Precision crosslinking and active-site modifications via site-specific protein engineering

Inaugural-Dissertation

Submitted in Partial Fulfilment of the Requirements for the Degree of

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Alec William Michels

Berlin, 2022

Doctoral studies were undertaken at the Leibniz Forschungsinstitut für Molekulare Pharmakologie from September 2017 to December 2022.

1st reviewer: Prof. Dr. Christian Hackenberger

2nd reviewer: Prof. Dr. Christian Freund

Date of Defense: 01.12.2022

Acknowledgements

This thesis represents the culmination of years of effort, both by myself and by the many, many people who have helped me along the way.

First, I would like to thank my supervisor Prof. Christian Hackenberger for the opportunity to work on these projects and for the support in achieving their successful conclusion. Your flexibility and guidance are directly responsible for this work, and I am grateful for the positive work environment you have created at the FMP which I benefited from immensely.

The entire Hackenberger group at the FMP provided constant and invaluable insight into the work I performed. Many also provided reagents and assistance in generating the peptides and proteins that I used in my experiments. Special thanks to Kristin, Ines, Jenny and Katta for going above and beyond with many hours of help and wonderful discussions.

My sister Audrey has been the greatest guidance and support a brother could ask for. I depend on you for your unique perspective and would be lost without it.

My parents Oren and Natalie, my mother Jacquie, my aunts Jody and Dina, by uncles Tom and Peter, my cousins Ruth, Sam and Lizzy, my inspirational grandfather Prof. Michels and of course all the rest of my large family located in America, have always been in my thoughts. Together you have provided the foundation for me to be successful in all that I do, and I think about you constantly with love and gratitude.

Justin Becker and the Hagelberg family welcomed me to Berlin many years ago and are responsible for giving me the confidence to start a life here. I feel lucky to be part of your family too.

My cat Bønne, ensconced even now on my lap, carried the weight of this thesis gracefully with her love and unparalleled fluff.

My Cal CoC friends Sean, Steph, Ripika, and Matt were invaluable for this work by introducing me to my love of chemistry. Our gin rummy games lasting late into the night in Kresge kept my spirits up throughout many challenges.

And finally, the contribution of my partner, Dr. Stine Vang Nielsen, simply cannot be conveyed with words. I am inspired by your knowledge and empathy and humbled by your trust. Your support meant and continues to mean everything to me.

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources or aids other than those indicated by me.

Intellectual property of other authors has been marked accordingly. I also declare that I have not applied for an examination procedure at any other institution and that I have not submitted the dissertation in this or any other form to any other faculty as a dissertation

Alec William Michels

Abstract

Protein engineering promises to enable rationally designed constructs composed of endogenous or enhanced protein components with properties far beyond the individual components themselves or the capabilities of raw chemical synthesis. Such constructs have the potential to revolutionize medicine, diagnostics, and fundamental research, but require additional sophisticated methods with high precision guided by accumulated knowledge in both chemistry and biology to yield homogenous products. This thesis therefore focuses on developing novel engineering strategies combining multiple existing biochemical methodologies that further push the boundaries of how chemical biologists can specifically manipulate complex protein targets.

First, a site-specific method for covalently and site-specifically attaching two proteins under biocompatible conditions was developed. The method combined the sortase-mediated installation of a peptide containing a cysteine-reactive functionality onto the N-terminus of an antigen binding protein followed by a subsequent chemoselective conjugation. Investigation of the promiscuity of sortase, a cysteine-dependent transpeptidase, for substrates containing cysteine-alkylating reagents suggested good tolerance of the sortase for such electrophiles. Multiple peptides each containing a sortase recognition sequence and a different cysteine-selective electrophile were therefore synthesized via solid-phase peptide synthesis (SPPS). Significant optimization of the sortase reaction was required to yield sufficient product for functional assays. Speeding up the sortase reaction using kinetically enhanced mutants was found to enhance the production of product material, likely by counteracting the concurrent alkylation of the sortase active site by the electrophilic peptide substrate. After successfully attaching cysteine-targeting moieties onto several antigen-binding protein, so-called “nanobodies” using sortase transpeptidation, site-specific conjugation to cysteine-containing proteins was optimized to generate useful yields of products with new functionalities. Conjugation was shown to be unsuccessful without prior purification of the sortase reaction, likely due to the presence of competing crosslinking partners and loss of the transpeptidation product over time in the presence of the sortase. Sortase-mediated electrophile installations were performed on nanobodies (with and without endogenous cysteines) and crosslinked to other nanobodies, GFP, and even intact IgG antibodies to yield bi-specific or mono-specific, fluorescent constructs combining the

properties of multiple intact proteins. Final conjugate yields were low after multiple purification steps. The function of both protein entities within all final constructs were confirmed using fluorescence microscopy.

In the second part, amber codon suppression was combined with the chemoselective Staudinger-phosphite reaction to enable the selective caging of nanobody active sites with a light-cleavable PEG group. Decaging was envisioned to progress via UV-induced photolysis of an ortho-nitrobenzyl polyethylene glycol (ONB-PEG) functionality, followed by hydrolysis to yield a final aniline. An azido-tyrosine was introduced at the active site of a GFP binding nanobody using amber suppression. To test the activity of the decaged product, TCEP reduction successfully converted the incorporated azide to the corresponding aniline, which was then subjected to a substrate binding assay. Unfortunately, the binding affinity of the decaged nanobody was only partially restored to that of the wild-type nanobody. The azide was subsequently functionalized with a light-cleavable ONB-PEG-phosphite via the Staudinger-phosphite reaction in aqueous buffer. According to a recent literature survey, this represents the first example of a Staudinger-phosphite reaction performed on a nanobody under biological conditions. Caging was shown to prevent GFP-nanobody interaction via a GFP binding assay utilizing the fluorescence-enhancing properties of the chosen nanobody. Cleavage of the ONB group was achieved using UV irradiation, and the resulting decaged construct was shown to possess partially restored GFP binding affinity.

In conclusion, two projects were pursued combining chemoselective chemistry with various site-specific protein modification strategies. The goal was to develop new methods benefiting from high selectivity, biocompatible conditions, and the potential to significantly alter or enhance the function of the targets. Despite challenges in identifying the conditions for optimal yield and purity, the projects successfully demonstrated the potential synergy of combining multiple protein engineering methodologies. These investigations support the claim that spatiotemporal control over protein activity and site-specific protein-protein conjugation can fundamentally enhance the usefulness of protein constructs and can be enabled in a straightforward manner using existing methods applied in tandem. Such results point towards future combinations of different protein engineering strategies that promise to carry the field of site-specific protein functionalization to new heights.

Zusammenfassung

Protein-Engineering verspricht, rational entworfene Konstrukte zu ermöglichen, die aus endogenen oder verbesserten Proteinkomponenten bestehen und die Eigenschaften aufweisen, die weit über die einzelnen Komponenten selbst oder die Möglichkeiten der reinen chemischen Synthese hinausgehen. Solche Konstrukte haben das Potenzial Medizin, Diagnostik und Grundlagenforschung zu revolutionieren, erfordern jedoch ausgefeilte Methoden mit hoher Präzision, um homogene Produkte zu generieren. Diese Dissertation konzentriert sich daher auf die Entwicklung neuartiger Strategien, die mehrere bestehende Methoden kombinieren, um die Grenzen der spezifischen Manipulation komplexer Proteinziele durch Proteiningenieure zu erweitern.

Zunächst wurde ein Verfahren zur kovalenten und ortsspezifischen Bindung zweier Proteine unter biokompatiblen Bedingungen entwickelt. Die Methode kombiniert die Sortase-vermittelte Installation eines Peptids, das eine cysteinreaktive Funktionalität enthält, mit einer anschließenden chemoselektiven Konjugation. Peptide, die mehrere cysteinselektive Elektrophile enthalten, wurden über Festphasen-Peptidsynthese (SPPS) synthetisiert und erfolgreich an mehrere antigenbindende Proteine gebunden. Die Untersuchung der Kompatibilität von Sortase, einer cysteinabhängigen Transpeptidase, mit Substraten die cysteinalkylierende Reagenzien enthalten ergab überraschende Ergebnisse, die auf eine gute Toleranz der Sortase für solche Substrate hindeuten. Dennoch war eine signifikante Optimierung der Sortase-Reaktion und der anschließenden Konjugation an Cystein erforderlich, um ausreichend Produkt für funktionelle Assays zu erhalten. Es wurde festgestellt, dass die Beschleunigung der Sortase-Reaktion unter Verwendung kinetisch verbesserter Mutanten die Ausbeuten erhöht, da wahrscheinlich der konkurrierenden Alkylierung des aktiven Zentrums der Sortase durch das elektrophile Peptidsubstrat entgegengewirkt wird. Es wurde auch gezeigt, dass die Vernetzung des Produkts der Sortase-Reaktion ohne vorherige Reinigung nicht erfolgreich war, wahrscheinlich aufgrund der Anwesenheit konkurrierender Vernetzungspartner und des damit verbundenen Verlusts des Transpeptidierungsprodukts in Gegenwart der Sortase. Sortase-vermittelte elektrophile Installationen wurden an exprimierten Nanokörpern (mit und ohne endogenen Cysteinen) durchgeführt, welche dann mit anderen Nanokörpern, GFP und sogar intakten IgG-Antikörpern vernetzt wurden, um bispezifische oder monospezifische, fluoreszierende

Konstrukte zu erhalten, die die Eigenschaften mehrerer intakter Proteine kombinieren. Die Ausbeuten waren nach mehreren Reinigungsschritten gering, aber die Funktion der individuellen Proteineinheiten aller dargestellten Konstrukte wurde durch Fluoreszenzmikroskopie bestätigt.

Als nächstes wurde die Amber-Codon-Unterdrückung mit der chemoselektiven Staudinger-Phosphit-Reaktion kombiniert, um das selektive Caging aktiver Zentren von Nanokörpern mit einer durch Licht spaltbaren Polyethylenglokol (PEG)-Gruppe zu ermöglichen. Es war vorgesehen, dass die Öffnung über eine UV-induzierte Photolyse des Orthonitrobenzyl (ONB)-PEG fortschreitet, gefolgt von einer Hydrolyse, um ein Anilin zu generieren. Um zu zeigen, dass damit ein funktioneller Nanokörper entstehen kann wurde zunächst in einem Modellsystem ein Azidotyrosin am aktiven Zentrum eines GFP-bindenden Nanokörpers unter Verwendung von Amber-Unterdrückung eingeführt. Durch direkte Reduktion wurde das entsprechende Anilin erfolgreich erzeugt, aber leider wurde damit nur Bruchteil der Bindungsaffinität des Nanokörper-Wildtyps erreicht. Trotzdem wurde der Nanokörper mit dem eingebauten Azidotyrosin anschließend auch mit einem lichtspaltbaren ONB-PEG-Phosphit über die Staudinger-Phosphit-Reaktion unter wässrigen Bedingungen funktionalisiert. Es wurde über einen GFP-Bindungsassay, unter Ausnutzung der fluoreszenzverstärkenden Eigenschaften des ausgewählten Nanokörpers, gezeigt, dass das Caging die GFP-Nanokörper-Wechselwirkung verhindert. Die Spaltung der ONB-Gruppe wurde dann unter Verwendung von UV-Bestrahlung erreicht, und es wurde gezeigt, dass das resultierende Konstrukt eine teilweise wiederhergestellte GFP-Bindungsaffinität besitzt.

Zusammenfassend wurden zwei Projekte verfolgt, die chemoselektive Reaktionen mit verschiedenen Proteinmodifikationsstrategien kombinierten, um neue Methoden zu entwickeln, die von hoher Selektivität, biokompatiblen Bedingungen und dem Potenzial profitieren, die Funktion der Ziele signifikant zu verändern oder zu verbessern. Trotz der aufgetretenen Herausforderungen bei Aufreingung und Ausbeute, demonstrierten die Projekte erfolgreich die potenzielle Synergie der Kombination mehrerer Protein-Engineering-Methoden. Solche Ergebnisse weisen auf zukünftige Kombinationen verschiedener Protein-Engineering-Strategien hin, die versprechen, das Gebiet der ortsspezifischen Proteinfunktionalisierung zu neuen Höhen zu führen.

Table of Contents

1. Introduction

1.1 Motivation

Human development can be measured by the sophistication of our tools. As our knowledge increases, so does our ability to create tools that better fulfill our needs. Natural proteins represent an opportune scaffold for tool development that has benefited from centuries of evolutionary improvement. For optimal usefulness, their naturally evolved abilities should ideally be molded towards a specific industrial or synthetic application. Mimicking evolution through non-rational improvements can be an option for guiding natural proteins towards becoming useful tools [1]. Some examples of this include functional screens involving libraries of millions of individual proteins subjected to ever more stringent selection criteria for a particular function, as well as directed evolution in which functions can be selected for in an iterative method similar to natural competition [2, 3]. These methods share a general strategy of manipulating protein platforms through gradual stochastic changes honed by some form of selection. Alternatively, one can employ rational design towards improving these tools for a specific function. However, this requires more advanced biochemical methods for enhancing their functions through specific chemical modifications. The advent of "new biotechnology" since the development of genetic manipulation has opened the door to modifying proteins to better serve as tools with incredible potential [4]. Better understanding has fueled a transition from screening-based stochastic modification and iteration strategies towards a more rational, design-based approach for improving protein species. As the link between structure and function becomes better understood, so too does the potential for rationally modifying and combining natural protein scaffolds to create powerful new tools.

Since the development of site-directed mutagenesis, significant work has been invested in developing chemical methods for enhancing the functionality of natural protein scaffolds based on a better understood link between structure and function [5]. Of particular interest is the work undertaken towards fusing different proteins to combine their properties, as well as the caging of protein active sites to confer spatiotemporal control over their function, much of which has been focused on antigen-targeting proteins. This work focuses on two different projects focusing on each of these strategies, respectively.

1.2 Antigen-targeting Proteins

Antigen-targeting proteins are a subject of intense focus because of their incredible natural specificity for certain targets. This makes them well suited for the diagnosis of disease, as well as the delivery of drugs to their intended target [6].

1.2.1 Antibodies

Antibodies are proteins that work by binding to their antigen target and then to another antigen target to form "coagulation" clumps [7]. IgG Antibodies contain a heavy chain and a light chain. Each IgG antibody has two heavy chains and two light chains, and each heavy chain and light chain has a variable domain and constant domains, giving the full antibody a constant domain and a variable domain. The constant domain (FC domain) is mostly conserved between antibodies and generally does not play a role in substrate recognition or binding, while the variable domain is responsible for substrate recognition and binding. Constant domains provide structure to an antibody, while variable domains determine substrate complementarity [8]. Variable domains of the heavy chains and light chains are composed of a “framework” domain and a “complementary determine region” (CDR). The framework domain is mostly constant between antibodies, while the CDR, which determines antigen binding, is hypervariable between antibodies.

Since each IgG antibody contains a light chain and a heavy chain variable domain (Vh and Vl), IgG antibodies bind to substrates in a "bi-dentate" mechanism. Interactions between antibody Vh/Vl domains and their targets occur through Van der Waals forces, hydrophobic interactions, and other non-covalent interactions. These are weak interactions that are reversible. As a result, inter-chain disulfides can be reduced and targeted for modification without compromising the overall structure and function of the antibody. This can yield many useful antibody conjugates, including bi-specific antibodies, which will be discussed later [9].

1.2.2 Nanobodies

Separating antibodies into their component domains represents an opportunity for generating molecules with antibody-like characteristics possessing stronger pharmacokinetic properties and accessibility. When a single Vh domain is expressed, the hydrophobic amino acids responsible for these interactions are solvent-exposed, presenting difficulties when single antibody domains are subjected to purification. The hydrophobic amino acids usually

engaged in hydrophobic interactions between the heavy and light chains instead aggregate and reduce solubility of the single domain. Single Vh domains from antibodies lack the substrate binding capacity of the VL domain, and thus display reduced substrate binding [10].

However, a class of IgG antibodies from *Camelidae* was isolated lacking light chains, possessing a single antigen binding domain (a VhH domain) [11]. The small size of this class of antibodies led to their pseudonym "nanobodies". Nanobodies differs from single Vh domain antibodies in two ways: first, the complementary determining region (CDR) of the VhH domain is larger, increasing the binding capacity of the nanobody and allowing them to bind their antigen in a convex paratope. Second, the hydrophobic amino acids present in IgG derived Vh domains are replaced with hydrophilic amino acids, improving solubility and stability of the nanobodies [10]. Of particular importance to this thesis are two specific nanobodies, known to not only bind specifically to GFP, but to also modulate the fluorescence of the target upon binding [12].

1.2.3 GFP-Binding Enhancer Nanobody

The first of the nanobodies to be introduced is the GFP-binding enhancer nanobody. Enhancer binding fixes Arg168 of GFP near to proton-donor His148. This conformation is stabilized by direct binding with Enhancer residues Tyr37 and Glu101. This rearrangement fixes His148 near the GFP chromophore hydroxyl group, which likely facilitates deprotonation of the fluorophore hydroxyl, stabilizing the phenolate anion of the chromophore. The stabilized phenolate thereby enhances absorption and fluorescence intensity. Interestingly, the enhanced fluorescence of eGFP is due to an S65T mutation that reduces the distance between His148 and the chromophore hydroxyl group, similar to the mechanism by which the nanobody enhances fluorescence [13]. Anti-GFP nanobodies were usually expressed as a genetic fusion to the PelB leader sequence. The PelB leader sequence is an amino acid sequence that leads to secretion of a translated protein into the periplasmic space. This enhances folding and disulfide bond formation which occurs in the periplasm. The PelB amino acid sequence is then cleaved in the periplasm, so the PelB sequence does not affect nanobody activity [14].

1.2.4 GFP-Binding Minimizer Nanobody

Minimizer binding occurs in a sidewise manner and is focused on interactions between GFP and CDR3. The minimizer targets the rigid and flat side of GFP and interacts with a smaller

GFP surface area than the enhancer, which supports the data suggesting that enhancer can displace minimizer from GFP but not the other way around.

For the minimizer, Arg168 of GFP interacts with Leu100 of the minimizer, which tilts Arg168 away from His148. Tilting away reduces the electrostatic forces felt by His148, causing it to be pulled away from the hydroxyl group of the chromophore (3.4 angstroms, compared to 3.5 in WT GFP). At this distance it is too far to stabilize the chromophore phenolate anion. The minimizer induced GFP conformation therefore likely stabilizes the neutral form of the chromophore phenol, which inhibits chromophore absorption at 470nm and thereby inhibits fluorescence.

The minimizer and enhancer nanobodies demonstrated the potential for not only the targeting, but also the biophysical manipulation of antigens by nanobodies *in vitro* and *in cellulo* [12]. Such characteristics of these nanobodies (i.e. target specificity and modulation) could be useful if combined with other constructs in a manner that preserves their function. In particular, modulating the properties of the substrate gives a straightforward experimental readout for binding. One can simply measure the change in fluorescence upon mixture of the nanobody and substrate to determine if binding is affected by any modifications made to the nanobody. Enabling modifications that broaden the potential of such functional proteins requires the application of protein engineering.

1.3 Protein engineering

The potential for proteins as functional agents for diagnostics and therapeutics is consistently being proven across a wide range of applications [15]. Early methods for improving such agents focused mainly on stochastic methods, i.e. random mutagenesis to initiate changes, followed by screening to identify improvements [3]. This method was necessary due to a missing link in understanding between structure/sequence and function in proteins. Understanding structure-function relationships in proteins has always had the ultimate goal of enabling protein engineers to design new proteins in the same way that architects can design new buildings: rationally, and with a specific purpose in mind [16].

Methods for modifying proteins face several challenges unique to their chemical properties. Relative to small molecules, polypeptides are larger and more chemically intricate. Indeed, the large size and complex structure of proteins can reduce the specificity of chemical

modification strategies while reducing their efficacy when applied as medical tools in living organisms [17, 18]. Complex structures represent a double edged sword, representing both the means by which proteins accomplish their function as well as a weakness that can be exploited by thermodynamic or entropic degradation [19]. To improve the structure-based activity of protein tools and enable new payload modification strategies, robust yet targeted engineering strategies are required that are compatible with the conditions favored by proteins and tolerate their complex chemical environment. Chemoselective chemistry can enable straightforward site-specific modifications, particularly when relevant protein targets contain a single accessible chemical handle. For example, the FDA-approved PEGylated protein interferon-alpha2a (PEGASYS) used to treat hepatitis C contains a single accessible lysine side chain that can be modified easily with a 40 kDa PEG chain to improve its stability over time in storage [20]. While chemoselective reactions targeting amino acid side chains or protein termini are biocompatible and often high-yielding, many substrates contain multiple accessible chemical targets. Unfortunately, chemoselective reactions are incapable of distinguishing between individual reactive groups, resulting in a broad range of substrate modifications that are difficult to characterize [21]. Chemoselective bioorthogonal reactions targeting specific noncanonical amino acids within a protein do overcome the challenge of product uniformity, but suffer from complex reaction schemes and expensive reagents like unnatural amino acids [22]. In comparison, chemoenzymatic reactions have the potential to site-specifically modify even complex proteins with a broad range of substrates in good yields and with fast reaction rates, while requiring only the introduction of an amino acid recognition motif, or even no recombinant engineering at all [23, 24]. Protein engineers therefore benefit from knowledge of all these methods and more, in addition to the ability to select the right method for the desired application. This section will provide an overview on the subset of protein engineering methods dedicated to modifying proteins to fulfill a particular design goal.

1.3.1 Site-directed mutagenesis

Ever since site-directed mutagenesis (SDM) enabled the specific replacement of individual residues (and eventually whole domains), engineers have inched closer to a sufficient understanding of structure and function to enable such rational design. As the first technology that enabled reliable deletion, insertion or substitution of specific residues in expressed

proteins, SDM enabled a rush of studies to identify the role in catalysis of individual residues [5]. For example, mutation of multiple residues in the bi-functional UDP-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase, an enzyme involved in sialic acid biosynthesis, identified several residues in spatially separated positions that were essential for both kinase and epimerase activity. This provided additional evidence that a single enzyme accomplished both functions using multiple active sites [25].

Early attempts were made towards applying SDM for rationally adjusting the specificity of antigen binding proteins. As an example of the potency of even single residue changes, the group of Wolfgang Höhne in Berlin were able to change the specificity of the anti-p24 antibody specific for a protein in the capsid of the HIV-1 virus from its native epitope [26]. The authors combined knowledge from early crystal structures of the epitope-antibody complex with sequencing data of the antibody fragment variable region to pick residues likely involved in unspecific interactions with non-homologous peptide variants and tested their hypotheses with SDM. Single residue substitutions were able to lower the affinity for the native epitope by a factor of 250.

SDM combined site-specificity with substrate control, but scientists were initially limited by a repertoire consisting only of the twenty canonical amino acids. Introducing non-canonical amino acids would later open the door to even more exciting applications but would require modifications both to the DNA code of the substrate as well as the tRNA aminoacylation machinery of the expression host.

1.3.2 Amber Suppression

When sequencing the genomes of extremophile archaea bacteria family *Methanosarcinacea* found in vents near the bottom of the ocean, several in-frame amber codons were found within genes that did not stop translation of the corresponding protein [27]. Eventually it was discovered that this micro-organism encoded a novel amino acid, pyrrolysine, at such positions [28]. Research starting from these observations lead to the discovery of a promiscuous tRNA synthetase capable of acylating not just pyrrolysine, but even non-natural lysine derivatives, thereby enabling genetically encoded non-natural functionalities [29]. These discoveries laid the foundation for a new development in protein engineering combining non-natural chemical functionalities with chemoselective chemistry, all built on the foundation of mutagenesis and genetic code expansion. Amber suppression could later

be expanded to specifically incorporate tyrosine derivatives using a tyrosyl-synthetase modified using site-directed mutagenesis [30]. Derivatives containing a wide variety of functional handles on the tyrosine side chain phenolic ring could be incorporated, some of which will be discussed later (Section 1.5). Together these examples demonstrate that amber suppression can be used to incorporate bioorthogonal handles at and only at specific positions within natural proteins, enabling modifications at these positions using chemoselective chemistry. Amber suppression can be advantageous for certain applications compared to alternative orthogonal reporter incorporation strategies like auxotrophic expression. The latter method relies on modified expression strains that lack the biosynthetic machinery for producing certain amino acids, enabling artificially added alternative to be incorporated in the place of the naturally encoded residue when that residue is absent in the expression culture [31]. Auxotrophic expression will lead to reporter incorporation at every position normally occupied by the deprived residue, which limits the options for restricting incorporation to a specific position. Amber suppression, by comparison, will achieve incorporation only at positions engineered to contain the amber stop codon. Nonetheless, auxotrophic expression can be useful when multiple handles are to be incorporated in a defined structure, such as when expression virus-like particles composed of repeating motifs of the same residues [32].

1.3.3 Chemoselective protein modifications

Amber suppression enabled the genetic encoding of non-natural functional groups into native proteins. Chemistry then supplied the tools to specifically target these functionalities via so-called “chemoselective chemistry” [33]. Such methods are characterized as high yielding and highly selective for a particular functionality. They should also work under mild conditions and without the need for protecting groups, to avoid unwanted modifications to the target biomolecule [34]. These chemical methods enable the specific attachment of small and large cargo at and only at a particular functional group, with minimal unspecific labelling despite the complex chemical environment found in most proteins. All that is combining the right chemistry with a properly incorporated bioorthogonal handle.

1.3.4 Targeting azides in proteins

Azides were added to the repertoire of non-natural genetically encoded reactive amino acids by Jason Chin at the turn of the millennium, although they had been previously incorporated

on glycans by Bertozzi and coworkers [35, 36]. Azides can function as bioorthogonal tags, modifiable via chemistry that can target it specifically while not participating in any natural processes [37]. In other words, the azide tag can be incorporated into a protein without affecting its natural function, before being used as a specific attachment point for cargoes ranging from small fluorophores to entire other proteins.

One of the most prolific methods for targeting azides in proteins is copper catalyzed azide-alkyne cycloaddition (CuAAC), also known as the “click” reaction [38]. The reaction is fast and irreversible, and happens between an azide, typically incorporated into a protein or peptide, and an alkyne-cargo to yield a triazole. Advantages of the method include the formation of an irreversible bond, and the use of small active groups that can be tolerated well by bulky protein secondary structures. The main disadvantage of “click” chemistry is the requirement for a copper (Cu) catalyst. Cu(I) oxidizes in water, which can cause oxidative damage to cells, limiting *in vivo* and *in cellulo* applications [39].

Cu toxicity can be circumvented by encouraging the reaction without a catalyst using ring strain as the driving force. Strain-promoted 1,3 dipolar cycloaddition between azides and strained alkynes such as dibenzocyclooctyne (DBCO) do not require a catalyst but benefit from irreversibility and short reaction times. DBCO reagents including fluorophores are often used to label living cells due to the high biocompatibility of the method [40]. Strain-promoted reagents suffer from the disadvantage of steric bulkiness and hydrophobicity. This requires the use of highly polar fluorophores to compensate, such as sulfated Cy5, which can cause solubility issues [41]. The reaction between azides strained alkynes is generally considered to be chemoselective, but side reactions between strained alkynes at high concentrations of thiols have been reported ([42]).

The Staudinger ligation can also enable bioconjugation to proteins without the need for a toxic catalyst. This reaction between an azide an aryl phosphine containing an electron withdrawing group results in an amide bond linking the two species and was popularized as a bioconjugation tool by Carolyn Bertozzi [43]. Mechanistically the reaction begins with an attack by the electron rich aryl phosphorous (III) species to the electron deficient azide before undergoing a rearrangement driven by the formation of molecular nitrogen to yield a P-N bonded intermediate. The phosphine species is synthesized with an electrophilic ester that will then trap the nitrogen of the intermediate to form an amide bond while reforming the

aryl phosphine. This reaction benefits from its fast speed and irreversible bond formation. It is also high yielding, can be done in water and in presence of other functional groups. Bioorthogonality is essentially guaranteed, since phosphorous(III) species are absent in living organisms [44]. However, the reaction can be hindered by P-N bond hydrolysis prior to amide bond formation, as well as challenging synthesis schemes for the aryl phosphine reagents.

Modifying the Staudinger ligation slightly to react phosphites instead of phosphines with aryl azides is termed the Staudinger-phosphite reaction, which forms stable phosphoramidate bonds without the need for an electrophilic trap [45]. This chemoselective azide functionalization tool has been used for site-specific glycosylation and PEGylation of azide-containing proteins and peptides, among other applications [46] [47, 48].

Finally, phosphonites can also be applied to Staudinger-type modifications of azido peptides and proteins [49]. The so-called Staudinger-phosphonite reaction enables the functionalization of azides in aqueous systems without the need for copper catalysis. When combined with CuAAC, alkyne phosphonites can even enable sequential azide couplings, showing the compatibility of the azide-targeting method with other similar strategies [50].

1.3.5 Targeting cysteines in proteins

Cysteine is one of the least frequently encoded amino acid residues at an average abundance of only 1.9%, while simultaneously possessing one of the most nucleophilic side chains [51, 52]. Its pKa of ~8.5 enables only minor changes in local protein environments to generate reactive sulfur species at physiological pH, and its intrinsically nucleophilic side chain is unique among amino acids. This makes it an excellent target for chemoselective chemistry. Drawbacks do exist to targeting cysteine, including its tendency to form unreactive disulfide bonds with other cysteine residues, and its tendency to undergo posttranslational oxidation, forming stable yet inert oxoforms [53]. These drawbacks have nevertheless failed to prevent many exciting recent developments in the field of cysteine-targeting chemistry [54].

Of particular importance as inspiration for this thesis was a recent development by our group incorporating ethynyl groups onto phosphonites to form an electron-deficient alkyne phosphonamidate following the Staudinger-phosphonite reaction [55]. Built off our developments in azide targeting reactions using the Staudinger-phosphonite reaction (discussed above), ethynyl phosphonite workflows start with the chemoselective introduction

of an ethynyl phosphonamidate at an azide-containing protein residue, followed by a subsequent attack by a cysteine residue at the resulting electron deficient alkyne. The cysteine conjugation proceeds under aqueous conditions, making it biocompatible, and showed excellent selectivity towards cysteine residues when applied to the bioconjugation of peptides, fluorophores, and probes to proteins. Vinylphosphonites can also be converted into cysteine-reactive vinyl phosphonamidates in an analogous manner, as can vinylphosphonothiolates, which will be discussed later in the context of protein-protein crosslinking [56, 57].

Faster reactions like the Michael addition can also be used to modify cysteine residues [58]. One thiol-targeting Michael acceptor is the maleimide, prized for its selectivity under aqueous conditions and rapid kinetics. Maleimides are often employed as probes for detecting cysteine-dependent oxidative PTMs [53]. They have also been used for fluorescent labelling, drug conjugation, peptide cyclization and hydrogel functionalization [33, 59]. While thiol-maleimide adducts are relatively stable, they are known undergo hydrolysis and unwanted thiol exchanges with other residues are free thiols like glutathione. They also occasionally undergo side reactions with amines, hindering their selectivity. Nonetheless they are a well-utilized component of the thiol modification toolbox and continue to be developed. Newer variants like the dibromo maleimide have been used, for example, as disulfide bridging reagents capable of maintaining protein structures under reducing conditions by preventing unwanted folding or disulfide bridge formation from free thiols [60]. Hydrolysis of the thiol conjugate has been addressed using self-hydrolyzing maleimides incorporating a basic amino group on a linker adjacent to the maleimide. This derivative undergoes hydrolysis rapidly at neutral pH but does not undergo a subsequent elimination reaction, preventing loss of the conjugate [61].

1.3.6 Targeting protein *N*-termini

Targeting protein *N*-termini can be invaluable, especially when attempting not to interfere with endogenous function or activity. One of the most common chemical strategies is to employ *N*-hydroxysuccinimde (NHS) esters, which react with primary amines to form an amide bond, but lack total site-specificity given the frequent prevalence of competing amines during labelling [52]. The *N*-terminal amine can be targeted by virtue of its adjacent amide bond, which lowers its pKa (6.0-8.0) relative to that of the lysine side chain amine (~10.5) [62].

This makes the terminal amine the primary reactive amine at neutral pH, enabling selective targeting [63]. For example, selective imine formation with an aldehyde followed by reduction at slightly acidic pH has been used to install *N*-terminal PEG-groups onto protein therapeutics [64]. Functional proteins like insulin have also been modified at their *N*-termini using alkyne-functionalized ketenes, enabling further functionalization via copper-catalyzed click reactions [65]. Chemoenzymatic methods, alternatively, offer total specificity at the *N*-terminus so long as the right enzyme is chosen.

1.3.7 Chemoenzymatic protein modifications

Enzymatic protein modification strategies benefit from inherent biocompatibility and straightforward reaction conditions, while still retaining high specificity and degrees of control. Specificity is conferred by the enzyme, with its unique structure determining the position at which it will bind to and catalyze modification reactions [66]. While not perfect, enzymes are recognized as being among the most site-specific protein modification methods available today [67, 68]. Such advantages enable chemoenzymatic modifications to both enhance the stability of and add new capabilities to functional proteins (Figure 1). For these reasons, chemoenzymatic protein modifications continue to be an active field of research [10].

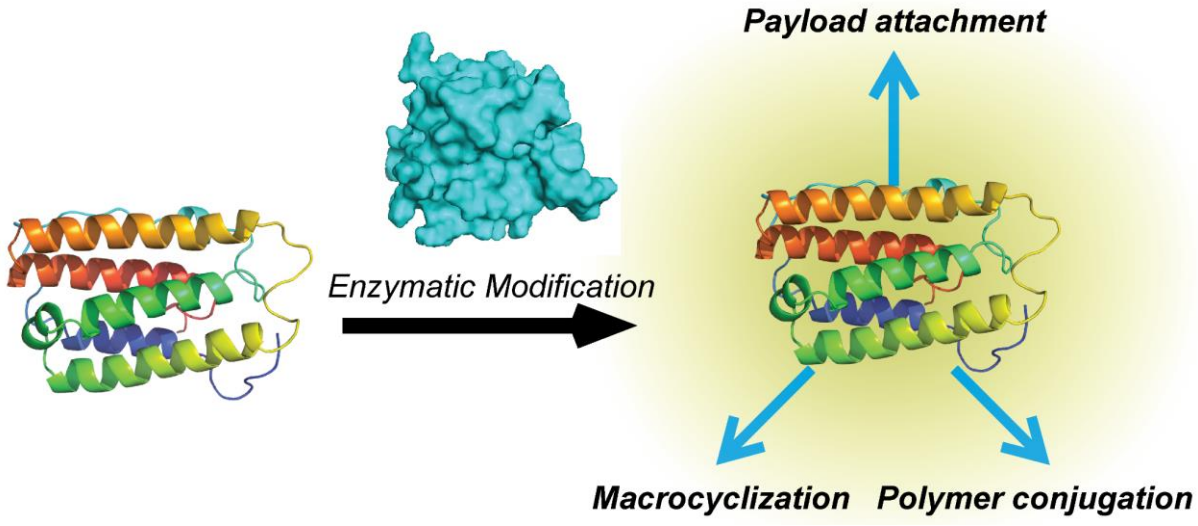


Figure 1: Chemoenzymatic engineering methods can stabilize and improve the capabilities of functional proteins.

