, *4*, 630–640.

(5)   Cal, P. M. S. D.; Bernardes, G. J. L.; Gois, P. M. P. Cysteine-Selective Reactions for Antibody Conjugation. *Angew. Chemie - Int. Ed.* **2014**, *53*, 10585–10587.

(6)   Toda, N.; Asano, S.; Barbas, C. F. Rapid, Stable, Chemoselective Labeling of Thiols with Julia- Kocie??ski-like Reagents: A Serum-Stable Alternative to Maleimide-Based Protein Conjugation. *Angew. Chemie - Int. Ed.* **2013**, *52*, 12592–12596.

(7)   Lyon, R. P.; Setter, J. R.; Bovee, T. D.; Doronina, S. O.; Hunter, J. H.; Anderson, M. E.; Balasubramanian, C. L.; Duniho, S. M.; Leiske, C. I.; Li, F.; *et al.* Self-Hydrolyzing Maleimides Improve the Stability and Pharmacological Properties of Antibody-Drug Conjugates. *Nat. Biotechnol.* **2014**, *32*, 1059–1062.

(8)   Simmons, R. L.; Yu, R. T.; Myers, A. G. Storable arylpalladium(II) Reagents for Alkene Labeling in Aqueous Media. *J. Am. Chem. Soc.* **2011**, *133*, 15870–15873.

(9)   Antos, J. M.; Francis, M. B. Transition Metal Catalyzed Methods for Site-Selective Protein Modification. *Curr. Opin. Chem. Biol.* **2006**, *10*, 253–262.

(10)   Yang, M.; Li, J.; Chen, P. R. Transition Metal-Mediated Bioorthogonal Protein Chemistry in Living Cells. *Chem. Soc. Rev.* **2014**, *43*, 6511–6526.

(11)   Cheng, G.; Lim, R. K. V.; Li, N.; Lin, Q.; Herman, J. G.; Graff, J. R.; Myohanen, S.; Nelkin, B. D.; Baylin, S. B.; Kaiser, E. T.; *et al.* Storable Palladacycles for Selective

Functionalization of Alkyne-Containing Proteins. *Chem. Commun.* **2013**, *49*, 6809.

(12)   Barder, T. E.; Biscoe, M. R.; Buchwald, S. L. Structural Insights into Active Catalyst Structures and Oxidative Addition to (Biaryl)phosphine - Palladium Complexes via Density Functional Theory and Experimental Studies. *Organometallics* **2007**, *26*, 2183–2192.

(13)   Lee, C.-F.; Liu, Y.-C.; Badsara, S. S. Transition-Metal-Catalyzed C□S Bond Coupling Reaction. *Chem. - An Asian J.* **2014**, *9*, 706–722.

(14)   Bong, D. T.; Ghadiri, M. R. Chemoselective Pd(0)-Catalyzed Peptide Coupling in Water. *Org. Lett.* **2001**, *3*, 2509–2511.

(15)   Korneeva, E. N.; Ovchinnikov, M. V.; Kostić, N. M. Peptide Hydrolysis Promoted by Polynuclear and Organometallic Complexes of palladium(II) and platinum(II). *Inorganica Chim. Acta* **1996**, *243*, 9–13.

(16)   Tilley, S. D.; Francis, M. B. Tyrosine-Selective Protein Alkylation Using π-Allylpalladium Complexes. *J. Am. Chem. Soc.* **2006**, *128*, 1080–1081.

(17)   Shafer, D. E.; Inman, J. K.; Lees, A. Reaction of Tris(2-Carboxyethyl)phosphine (TCEP) with Maleimide and α-Haloacyl Groups: Anomalous Elution of TCEP by Gel Filtration. *Anal. Biochem.* **2000**, *282*, 161–164.

(18)   Gorin, G.; Martic, P. A.; Doughty, G. Kinetics of the Reaction of N-Ethylmaleimide with Cysteine and Some Congeners. *Arch. Biochem. Biophys.* **1966**, *115*, 593–597.

(19)   Gilbreth, R. N.; Koide, S. Structural Insights for Engineering Binding Proteins Based on Non-Antibody Scaffolds. *Curr. Opin. Struct. Biol.* **2012**, *22*, 413–420.

(20)   Kung, K. K.-Y.; Ko, H.-M.; Cui, J.-F.; Chong, H.-C.; Leung, Y.-C.; Wong, M.-K. Cyclometalated gold(III) Complexes for Chemoselective Cysteine Modification via Ligand Controlled C-S Bond-Forming Reductive Elimination. *Chem. Commun. (Camb).* **2014**, *50*, 11899–11902.

(21)   Arora, N.; Leppla, S. H. Fusions of Anthrax Toxin Lethal Factor with Shiga Toxin and Diphtheria Toxin Enzymic Domains Are Toxic to Mammalian Cells. *Infect. Immun.* **1994**, *62*, 4955–4961.