Creative applications of this technology span from payload attachment to crosslinking and can have profound effects on the function of the modified protein. Payload attachment can be used to generate protein-drug conjugates, antigen presenters, or biological imaging

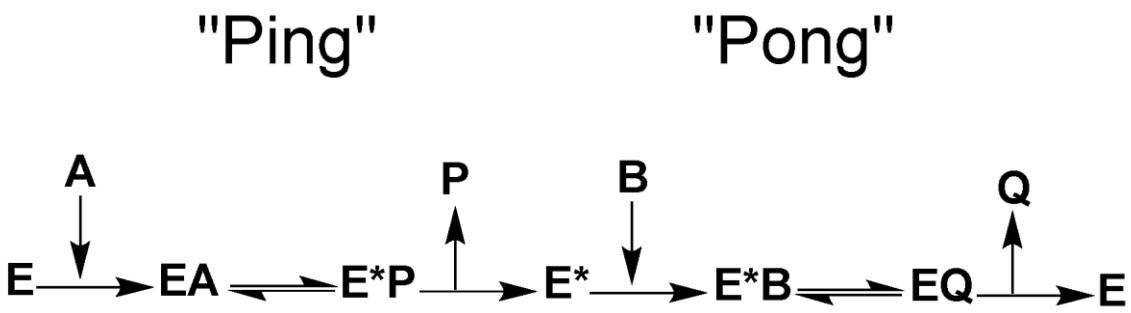
agents, among many other applications [67]. Beyond attaching drug payloads to protein scaffolds, chemoenzymatic reactions can also be used to improve the stability and efficacy of protein and peptide therapeutics. One method for achieving this is through enzyme-catalyzed cyclization, for instance by generating a covalent linkage between the *N* and *C* termini. The resulting circular protein is more resilient to thermal instability and recognized less frequently by proteasomal degradation pathways [69]. Chemoenzymatic methods enable such modifications with good specificity, all while utilizing biocompatible conditions, making them ideal protein modification tools [70].

1.3.8 Sortase

By far the most prolific method for this application is the sortase enzyme, which catalyzes the transpeptidation of peptide substrates containing a short recognition motif (LPXTG) onto the *N*-termini of proteins of peptides, most often preceding a poly-glycine motif [71]. Originally purified and characterized in 1999, sortase was first known only as a bacterial protein involved in anchoring proteins to the surface of *Staphylococcal* bacteria [72]. Its role in enabling host recognition and infection by modifying the bacterial cell surface made it an ideal drug target [73], and indeed several small molecule inhibitors have been discovered [74]. A year after its discovery it was applied for the *in vitro* transpeptidation of synthetic peptides [75], and shortly after this its broad applicability to proteins and peptides was established [76].

1.3.8.1 Sortase Enzyme Kinetics

Enzymes that catalyze reactions involving two or more substrates can proceed through sequential or non-sequential mechanisms. The non-sequential mechanism, such as that employed by the Sortase, differs from the sequential mechanism in that it does not require both substrates to bind to the enzyme prior to the reaction for successful product formation. This non-sequential mechanism is often referred to as a “ping-pong” or double-displacement mechanism, because the enzyme bounces back and forth between a standard and intermediate state as it associates and dissociates with the first of its substrates (Scheme 1) [77]. The ping pong mechanism has two distinct reactions with two distinct substrates (A, B). It is defined by the hallmark feature that the first reaction and the release of its product (P) occur before the second substrate binds and the second reaction begins to yield its product (Q).



Scheme 1: "Ping-Pong" reaction schematic.

In the case of the sortase, the first substrate to bind is the LPXTG-containing species, leading to the formation of an acyl-enzyme intermediate and eventually the release of the first product, containing the fragment downstream of the threonine residue [78, 79]. Following this first reaction the second substrate binds (containing usually a glycine or alanine residue at the *N*-terminus, depending on the sortase variant) and intercepts the acyl-enzyme intermediate, leading to release of the final transpeptidation product [80]. The rate-limiting catalysis step is the formation of the acyl-enzyme intermediate via attack of the active-site cysteine between the threonine and glycine residues of the recognition motif [81].

1.3.8.2 Advantages of sortase

Sortase offers several advantages as a tool for modifying proteins. Sortase offers high selectivity for its recognition sequence, enabling highly selective modifications [82]. Fortunately, sortases from different micro-organisms, or variants created using directed evolution, can recognize different recognition sequences or different nucleophiles [83, 84]. This can enable different sortase enzymes to be employed in parallel for the site-specific assembly of multiple peptide or protein fragments [85]. A recent method has been introduced that enables multi-fragment assembly without the need for multiple sortase variants by taking advantage of years of studies into the sortase active site structure. This strategy utilizes so-called ligation-site switching, in which a sortase recognition site is chemically converted from an “on” position to an “off” position in a sequential manner to enable specific ligations, one after the other [86]. The method relies on the precise nature of the substrates tolerated by the sortase, both in the chemical nature of P4 residue (leucine) of the recognition sequence AND the nucleophilic residue attacking the acyl-enzyme intermediate. Such specificity meant that Schwarzer et al. could apply chemoselective chemistry to protect and deprotect each

motif separately to achieve four-fragment assemblies without significantly affecting their target and using only standard SPPS protecting groups.

The sortase enzyme is also amenable to directed evolution. Scientists have been able to raise calcium-independent variants, widening the range of buffers applicable to the method [87]. Others have applied yeast display to generate 140-fold kinetically enhanced variants [88]. Yeast display can even direct sortase to drastically prefer sequences found in endogenous proteins, as has been done with the amyloid-β protein, proving that tailored sortase variants can be evolved for specific applications [89]. It has even been shown that lysine residues and other amines participate in sortase transpeptidation, expanding the potential application of the enzyme greatly [90, 91]. Simplifying the generation of sortase nucleophiles can also be done by employing the TEV protease, which generates a terminal glycine residue following cleavage, thereby converting cleaved species into sortase substrates [92].

1.3.8.3 Disadvantages of sortase

First and foremost among the disadvantages of sortase is the comparatively low catalytic efficiency, which is primarily the result of a weak binding affinity for the LPXTG recognition motif [93]. K_m values for the LPXTG motif during acylation are reported to range from 141uM to 8.7mM [77, 84, 94]. This is a particular issue for sortase applications targeting LPXTG-containing protein *C*-termini, since the protein is typically a precious reagent and therefore present in low abundance [95]. Utilizing *N*-terminal amines as nucleophiles in the transpeptidation reaction and LPXTG-peptides as electrophiles can circumvent this issue since the peptide can be easily synthesized and added in excess, but this can bring other issues including high DMSO concentrations and side reactions if the peptides contain reactive handles [96].

Compared to traditional small molecule crosslinkers such as the NHS-maleimide SMCC, the crosslinking motif generated by sortase is significantly larger, containing four amino acid residues in addition to the cysteine-targeting component. Like the maleimide, sortase transpeptidation has also been shown to be reversible, due to the formation of a new recognition motif in the product. Prolonged incubation with the sortase can therefore lead to poor yields or even the entire loss of the product [71]. Fortunately, purification strategies utilizing *C*-terminal Histidine tags cleaved during the transpeptidation reaction enable

purification of the sortase and peptide fragments from the reaction, thereby mitigating this issue [97].

1.3.8.4 Sortase applications

Straightforward and well-defined reaction conditions enable sortase to be applied in combination with other protein modification workflows to enable a wide variety of modifications [98]. POEGMA was polymerized in situ on the initiator moiety (Br) installed site-specifically on the *C*-terminus of the diabetes peptide drug exendin-4 (Figure 2A). Application of the same strategy to interferon-alpha by Liu et al. [12] followed by polymerization-induced self-assembly enabled the formation of polymer-conjugated interferon-alpha micelles (Figure 2B). Micelle formation resulted in therapeutics with enhanced efficacy, demonstrating over 20-fold higher *in vitro* bioactivity and an almost 2-fold higher *in vivo* half-life in mice likely due to its large size preventing renal clearance. Essential to this strategy was the enzyme-enabled site-specific and homogenous incorporation of the polymer initiator to the *C*-terminus of interferon-alpha, which prevented the reduction of bioactivity often observed with nonspecific polymer conjugation to protein therapeutics while still accessing the improved pharmacokinetics of polymerization-induced self-assembly.

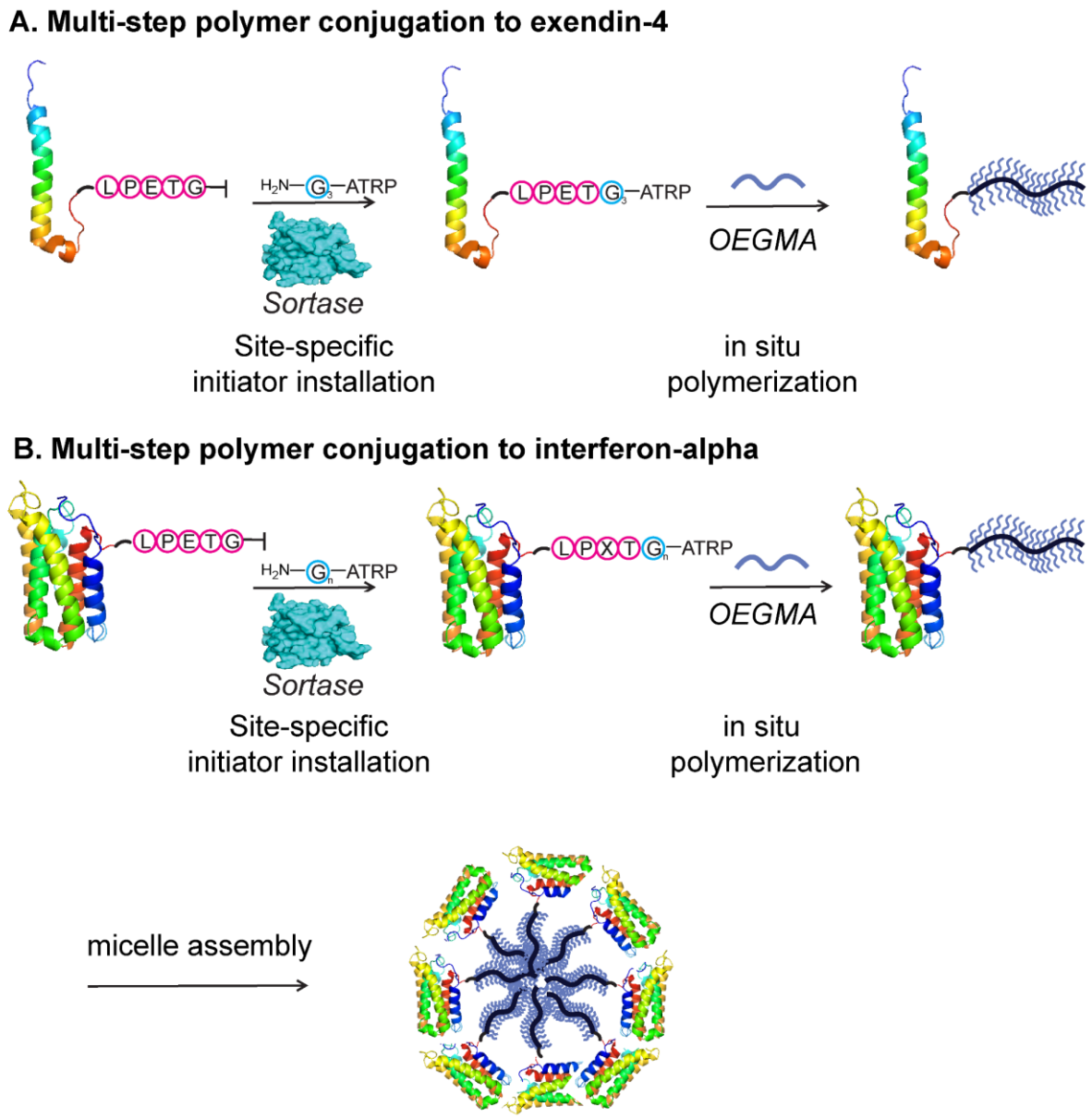


Figure 2: Sortase enables site-specific installation of reactive polymerization initiators.

A particularly promising strategy reported by Hou et al. [99] combines sortase-enabled macrocyclization with cysteine-selective native chemical ligation (NCL) (Figure 3). In this two-step protocol, interferon was recombinantly expressed with a *C*-terminal sortase recognition motif and an *N*-terminal cysteine residue before reacting with a polyglycine peptide functionalized with twenty PEG3-Glu residues and a single phenyl thioester for NCL to cysteine. This first chemoselective conjugation step between the engineered cysteine of interferon and the peptidic thioester occurred over several hours at pH 7.4 in PBS to yield selective ligation of the polyglycine-PEG peptide to the interferon *N*-terminus. Following this

conjugation, a sortase-catalyzed intramolecular cyclization was performed in 30 minutes with catalytic amounts of sortase to yield a macrocyclic, PEGylated interferon variant. Multiple purification steps using NiNTA and SEC chromatography were required to yield the pure circular product, suggesting incomplete conversion during the cyclization reaction. However, subsequent characterization of the cyclized, PEGylated construct in direct comparison with just-PEGylated variant showed improved protease resilience and thermal stability *in vitro*, as well as improved circulation half-life and tumor retention in an *in vivo* mouse model. Altogether, the authors demonstrated significant *in vivo* biomedical advantages for chemoenzymatically macrocyclized protein constructs using a method requiring minimal engineering of the protein substrates.

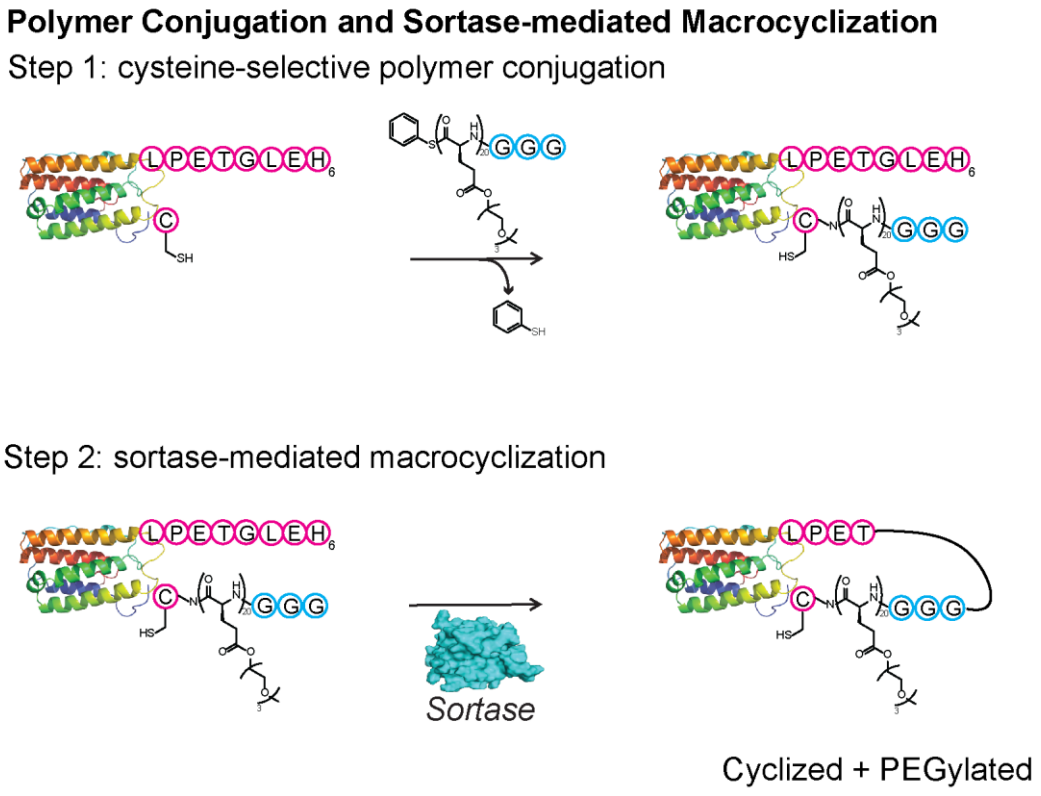


Figure 3: Combining sortase modifications with chemoselective chemistry [99].

Popp et al. were able to PEGylate and cyclize interferon using two sequential sortase reactions by taking advantage of the unique nucleophiles accepted by the different sortase variants [100]. An amino-oxy-containing alanine peptide with a sortase recognition motif was first selectively ligated to the cytokine *C*-terminus using a sortase variant specific for alanine nucleophiles. Macrocyclization was then achieved using a standard glycine-specific Sortase variant, followed by a chemoselective oxime ligation with an aldehyde-modified PEG moiety

to attach the polymer (Figure 4). PEGylation and cyclization together proved to have an even greater cytokine-stabilizing effect than circularization alone or PEGylation alone, with the construct demonstrating both resilience to boiling and significantly improved circulatory half-life in mice compared to the linear, non-PEGylated variant. This example demonstrates not only the broad applicability of sortase, but also its compatibility with other modification paradigms including chemoselective polymer conjugation.

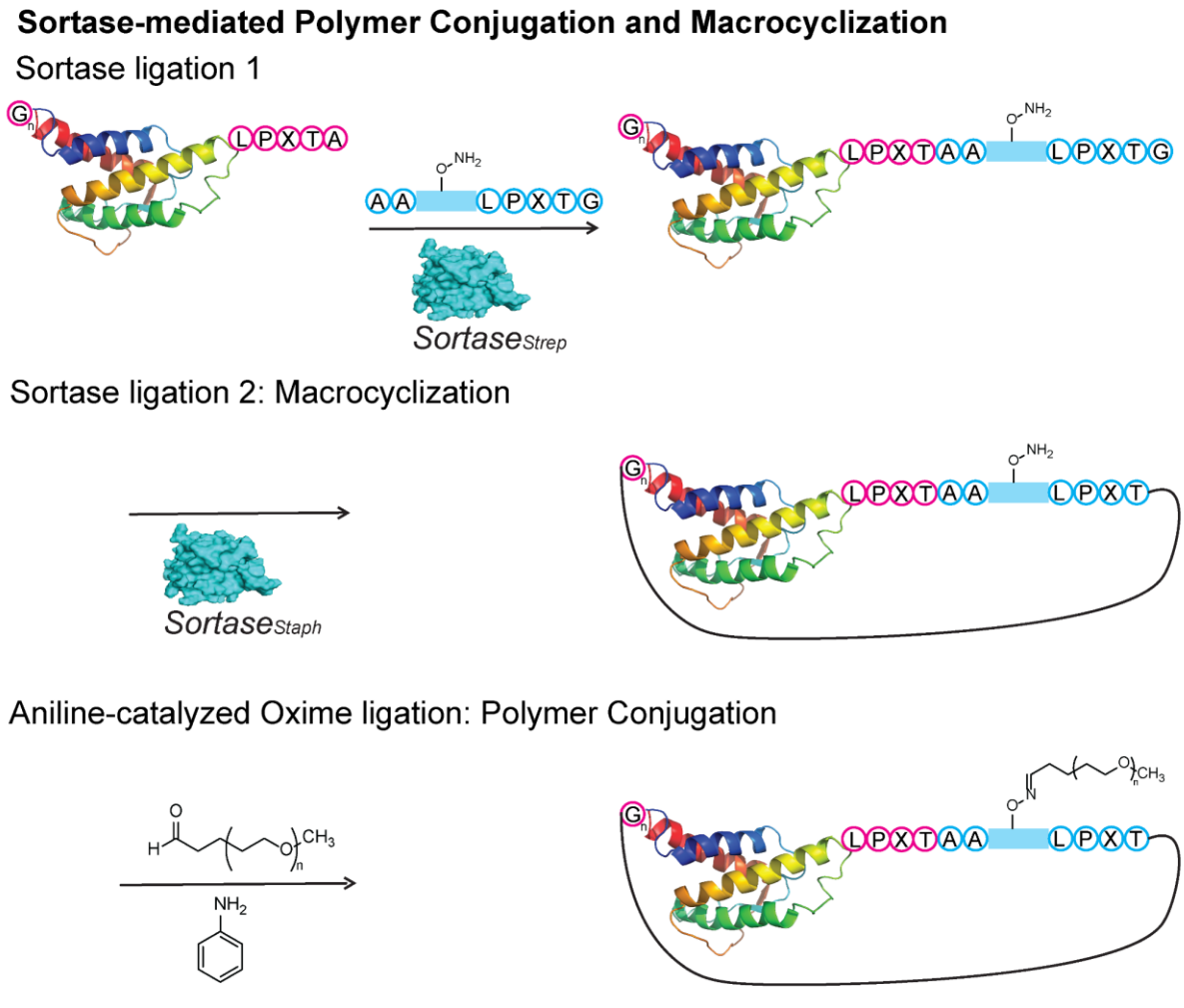


Figure 4: Dual sortase modification of a single protein substrate [100].

1.4 Protein-protein conjugation

Protein-protein conjugation is an established strategy for generating novel protein species by attaching individual proteins to one another (Figure 5). In this way, protein expression is decoupled from protein conjugation, giving scientists the power of modularity when imagining novel multi-protein constructs so long as an appropriate conjugation strategy is chosen for the individual species being linked. Choosing an appropriate strategy is dependent

upon the characteristics of the target proteins. Successful strategies will take advantage of the presence of certain inherent functionalities, such as accessible termini, nucleophilic residues, or unique peptide sequences recognized by specific enzymes. The chosen method will also depend upon the degree of specificity required for the desired application [101].

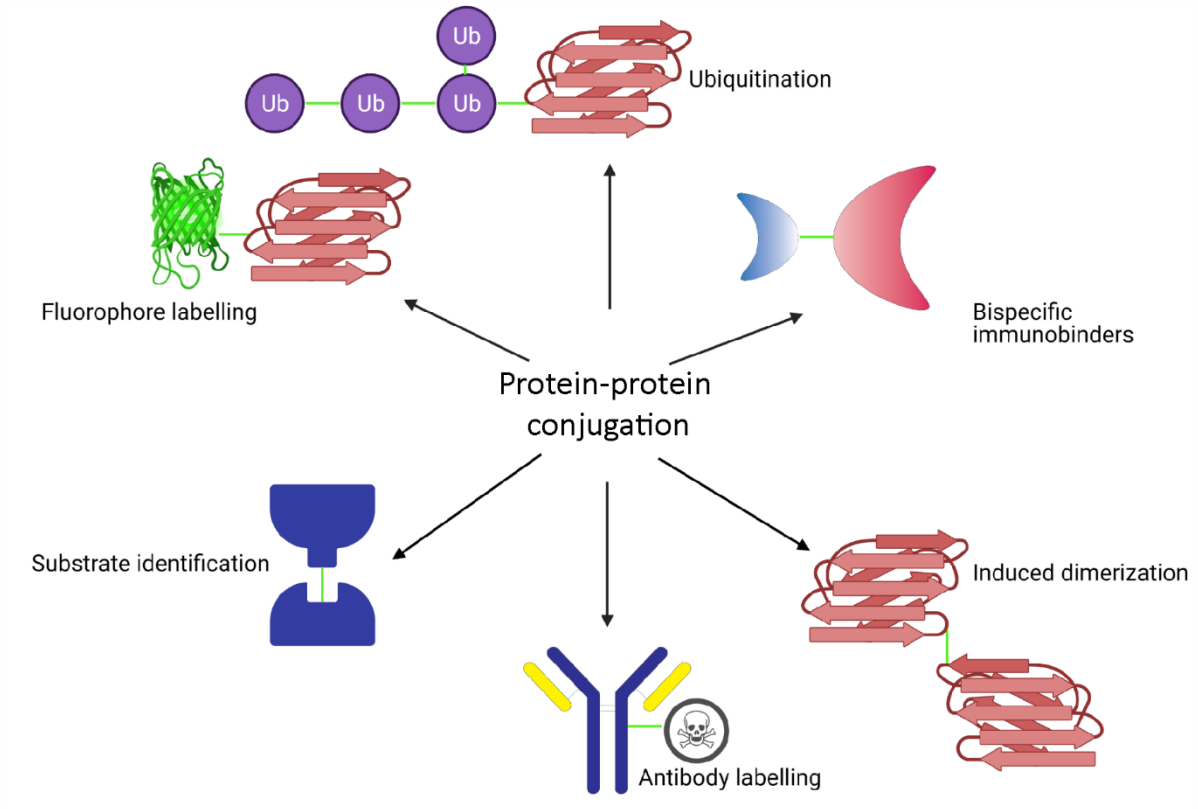


Figure 5: Protein-protein conjugation applications.

When low specificity is required, heterogenous protein conjugation, also known as unspecific protein crosslinking, can be used (Figure 6, left). Unspecific crosslinking reagents such as formaldehyde stochastically generate covalent linkages between essentially any protein without any specificity, and has significant applications in fixing samples for microscopy, mass spectrometry and storage, among other uses [102]. Crosslinking for chemical fixation is characterized by a non-biased system that comprehensively crosslink an entire sample, and do not suffer any downsides from poor specificity because uniformity is desired. The use of chemoselective chemistry (discussed above) can significantly improve specificity by targeting specific functional groups during covalent bond formation [103]. One popular iteration of this approach is Succinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (SMCC), a bispecific linker combining NHS-mediate primary amine targeting with maleimide-mediated

thiol targeting [104]. SMCC enhances specificity by primarily targeting lysine and cysteine residues for covalent linkage, but still yields a heterogenous product due to the highly variable availability of such residues in various protein targets (Figure 6, right) [105].

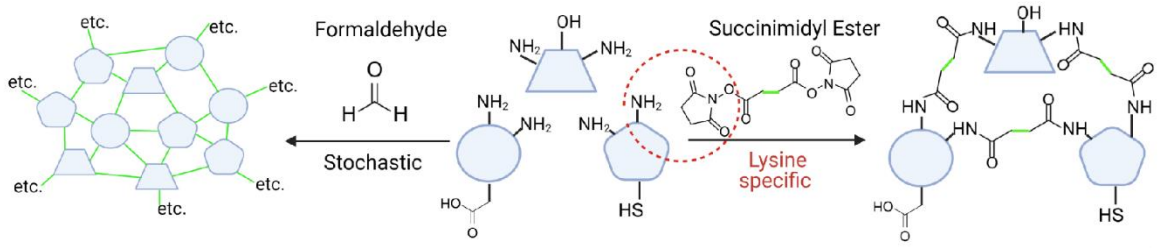


Figure 6: Protein crosslinking yielding heterogenous products.

Enzymatic reactions are often well suited to the chemical attachment of proteins to each other, and often benefit from the site-specificity associated with such methods. Sortase, for example, has been used to link proteins containing engineered sortase recognition motifs via native peptide bonds [106]. Despite being straightforward to apply, the method does require the genetic incorporation of a sortase recognition motif into the protein being targeted. Sortase has also been used to generate chimeric proteins by incorporating "click" handles at either termini, thereby generating *N*-*N* or *C*-*C* chimeric fusions [107]. However, this method is even less convenient, requiring each protein partner to possess a terminal, recombinant recognition sequence and each be modified by a sortase reaction to possess the required "click" handle before protein fusion.

Cysteines can also be targeted chemoenzymatically to enable protein-protein conjugation, as shown by recent work from the Francis group demonstrating that tyrosinase can convert tyrosine into an cysteine-reactive *o*-quinones [108]. The tyrosine-cysteine reaction proceeds quickly and under mild conditions and can be followed up by any of several previously discussed cysteine-targeting chemoselective chemistries. Vinylphosphonites are another class of compounds, recently introduced by our group, that can be installed onto proteins chemo-selectively at cysteines engaged in reactive disulfide bonds [57]. Once installed, the resulting phosphonothiolate presents an electron deficient vinyl group that can be targeted by nucleophilic thiols from the conjugate partner, thereby enabling protein-protein conjugation. The method benefits from high specificity and the requirement of only a small chemical entity linking the two proteins.

1.4.1 Antibody Conjugation

IgG antibodies can be engineered to generate therapeutic agents capable of targeting specific antigens, but represent a large and complex chemical environment with many potential modification sites. Therefore, targeted chemistries are required to generate homogenous and well-defined therapeutic agents from such a complex natural product. Methods range from site-specific options using recombinant IgG starting materials containing, for example, chemoenzymatic recognition sequences, to chemoselective conjugations that selectively target specific residues or functional motifs. Several reviews are available on methods for generating such constructs, with wide ranging examples encompassing different attachment chemistries, linkers, and degrees of specificity [6, 8, 109-111].

Sortase reactions have for example been applied to the generation of site-specific full-length antibody-drug conjugates [112]. One such strategy has been to recombinantly introduce a sortase recognition sequence into the Fab itself, applying the sortase to ligate a small probe at that precise location [113]. Direct counterparts of commercially-available ADCs Adcetris and Kadcyla have been produced chemoenzymatically using a similar method [114]. In this example, the C-terminus of both the heavy and light chain segments of the IgG antibody scaffold were genetically modified with sortase recognition motifs, while the toxic payload was modified with a poly-glycine motif, enabling site-specific installation of the drug payload onto the antibody ().

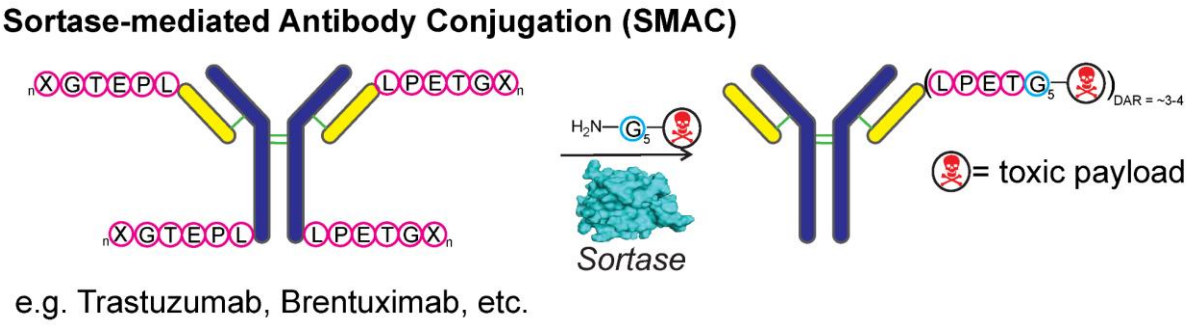


Figure 7: Sortase-mediated antibody conjugation[114] .

More recent examples of site-specific chemoenzymatic IgG manipulation have been accomplished on native IgG antibodies without genetic engineering [115]. One example used proximity-induced transpeptidation with engineered sortase variants to specifically target

two lysine residues in native human IgG1, one on the heavy chain and one on the light chain [116].

Chemoselective chemistry represents another option when designing a workflow for IgG antibody conjugations, with varying degrees of selectivity depending on the method employed. For example, alkyne phosphonamidates have been shown to be particularly well suited for the selective attachment of drug payloads to reduced disulfides in antibodies [117]. When attaching a cysteine-containing protein to various amines in the IgG antibody in situations where a homogenous product is not required, SMCC is an ideal method (Figure 8).

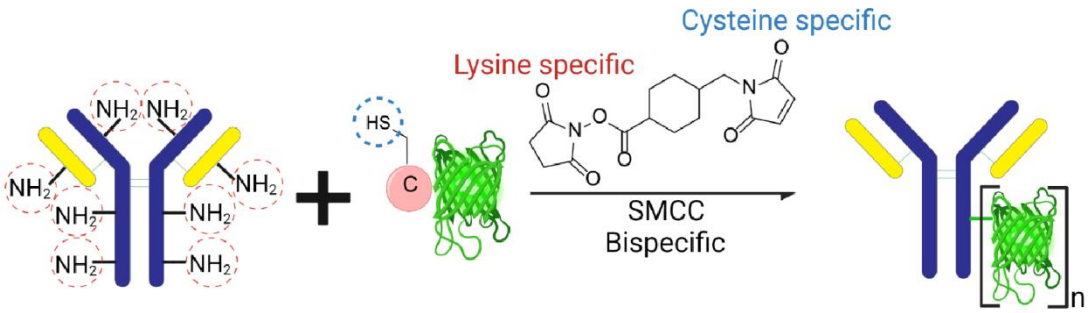


Figure 8: SMCC and its application towards antibody conjugation.

1.5 Active-site caging and photo-deprotection

Before its proof-of-principle demonstration in 1998 with 4-fluoro-L-phenylalanine, it had been a long standing goal of protein engineers to utilize endogenous protein translation for incorporating useful, non-natural chemical groups into proteins [118]. From this foundation, extensive work was done to expand the repertoire of incorporable functionalities [119]. Bertozzi expanded bioorthogonal chemistry significantly upon demonstrating the successful incorporation of azide-containing amino acids (azidohomoalanine) using methionine auxotrophs, enabling the selective installation of a chemical group targetable by the Staudinger ligation [120]. Engineers were eventually able to combine Bertozzi's work with azides together with the genetic code expansion to enable the site-specific incorporation of azide-derived lysine residues translationally into proteins [121]. Shortly after this, our own group demonstrated the potential for the Staudinger-phosphite reaction to site-specifically phosphorylate [45] and PEGylate [47] aryl azide-containing proteins. The latter application combined with site-specific azide incorporation enabled the installation of chemical "cages" at potentially any translatable position within a protein [122]. Protein cages mask the function

of a particular entity, (i.e. “cage”), until such point as some intervention enables the removal of the cage, thereby rendering the entity functional (i.e. “decaging”) [123]. Cleavable PEG groups are often used as a protein cage due to its high dispersity, which sterically blocks access to desired domains of a protein [124]. However, other functionalities can obviate the need for PEG, such as ortho-azideobenzyloxycarbonyl lysine, which is bulky enough to block nanobody bonding sites and removable using a small-molecule triggered Staudinger reduction [125].

Around the time of the described azidohomoalanine experiments, the mechanism of a photo-deprotection reaction involving chemical groups with structural similarity to some amino acid side chains was being elucidated [126]. These O-alkylated nitrophenyl compounds had by this time become widely used among biochemists as photo-labile “caging” groups [127]. Photo-decaging, or photolysis of the caging group, proceeds via the photo-tautomerization of the nitro group (Figure 9). In the basic example shown involving caged ATP, tautomerization can be seen to lead towards a cyclic benzisoxazoline intermediate, which undergoes rearrangement to yield the benzylic ketone and the free nucleotide. Other intermediates have been observed via spectroscopic monitoring of the reaction, suggesting the reaction mechanism may in fact be more complicated and dependent upon the entity being caged [127]. Nevertheless, photo-labile groups enabled light, with its high degree of control and tissue penetration, to function as the decaging initiator even within living cells [128]. The structural similarity of nitrophenyl compounds to common side chains eventually lead to the discovery of a genetically encoded photo-caged tyrosine derivatives [129]. Nitrophenyl-caged tyrosine derivatives could now be incorporated with site-specificity using genetic code expansion [130]. It should be noted that at least one application of such technology to protein nanobodies is known to have been reported [131].

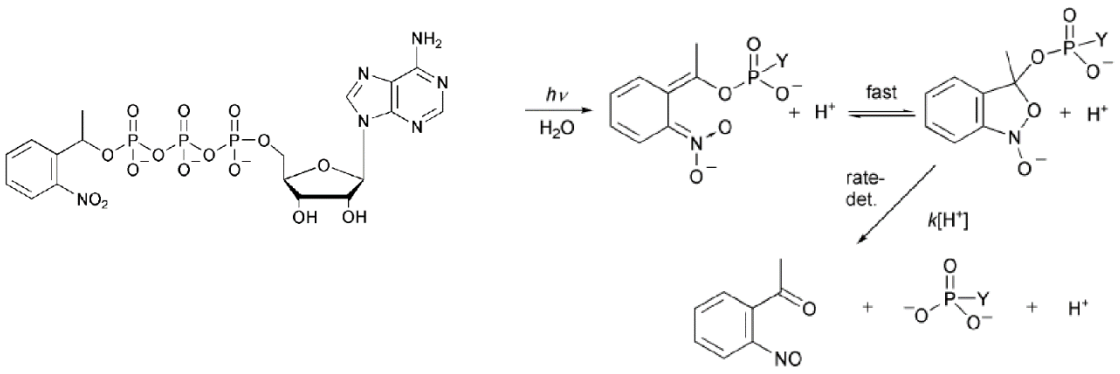


Figure 9: ONB photo-deprotection mechanism using caged ATP as an example. [123]

These examples represent two distinct pathways to enable protein caging: first, the direct installation of a cage, and second, the direct installation of a chemoselective functionality (i.e. azide) followed by the specific installation of a cage. Previous work in our group lead to the creation of a technology attempting to combine both methods: a photo-labile azide-containing lysine derivative that could benefit from site-specific incorporation and photo-deprotection as well as modular attachment of additional caging groups like PEG (Figure 10). This tool could therefore be installed at theoretically any position within a protein, before being further modified via a Staudinger ligation to contain a bulky caging functionality, all of which would be removed under UV irradiation to yield a native lysine.

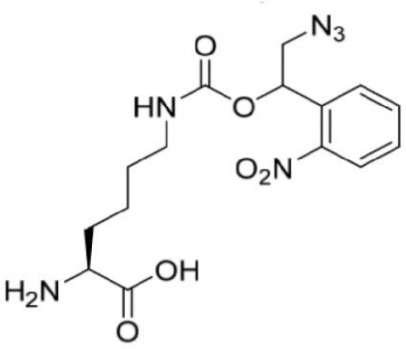


Figure 10: Lysine derivative bearing an azide on a photo-cleavable orthonitrobenzyl moiety. Synthesized by Oliver Reimann [132].

2. Research Objectives

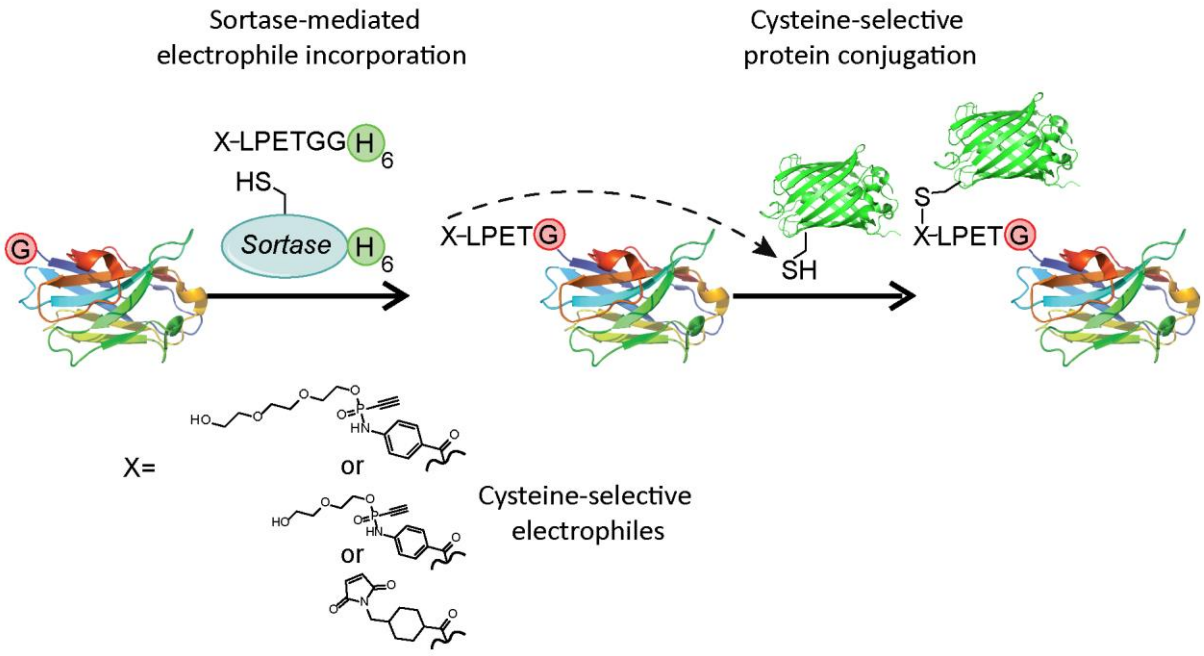
Recombinant proteins represent an attractive engineering target due to their sophisticated catalytic capabilities. However, the complexity of the protein chemical environment makes precision modifications challenging. Screening can be used to iteratively improve existing proteins but does nothing to enable engineers to implement rational modifications on the protein. Proteins can be identified from nature or designed rationally to target specific protein sites, enabling site-specific modifications under biocompatible conditions. Chemical protein engineering methods have also achieved modifications with high specificity. Current protein engineering methods often target either endogenous residues like cysteine or lysine, or bioorthogonal moieties like azides, each of which present both advantages and disadvantages from a bio-engineering standpoint. Endogenous residues are naturally occurring, offering chemical tractability in a single step. Such methods also benefit from modularity, because they can be applied identically to multiple protein substrates so long as the desired residues is present, and a compatible modification reaction is available. However, selectively targeting endogenous residues can be challenging to accomplish site-specifically for most substrates, based on the prevalence of the targeted residue within the protein in undesired locations. Incorporating orthogonal moieties bypasses the need to rely on endogenous residues for chemical targeting, which can circumvent the site-specificity challenge and enable precise targeting of the desired modification.

The work of this thesis applies both native and bio-orthogonal residue targeting chemistry for the precise modification of recombinant proteins. The first objective demonstrates the ability for a chemoenzymatic strategy to site-specifically incorporate chemical groups targeting native cysteine residues for protein-protein conjugation. The second objective uses codon expansion to site-specifically incorporate bio-orthogonal groups for the targeted caging and photo-deprotection of antigen binding proteins.

2.1 Objective 1: Sortase-enabled installation of cysteine-selective electrophiles for protein-protein conjugation

Novel electrophilic peptides containing cysteine-selective reactive handles and sortase recognition motifs should be synthesized and characterized, along with compatible protein substrates to be modified at their *N*-termini. Given the intention to use the sortase to

manipulate cysteine-selective reagents, the tolerance of the cysteine-dependent sortase active site for such reagents must be confirmed, and the reactivity of the cysteine-reactive handle following the sortase reaction must be verified. In addition, the two-step nature of this conjugation workflow may require additional purification after the sortase reaction to prevent side reactions. The known prevalence of reverse-transpeptidation by the sortase will also pose a challenge to yields and should be optimized in the context of the subsequent conjugation step. A comparison between Michael acceptor (i.e. maleimide) and non-Michael acceptor (i.e. PEG-functionalized alkyne phosphonamidate) cysteine-selective electrophiles should be performed to evaluate their effect on sortase incorporation and overall conjugation yields. Finally, the utility of the method should be demonstrated to generate relevant and biologically useful protein products that preserve the activity of the substrates and show novel functionality *in cellulo* (Scheme 2).

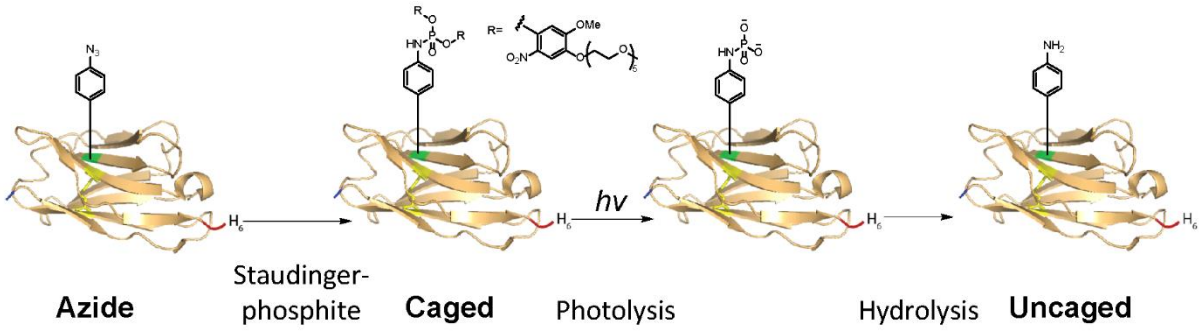


Scheme 2: Sortase-mediated cysteine-selective electrophile installation and crosslinking outline

2.2 Objective 2: Generation of azide-containing nanobodies and subsequent functionalization for caging and photo-deprotection

The general goal of this objective is to add spatiotemporal control to antigen binding protein-based tools by caging the active site with a photo-labile protecting group that can be removed using UV irradiation at a time and position of choosing. Protection of a nanobody active site by amino acid caging will first require the identification of a residue required for antigen

recognition via targeted mutagenesis. Mutant candidates should be characterized via an *in vitro* assay and then modified site-specifically with a reactive azide handle using amber suppression. Demonstration of the reactive handle should be accomplished via chemoselective modification via the Staudinger-phosphite reaction. A bulky and light-cleavable caging reagent built upon a phosphonamidate scaffold and capable of chemoselective installation on the engineered azide should be synthesized. PEG should be used = due to its bulky yet biocompatible characteristics to block the nanobody active site most effectively without inducing immunogenicity. Photo-lability of the PEG group should be conferred by an ONB functionality that can be cleaved upon UV irradiation, and its potential as a photo-cleavable caging group for azide-containing nanobodies should be characterized (Scheme 3).



Scheme 3: Pathway of UV-induced deprotection of the Staudinger-modified binding interface azide.

3. Results and Discussion

3.1 Objective 1

3.1.1 Sortase alkylation

One of the first questions set out to be investigated was the effect incubation with several general cysteine-selective reagents on sortase transpeptidation. This was of primary importance due to the essential sortase active site cysteine (Figure 11).

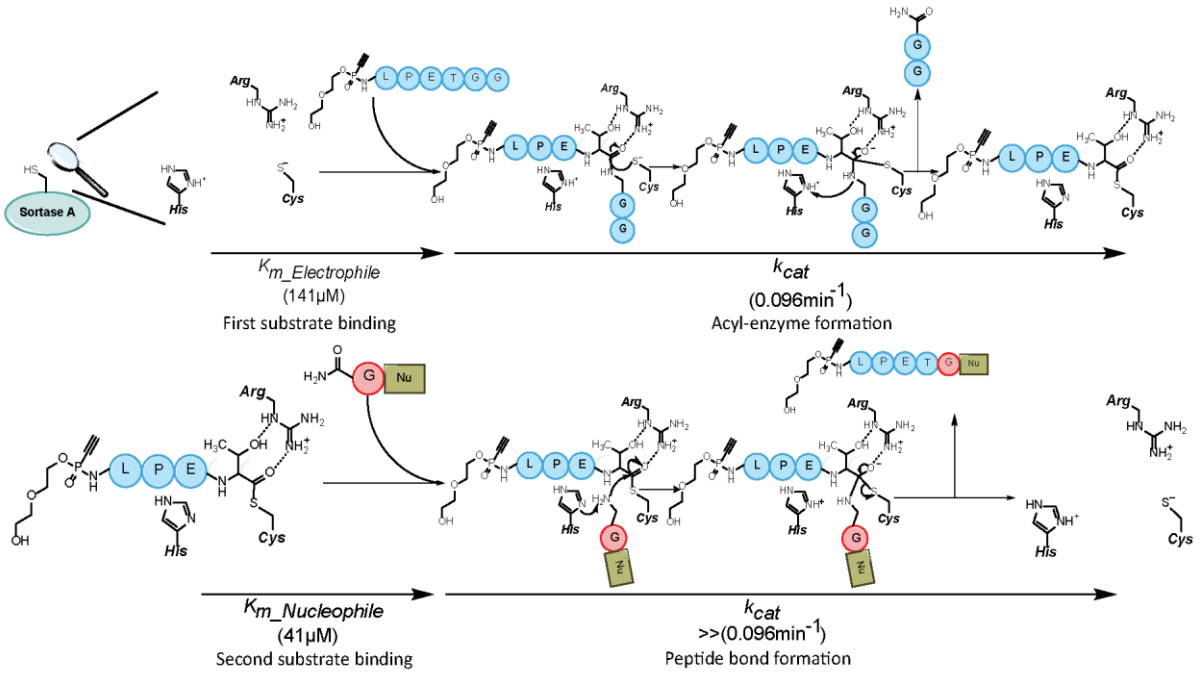
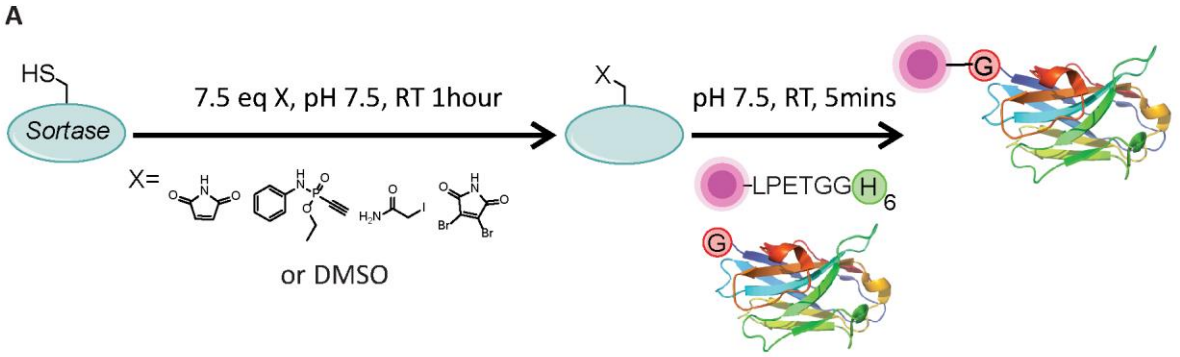
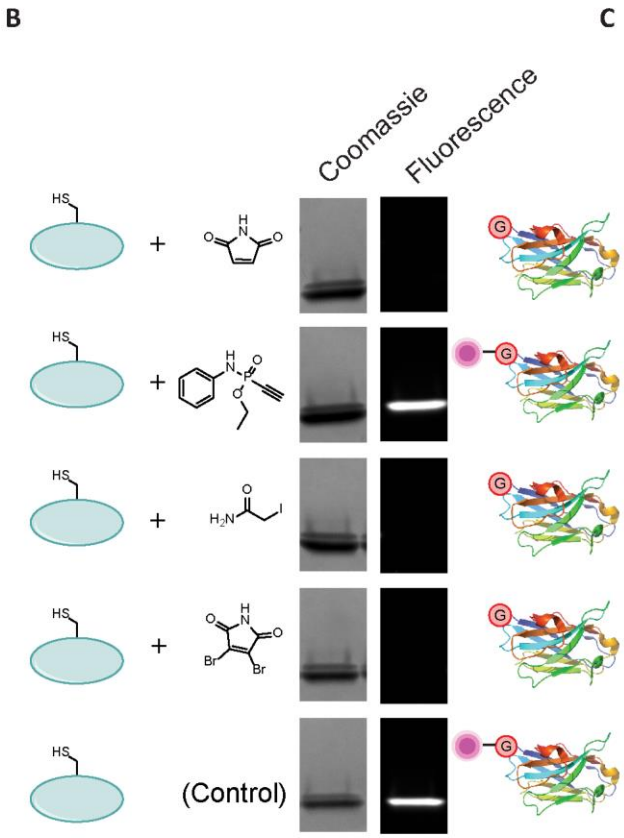


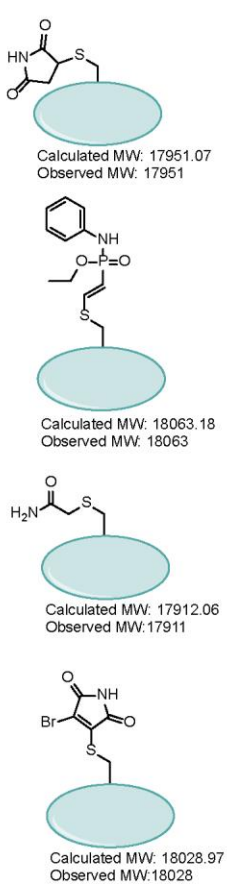
Figure 11: Sortase reaction mechanism involving a cysteine-selective electrophile-containing LPETGG peptide undergoing transpeptidation with a glycine species (i.e. protein N-terminus). Cys-His-Arg active-site catalytic triad is highlighted. Figure adapted from [133]. *Kinetic parameters from [77].*

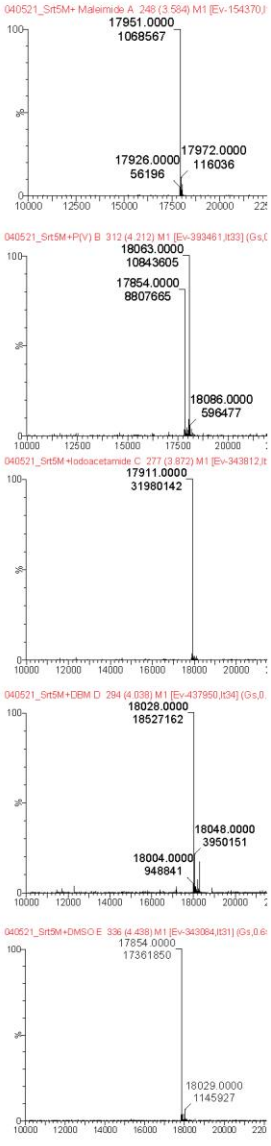
Transpeptidation efficiencies before and after incubating sortase with small molecule thiol alkylators were evaluated to determine the alkylators' effects on sortase activity. Incubation with the small molecules was done at RT, pH 7.5 for 1 hour using 7.5 equivalents of alkylator followed immediately by a fluorescent labelling reaction (Figure 12A). Alkylator examples were chosen from a wide range of cysteine-reactive moieties, including a Michael acceptors (maleimide), unsaturated electrophiles (alkyne phosphonamidate), and two halogen-containing electrophiles (iodoacetamide and dibromo maleimide) [134]. Installation of a fluorophore onto a protein using the sortase following incubation with the small molecules was quantified using SDS-PAGE via in-gel fluorescence and Coomassie band intensity analysis (Figure 12B, D). The degree of modification of the sortase with the alkylators was determined

using intact protein MS (Figure 12C). After incubation with the cysteine-targeting electrophiles, a single peak with a mass corresponding to alkylation of the sortases' only cysteine residue was observed for all reactions except for the aryl-modified alkyne phosphonamidate, which instead showed two peaks with masses corresponding to the unreacted and phosphonamidate-modified sortase variants. From this data it was concluded that the cysteine electrophiles specifically modify the sortase only at its single cysteine residue, and that the degree of modification varies based on the chosen electrophile, with the phosphonamidate incubation resulting in incomplete alkylation under these conditions. No labelling activity was observed for any sortase that had been shown to be fully converted by the cysteine electrophiles (i.e. incubated with maleimide, iodoacetamide or dibromomaleimide), whereas normal labelling activity was observed for the sortase showing incomplete conversion by the phosphonamidate. We inferred from this that the transpeptidation activity of the sortase could be completely impaired by the presence of highly reactive cysteine-selective electrophiles under these conditions, but that this impairment could be prevented if the degree of active site alkylation at the time of the labelling reaction was incomplete.


A
HS
Sortase
7.5 eq X, pH 7.5, RT 1hour
X=
or DMSO
X
pH 7.5, RT, 5mins
LPETGG
H
6
G


B
Coomassie
Fluorescence
HS
(Control)
G


C
Calculated MW: 17951.07
Observed MW: 17951
Calculated MW: 18063.18
Observed MW: 18063
Calculated MW: 17912.06
Observed MW:17911
Calculated MW: 18028.97
Observed MW:18028


17951.0000
1068567
17926.0000
56196
17972.0000
116036
18063.0000
10843605
17854.0000
8807665
18086.0000
596477
17911.0000
31980142
18028.0000
18527162
18048.0000
3950151
18004.0000
948841
17854.0000
17361850
18029.0000
1145927

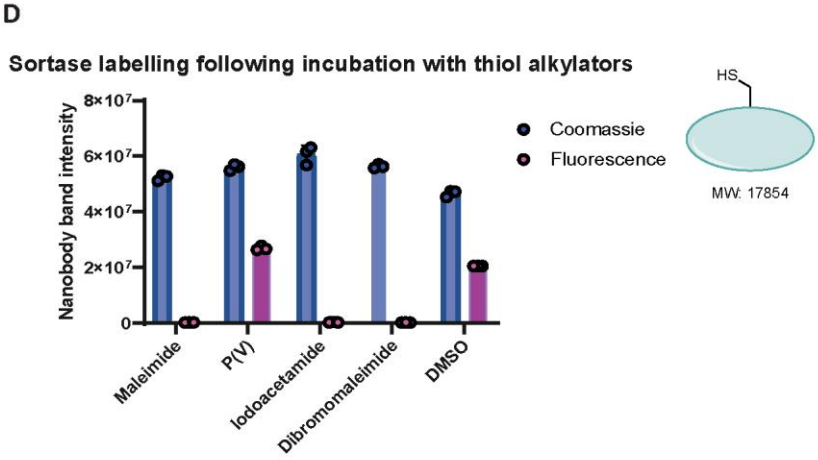

D
Sortase labelling following incubation with thiol alkylators
Nanobody band intensity
8×10⁷
6×10⁷
4×10⁷
2×10⁷
0
Maleimide
P(V)
Iodoacetamide
Dibromomaleimide
DMSO
Coomassie
Fluorescence
HS
MW: 17854

Figure 12: Sortase alkylation and its effect on transpeptidation activity. A: Reaction schematic with alkylation followed by transpeptidation and SDS-PAGE analysis. B: SDS-PAGE with coomassie (right) and in-gel fluorescence (left) of transpeptidation reaction following incubation with the indivated small molecule. C: Deconvoluted ESI-MS of sortase following incubation with the indicated small molecule. D: quantification of coomassie and fluorescence signal intensity of the band corresponding to the labelled nanobody in each sortase reaction following incubation with the indicated electrophile.

3.1.2 Testing different glycine-containing nanobody substrates

The sortase variants used in this work require an *N*-terminal glycine for efficient labelling. Various factors can affect the degree of labelling, including steric accessibility of the glycine, which must enter the active site of the sortase to intercept the activated acyl-enzyme intermediate [77]. To make this method as accessible as possible, an investigation was undertaken to determine if a single glycine (rather than the typical polyglycine motif) generated by endogenous methionine aminopeptidase (MAP) activity could act as a suitable sortase substrate. An additional investigation was done to determine if protease treatment would also be suitable to render an accessible glycine. Protease treatment is a common part of many recombinant expression protocols, as it is well suited to remove *N*-terminal purification tags used for isolation of the protein from bacterial cultures while simultaneously generating an *N*-terminal glycine residue [92]. It should be noted that, without additional recombinant engineering, such a protease-driven glycine generation strategy would only result in the formation of a single *N*-terminal glycine. Previous reports have shown that increasing the quantity of glycine residues at the *N*-terminus correlates with improved sortase labelling [135]. It would therefore follow that recombinant insertion of additional glycine residues downstream from the protease cleavage site would improve overall sortase labelling yields, a strategy which should be pursued in future investigations. This investigation chose instead to highlight the sufficiency of TEV and Thrombin recognition sequences for generating sortase recognition motifs without additional recombinant engineering. A discussion of initial investigations into polyglycine-containing nanobody substrates can be found in section 3.1.2.3.

3.1.2.1 Glycine generation via endogenous MAP activity

AntiEGFR nanobody sortase substrates were expressed from a commercial plasmid containing a fortuitous glycine residue directly following the start codon methionine (Figure 13A). Expression, endogenous cleavage of the *N*-terminal methionine by MAP and NiNTA purification were followed by TEV cleavage to remove the *C*-terminal purification tag and reverse NiNTA to remove the protease. Purity and identity of the expressed nanobody was

confirmed via intact protein MS (Figure 13B). As an initial test, an impure fraction from the nanobody purification following TEV cleavage was tested in a standard sortase labelling experiment involving the installation of a fluorophore-containing peptide (Figure 13C). Labelling was efficient and was only observed in the presence of sortase, nanobody and peptide, and self-labelling of the Sortase was observed.

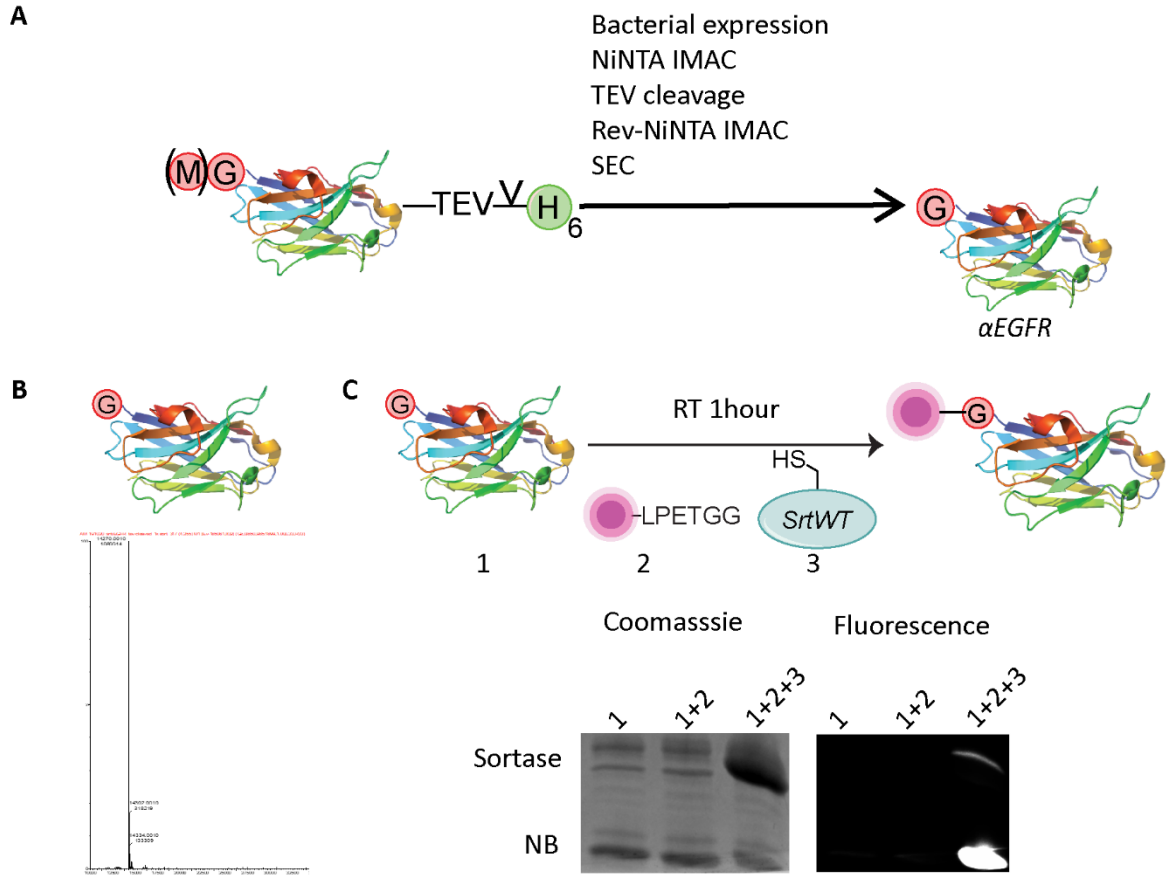


Figure 13: Expression and purification of antiEGFR nanobody sortase substrate. A: Purification schematic. B: Deconvoluted ESI-MS of purified nanobody. C: Reaction check using TAMRA-containing LPETGG peptide for N-terminal fluorophore installation. Analysis by SDS PAGE with coomassie (left) and in-gel fluorescence (right).

3.1.2.2 Glycine generation via protease digestion

As an alternative strategy, sortase substrate generation by standard thrombin cleavage (which generates an *N*-terminal glycine when used for *N*-terminal cleavage protocols) was tested on the recombinant anti-GFP nanobody GBP1 [12]. First, GBP1 was cloned into the pET28a vector to confer an *N*-terminal hexa-histidine purification tag separated by a thrombin cleavage sequence (Figure 14A). Expression/purification was carried out in a comparable manner as for the anti-GFP nanobody described above, although it should be noted that the thrombin cleavage both removes the purification tag AND generates the sortase recognition

site, whereas the TEV cleavage only removes the purification tag. MS confirmed the identity of the protein again (Figure 14B). Fluorescent labelling was observed by the sortase, in addition to sortase self-labelling as observed before (Figure 14C).

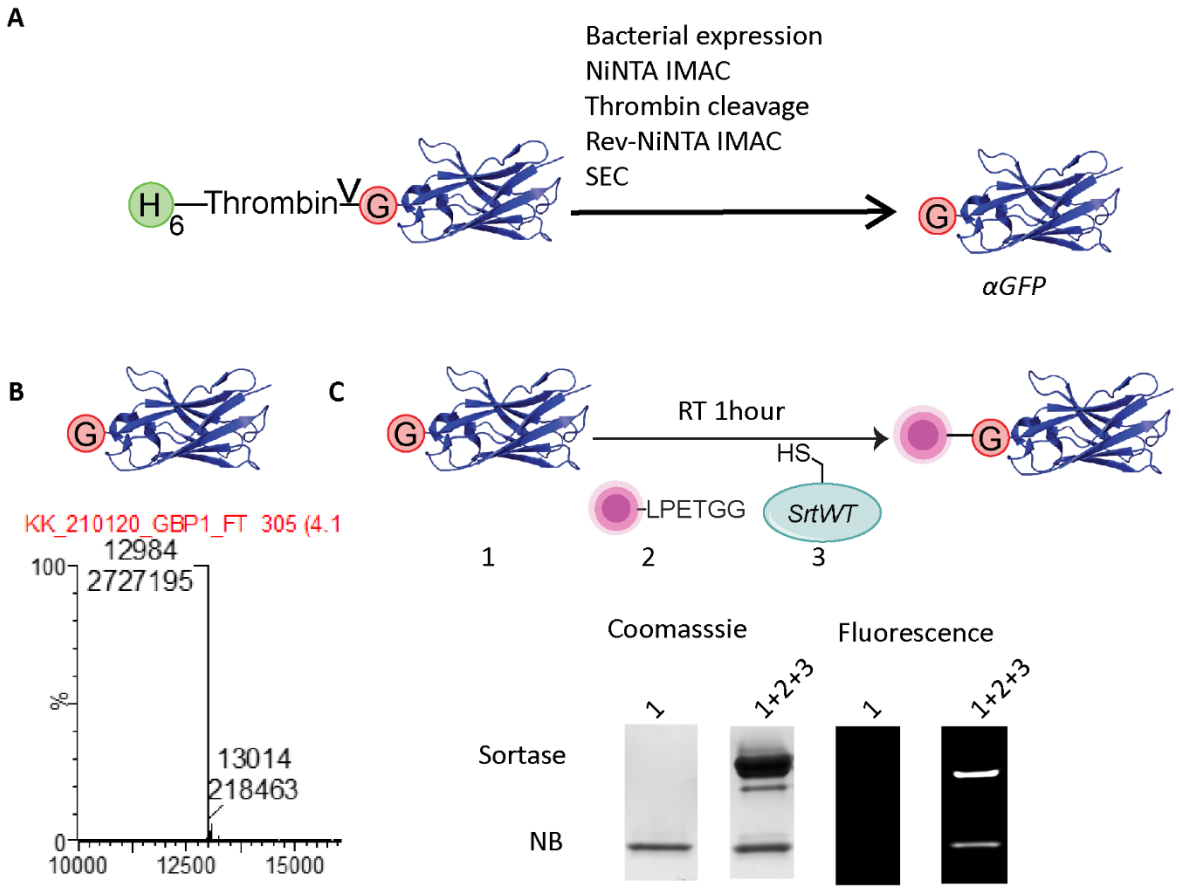


Figure 14: Expression and purification of antiGFP nanobody sortase substrate. A: Purification schematic. B: Deconvoluted ESI-MS spectrum of purified product. C: Reaction check using TAMRA-containing LPETGG peptide for N-terminal fluorophore installation. Analysis by SDS PAGE with coomassie (left) and in-gel fluorescence (right).

3.1.2.3 Effect of PelB on MAP activity and glycine accessibility

Initially, a GBP1 variant was recombinantly engineered to contain an *N*-terminal polyglycine tag downstream of the initiating methionine, following literature examples of enhanced sortase yields on the *N*-terminus with poly-glycine motifs [96]. Due to poor soluble protein yields, the entire construct (including its *N*-terminal methionine) was cloned into the pTrcHis vector downstream of a PelB leader sequence (Figure 15A). The PelB sequence signals for translated proteins to be translocated to the periplasmic space. The reducing environment of the periplasm enables proper disulfide formation and enables cleavage of the PelB sequence to yield the native target protein. Surprisingly, the expressed nanobody showed a mass corresponding to the intended sequence including the *N*-terminal methionine when analyzed by intact protein MS (Figure 15B). The resulting nanobody experienced low transpeptidation

efficiency as observed in the fluorescent labelling experiment, which was puzzling at the time of its discovery (Figure 15C). As a result, the other strategies discussed above were explored to solve the problem instead of simply removing the problematic methionine residue recombinantly. The protease strategy also offered the benefit of being a more modular approach for producing compatible substrates, which was thought to make the method more applicable.