(22)   Chari, R. V. J.; Miller, M. L.; Widdison, W. C. Antibody–Drug Conjugates: An Emerging

Concept in Cancer Therapy. *Angew. Chemie Int. Ed.* **2014**, *53*, 3796–3827.

(23)   Sun, M. M. C.; Beam, K. S.; Cerveny, C. G.; Hamblett, K. J.; Blackmore, R. S.; Torgov, M. Y.; Handley, F. G. M.; Ihle, N. C.; Senter, P. D.; Alley, S. C. Reduction–Alkylation Strategies for the Modification of Specific Monoclonal Antibody Disulfides. *Bioconjug. Chem.* **2005**, *16*, 1282–1290.

(24)   Bird, G. H.; Gavathiotis, E.; LaBelle, J. L.; Katz, S. G.; Walensky, L. D. Distinct BimBH3 (BimSAHB) Stapled Peptides for Structural and Cellular Studies. *ACS Chem. Biol.* **2014**, *9*, 831–837.

(25)   Verdine, G. L.; Hilinski, G. J. Stapled Peptides for Intracellular Drug Targets. *Methods Enzym.* **2012**, *503*, 3–33.

(26)   Lau, Y. H.; de Andrade, P.; Wu, Y.; Spring, D. R. Peptide Stapling Techniques Based on Different Macrocyclisation Chemistries. *Chem. Soc. Rev.* **2015**, *44*, 91–102.

(27)   Campagne, J.-M. Writing the Macrocycle Manual. *Nat. Chem. Biol.* **2014**, *10*, 693.

(28)   Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L. A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013**, *135*, 5946–5949.

(29)   Pentelute, B. L. *New Chemical Methods for the Synthesis of Proteins and Their Application to the Elucidation of Protein Structure by Racemic Protein Crystallography*; 2008.

(30)   Sheldrick, G. M.; IUCr. Phase Annealing in SHELX-90: Direct Methods for Larger Structures. *Acta Crystallogr. Sect. A Found. Crystallogr.* **1990**, *46*, 467–473.

(31)   Sheldrick, G. M.; IUCr; S., W. K.; J., D. E.; S., D. M.; P., P. J.; G., G. M.; E., M. P. R.; A., D. L. R. M.; M., S. G. A Short History of *SHELX*. *Acta Crystallogr. Sect. A Found. Crystallogr.* **2008**, *64*, 112–122.

(32)   Müller, P. Practical Suggestions for Better Crystal Structures. *Crystallogr. Rev.* **2009**, *15*, 57–83.

(33)   and, W. M. S.; DeShong, P. Preparation and Palladium-Catalyzed Cross-Coupling of Aryl Triethylammonium Bis(catechol) Silicates with Aryl Triflates. **2004**.

(34)   Dickinson, B. C.; Huynh, C.; Chang, C. J. A Palette of Fluorescent Probes with Varying

Emission Colors for Imaging Hydrogen Peroxide Signaling in Living Cells. *J. Am. Chem. Soc.* **2010**, *132*, 5906–5915.

(35) Lee, H. G.; Milner, P. J.; Buchwald, S. L. An Improved Catalyst System for the Pd-Catalyzed Fluorination of (Hetero)Aryl Triflates. *Org. Lett.* **2013**, *15*, 5602–5605.

(36) Drew, D.; Doyle, J. R.; Shaver, A. G. Cyclic Diolefin Complexes of Platinum and Palladium. In; John Wiley & Sons, Inc.; pp. 346–349.

(37) McAtee, J. R.; Martin, S. E. S.; Ahneman, D. T.; Johnson, K. A.; Watson, D. A. Preparation of Allyl and Vinyl Silanes by the Palladium-Catalyzed Silylation of Terminal Olefins: A Silyl-Heck Reaction. *Angew. Chemie Int. Ed.* **2012**, *51*, 3663–3667.

(38) Simon, M. D.; Heider, P. L.; Adamo, A.; Vinogradov, A. A.; Mong, S. K.; Li, X.; Berger, T.; Policarpo, R. L.; Zhang, C.; Zou, Y.; *et al.* Rapid Flow-Based Peptide Synthesis. *Chembiochem* **2014**, *15*, 713–720.

(39) Liao, X.; Rabideau, A. E.; Pentelute, B. L. Delivery of Antibody Mimics into Mammalian Cells via Anthrax Toxin Protective Antigen. *Chembiochem* **2014**, *15*, 2458–2466.

(40) Spangler, J. B.; Manzari, M. T.; Rosalia, E. K.; Chen, T. F.; Wittrup, K. D. Triepitopic Antibody Fusions Inhibit Cetuximab-Resistant BRAF and KRAS Mutant Tumors via EGFR Signal Repression. *J. Mol. Biol.* **2012**, *422*, 532–544.

(41) Rabideau, A. E.; Liao, X.; Pentelute, B. L. Delivery of Mirror Image Polypeptides into Cells. *Chem. Sci.* **2015**, *6*, 648–653.