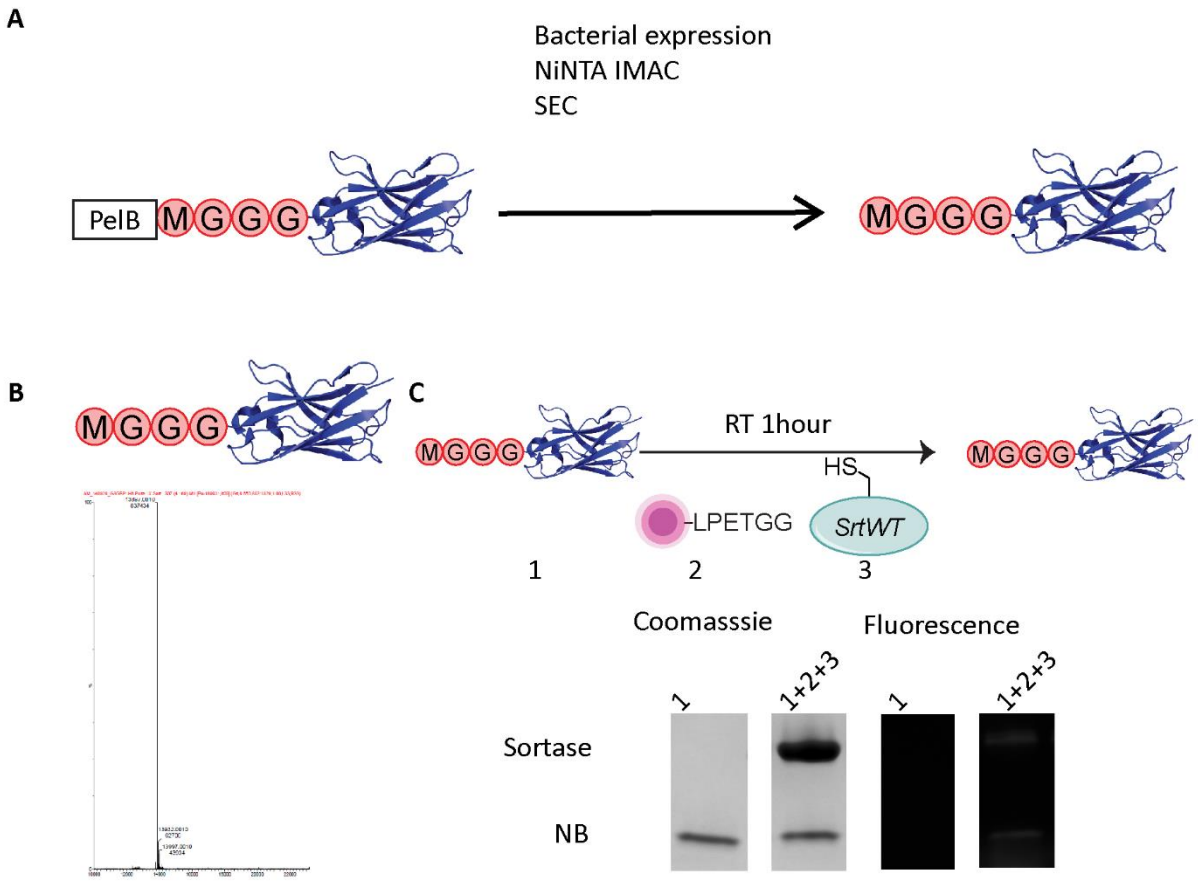


Figure 15: PelB prevents cleavage of methionine, preventing sortase recognition. A: Purification schematic. B: Deconvoluted ESI-MS spectrum of purified product. C: Reaction check using TAMRA-containing LPETGG peptide for N-terminal fluorophore installation. Analysis by SDS PAGE with Coomassie (left) and in-gel fluorescence (right).

3.1.3 Investigating and optimizing the sortase reaction conditions

Reactions involving the sortase variants used for these investigations always involved three components: an LPETG-containing peptide functionalized with a payload, a glycine containing protein of interest, and the sortase itself, suspended in a Tris buffer containing calcium and magnesium. Once the glycine-containing substrates had been generated, work turned towards identifying and optimizing sortase reaction conditions that enabled the installation of cysteine electrophiles.

3.1.3.1 Comparing SrtWT and Srt5M sortase variants

Time course measurements were performed in which the two varieties of sortase were compared regarding their ability to install a TAMRA functionality onto both anti-GFP anti-EGFR and nanobodies (Figure 16A). Strangely, a double band was observed near the molecular weight of both nanobodies following the sortase reactions (Figure 16B, C). Given the observed MW weight shift between bands and the increased fluorescence intensity at higher MWs, the signals likely correspond to multiple peptide-nanobody adducts, but these species could not be detected by intact protein MS. The results for both nanobodies were similar, although the fluorescence signals of the antiEGFR transpeptidation products were slightly higher, suggesting a more efficient sortase reaction. For this reason, the antiEGFR nanobody was the primary choice for future applications. Reducing the equivalents of sortase led to more efficient labelling for both variants, likely by reducing the prevalence of the reverse labelling reaction. The peptide was present in large excess for both conditions to drive the reaction to completion despite the more demanding nature of labelling intact protein *N*-termini with the sortase, which is known to be made difficult by the need for the glycine nucleophile to penetrate the compact sortase active site [96]. Due to thrombin-mediated removal of the *N*-terminal H6 tag for the wild-type sortase rendering a glycine accessible, self-labelling was observed for that variant, whereas the pentamutant lacked the accessible glycine and therefore did not label itself (Figure 16B, C).

The pentamutant sortase (Srt5M) was the most efficient variant, achieving the strongest product signal intensity in only five minutes. This result confirmed the pentamutant's previously reported kinetic acceleration [88]. With this data viewed in the context of the previously described sortase alkylation experiment, it was reasoned that any protein labelling application of the sortase that relied upon its active-site cysteine would be hampered by the presence of cysteine-reactive electrophiles. A competition mechanism was envisioned, by which the cysteine electrophile would alkylate and inhibit the sortase concurrently with its labelling activity. In this case, a faster reaction was believed to be advantageous, since more labelling would be accomplished before the sortase was rendered fully inactive by the excess quantities of electrophilic peptide present in the reaction.

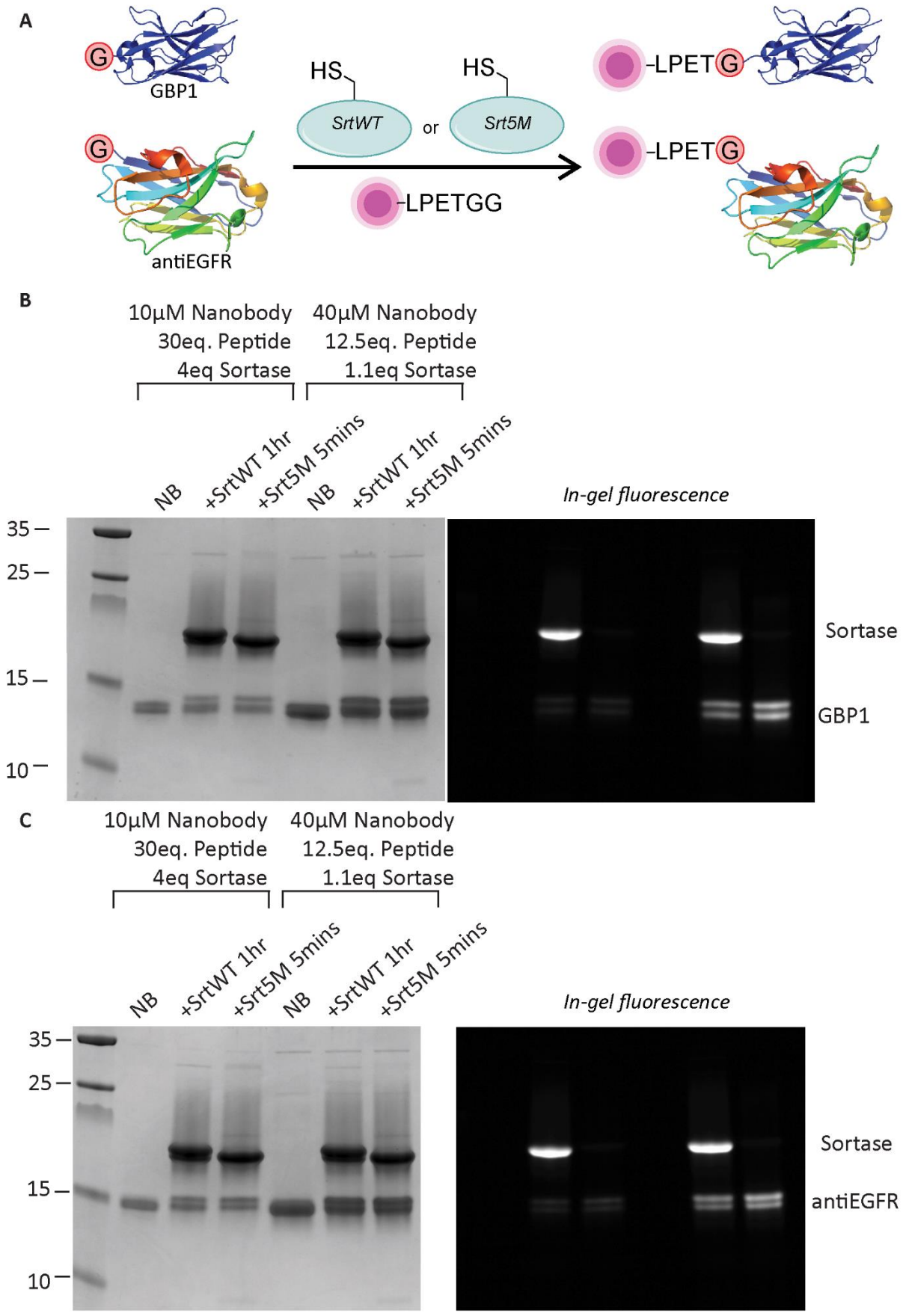


Figure 16: Comparing the wild-type and pentamutant sortase variants. A: Reaction schematic showing fluorophore installations onto two different nanobodies using either the WT or pentamutant sortase. B: SDS-PAGE analysis with coomassie staining and in-gel fluorescence for the GBP1 nanobody modified with a fluorophore using both sortase variants under two

different conditions. C: SDS-PAGE analysis with coomassie staining and in-gel fluorescence for the antiEGFR nanobody modified with a fluorophore using both sortase variants under two different conditions.

3.1.3.2 Adjusting the equivalents of sortase and the reaction conditions

Different equivalencies and reaction durations were tested to determine the optimal reaction conditions for the sortase pentamutant (Figure 17). Reactions conditions were assayed using a labelling reaction between a fluorescent peptide and a glycine-containing antiEGFR nanobody (Figure 17A). Reactions were analyzed by SDS-PAGE and in-gel fluorescence (Figure 17B). Regardless of the equivalents of sortase employed, reactions durations longer than 1 hour resulted in a complete loss of product, likely due to the dominance of the reverse labelling reaction. The addition of 2 equivalents of sortase resulted in less product formation than 1.1 equivalents, which can be seen by observing the presence of the red feedback indicating signal exceeding the detector limit in the fluorescence image with 1.1 equivalents of Srt5M as well as its absence with 2.2 equivalents of Srt5M. This demonstrated another advantage of the catalytically enhanced variant beyond just shortening reaction times; by requiring less sortase, the pentamutant enables simpler purification for downstream applications of the reaction product.

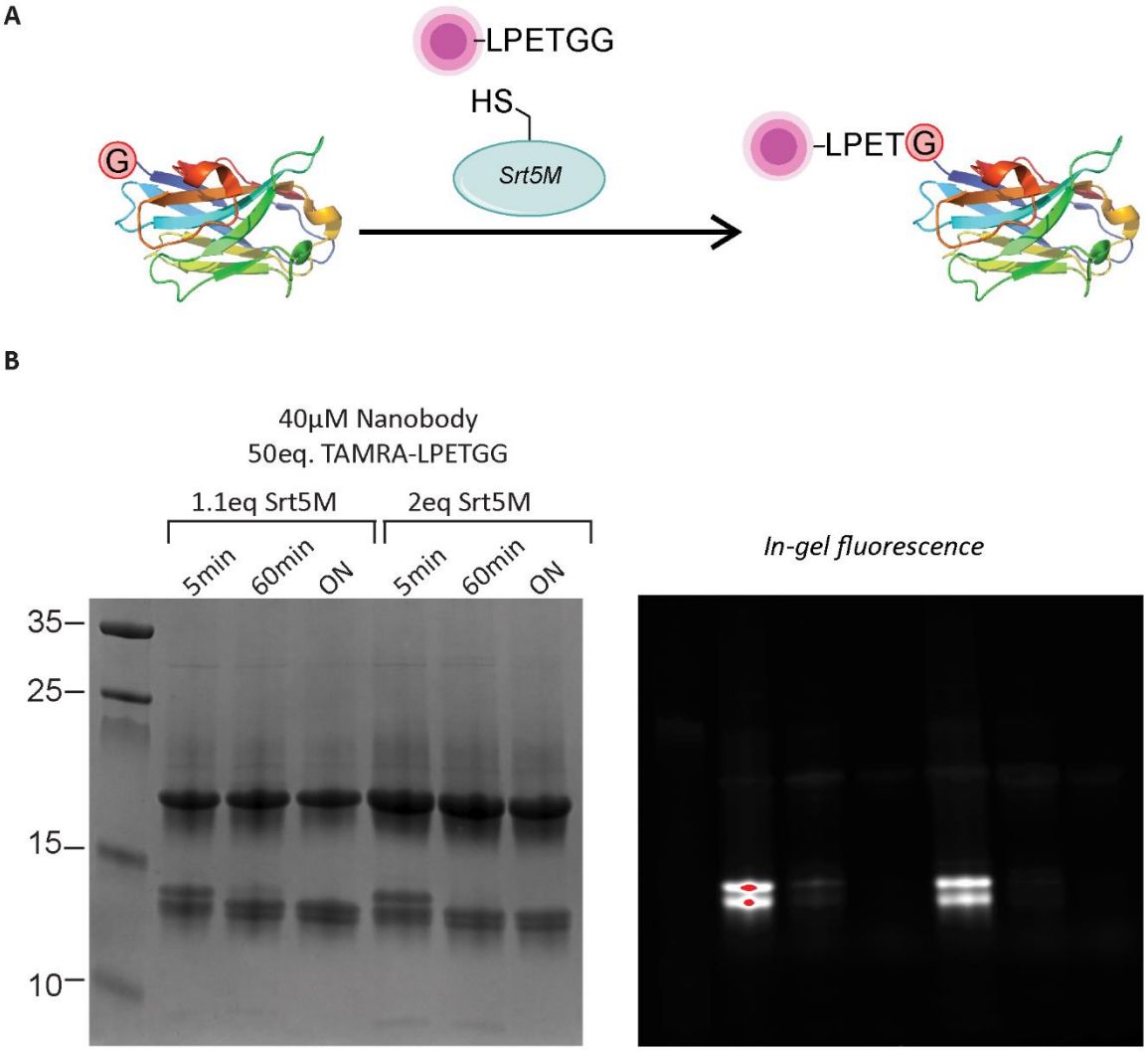


Figure 17: Investigation of different sortase equivalents on transpeptidation efficiency. A: Reaction schematic. B: SDS-PAGE analysis of reaction following transpeptidation. Coomassie stain (left) and TAMRA in-gel fluorescence (right).

From these results the optimal amount of Srt5M relative to nanobody (1.1eq) for good yields while still enabling complete removal before downstream crosslinking was identified (see 3.1.5). A large excess of peptide (20eq) was selected rationally based on the reported benefits of excess LPETG species when attempting to label protein N-termini with good yields [136]. This choice was further supported by the rational that peptide material was simple to synthesize (i.e. not precious) and straightforward to remove from the reaction when combined with a C-terminal His tag (see section 3.1.5). These optimized conditions were not used for the experiments discussed below using non-His tag containing peptides but were used for all experiments involving the His-tagged peptides.

3.1.3.3 Comparing various peptide payloads for incorporation efficiency

Different peptides with functional modules were compared for their incorporation efficiency onto the protease-treated GBP1 nanobody using sortase with the optimized reaction conditions of 1.1eq sortase and 20eq. peptide (Figure 18A). These modules included fluorophores, electrophiles, purification tags and PEG chains, as well as several combinations of different functionalities, all attached to an LPETGG sortase recognition motif (Figure 18B). Labelling reactions were analyzed by lithium dodecyl sulfate (LDS)-based gel electrophoresis to resolve the labelled from the unlabeled nanobodies for quantification purposes (Figure 18C). LDS is a variant of SDS-PAGE that forms smaller micelle adducts with sample proteins, increasing the potential resolution within the acrylamide gel matrix. While some variation in labelling efficiency was observed, most peptides were incorporated to approximately the same degree regardless of functionality under the conditions tested, as can be seen in the consistent Coomassie intensity of the product band for all reactions. Electrophile-containing peptides were incorporated to a lesser extent than their no electrophile-containing counterparts, but a significance test to prove if the observation is caused by the electrophile was not performed. The only exceptions were the peptides functionalized with PEG chains containing free amines (8 and 10), which showed significantly lower incorporation efficiencies. This is most likely due to the free amine, which is a nucleophile capable of intercepting the acyl enzyme intermediate formed by the sortase and the LPETGG species. This side reaction would compete with the labelling reaction by preventing the glycine-containing protein nucleophile from being labelled. From this experiment it was concluded that electrophilic payloads could be installed onto nanobodies using the sortase under these optimized conditions, although the functionality of these installed electrophiles had yet to be demonstrated.

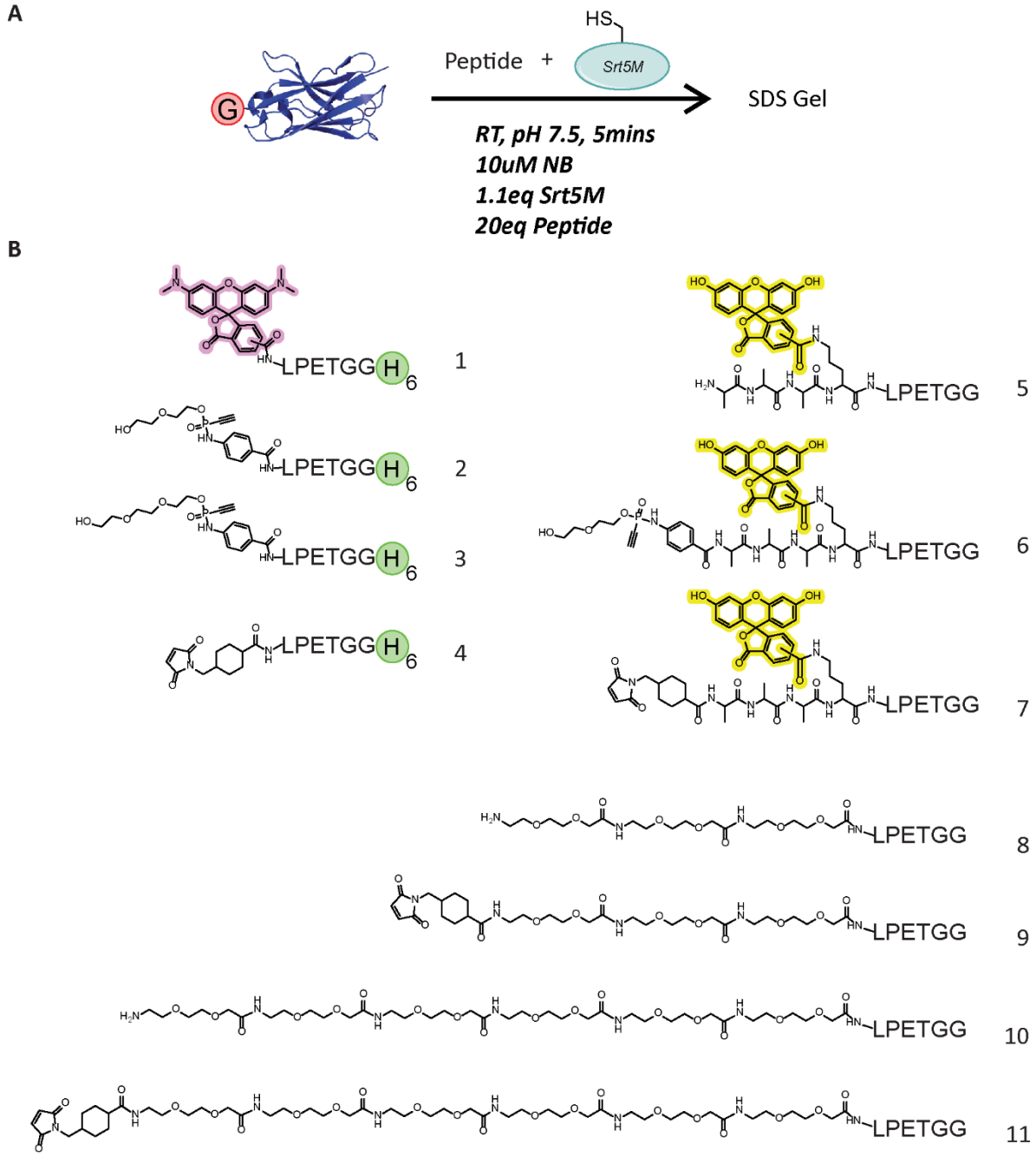

A
HS
Peptide +
Srt5M
SDS Gel
RT, pH 7.5, 5mins
10uM NB
1.1eq Srt5M
20eq Peptide
B
LPETGG H6
1
2
3
4
LPETGG
5
6
7
8
9
10
11
C
DMSO
1 2 3 4 5 6 7 8 9 10 11
Sortase+Peptide
Sortase
NB-Peptide (product)
NB
Srt5M, 5min, antiGFP NB
Product coomassie Intensity
Unmodified

Figure 18: Effect of peptpide functionalities on sortase incorporation efficiency. A: Reaction schematic. B: Structures of LPETGG peptides and their cargo functionalities. C: SDS-PAGE analysis of nanobody labelling reactions with the indicated peptide substrate (left) and quantification of product band coomassie signal intensity (right).

3.1.4 Generating thiol-containing protein crosslinking partners

Precise crosslinking in the complex environment of a protein requires chemical strategies capable of targeting specific functional groups. In our case, cysteine was chosen as the targeted group, due to its relatively low abundance and reactivity as a nucleophile under basic conditions [33]. To establish this protocol in the context of the sortase-enabled installation of cysteine-targeting electrophiles, model proteins were generated containing engineered cysteine residues at precise positions via a variety of methods. For example, an eGFP variant used frequently in our lab containing a single accessible cysteine was generated via simple site directed mutagenesis (eGFPC70MS147C) by Kristin Kemnitz-Hassanin (Figure 20A). Nanobodies containing engineered cysteines were also generated, either via site-directed mutagenesis (i.e. GBP1AAAC, containing a *C*-terminal cysteine residue separated by a 3-alanine spacer) (Figure 20B) or via expressed protein ligation (EPL) with cysteine, as in the case for the cysteine-containing antiEGFR nanobody (Figure 19, Figure 20C). In addition, IgG antibodies were reduced with TCEP to convert their inter-chain disulfides into cysteine residues for chemoselective crosslinking (Figure 20D). These proteins were subjected to crosslinking reactions with nanobodies modified with electrophiles using the sortase. Please note that an additional cysteine-containing super-folder GFP variant was generated by the author for this study, but was used to investigate the applicability of sortase-mediated electrophile installation on cysteine-containing protein substrates for subsection protein-protein conjugation (Section 3.1.7).

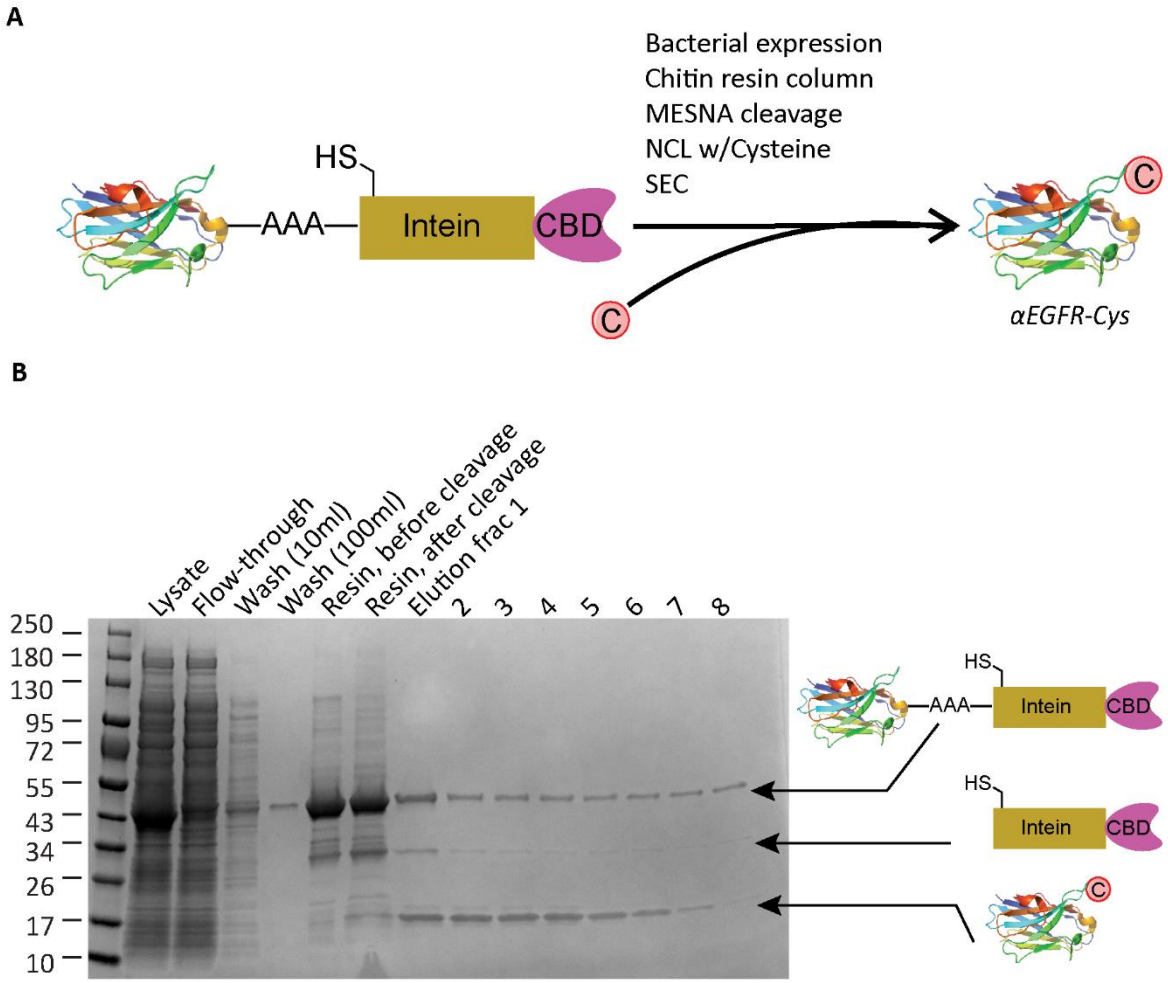


Figure 19: Expression and purification of thiol-containing nanobody for protein-protein conjugation. A: Purification schematic. B: SDS-PAGE of various samples taken during purification of the thiol-containing antiEGFR nanobody.

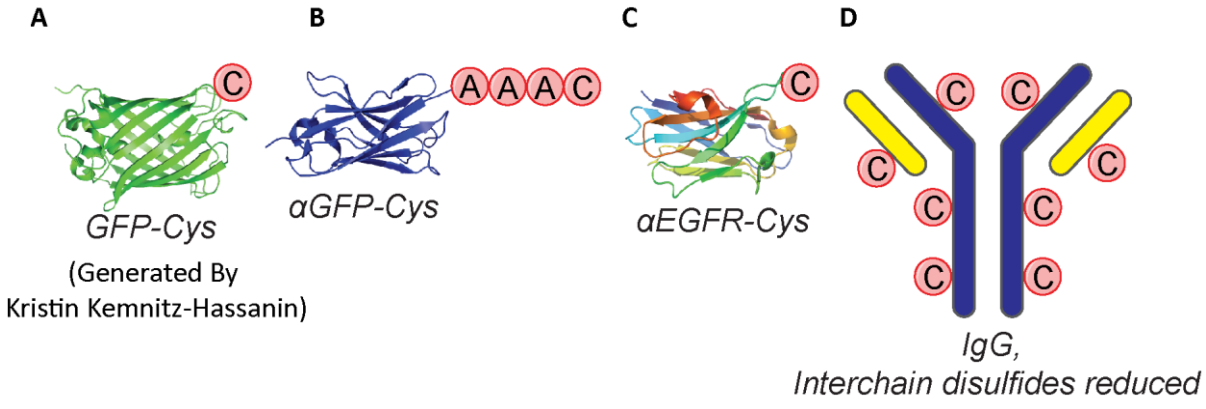


Figure 20: Cysteine-containing protein conjugation substrates used in this work.

3.1.5 Crosslinking optimization and yield determination

Initial attempts at utilizing installed electrophiles for chemoselective crosslinking were unsuccessful. A first attempt which simply combined the sortase reaction mixture with a thiol-

containing GFP without any prior purification showed no specific targeting of the installed electrophile to the desired thiol, despite ESI-MS data showing good conversion to the electrophilic sortase product (Figure 21). The presence of fluorescent, high MW bands designated in the figure as “Unspecific Crosslinking” for both GFP variants, one of which lacks an accessible cysteine, indicates that the reaction products are likely unspecific multimers of sortase-labelled GFP. The identity of these multimers was not determined definitively, but they occurred in the presence and absence of an accessible cysteine on the eGFP crosslinking partner (suggesting they are not the result of cysteine-specific crosslinking, i.e. unspecific) and only when all the components of the sortase reaction were added. Therefore, they are most likely the product of a sortase reaction resulting in the formation of high MW conjugates between eGFP and/or nanobody monomers. These results suggested that purifying the sortase reaction prior to crosslinking would be necessary to prevent the formation of unspecific sortase products when multiple protein species are present.

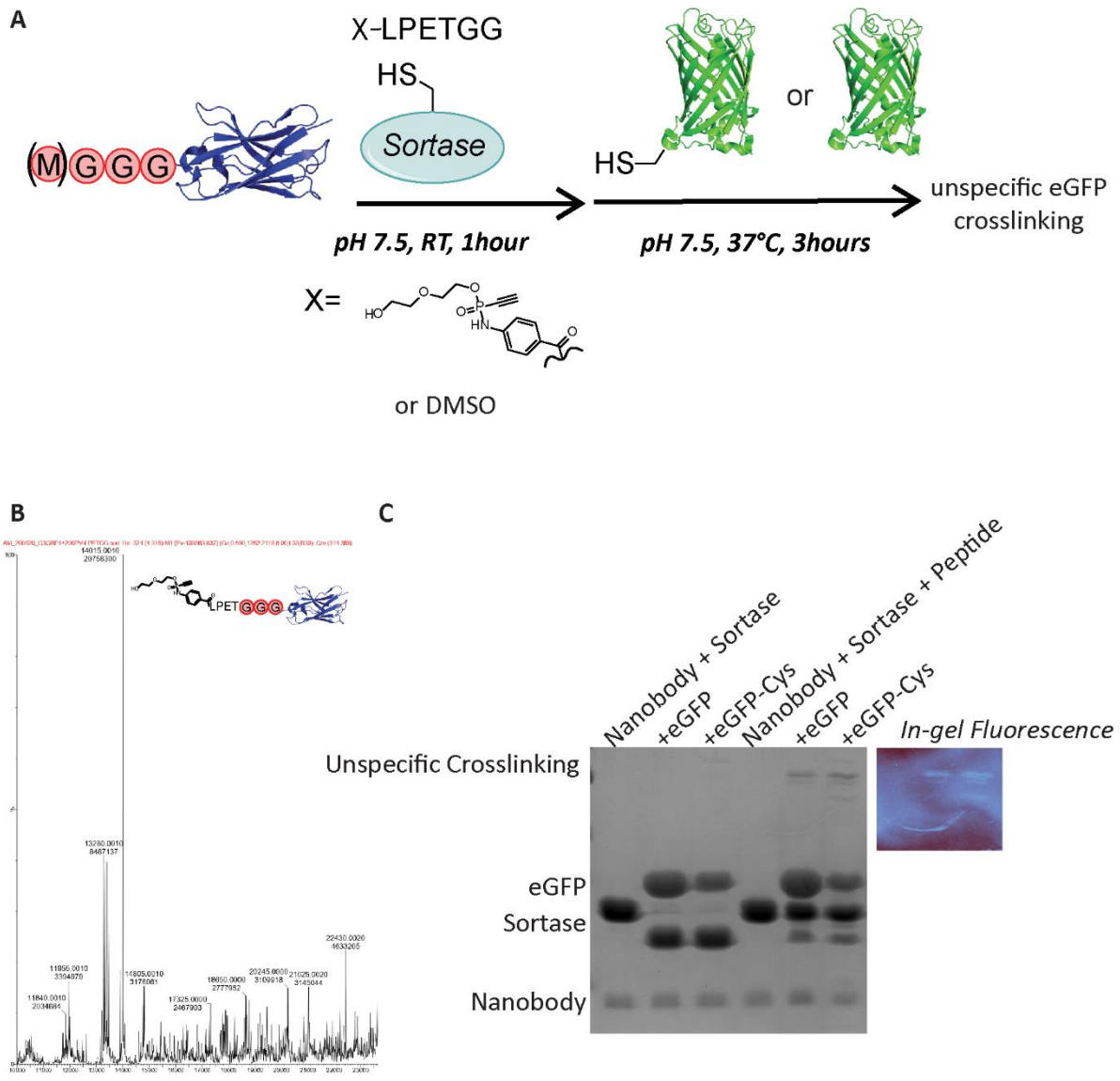


Figure 21: Unspecific crosslinking without purification of the sortase reaction. A: schematic of reaction. B: ESI-MS of sortase product following 1-hour transpeptidation. C: SDS-PAGE of crosslinking reactions involving unlabeled left 3 lanes) or electrophilic (right three lanes) anti-GFP nanobody.

Purifying the sortase was accomplished using a histidine-tagged sortase variant in conjunction with NiNTA chromatography (Figure 22). Peptide removal was accomplished using gel filtration at this stage, taking advantage of the size difference between the protein and peptide species. Initial experiments were performed using the WT sortase and demonstrated that the sortase could be fully removed from the reaction using affinity chromatography.

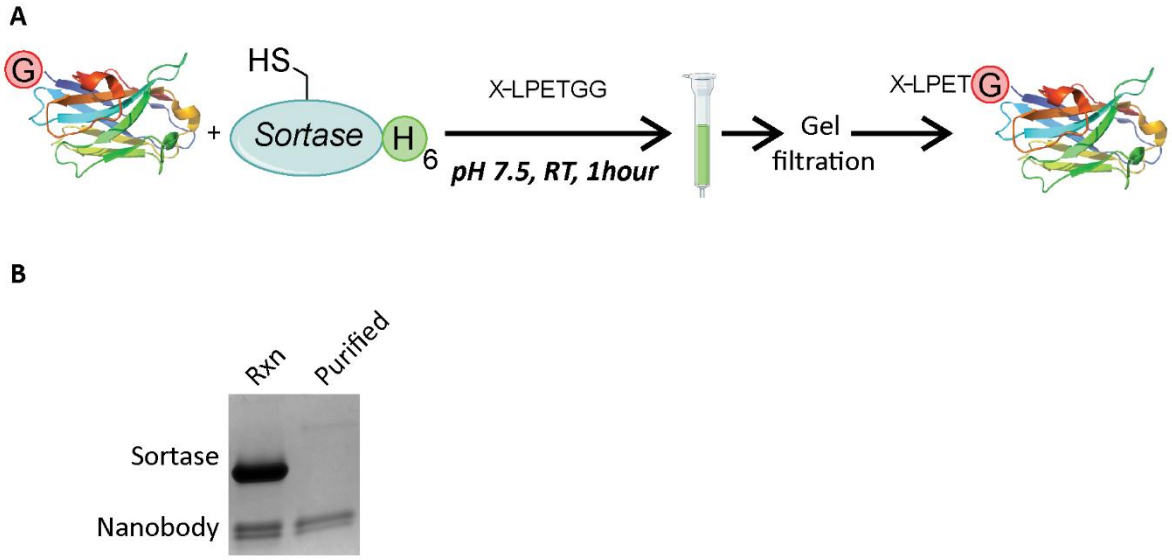


Figure 22: Affinity purification of sortase from the reaction following transpeptidation. A: Schematic highlighting the removal of sortase via affinity chromatography followed by peptide removal via gel filtration. B: SDS-PAGE with Coomassie staining of the reaction before and after the purification protocol.

Removing the sortase with affinity chromatography after the sortase reaction enabled the installation of both phosphonamidate (Figure 23A) and maleimide (Figure 24A) cysteine-selective electrophiles onto nanobodies, as well as their successful characterization by ESI-MS (Figure 23B, Figure 24B). While the product had the strongest signal intensity, full conversion to the electrophilic species could not be achieved. This result is not surprising in the context of the sortase reaction, which is recognized as having a moderate to low catalytic efficiency, partly due to the low K_m of the LPETG motif (~8.7mM) [93]. As discussed previously, the presence of an electrophile on the peptide was not observed to significantly change the degree of incorporation of the electrophile during the sortase reaction (Figure 18). The exact sortase reaction conditions in these initial experiments varied from experiment to experiment, but generally used ~10uM glycine-containing nanobody, excess peptide (~300uM), and excess sortase (~40uM).

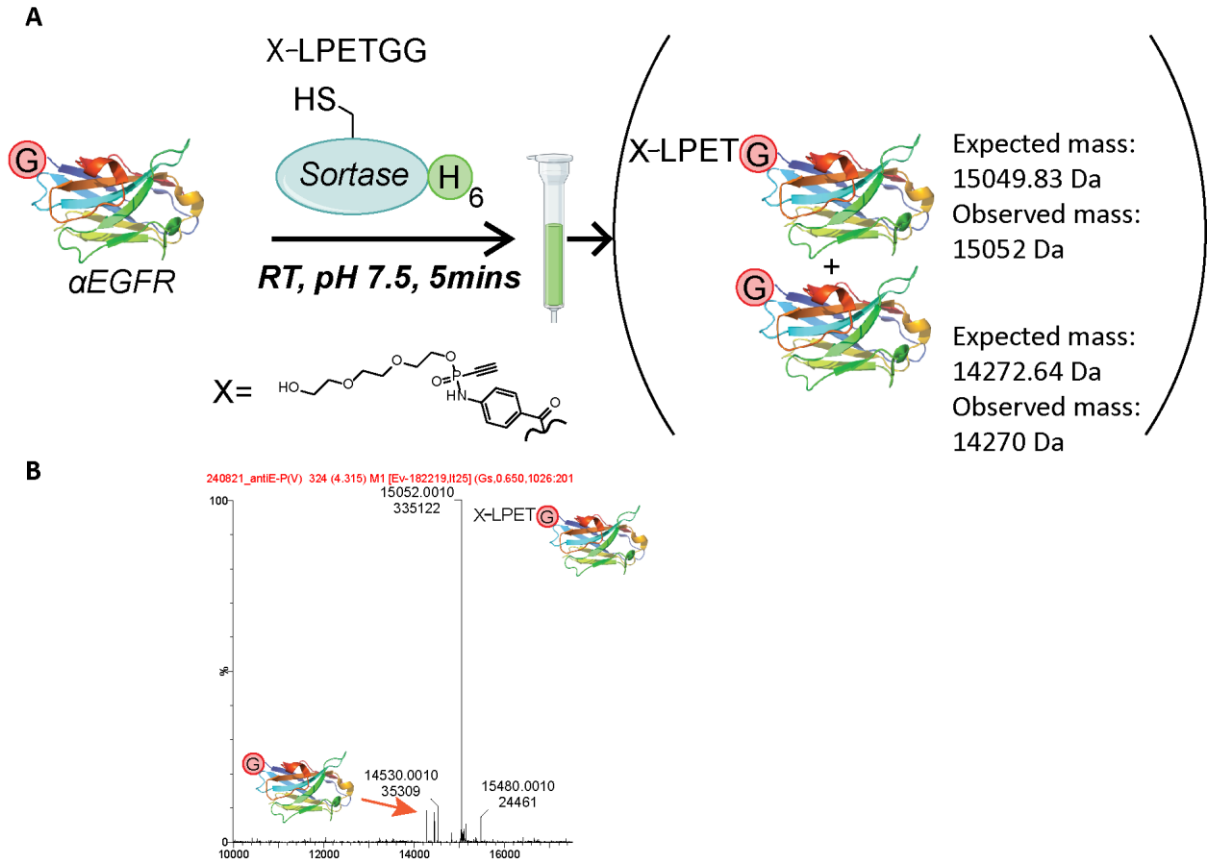


Figure 23: Installation of PEG_3-Alkyne-P(V) onto a nanobody using sortase with subsequent purification of the sortase. A: Reaction schematic with single-step purification. B: ESI-MS of purified product showing both converted and unconverted nanobody.

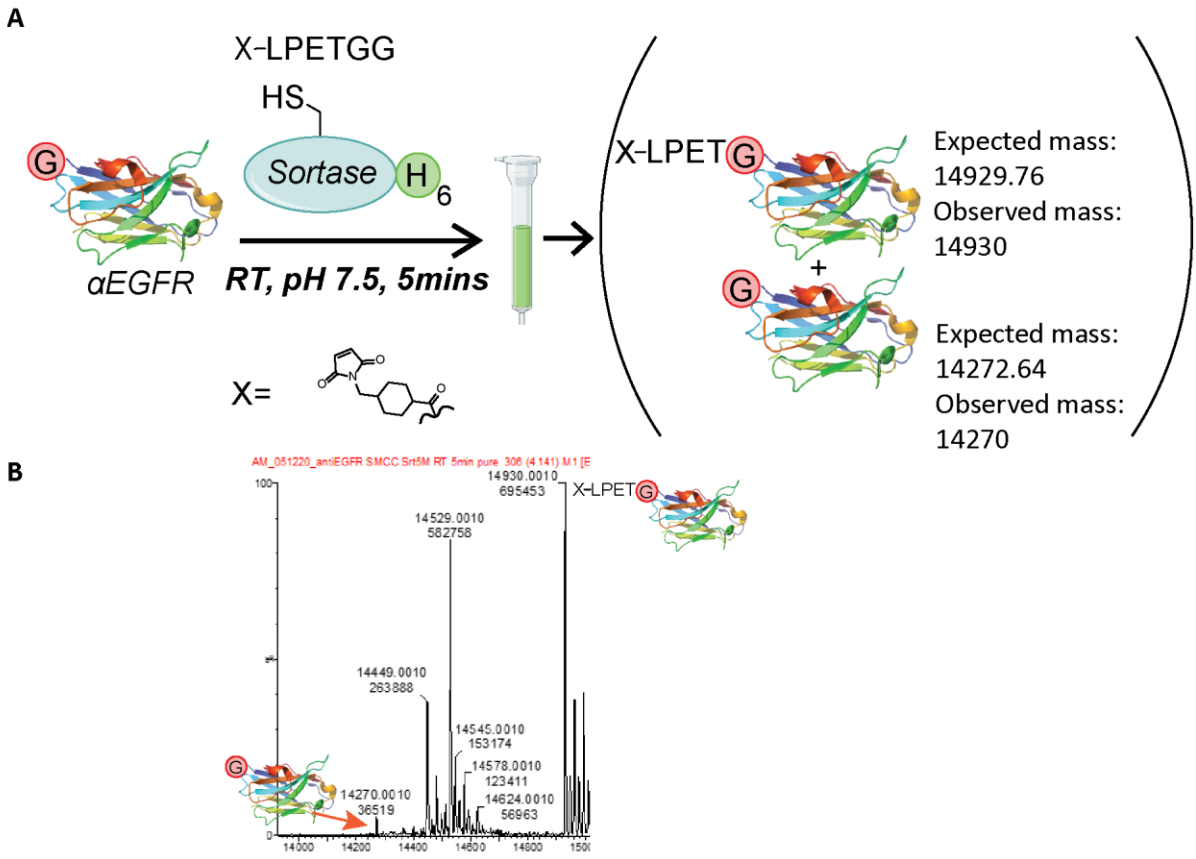


Figure 24: Installation of a maleimide onto a nanobody using sortase with subsequent purification of the sortase. A: Reaction schematic with single-step purification. B: ESI-MS of purified product showing both converted and unconverted nanobody.

With electrophilic nanobodies in hand, the focused turned towards optimizing the crosslinking component of the method to generate pure, crosslinked product. The first concern was making sure that the thiol on the cysteine-containing species was accessible during crosslinking. Cysteines often spontaneously oxidize to form disulfides, either within a protein or between two different cysteine-containing proteins [137]. Disulfide reduction can therefore be required to render the cysteines accessible to the sortase-installed electrophiles for chemoselective crosslinking, for example through thiol exchange with small molecule thiols like TCEP [138]. Indeed, experiments using the antiGFP nanobody derivative engineered with an artificial cysteine showed that crosslinking only occurred after reduction with 16eq of TCEP for 2 hours (Figure 25).

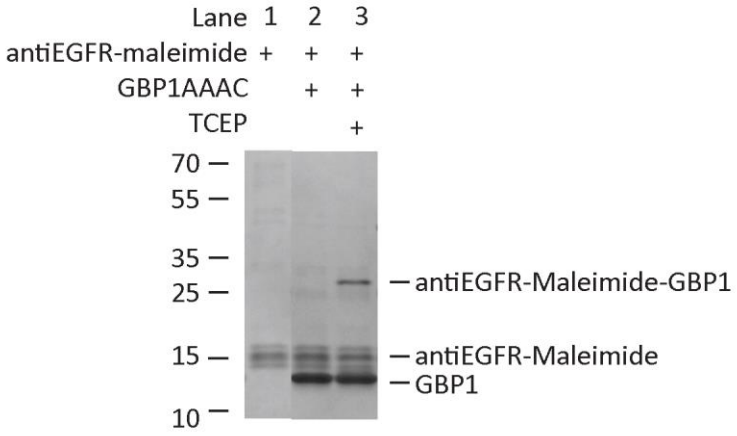


Figure 25: Reduction of thiol-containing antiGFP nanobody "GBP1" is required for chemoselective crosslinking.

Different equivalents of thiol-containing antiGFP nanobody relative to the electrophilic antiEGFR nanobody sortase product were tested in several reactions that were analyzed by SDS-PAGE with Coomassie staining. The Coomassie intensity of the crosslinked product band was quantified and compared between reactions to determine ideal conditions (Figure 26). The addition of 4-fold excess of cysteine-containing nanobody resulted in less crosslinked product than a 4-fold excess of electrophilic nanobody, suggesting that the main determinant in crosslinked product yield was the quantity of electrophilic nanobody product used. Indeed, using a larger amount of electrophilic nanobody (40uM) with only 0.5 equivalents of thiol containing nanobody resulted in the highest conjugate Coomassie intensity. Such a result is supported by previous ESI-MS data showing that the sortase reaction does not fully convert the nanobody to the electrophilic product. Combined with the results of this equivalency screen, it became apparent that the sortase reaction is converting at most 50% (but probably less) of the nanobody to the electrophilic product, and therefore at most 50% of the measured concentration of recovered nanobody sortase product is actually electrophilic. A strategy was therefore adopted to use sortase conditions and purification methods that optimized for the highest amount of recovered nanobody possible, and to then apply only half as much thiol containing protein during the crosslinking step. Based on these results, such a strategy would have negligible effect on overall product yields while simplifying the purification of the final product.

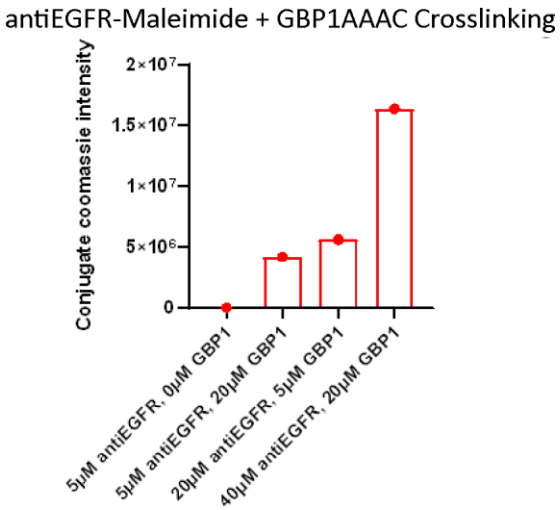


Figure 26: Effect of different equivalents of proteins in cysteine-selective crosslinking reaction on crosslinked product yields.

Specific targeting of the engineered cysteine during crosslinking was hypothesized to be challenged by the competing effects of the excess electrophile-containing peptide as well as peptide-modified sortase (Figure 28A). An experiment was designed to test this hypothesis using a bifunctional fluorescent and electrophilic peptide substrate that would enable tracking of the peptide over the course of the method (Figure 27). LPETGG peptides containing an N-terminal fluorescein linker with or without an N-terminal maleimide were installed onto the antiEGFR nanobody using the sortase and purified with gel filtration and IMAC (Figure 27A). The resulting products were reacted with a cysteine-containing antiGFP nanobody that was either used directly from the freezer or reacted for 2 hours with 16eq TCEP (reduced) to interrupt the spontaneously forming disulfide-linked nanobody dimers, rendering their engineered cysteines accessible to the maleimide. Fluorescein enabled tracking of the peptide and peptide-modified products via SDS-PAGE with in-gel fluorescence, confirming that the antiEGFR nanobody as successfully modified with both peptide substrates (Figure 27B). In-gel fluorescence showed a single major fluorescent band for the antiEGFR nanobody “1” with no electrophile in the presence and absence of the thiol-containing nanobody, indicating no reaction. Interestingly, the maleimide-containing nanobody “2” showed high-MW species even before the addition of the thiol-containing nanobody. These unspecific electrophile-dependent multimers will be discussed in a later section (3.1.6). As expected, a clear product band was only observed in the presence of both the electrophilic antiEGFR nanobody “2” and the reduced, cysteine-containing antiGFP nanobody “3”. Another

band appeared only in the reaction with these species, with a MW just slightly larger than the monomeric antiGFP nanobody. In-gel fluorescence confirmed that this band was peptide-modified antiGFP nanobody, and the fact that it only occurred after TCEP reduction and with the maleimide-containing peptide strongly suggests that the modification occurred at the engineered cysteine. From this result it was concluded that excess electrophilic peptide present in the crosslinking reaction despite the gel filtration partially sequesters the thiol-containing partner, most likely lowering the yield of crosslinked product and necessitating a more stringent peptide removal process. The large excess of peptide used to drive the sortase reaction to completion was calculated to exceed the binding capacity of the 0.5ml ZEBA spin columns used for the cleanup before crosslinking, resulting in incomplete removal of the peptide by gel filtration and the observed undesired thiol sequestration. Nevertheless, the crosslinked product formation was shown to clearly depend on both the presence of a cysteine-selective electrophile and a reduced, accessible thiol-containing partner, providing compelling evidence that the intended cysteine-selective crosslinking was occurring.

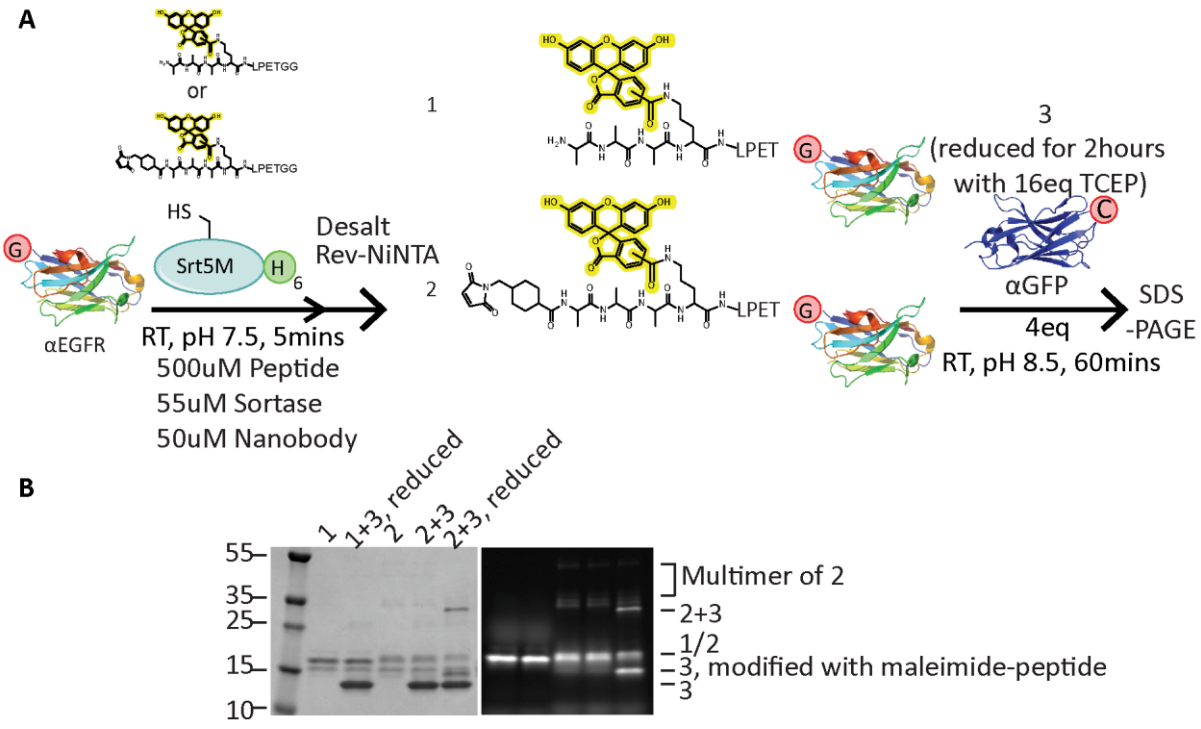


Figure 27: Excess electrophilic peptides react with reduced cysteine-containing proteins during crosslinking. A: Reaction schematic. B: SDS-PAGE with Coomassie staining (left) and in-gel fluorescence (right).

The successful sortase reaction purifications described in Figure 22, Figure 23 and Figure 24 by IMAC lead to the idea that the problems of unspecific crosslinking (Figure 21) and thiol sequestration (Figure 27) could be elegantly solved in a single step by further application of

the purification strategy to encompass the excess electrophilic peptide. A solution was envisioned to enable removal of both the excess peptide and the sortase in a single chromatography step by installing histidine purification tags on both the sortase and the peptide, inspired by previous work using His-tagged sortase variants [95] and LPXTG peptides with His tags downstream of recognition sequence [97]. Following the transpeptidation reaction, a single NiNTA column could then be used to selectively target both species (Figure 28B).

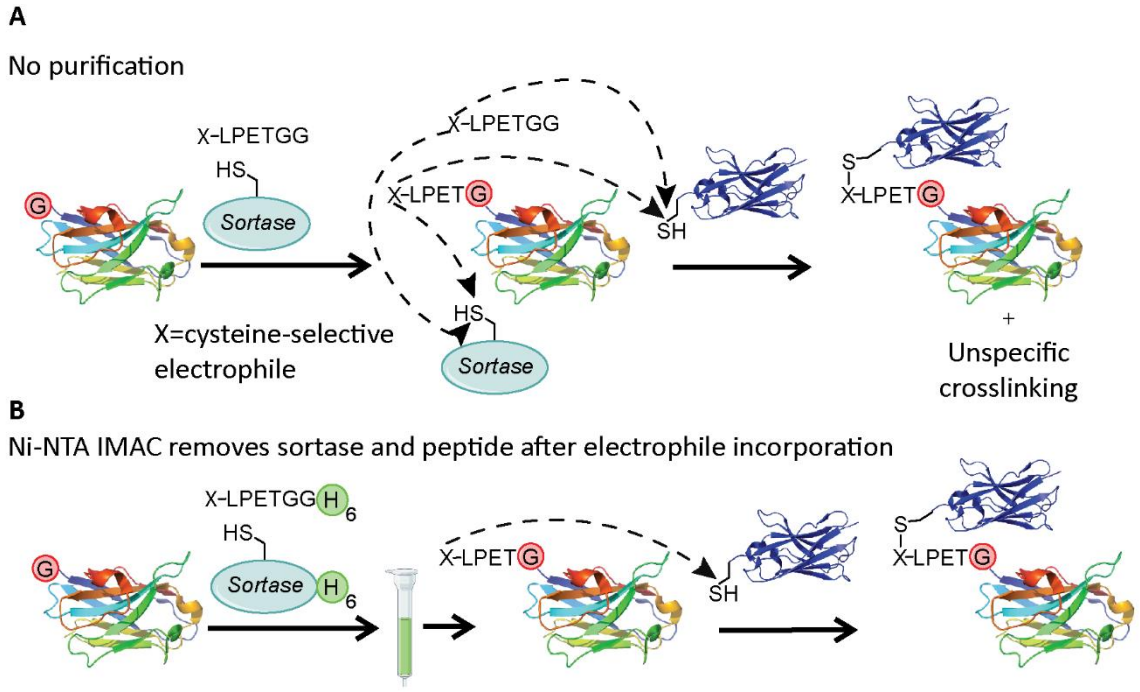


Figure 28: A: Rational for purification of sortase reaction prior to crosslinking. B: Strategy for single-step removal of peptide and sortase following transpeptidation.

Histidine-tagged LPETGG peptides containing an *N*-terminal PEG_2-alkyne-phosphonamidate (Figure 29A), PEG_3-alkyne-phosphonamidate (Figure 30A) or maleimide (Figure 31A) were therefore synthesized and characterized as sortase substrates for *N*-terminal nanobody labeling using the optimal conditions previously described (Section 3.1.3.2). All electrophilic peptides could be installed successfully, and all electrophilic nanobody intermediates could be purified and characterized by ESI-MS, which showed either no or only trace amounts of detectable sortase or peptide relative to the nanobody after purification (Figure 29B) (Figure 30B) (Figure 31B).

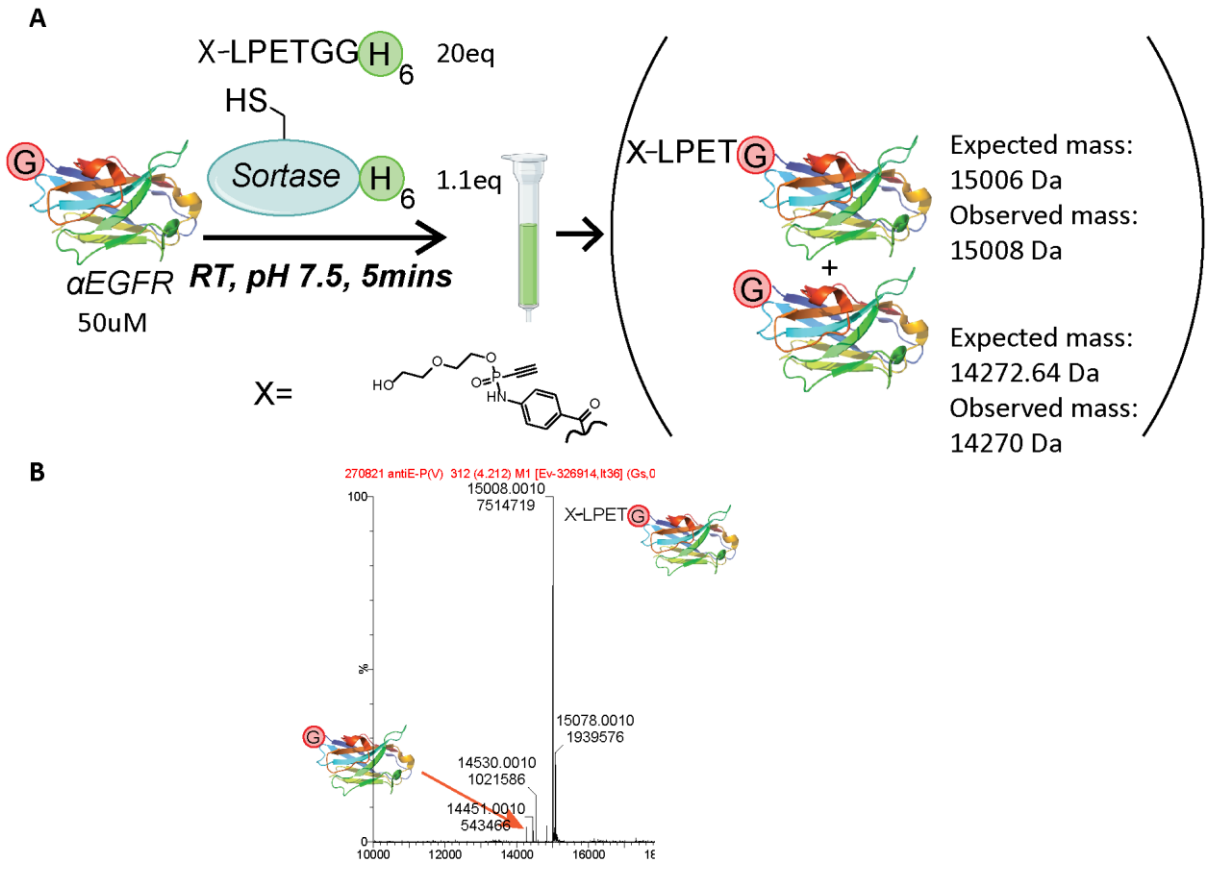


Figure 29: Installation of PEG_2-Alkyne-P(V) onto a nanobody using sortase with subsequent purification of the sortase and peptide. A: Reaction schematic with single-step purification. B: ESI-MS of purified product showing both converted and unconverted nanobody.

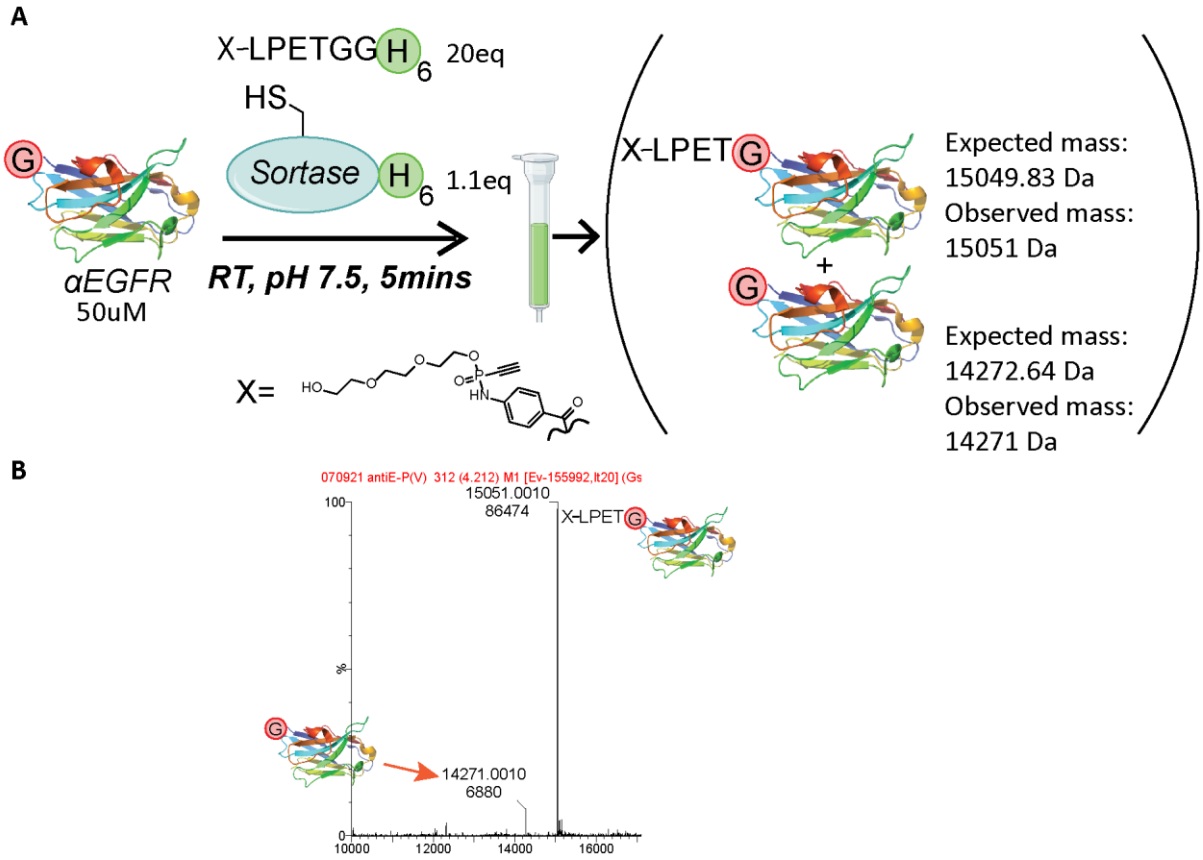


Figure 30: Installation of PEG_3-Alkyne-P(V) onto a nanobody using sortase with subsequent purification of the sortase and peptide. A: Reaction schematic with single-step purification. B: ESI-MS of purified product showing both converted and unconverted nanobody.

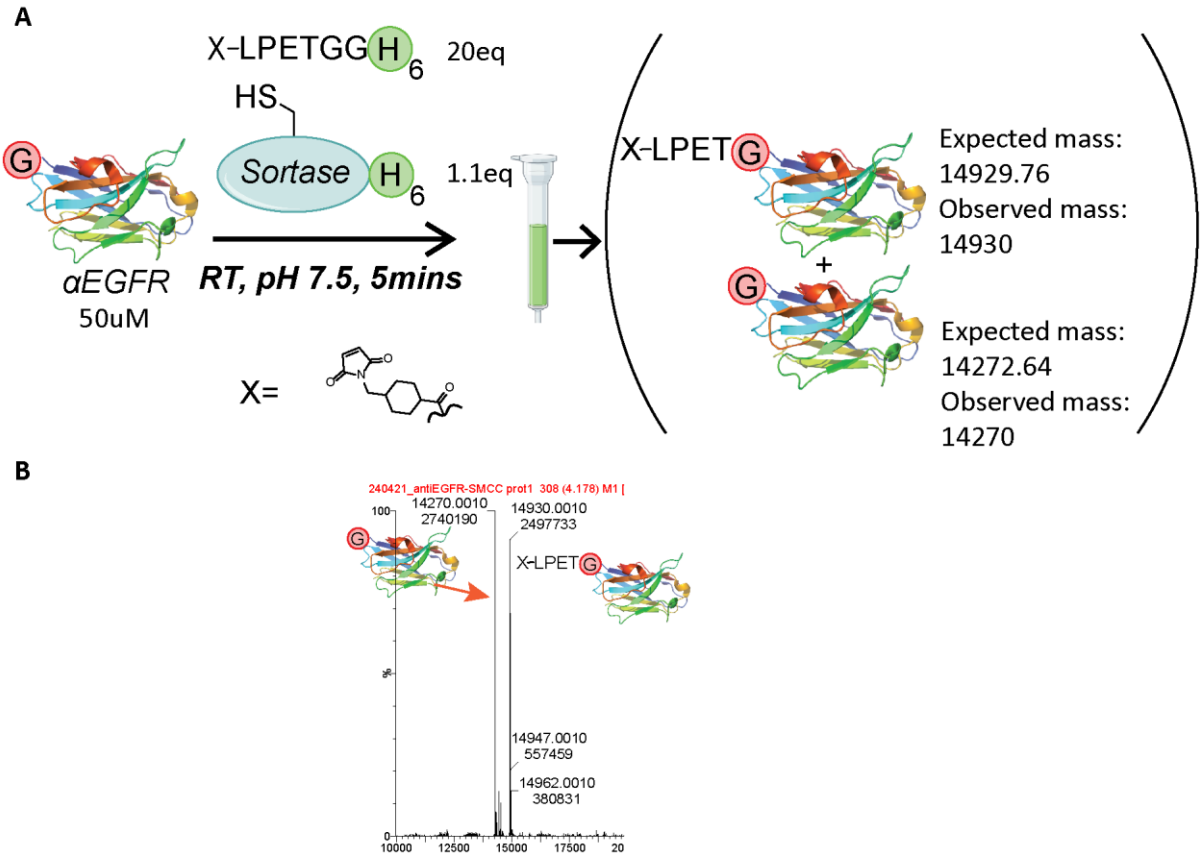


Figure 31: Installation of maleimide onto a nanobody using sortase with subsequent purification of the sortase and peptide. A: Reaction schematic with single-step purification. B: ESI-MS of purified product showing both converted and unconverted nanobody.

A comparison between the single-step IMAC-based peptide and sortase removal protocol and the two-step IMAC and gel filtration protocol was performed, to determine which resulted in better overall crosslinked product yields. Each sortase reaction was performed under the same conditions just described, with the only difference being the peptide used and the sortase purification method employed. In protocol 1, the new single-step IMAC purification was performed on an antiEGFR nanobody "1" with the maleimide-containing His-tagged LPETGG peptide, while in protocol 2 the old two-step IMAC and SEC purification protocol was performed with the non-His tagged peptide (Figure 32A). Both products "2" were diluted to 40uM in an 11ul reaction volume and mixed with 0.5eq of thiol-containing antiGFP nanobody, then subjected to crosslinking for 30 minutes at room temperature before being boiled in SDS-PAGE buffer. Crosslinking reactions were analyzed by SDS-PAGE with Coomassie signal intensity quantification of the band corresponding to the crosslinked product "3" (Figure 32B). SDS-PAGE enabled unambiguous determination of the product band for Coomassie quantification, which showed a greater than 2-fold increase in product intensity for the new,

single-step purification method (protocol 1) compared to the previous method (protocol 2). A double band was observed near the product MW for both protocols, the identity of which will be discussed later (Section 3.1.6).

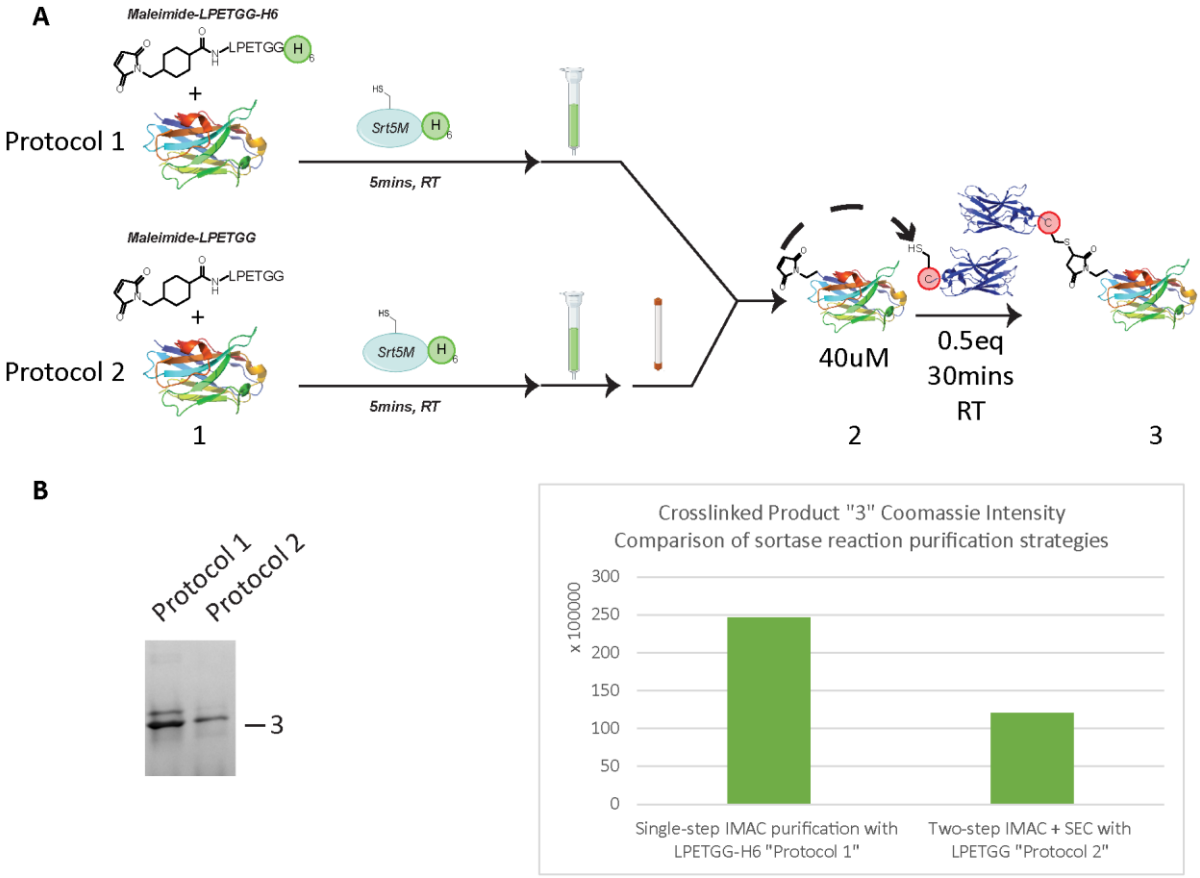


Figure 32: Comparison between single-step and two-step sortase reaction purification and the effect on crosslinked product yields. A: Reaction schematic. B: SDS-PAGE with Coomassie staining of the SEC-purified crosslinked product for both protocols (left) and Coomassie intensity-based quantification of the band corresponding to the crosslinked product "3".

To further optimize the crosslinking conditions and to generate a functional construct, a histidine tag-containing LPETGG peptide functionalized with a maleimide was applied to the generation of an antiEGFR nanobody-GFP conjugate (Figure 33A). Time course measurements were done during the crosslinking step, which showed that the maximum conjugate yield was attained after approximately 5 minutes incubation, but no product loss was observed for up to 30 minutes, and so reactions were typically incubated overnight (Figure 33B). Purity and identity of the product after SEC were confirmed by SDS-PAGE and ESI-MS (Figure 33C). To confirm the functionality of both proteins following the conjugation protocol, fluorescence microscopy was performed using EGFR+ (A549) and EGFR- (HEK293T) cells (Figure 33D). Fluorescence was observed in the GFP channel for cells treated with the construct, suggesting

that GFP remained functional in the final product. Cell staining by the construct was localized to the cell membrane and was observed for the EGFR+ cells but not for the EGFR- cells, indicating that the nanobody retained its specificity in the final product.

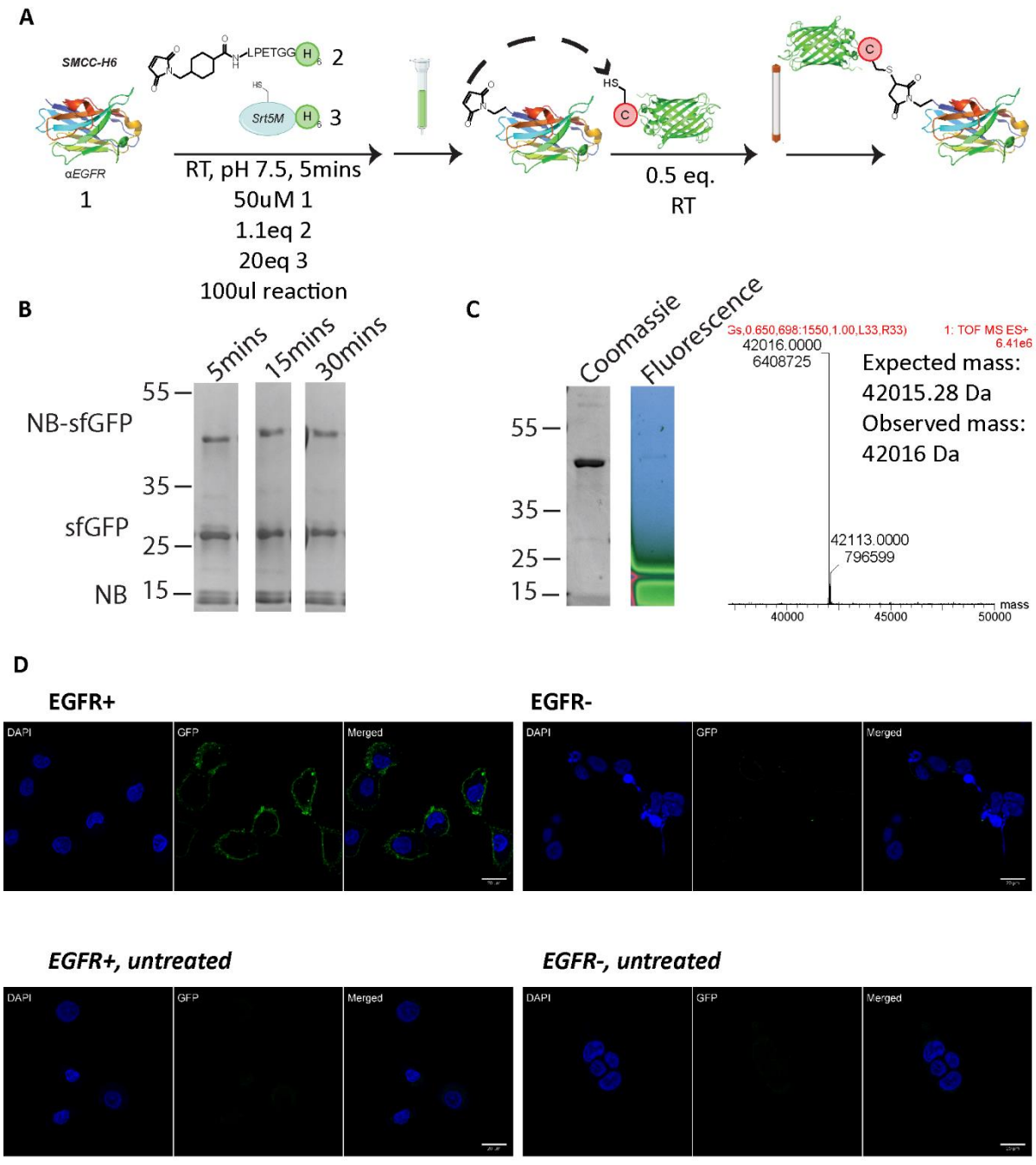


Figure 33: Nanobody-Maleimide-GFP conjugation using optimized conditions. A: Reaction schematic. B: Time course measurements from the crosslinking step subjected to SDS-PAGE and Coomassie staining. C: Crosslinked construct, purified by SEC and subjected to SDS-PAGE and ESI-MS. D: Fluorescence microscopy using purified construct with EGFR+ (A549) and EGFR- (HEK293T) cells.

Using the optimized sortase reaction conditions at small (100ul) volumes followed by the described single-step IMAC purification and crosslinking for 18 hours initially produced overall

crosslinked product yields 2.1% for the maleimide functionality and 1.2% for the phosphonamidate (Figure 34).

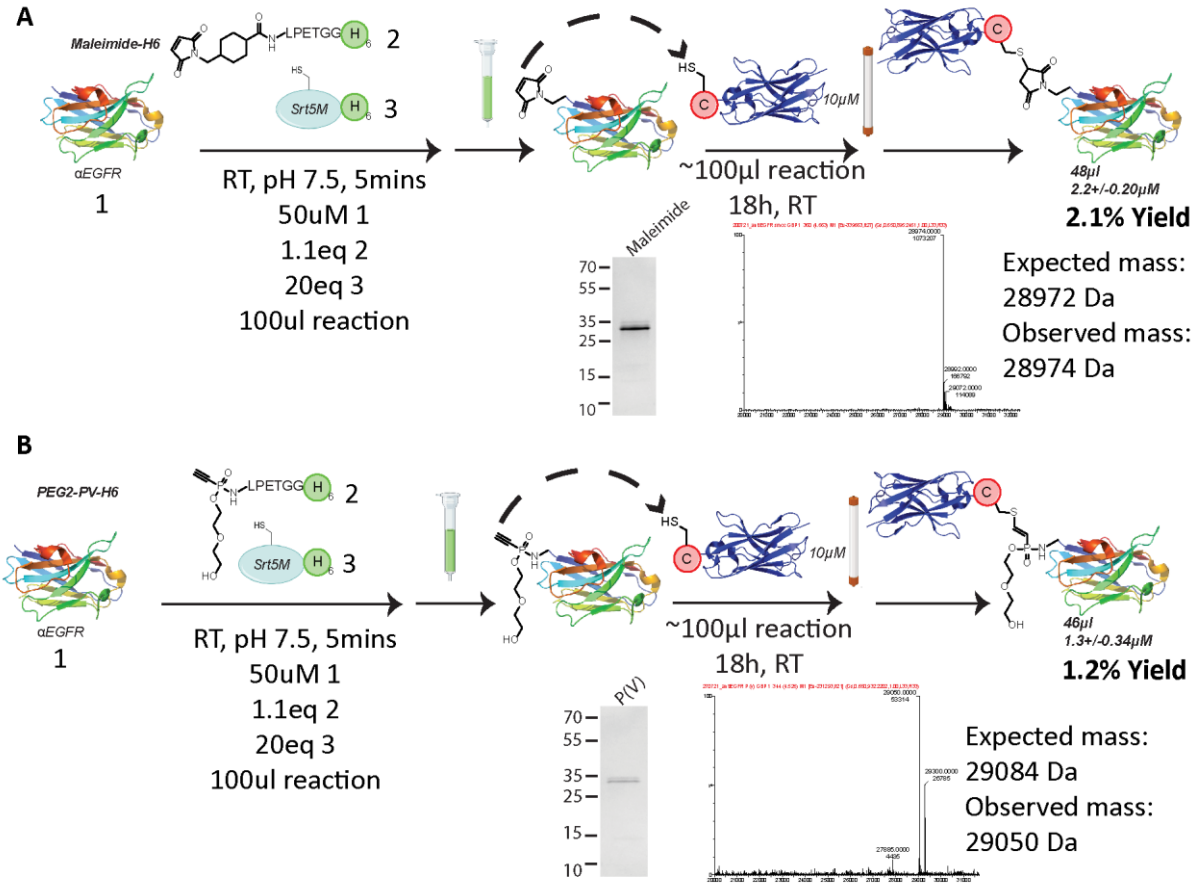


Figure 34: Initial yield determination experiments using A: maleimide and B: PEG2-P(V) as the cysteine-selective electrophile. SDS-PAGE with Coomassie staining and ESI-MS analysis were performed on the final, purified crosslinked product.

Yields of the crosslinked product were determined as a function of the quantity of nanobody input in the sortase reaction, with concentrations determined spectrophotometrically via measurement at 280nm (Equation 1).

$$\%\ yield = \frac{[Crosslinked\ Construct]}{[Nanobody\ in\ Sortase\ reaction]}$$

Equation 1: Calculating the crosslinked product yield.

Further optimization was performed to improve the yields. First, the sortase reaction was scaled up to 700ul from 100ul to reduce the proportion of material lost during purification steps. In addition, the equivalents of TCEP used to disrupt the naturally occurring disulfide bonds between cysteines in the cysteine-containing conjugation partner had to be adjusted. Lowering the equivalents of TCEP for the reduction prior to crosslinking and then subjecting

the reaction mixture to a large TCEP reduction immediately prior to SEC purification significantly improved the overall yields to between 4 and 6% depending on the electrophile (Figure 35). The final conditions used combined the 5-minute optimized sortase reaction conditions purified in a single IMAC step with an overnight crosslinking reaction to ~0.5eq of a TCEP-reduced thiol-containing nanobody before a final SEC purification to yield the pure product (Figure 35A). Characterization of the pure products was performed using SDS-PAGE and ESI-MS (Figure 35B). Yields were calculated from spectrophotometric measurements performed at 280nm using Equation 1.

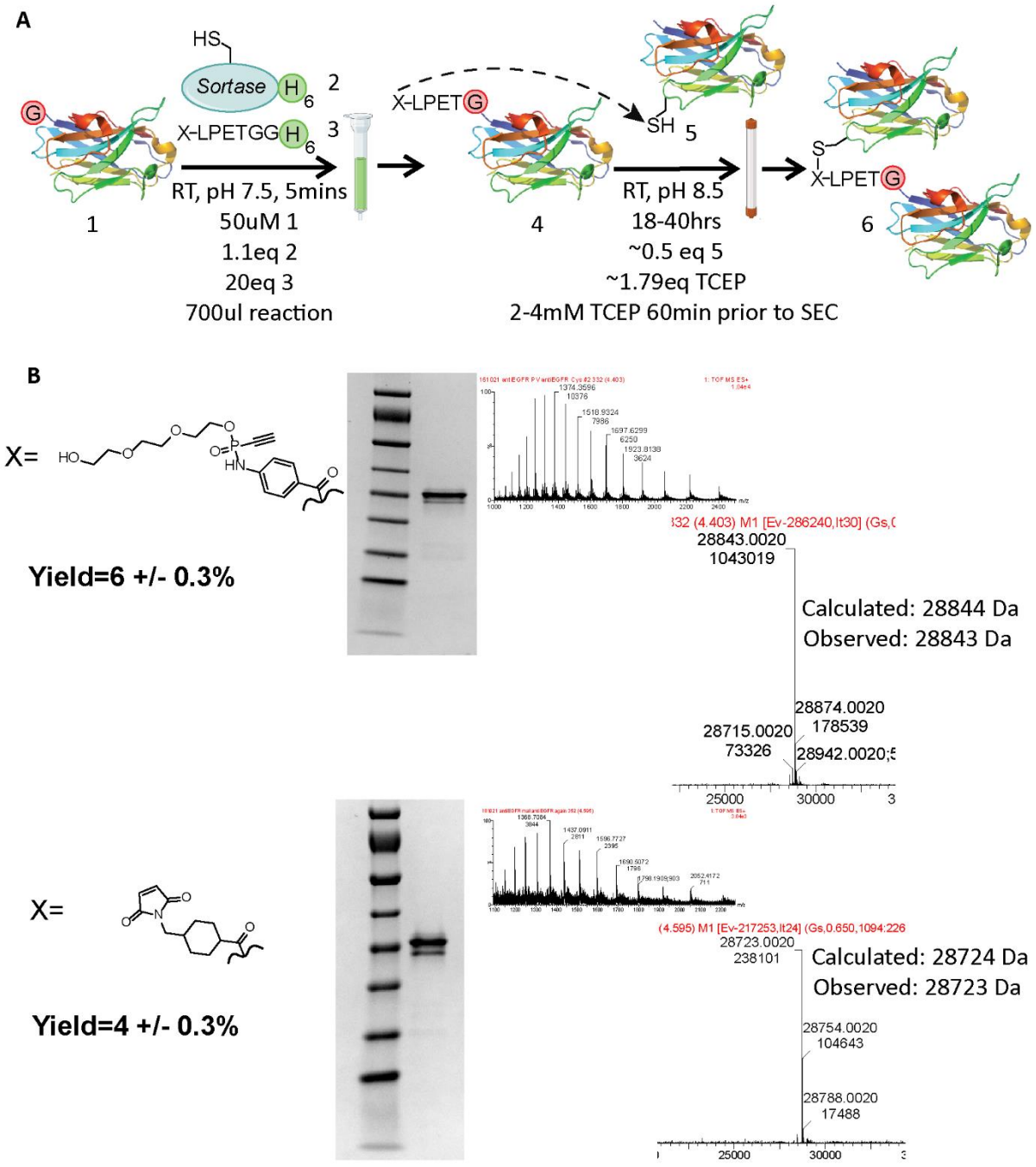


Figure 35: Crosslinking yield determination. A: Schematic for crosslinking protocol. B: Yield for nanobody dimer 6 using PEG3-Alkyne-P(V) and Maleimide as electrophiles, respectively. Characterization of the final crosslinked product by SDS-PAGE with Coomassie staining, as well as intact protein ESI-MS with both fragmentation pattern and deconvoluted spectra (right) displayed.

Although the desired product is the only species clearly identified by ESI-MS, multiple SDS-PAGE bands can be seen in the purified product formed during the experiment portrayed in Figure 35. Given the stringent IMAC purification undertaken to remove the sortase reaction components prior to crosslinking, it was evident that the identity of the unwanted band corresponded to a species lacking a HIS-tag, i.e. a nanobody. The MW of the unknown band

closely resembled the desired nanobody-linker-nanobody dimer **6**, which lead to the hypothesis that the species corresponded to a nanobody dimer likely resulting from installation of the reactive electrophile. An experiment was therefore devised using the bifunctional fluorescent and electrophilic peptides to try and reproduce the nanobody dimer after installation of the electrophile (Figure 36). Maleimide installation successfully reproduced the expected nanobody dimer for the antiEGFR and antiGFP_enhancer nanobodies, but not for the antiGFP_minimizer. Further investigation revealed a likely mechanism for this dimer formation that depends on the nanobody internal disulfide bond (discussed further in the next chapter, 3.1.6), which suggests that the minimizer nanobody expressed under these conditions possesses a more resilient or faster-forming internal disulfide. The phosphonamidate resulted in neither an alkylated sortase nor a nanobody dimer, which was expected given the short reaction time of only five minutes. This experiment confirmed that the installation of cysteine-reactive electrophiles on the nanobodies used in this study resulted in the undesired formation of a nanobody-nanobody dimer of a similar size to the desired nanobody-nanobody crosslinking product and therefore likely explains the double band observed in the purified products in Figure 35. A further discussion on this topic of electrophile-dependent nanobody multimerization can be found in the next chapter.

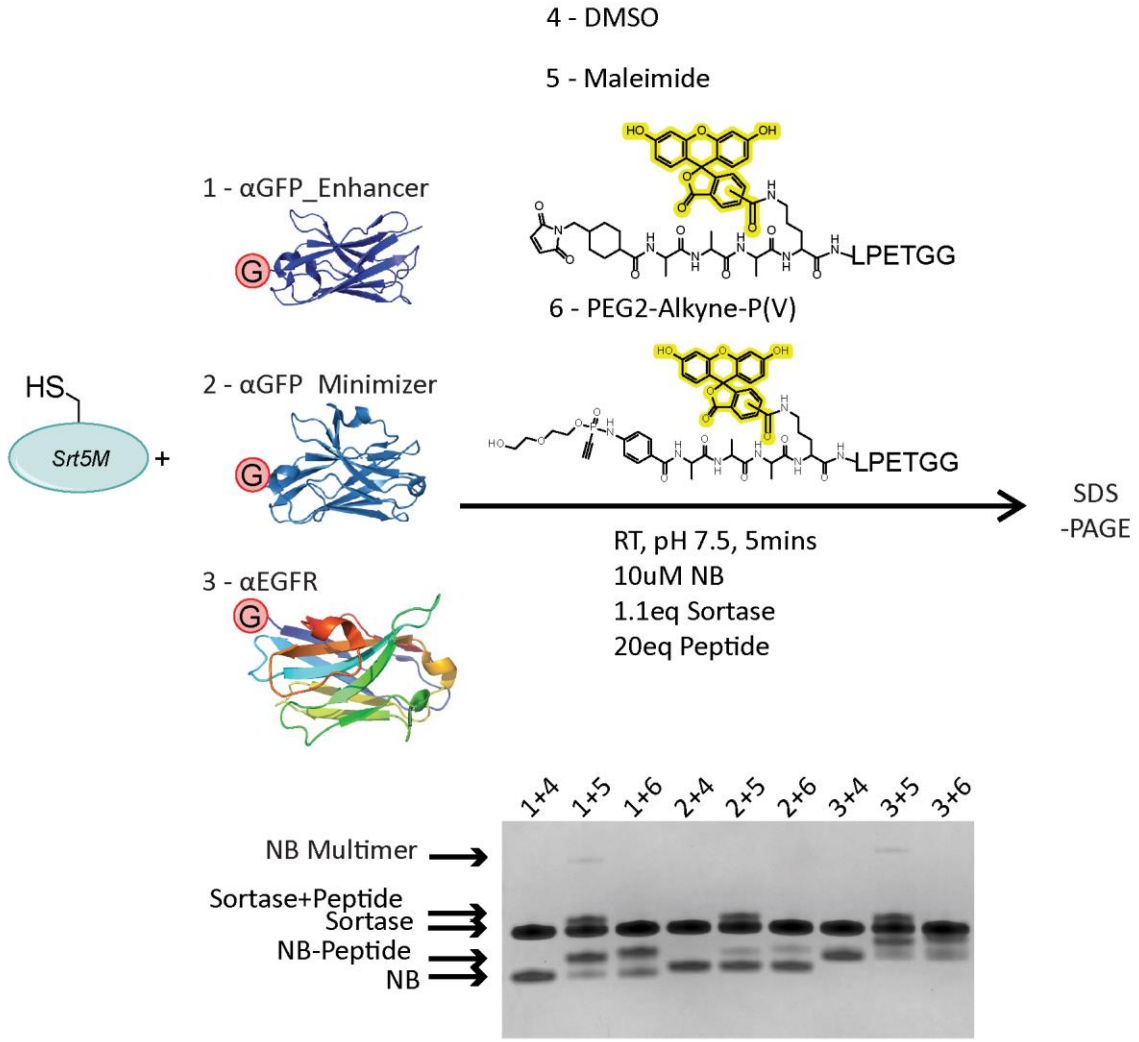


Figure 36: nanobody multimerization observed five minutes after installation of the maleimide.

3.1.6 Multimer formation

SDS-PAGE analysis of the purified crosslinking products, as well as previously discussed experiments analyzing the electrophilic nanobody intermediates revealed the formation of multiple high-MW species even without the addition of a thiol-containing species (Figure 35B and others). These species have been described as “multimers” and represent a cysteine electrophile-dependent effect observed after installation onto a cysteine-containing protein like a nanobody (although nanobody cysteines are typically engaged in internal disulfide bonds). To investigate this multimerization further, bifunctional fluorescent and electrophilic peptide were synthesized and employed to determine the factors that lead to formation of the unwanted species (Figure 37A). The installation of a maleimide onto the GBP1 nanobody using the optimized five-minute sortase protocol resulted in the formation of a high MW

species, whereas the P(V) and no-electrophile peptides did not (Figure 37B). Fluorescence signal from the species indicated it the presence of the peptide, and the gel shift indicated that it was likely a nanobody multimer. The fact that the species only formed in the presence of a highly reactive electrophilic peptide strongly suggested that it was the result of cysteine-selective crosslinking. A P(V) multimer was not observed in this case likely because the samples were only allowed to incubate for five minutes prior to boiling and analysis by SDS-PAGE, and P(V) cysteine targeting requires longer reaction times [55].

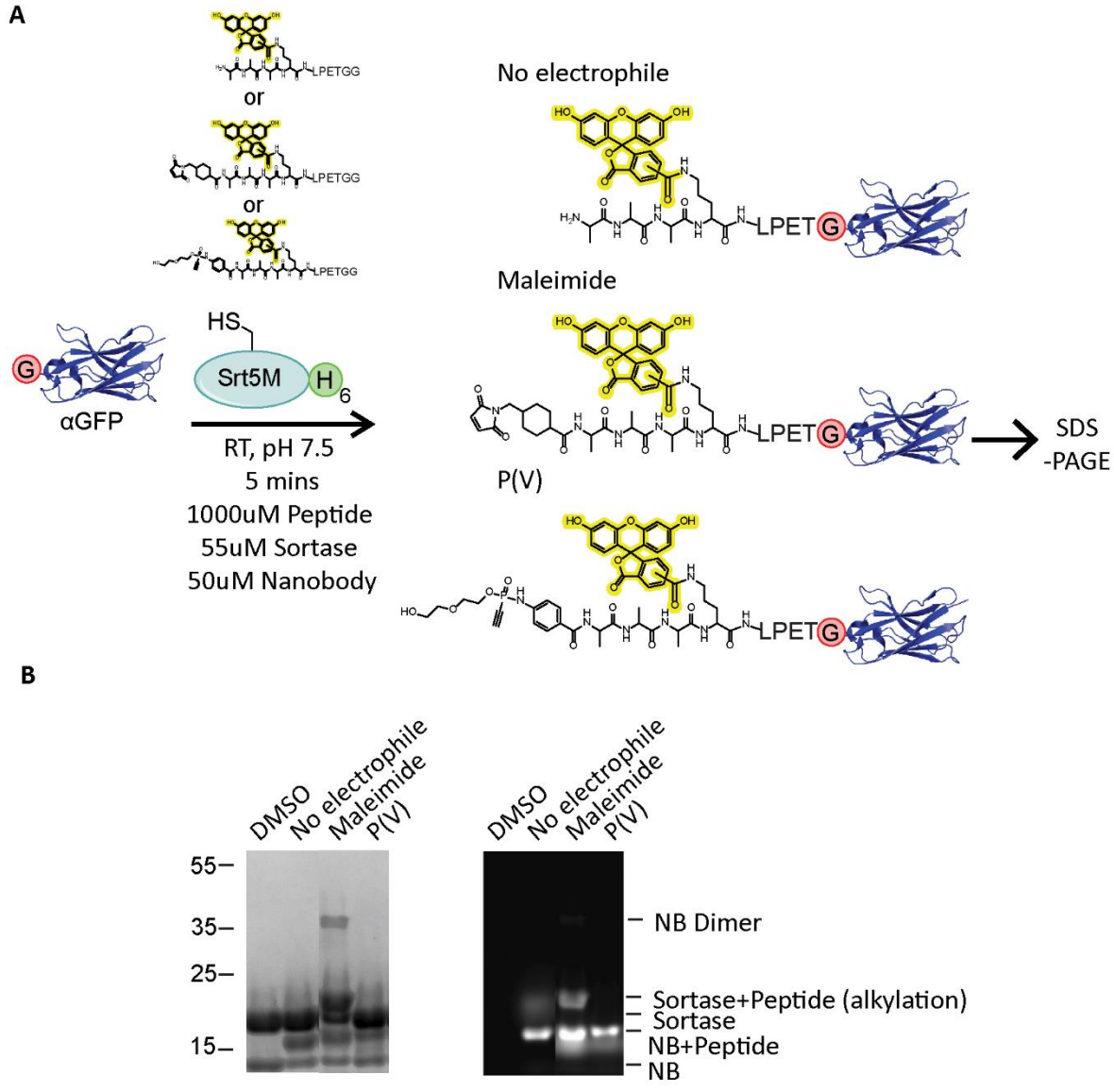


Figure 37: Multimer formation after electrophile incorporation. A: reaction schematic with bifunctional fluorescent/electrophilic peptides. B: SDS-PAGE with Coomassie (left) and in-gel fluorescence (right) following the sortase reaction depicted above.

Further investigation on the unwanted multimer was performed using ESI-MS following incorporation of the non-fluorescent maleimide peptide onto the GBP1 nanobody and subsequent purification (Figure 38A). SDS-PAGE analysis showed the formation of multiple high MW species after installation of the maleimide (Figure 38B). ESI-MS conclusively showed the presence of unmodified and modified nanobody, as well as a modified nanobody dimer, strongly suggesting that the high MW species is the result of electrophile-enabled crosslinking to other nanobodies (Figure 38C). Given the reactivity of the maleimide towards thiol nucleophiles, a mechanism involving the internal nanobody disulfide was imagined to be responsible for multimer formation. Nanobodies were chosen for this project based on the presumed non-reactivity of their oxidized internal disulfides, so such an effect was unanticipated,

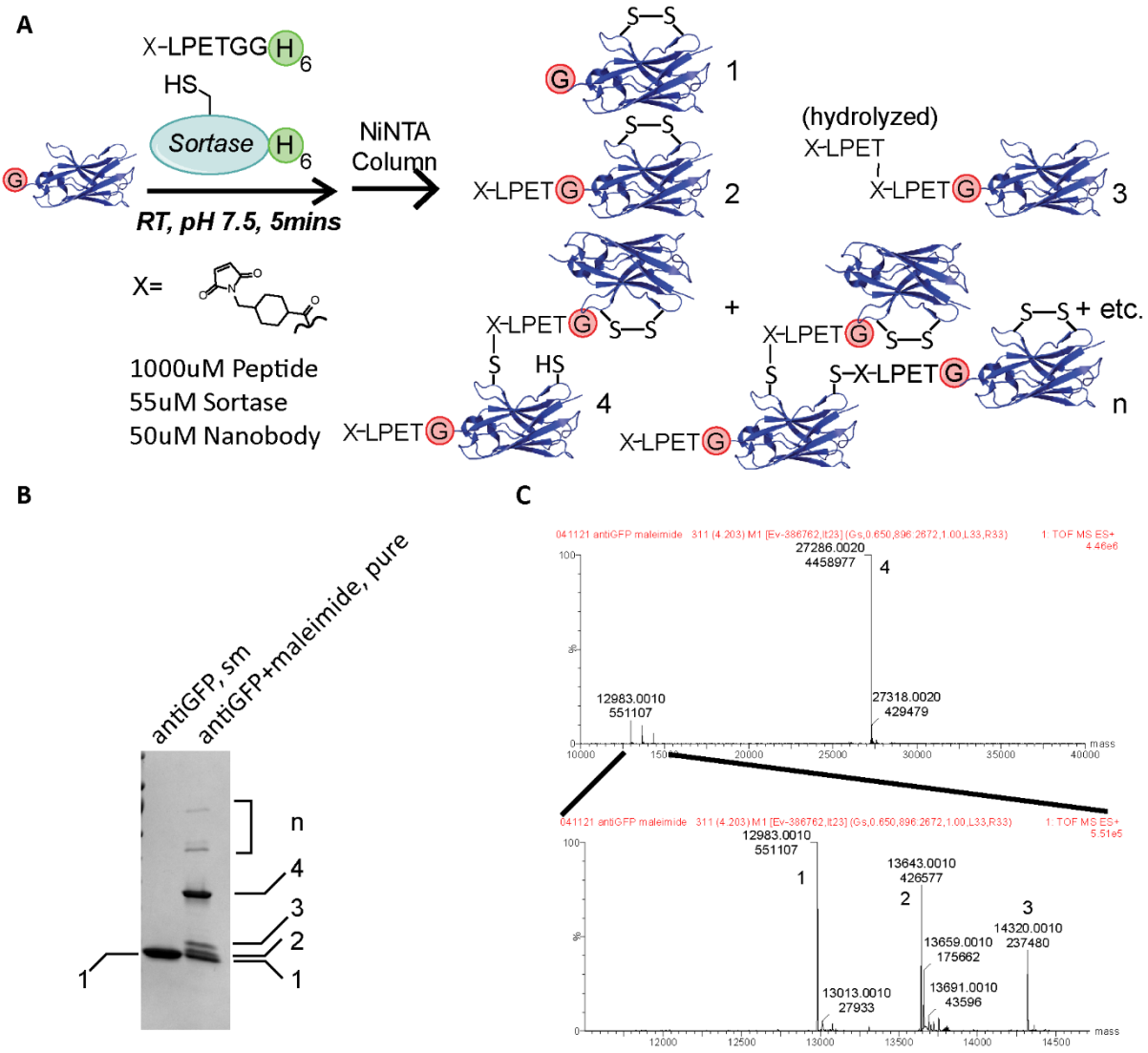


Figure 38: Multimer investigation of purified sortase product via ESI-MS. A: Reaction schematic. B: SDS-PAGE and Coomassie staining of starting material (SM) and purified sortase product. C: ESI-MS deconvoluted spectrum of purified sortase product.

ESI-MS measurements of two different nanobodies, both expressed without a PelB leader sequence, showed the presence of multiple +32Da mass adducts, with the number of adducts corresponding exactly with the number of cysteines present (Figure 39A, B). The antiEGFR nanobody contained only two cysteines and demonstrated two +32 adducts, while the GBP4 nanobody contained four cysteines and demonstrated four +32 adducts. The mass adducts matched the published mass for sulfinic acid, a post translational modification reported in bacteria that converts cysteine residues into a stable oxoform (Figure 39C) [139] [140]. Sulfinic acids are known to be reactive to Michael acceptors such as the maleimide, suggesting that these sulfinic acid-containing nanobodies were reacting with incorporated electrophiles to form the observed multimers (Figure 39D) [141]. In addition, the resulting sulfonyl succinimide was shown to be unstable at even slightly basic pHs after 60 minutes, explaining

why only thiosuccinimides and not sulfonyl succinimides were observed by ESI-MS (Figure 38). From this data, and the recollection that these nanobodies were expressed without a PelB leader sequence, it was concluded that prolonged incubation in the cytosol without periplasmic export leads to oxidation of the nanobody cysteines forming highly reactive sulfinic acid moieties instead of a disulfide(s) (Figure 40), which in turn leads to multimerization after installation of the electrophiles used for crosslinking in this study.

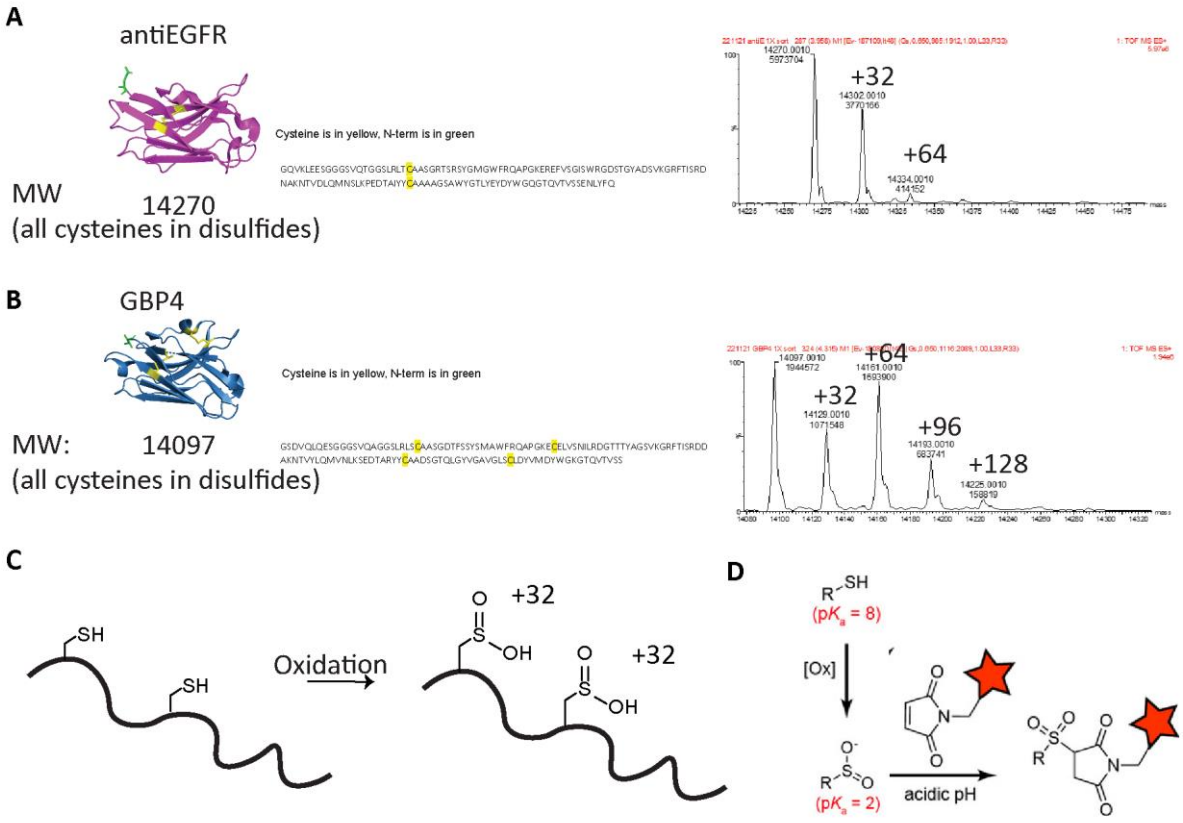


Figure 39: Nanobody disulfide oxidation. A: antiEGFR nanobody PDB file (4KRL) rendered using Pymol, amino acid sequence with two cysteines highlighed, and ESI-MS showing two +32 mass adducts. B: GBP4 nanobody PDB file (4GBP) rendered using Pymol, amino acid sequence with four cysteines highlighted, and ESI-MS showing four +32 mass adducts. C: Schematic showing hypothesized oxidation of nanobody cysteines to form sulfinic acids. D: Reactivity of sulfinic acids towards maleimides [141].

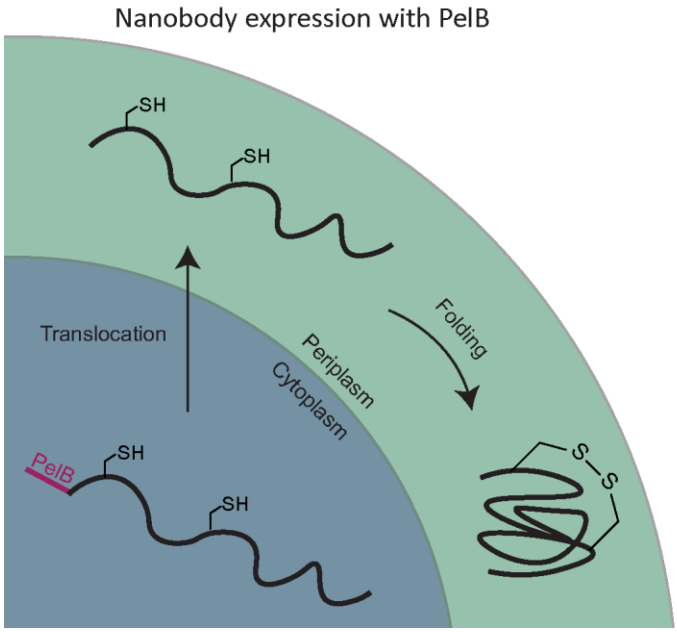


Figure 40: PelB translocation is hypothesized to enable proper disulfide formation, while its absence leads to stable cysteine oxoforms.

3.1.7 Applications for generating useful protein constructs

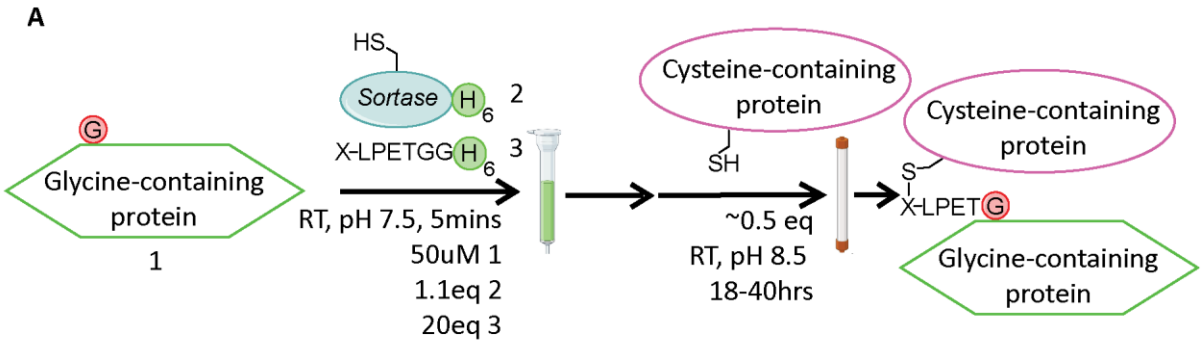


Figure 41: Optimized method conditions without TCEP reduction prior to SEC purification of the final product.

We sought to apply this method towards generating protein constructs that could be verified to retain the function of their component parts. The reaction conditions used were those found to be optimal according to the previous discussion (section 3.1.5), but the TCEP reduction prior to SEC purification was not performed for these constructs as this step was a later innovation motivated by the need for ultra-pure products during final yield determinations (Figure 42). The final products therefore contained varying proportions of unreacted monomeric nanobody (as well as the multimer discussed in section 3.1.6). First, a bispecific nanobody construct was generated by installing a maleimide onto an antiEGFR nanobody (multi-colored) and conjugate it to a cysteine-containing antiGFP nanobody, GBP1

(blue) followed by SEC purification (Figure 42A). The overall yield calculated using Equation 1 was 9.4%, but this was based off of a spectrophotometric measurement of a sample that was shown to be impure by SDS-PAGE (Figure 42B) and therefore does not accurately reflect the yield of the desired product. ESI-MS of the purified construct (Figure 42C) confirmed the identity of the bispecific nanobody. A microscopy assay involving cells both positive and negative for EGFR to be treated and untreated with GFP was devised to confirm that both nanobodies retained their target specificity following the protocol (Figure 42D). Microscopy showed clear membrane staining for the EGFR+ cells in the presence of GFP and the construct (Figure 42E) while the EGFR- cells showed no membrane staining even in the presence of construct and GFP (Figure 42F), confirming the efficacy of both nanobodies following the application of the method.

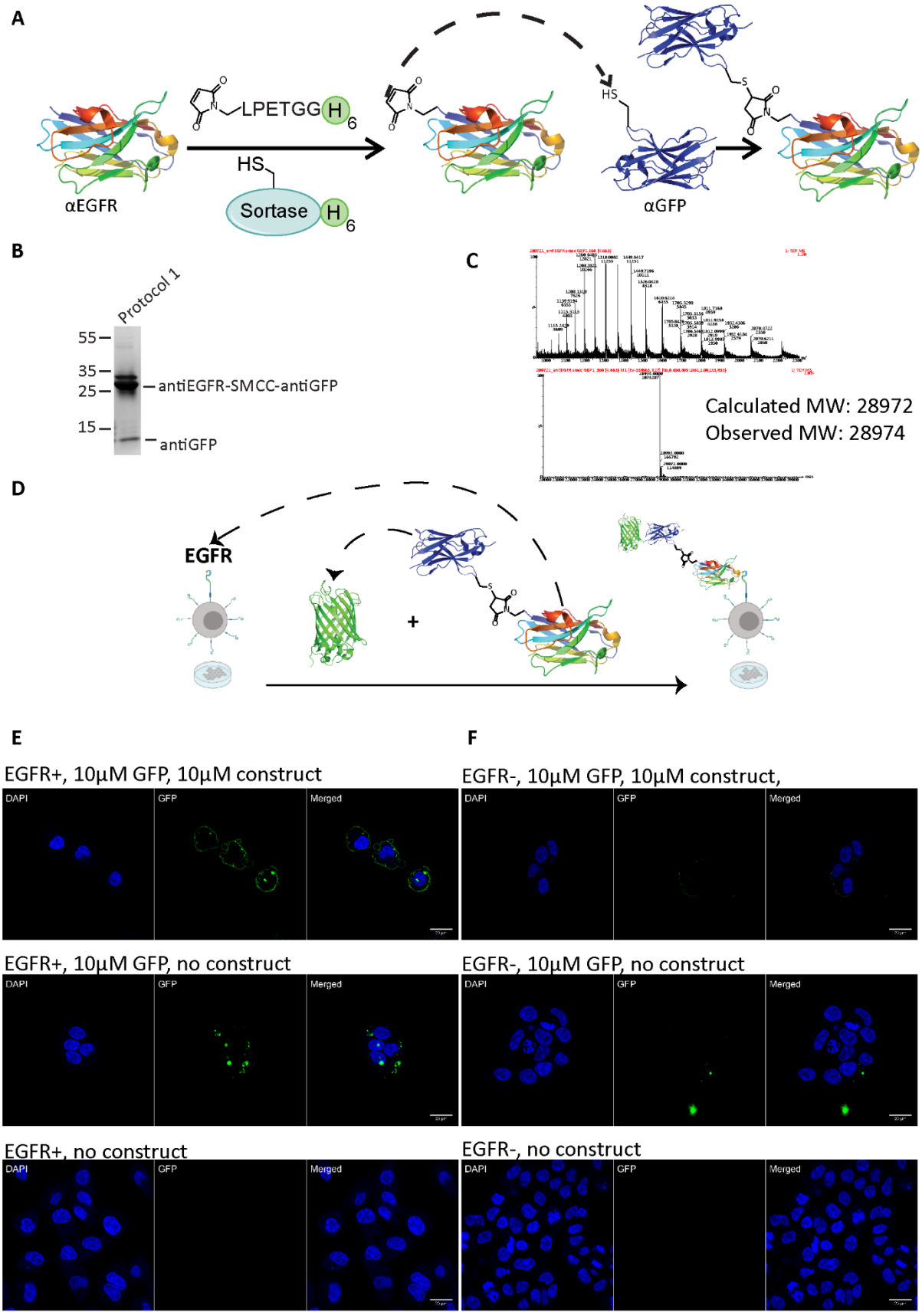


Figure 42: Bispecific nanobody generation. A: Reaction schematic for generating construct. B: SDS-PAGE of purified construct, Coomassie stained. C: ESI-MS of purified construct. D: Schematic of fluorescence microscopy assay showing targeting of GFP

to EGFR+ cells. E: Multichannel confocal microscopy images of EGFR+ (A549) cells treated with the construct in the presence or absence of GFP, or untreated. F: Multichannel confocal microscopy images of EGFR- (HEK293T) cells treated with the construct in the presence or absence of GFP, or untreated.

The scope of the method was widened through its application towards IgG antibody modification. First, a super-folder GFP variant was modified with a maleimide using the optimized sortase method. In parallel the IgG antibody drug Trastuzumab (Herceptin, trademarked) was reduced using TCEP and then subjected to overnight crosslinking with the modified GFP followed by SEC purification (Figure 43A) and analysis by SDS PAGE (Figure 43B). Conversion to the desired GFP-antibody conjugate could be confirmed using ESI-MS (Figure 43C). Overlapping the multichannel confocal fluorescence microscopy images from stained HER2+ cells showed clear membrane localization of GFP when treated with the construct (Figure 43D). Membrane staining was absent in the stained HER2- cells (Figure 43E), proving that the antibody retained its specificity following the protocol. The same experiment was performed using a cysteine containing GFP variant with comparable results (Figure 44A). For this experiment, the purification of the sortase product was done at pH 7.5 to prevent side reactions between the cysteine and the installed maleimide, followed by rebuffering to pH 8.5 immediately prior to crosslinking. This demonstrated that the sortase installation method can tolerate the presence of additional cysteine residues under these conditions.

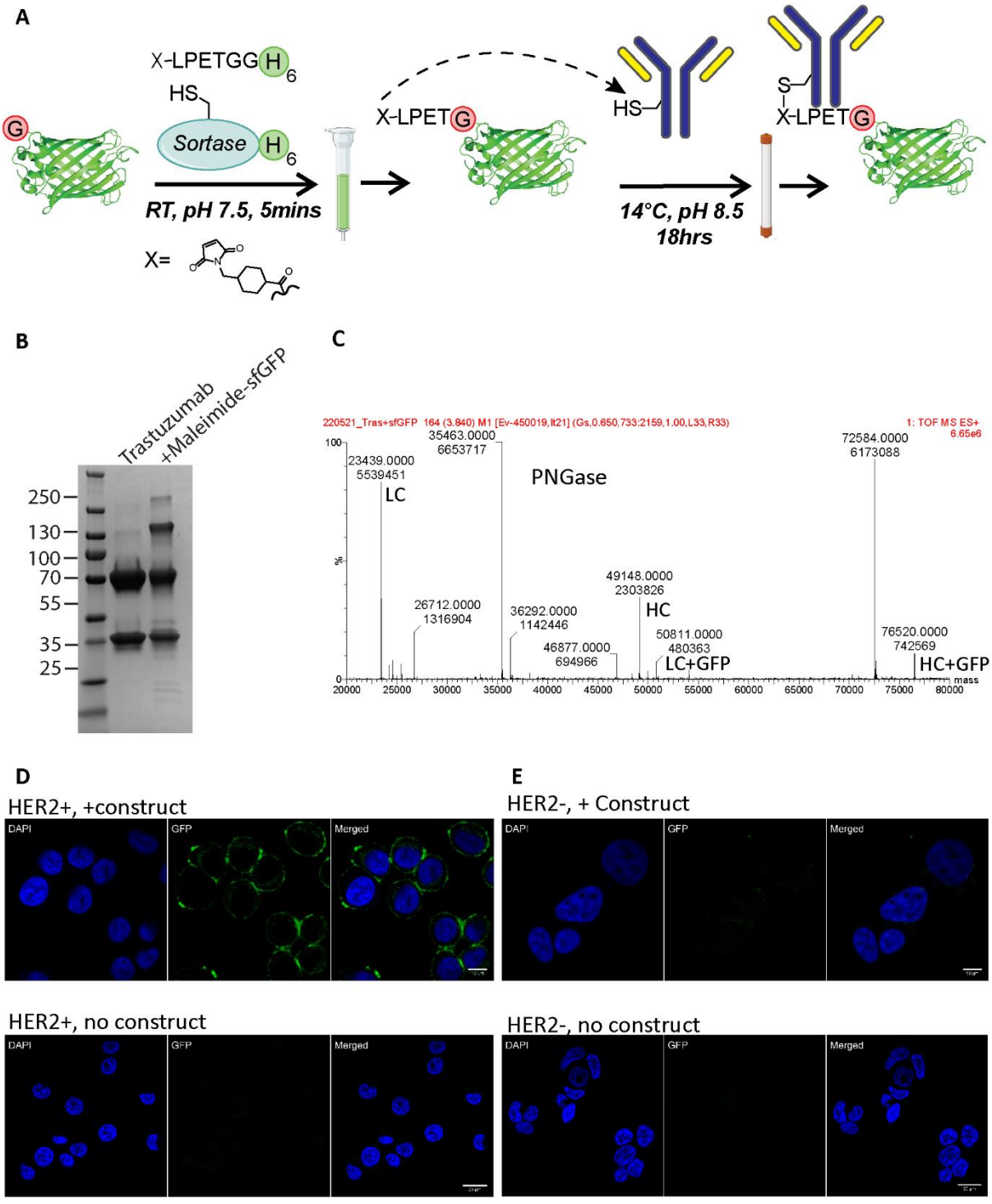


Figure 43: IgG crosslinking with GFP. A: Reaction schematic. B: SDS-PAGE and Coomassie staining of purified starting material and product. C: ESI-MS of purified product. D: Multichannel confocal fluorescence microscopy of HER2+ (SKBR3) cells treated with 5ug/ml construct. E: Multichannel confocal fluorescence microscopy of HER2- (MDA-MB-468) cells treated with 5ug/ml construct.

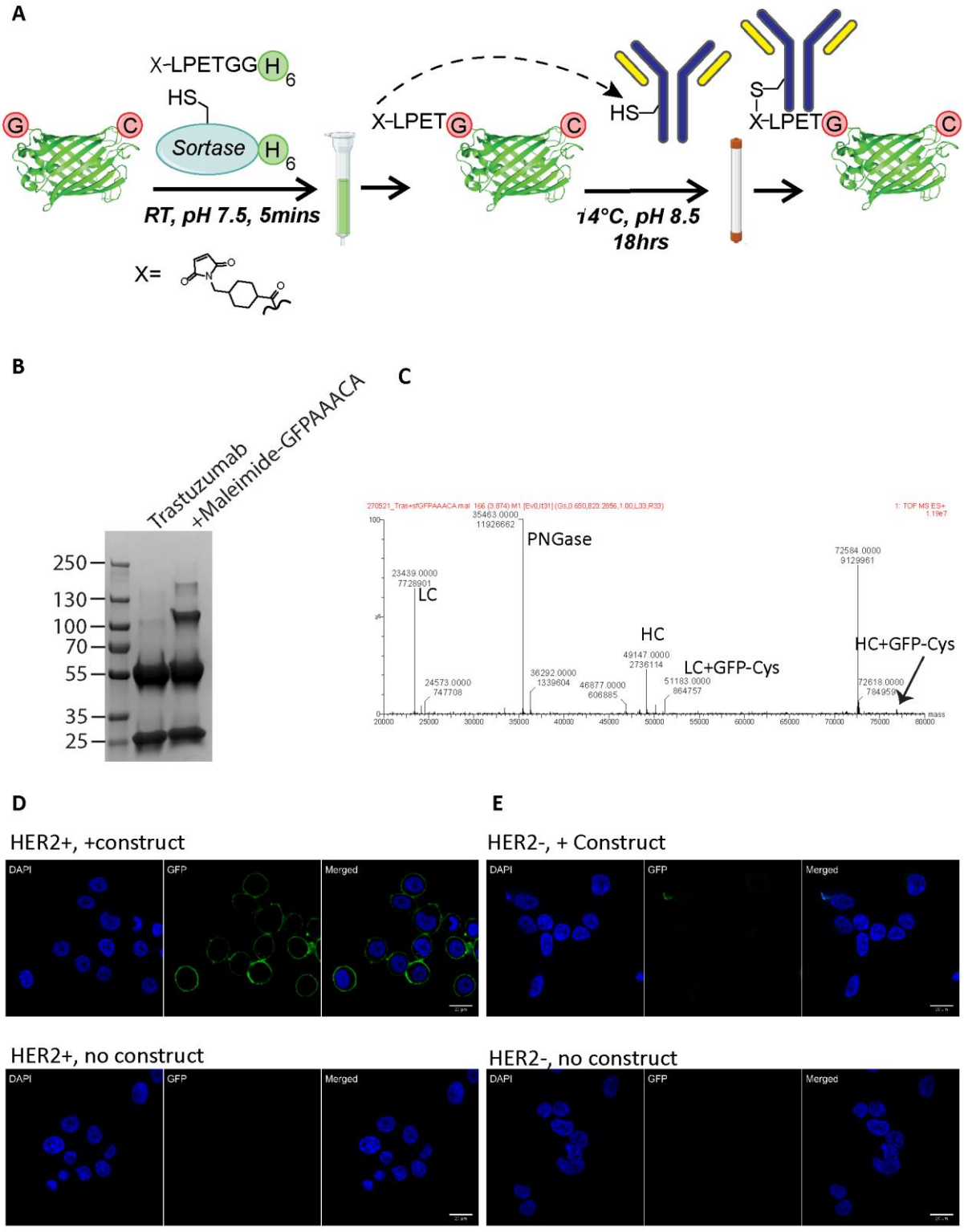


Figure 44: IgG crosslinking with GFP-Cys. A: Reaction schematic. B: SDS-PAGE and Coomassie staining of purified starting material and product. C: ESI-MS of purified product. D: Multichannel confocal fluorescence microscopy of HER2+ (SKBR3) cells treated with 5ug/ml construct. E: Multichannel confocal fluorescence microscopy of HER2- (MDA-MB-468) cells treated with 5ug/ml construct.

3.2 Objective 2

3.2.1 Identifying potentially critical nanobody residues for substrate binding

The general objective of this project was to replace a residue involved in substrate binding with a functional group that interrupts substrate binding, which could then be released on demand using photo-deprotection. Such a modification would add a degree of temporal control to nanobody function. This was envisioned to be accomplished one of two ways: either with the initial introduction of a bioorthogonal handle (like an azide) followed by chemoselective installation of the photo-labile caging functionality, or via the direct incorporation of a photo-labile protected amino acid derivative such as the one discussed previously which could then be further functionalized with a bulky caging group (Figure 10).

Our initial goal was to identify residues within a nanobody that were critical for substrate interaction which could be mimicked by the decaging product of our intended caging group. This would enable the decaged product to retain its native function. Lysine residues were the initial choice, given the positive charge and ammonia-like structure of the side chain, which was hypothesized to behave similarly to the decaging product of the photocleavable azide-containing lysine derivative discussed previously (Figure 10). Unfortunately, all lysine residues were shown to be spatially separated from the binding interface and therefore were not pursued further for this project (Figure 45).

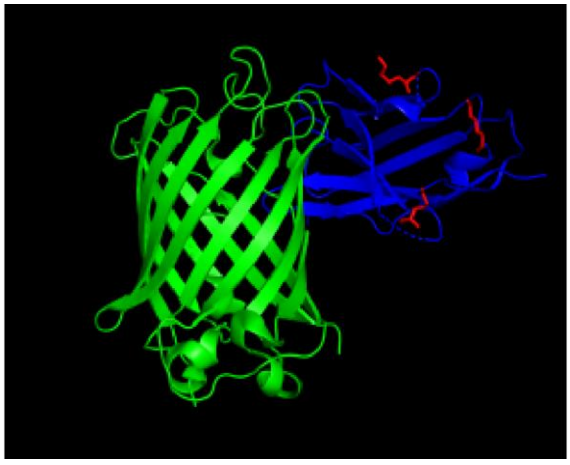


Figure 45: GBP1 in complex with GFP. GFP in green, GBP1 in blue. Lysine residues highlighted in red.

After this observation, arginine was proposed as an alternative to lysine, given the chemical similarities of their side chains. Both lysine and arginine are basic amino acids containing nitrogenous side chains, meaning that at physiological pH they tend to possess the same

positive charge and hydrogen bonding profile [142]. Structure analysis showed that several arginine residues were localized near the binding interface (Figure 46). Work from the Deiters group encoding photocleavable tyrosine derivatives using codon expansion also guided our focus towards the inclusion of tyrosine residues and residues that could be replaced by tyrosine on the list of candidates for caging.

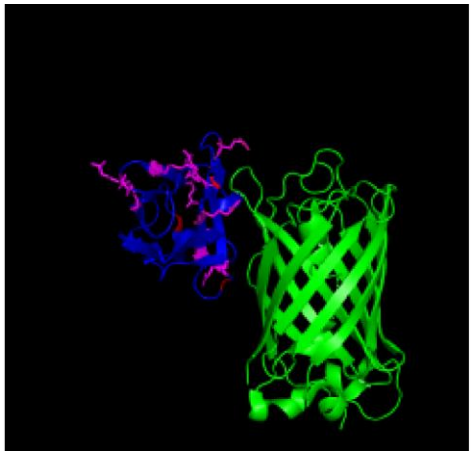


Figure 46: GBP1 in complex with GFP. GFP in green, GBP1 in blue. Arginine residues highlighted in pink.

A previous study on GFP binding nanobodies, in which the crystal structure of the nanobody-substrate binding complex was solved, emphasized the involvement of electrostatic interactions at the interface between the nanobody and its substrate (Figure 47). Residues were therefore selected based both on their participation in electrostatic substrate interactions and on their ability to be replaced with a photo-caged cognate using amber suppression. With these criteria, the best residue candidates were identified to be R35, R57, Y37 and Y118.

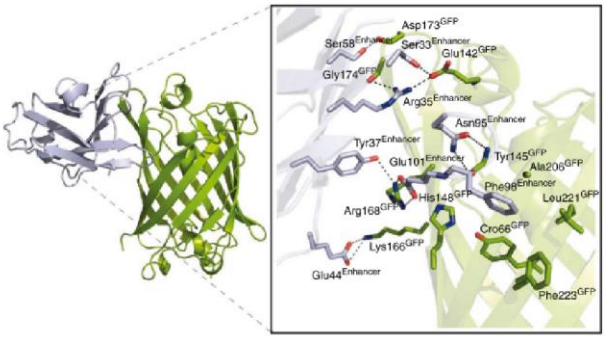


Figure 47: Crystal structure of GBP1 in complex with GFP. Residues involved in electrostatic binding interactions are highlighted [12].

A plan was envisioned involving the systematic expression of a panel of mutant nanobodies, followed by characterizing their GFP binding in a fluorescence readout assay. It was intended to begin with the residues identified above as likely involved in electrostatic substrate interactions. Initial experiments with replacement mutants generated at position R35 using a plate-reader based fluorescence readout assay measuring the % enhancement of GFP fluorescence after exposure to the nanobody showed that tyrosine was able to almost restore the WT-levels of GFP enhancement, whereas isoleucine could not (Figure 48). This data suggested that binding between the GBP1 nanobody and GFP was at least partially retained when arginine 35 was replaced with tyrosine.

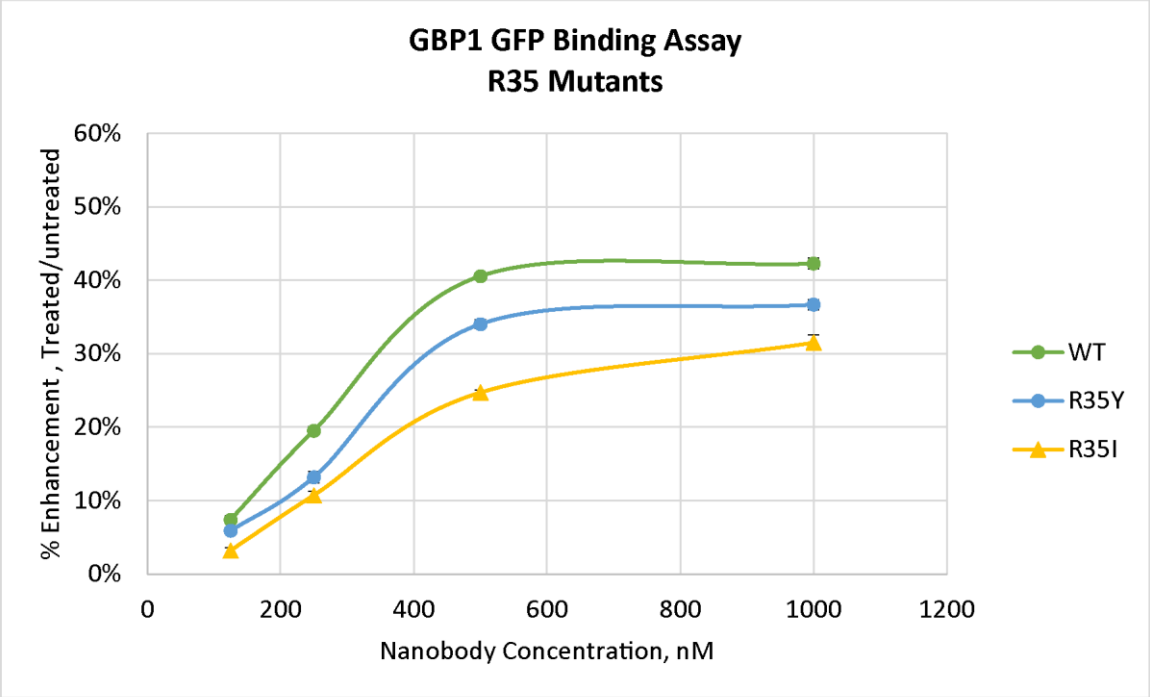


Figure 48: GBP1 GFP binding assay for two replacement mutants at position R35.

Additional experiments were planned with further substitutions at additional locations. However, shortly after this point it was discovered that the R35 residue had previously been mutated and replaced by an alanine with a resulting impairment to binding [143]. Armed with this knowledge, we were then confident to proceed directly to replacing the R35 codon with an amber stop codon (TAG) that could be uniquely translated to encode an azidophenylalanine residue bearing a bioorthogonal attachment point. Azidophenylalanine would be reduced to an aniline upon photo-deprotection of the eventual caging group, yielding a species that presented similar polarity and hydrogen bonding patterns as tyrosine.

Based the discussed results suggesting the suitability of tyrosine as an arginine replacement at position 35 (Figure 48), it was decided that this strategy made the best use of the expertise of our collaboration partners in the Rubini group at incorporating this tyrosine analogue. Mutagenesis was accomplished using the NEBase Changer site directed mutagenesis kit and was confirmed via sequencing. The amber codon was introduced into an otherwise wild type GFP enhancer nanobody (GBP1) bearing a *C*-terminal HIS tag for purification.

We decided to utilize a caging technology previously developed in our group involving PEG-containing light-cleavable ortho-nitrobenzyl (ONB) groups arranged around a phosphorous (III) scaffold to form a ONB-PEG-phosphite [45, 47]. The phosphite offered the two-fold advantage of presenting a bulky caging group likely to impair substrate binding, while also enabling photo-removal of the large PEG groups. The phosphite could also be conjugated specifically to azides using the chemoselective Staudinger-phosphite reaction, necessitating the introduction of azidophenylalanine at the desired R35 caging position.

3.2.2 Azide incorporation via amber suppression

Azides fulfill the criteria of bio-orthogonality, enabling chemoselective functionalization despite the presence of many endogenous functional groups within the nanobody [120]. Well-established methods exist for incorporating azides into proteins, including the use of amber codon suppression in conjunction with a promiscuous tRNA synthetase compatible with azide-containing amino acids such as azido-tyrosine [29]. This strategy has the advantage of total specificity over the incorporation position but can be technically challenging. Unnatural amino acids can experience poor solubility in liquid media, and high yields of the desired protein can come at the expense of the health of the expression host. The mutagenized GBP1 nanobody was sent to our collaborator Marina Rubini, who expressed the azide-containing variant via amber suppression in her lab and then supplied the material to us for further modification and testing (Figure 49).

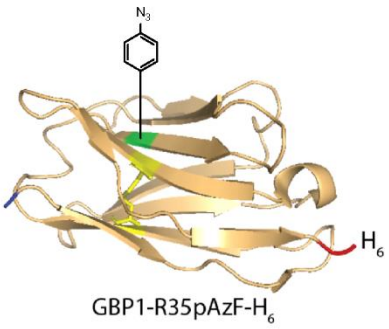


Figure 49: GBP1 nanobody modified at position R35 to contain an

3.2.3 Azide reduction and GFP binding assay

To enable photo control of the nanobody-GFP interaction, the product of the photolysis reaction must retain its binding capabilities. Photolysis of the caged azido-phenylalanine yields an aniline derivative, and so an experiment was designed to directly convert the incorporated azide into the aniline to test for its substrate binding capacity (Figure 50).

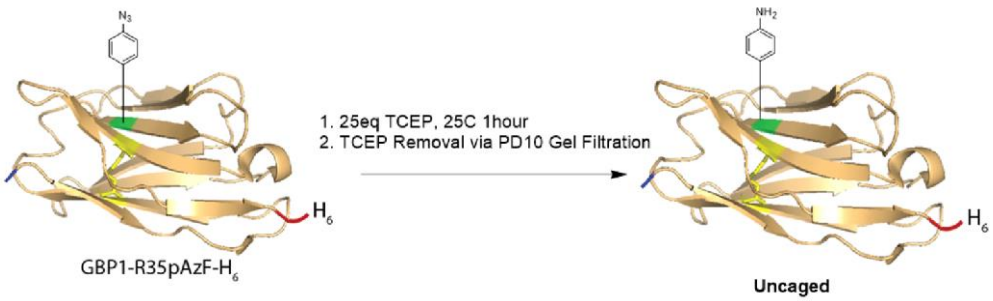


Figure 50: Reaction schematic for the reduction of the incorporated azide to the final de-caging aniline product using TCEP.

Reduction was performed using TCEP, which was removed via gel filtration (Figure 51). TCEP reduction was able to fully convert all incorporated azide to the corresponding aniline.

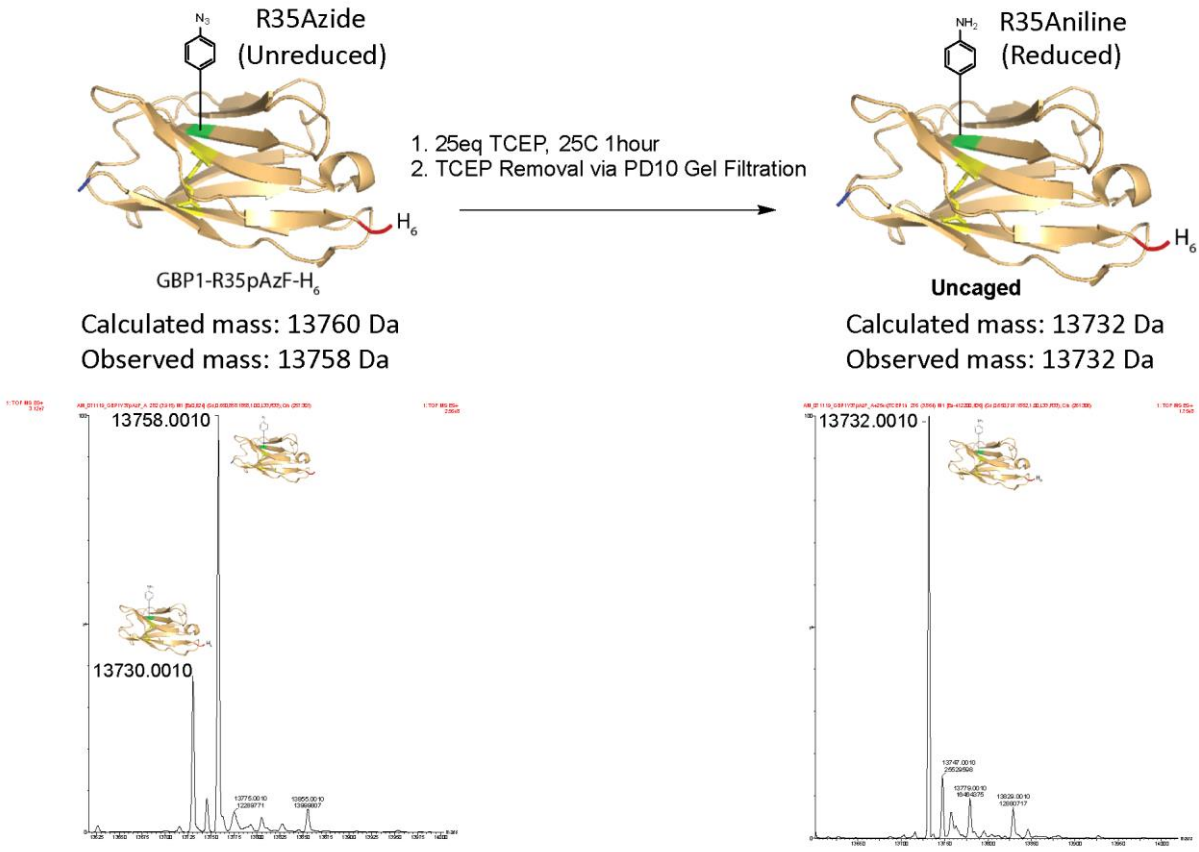


Figure 51: TCEP reduction of GBP1R35pAzF to the corresponding aniline. Annotated ESI-MS spectra before (left) and after (right) reduction are shown.

Fluorescence enhancement was chosen as a readout for binding between the enhancer nanobody GBP1 and its substrate GFP. This was done to take advantage of the endogenous enhancement properties of nanobodies. While this was a convenient readout, it had suffered from a lack of sensitivity, because the enhancer only increases GFP fluorescence by a factor of 1.6, which can make detection of the enhancement challenging [12]. Initial binding assays were performed on a plate reader in a 96-well plate, but due to insufficient sensitivity observations were subsequently shifted to the cuvette scale using a tabletop fluorimeter (JASCO).

GFP binding assays were performed in triplicate using both the reduced and non-reduced samples, compared to WT GBP1 (Figure 52). The hope was to observe that, following TCEP reduction, full WT-level GFP enhancement could be restored, which would imply that the nanobody had regained its ability to fully bind to its substrate. Unfortunately, the aniline was only observed to partially recover its binding capability compared to the WT. The azide did

show some binding reduction, supporting the hypothesis that additional caging at that position would reducing binding further.

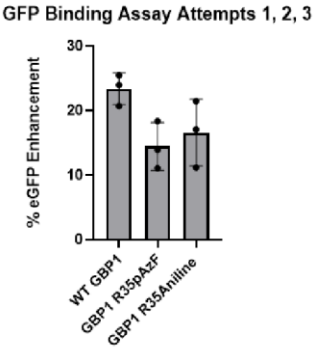


Figure 52: GFP binding assay before and after TCEP reduction of GBP1R35pAzF to the aniline, compared to WT GBP1.

3.2.4 ONB-PEG-Phosphite Synthesis

To functionalize the incorporated azides and confer photo-control to the nanobodies, we chose to employ the Staudinger-phosphite reaction, for which we required a PEG-substituted ortho nitrobenzyl-linked phosphite. The reaction between the phosphite and the azide would result in a photo-labile phosphoramidate. Inspiration for the synthesis scheme of the photo-cleavable, pegylated phosphite was drawn from previous work in our group [45] (Figure 53).

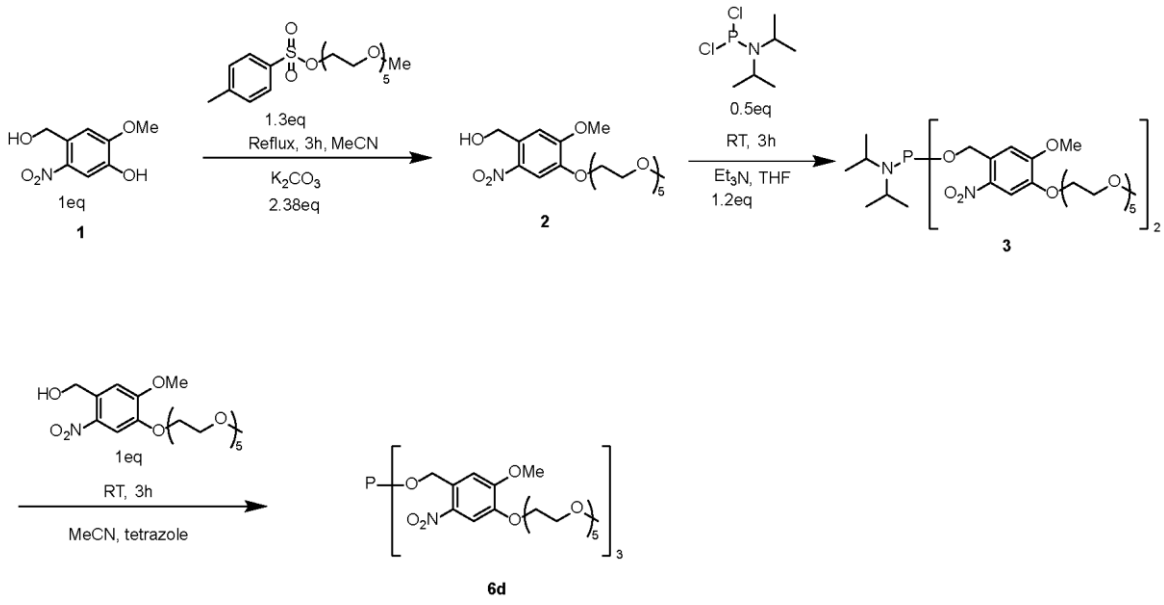


Figure 53: Single-step ONB-PEG-Phosphite synthesis previously reported from our group [45].

As a precursor step, the pentaethylene glycol toluene-p-sulfonate had to first be synthesized via an overnight reaction in acetonitrile between toluene sulfonyl chloride and pentaethylene glycol (Figure 54). Conversion to the desired product was confirmed using 1H NMR (Spectrum 16).

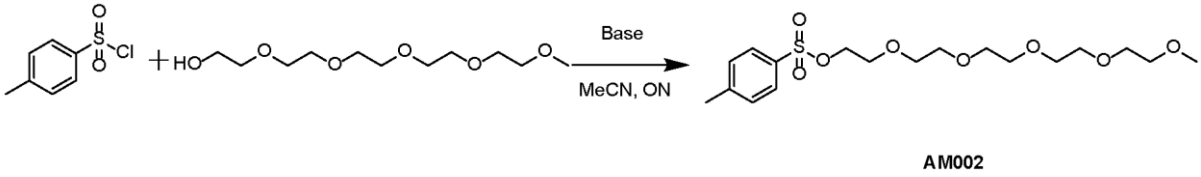


Figure 54: Pentaethylene glycol toluene-p-sulfonate AM002 precursor synthesis scheme.

Based on the advice of my colleagues suggesting an improvement in the yield, we adjusted the synthesis scheme to take place in two steps: synthesis of the pegylated ONB alcohol AM003, followed by generation of the phosphite AM004 using symmetrical phosphorous trichloride (Figure 55).

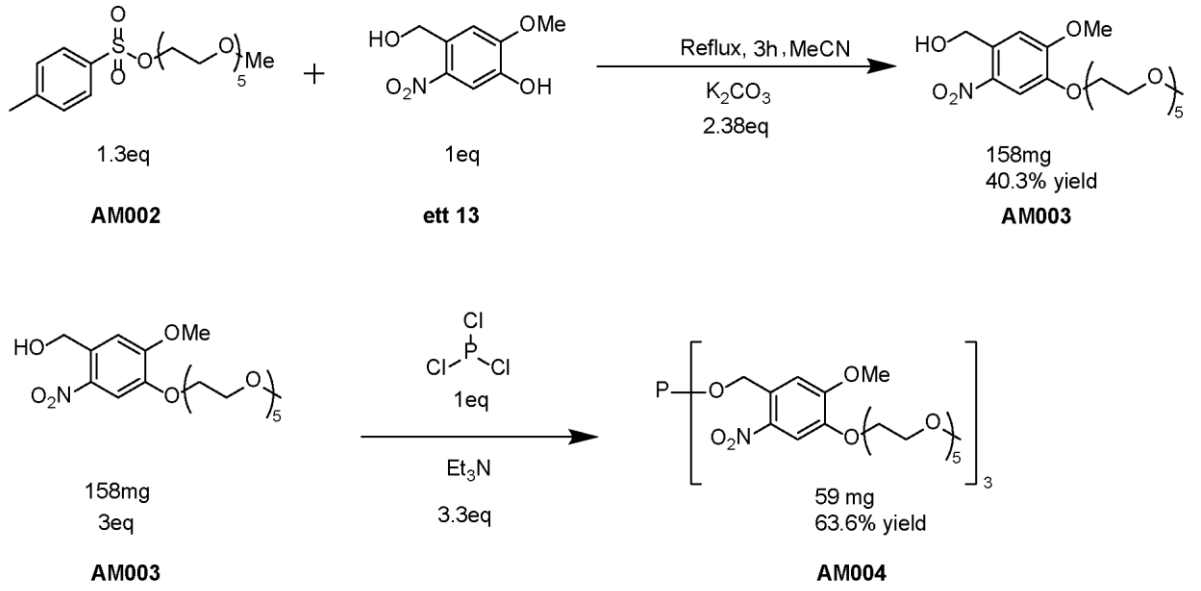


Figure 55: Two-step synthesis scheme for light cleavable, PEGylated ONB phosphite employed in this thesis.

Phosphorous NMR in deuterated acetonitrile of the purest fraction following silica purification of AM004 showed a clean peak at the appropriate shift for a P(III) species (Spectrum 19). There are a few other phosphorous peaks that could correspond to oxidized or hydrolyzed phosphites, but they are much smaller than the primary peak. Proton NMR showed the proper spectra for the phosphite, but also the presence of a larger amount of

unreacted alcohol AM003 (Spectrum 18). In particular, the two aromatic protons in the 8 ppm region shifted right in the phosphite compared to the alcohol aromatic protons. These protons were integrated to a value of 1, but the corresponding alcohol protons in that region showed integrations of ~1.3. in addition to other integrated peaks, this indicates that the alcohol is not only present in this fraction but is present in a larger quantity than the phosphite. Despite containing a mixture of phosphite and unreacted alcohol, this fraction was deemed adequate for the Staudinger phosphite reactions, as these contaminants lacked the reactivity towards the azide which was the hallmark of the desired phosphite.

3.2.5 Staudinger-phosphite azide functionalization

With the phosphite on hand and a stock of azide-containing nanobody from our collaborator Marina Rubini, the next step was to confer photo-control over the substrate binding by attaching the light-cleavable PEG group to the azide (Figure 56).

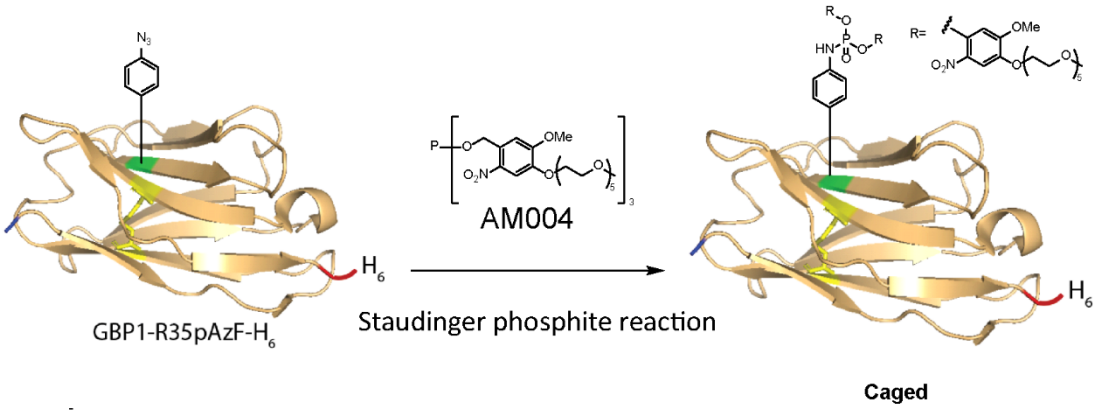


Figure 56: Staudinger-phosphite reaction between the incorporated azide and the light-cleavable PEG phosphite AM004.

An initial reaction was attempted using 2100eq of phosphite relative to the nanobody (Figure 57A). ESI-MS of the purified product showed the desired mass, in addition to several +433Da adducts matching the mass of AM003, the co-eluted alcohol precursor to the desired phosphite (Figure 57B).

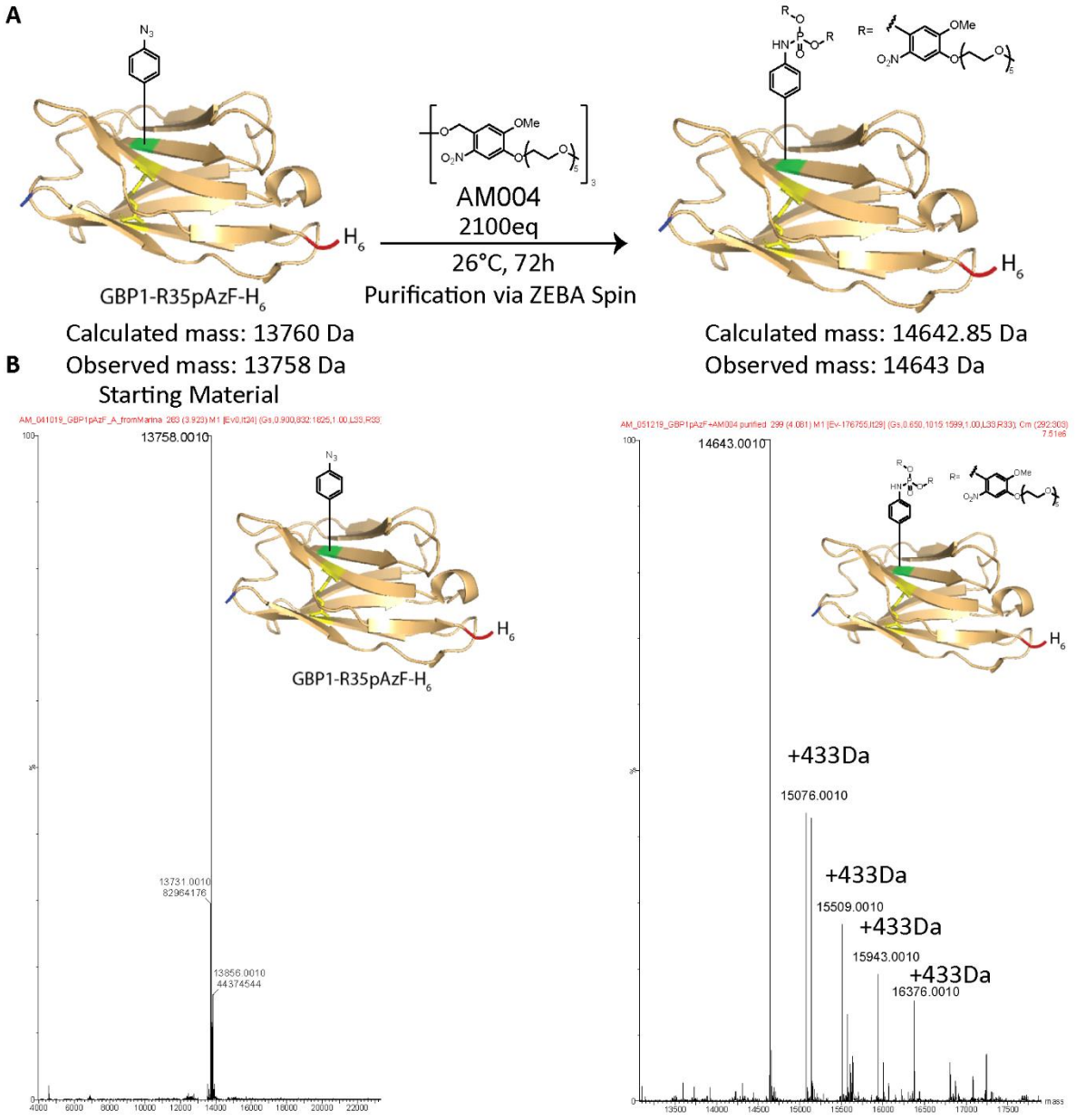


Figure 57: Initial conditions for the Staudinger-phosphite reaction employed to photocage the GBP1 nanobody. A: Reaction schematic. B: ESI-MS, deconvoluted, of the starting material (left) and purified product (right)

Time course measurements were taken over 72 hours analysis by SDS-PAGE (Figure 58). The product was detectable at the first time point, and only background levels of the unmodified nanobody could be detected via Coomassie staining even after only 24 hours.

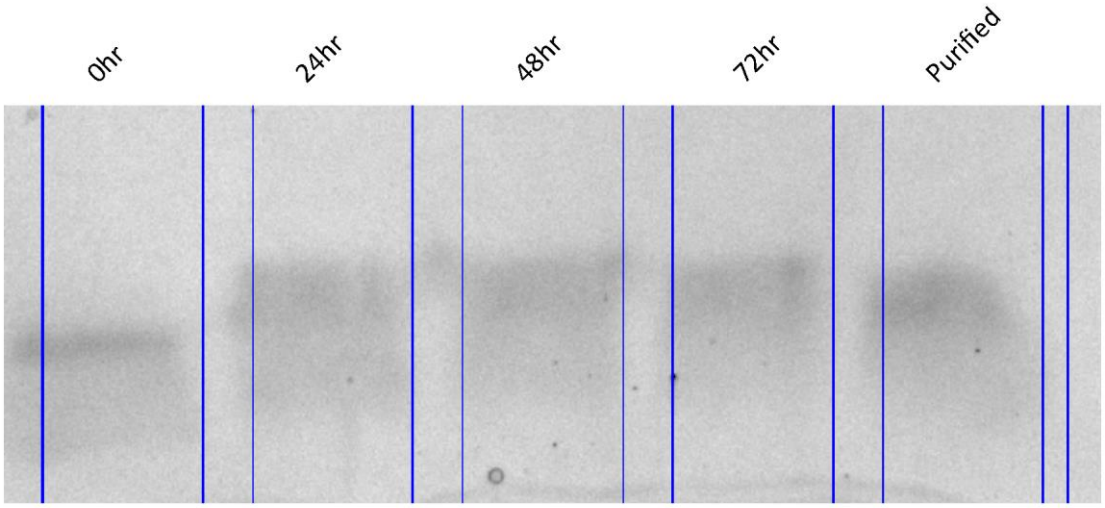


Figure 58: Time course samples from the 2100eq phosphite reaction analyzed by SDS-PAGE with Coomassie staining.

To prevent the adduct from forming, the reaction was repeating using only 100eq of phosphite and stopped after only 24 hours (Figure 59A). With these new conditions, the adducts could no longer be detected by ESI-MS (Figure 59B). The only other species detected matched the mass of the aniline-containing nanobody formed by reduction of the incorporated azide.

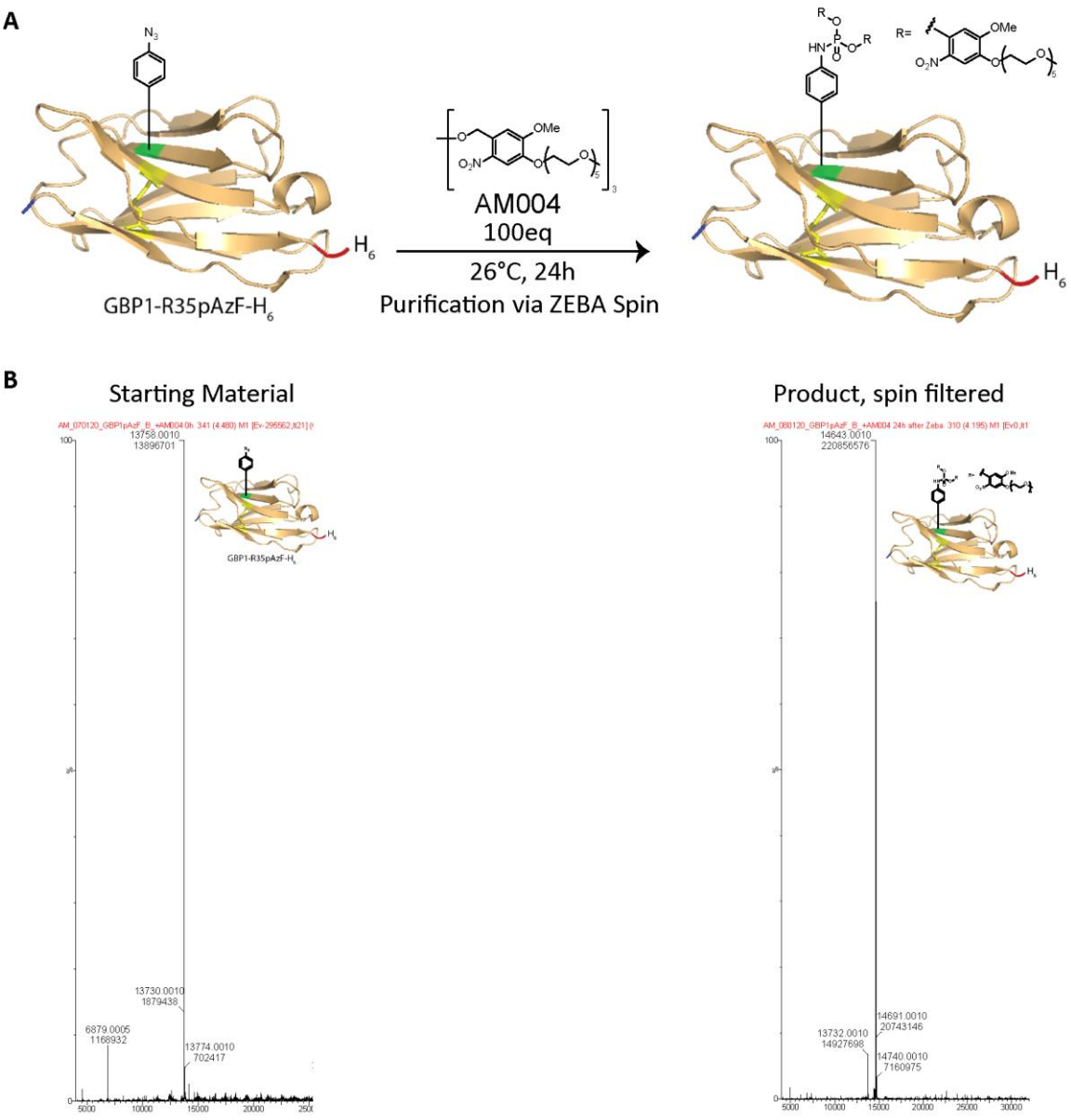


Figure 59: Staudinger phosphonite reaction with revised conditions to prevent adduct formation. A: Reaction schematic. B: ESI-MS, deconvoluted, of the starting material (left) and purified product (right)

Time course measurements detected the product by SDS-PAGE after only 4 hours with apparently full conversion (Figure 60).

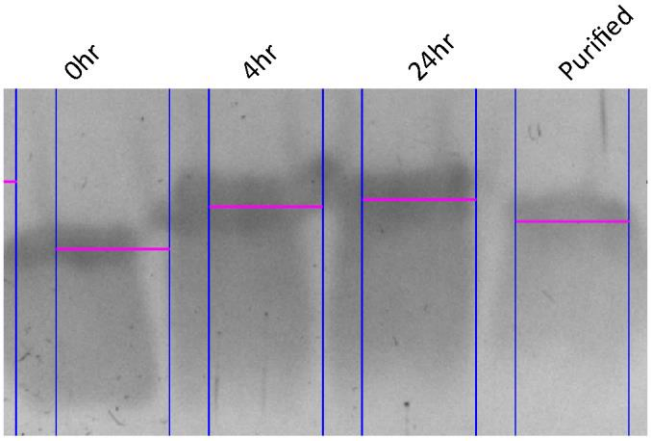


Figure 60: Time course samples from the 100eq phosphite reaction analyzed by SDS-PAGE with Coomassie staining.

3.2.6 Photo-deprotection and GFP binding assays

Photocontrol of the nanobody was enabled via the UV-induced photolysis of the ortho-nitrobenzyl groups to yield a phosphor-tyrosine analogue that gets hydrolyzed to form an aniline group (Figure 61). UV deprotection was performed at 365nm and monitored by ESI-MS. 15 minutes of irradiation proved unable to yield any detectable deprotection, while 60 minutes successfully consumed all ONB groups to yield two major species with masses matching the aniline and the azide. Given the irreversibility of the Staudinger reaction, it is unlikely that the azide resulted from the UV irradiation. One explanation is that the detected azide represents a population not modified by the Staudinger phosphite reaction, which only became detectable after deprotection. As controls, WT and R35pAzF GBP1 were both irradiated and monitored via ESI-MS (Spectrum 2, Spectrum 3).

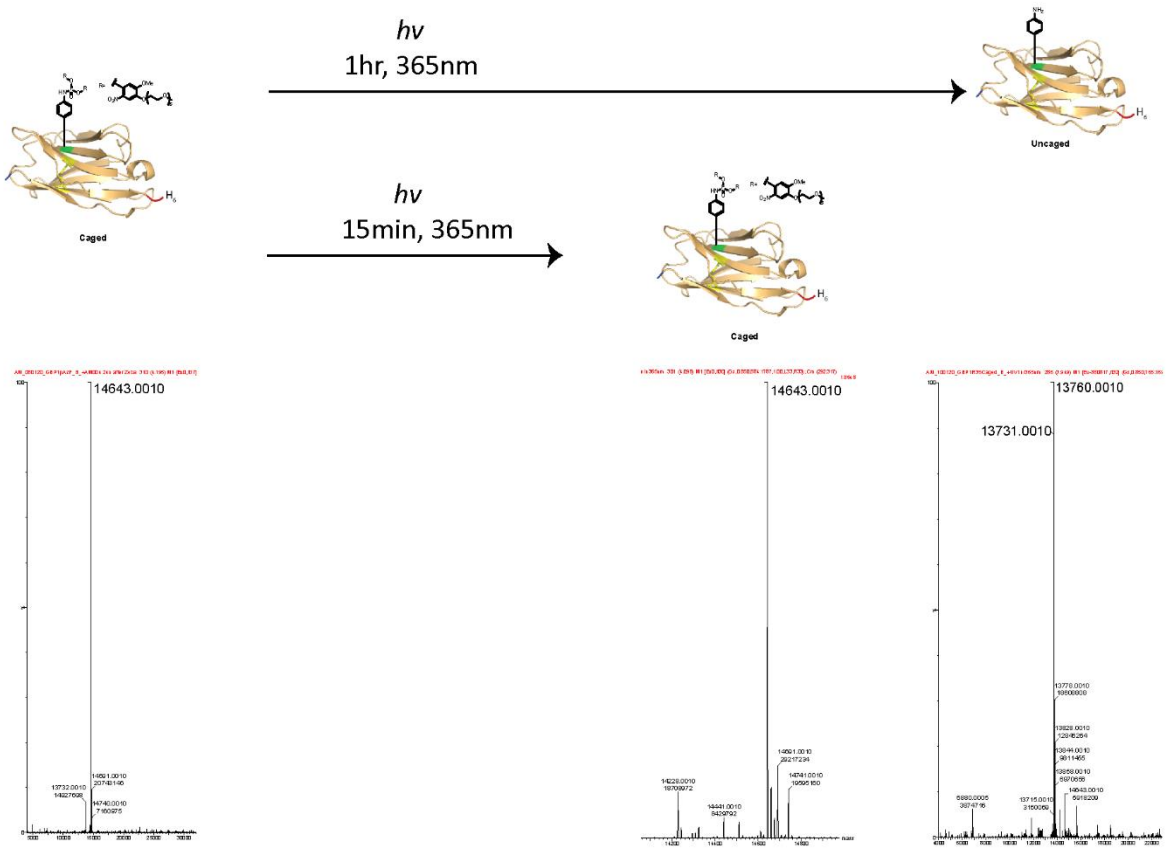


Figure 61: ESI-MS monitoring of UV-induced photo-deprotection of the GBP1R35Caged nanobody.

Following UV deprotection, a GFP binding assay was performed using both GBP1 WT and R35Caged variants (Figure 62). In the absence of UV irradiation, WT GBP1 showed up to 4% GFP fluorescence enhancement at high nanobody concentration, while the R35Caged showed no enhancement at any concentration. After UV irradiation, both variants showed comparable GFP fluorescence enhancement at high nanobody concentration, suggesting that near-WT binding capabilities could be restored using UV irradiation. UV irradiation is not expected to be responsible for the negative enhancement observed at low irradiated WT nanobody concentrations, since the WT nanobody should not react significantly to UV irradiation. Instead, this phenomenon can be traced directly to the lower-than-expected fluorescence enhancement of the WT nanobody (maximum ~4%), especially compared to the results obtained on the plate reader showing 40% enhancement or more (Figure 48). With such low enhancement, minor fluctuations in signal due to handling could not be unambiguously distinguished from biophysical effects of nanobody-substrate interactions. Conclusions on the efficacy of the caging and decaging are instead mostly based on the overall trends of the replicated measurements performed at a series of nanobody concentrations

confirming that that the cage prevented nanobody-induced fluorescence enhancement until the point at which UV irradiation removed the cage.

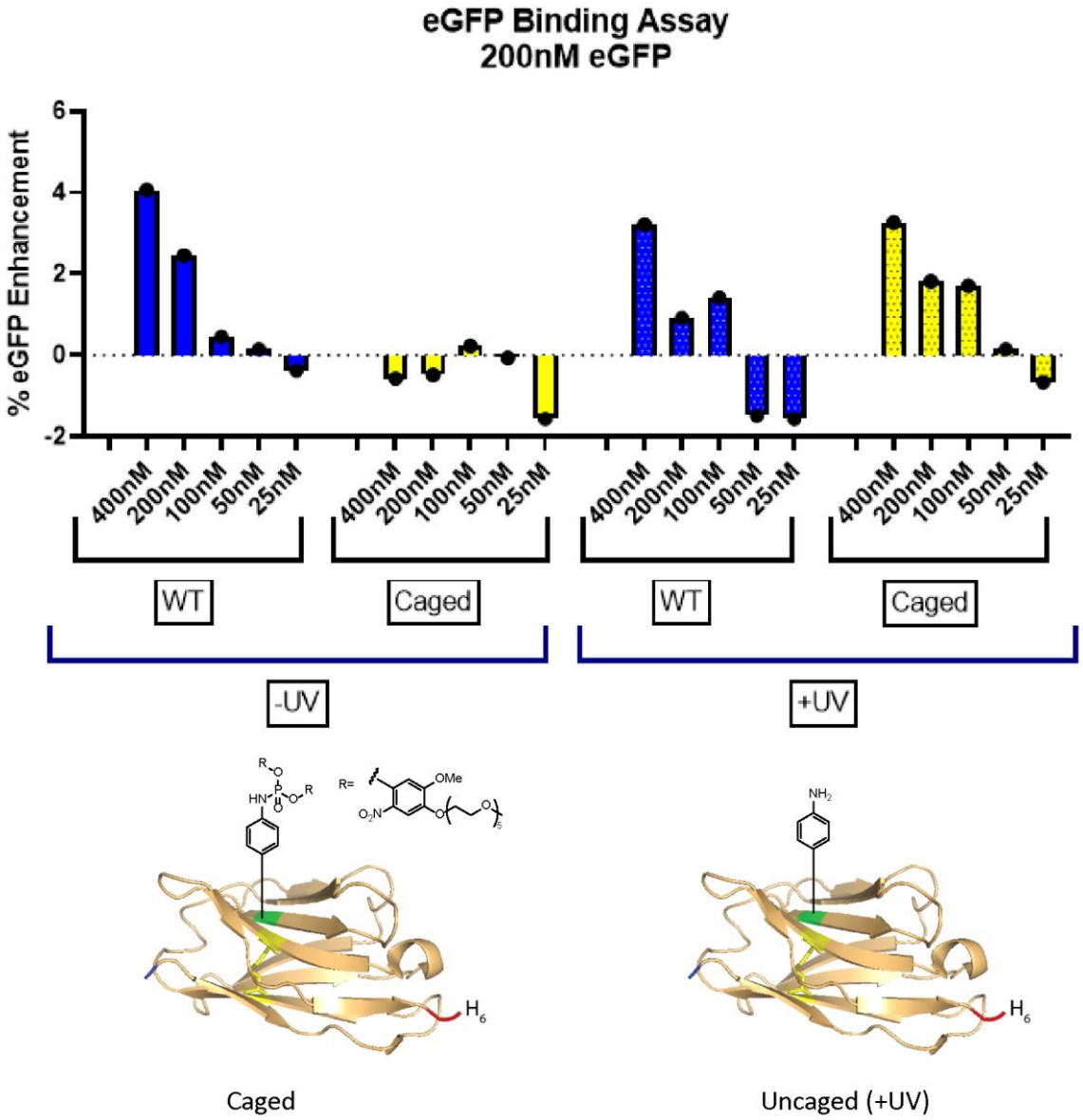


Figure 62: GFP binding assay before and after UV irradiation of both WT and R35Caged GBP1 variants. The experiment was performed once at room temperature in PBS, with triplicate measurements performed in batches to generate the reported average % GFP enhancements with error. At the beginning and end of each batch, the eGFP alone was measured to account for any photobleaching incurred during the batch measurement.

4. Outlook and Conclusion

4.1 Objective 1: Sortase-enabled installation of cysteine-selective electrophiles for protein-protein conjugation

The aim of this thesis objective was to demonstrate that the sortase could be added to the list of methods capable of installing chemoselective functionalities onto proteins site-

specifically for generating well-defined protein conjugates. Chemoenzymatic introduction of cysteine electrophiles represents a rare combination among the pantheon of protein-protein conjugation tools; the method offers site-selectivity for attachment points to both proteins and can be performed under certain conditions in minutes instead of hours. Thanks to the use of LPETGG-containing peptides, the sortase reaction requires little to no modification of the protein substrate and allows for the installation of a nearly limitless variety of electrophilic functionalities in a modular fashion. While Sortase is most commonly used in the literature for the *C*-terminal modification of proteins [95], the *N*-terminal targeting strategy employed here facilitated both modularity of the installed functionality as well as the addition of large excesses of peptide, which helped counteract the kinetic deficiencies of the enzymes [95]. A 20-fold excess of peptide combined with a suitable purification strategy for removing the peptide, likely pushed the reaction further to completion within the short (five minute) reactions employed, achieving good conversion while avoiding unspecific side products. Adding a His-tag to the peptide and sortase, while not original to this work, was particularly well suited to this application. Removal of the sortase and peptide could be done in a single step, which proved to be essential for high-yielding cysteine conjugation reactions.

Sortase can be, and has been, used to generate chimeric proteins in a variety of ways, some of which have been previously described in this dissertation [81, 106, 107]. The purpose of these strategies, like that of the method described in this dissertation, is to generate homogenous, specifically linked protein conjugates with complete control over the position and manner of the covalent linkage. Some strategies involve either the direct fusion of two proteins in a single sortase step, which requires at least one to contain a recombinant recognition sequence (while the other contains at least an endogenous, *N*-terminal glycine or alanine). This accomplishes the desired effect but requires laborious recombinant engineering of a target protein that might be drastically altered as a result. Another strategy involves the installation of compatible “click” handles on each of two conjugation partners. This also achieves the desired effect but requires both dual recombinant engineering and two sortase reactions prior to fusion. The method described herein improves upon these designs by requiring only one sortase reaction and no recombinant engineering on the target protein if a cysteine is present. Modularity is also a major advantage of this method. Since in principle any functionality can be attached to the LPETGG peptide, different effects can be achieved

without significantly altering the method, as was demonstrated with the bifunctional electrophilic/fluorescent peptide substrates.

The specific installation of a cysteine-selective electrophile directly enables fusions to cysteine-containing proteins, which can be endogenous or otherwise recombinantly introduced. As a result, the method is not limited to *N* and *C*-terminal fusions. As was demonstrated with the IgG antibody conjugation example, proteins with *N*-terminal electrophiles could be crosslinked to cysteine residues located within the secondary structure of a conjugation partner. If multiple cysteines are endogenously present (such as for the IgG antibodies), crosslinking would be unspecific (but still site-selective) using this method. Nonetheless, the method described here provides numerous advantages in time saved and reduced recombinant engineering if compatible (i.e. cysteine-containing) targets are chosen.

Downsides do still exist, however – the method is limited to linking cysteine residues to *N*-termini and does suffer from limited yields due mainly to the purification of the sortase reaction. When viewed as a proof-of-principle that cysteine-selective electrophiles are tolerated by cysteine transpeptidases, however, these shortcomings are less significant. Moreover, the use of the site-specific sortase reaction presents distinct advantages over semi-selective N-terminal crosslinking methods like **SMCC** due to challenge of specifically targeting the N-terminus. SMCC-based applications targeting the N-terminus must be constantly aware of side reactions with other residues containing nucleophilic amine side chains, whereas the sortase can reliably target the N-terminus with little regard for neighboring residues. It should be noted that some N-terminal selectivity with SMCC can be achieved by utilizing the fairly unique pKa of the N-terminal aliphatic amine, which is slightly less basic than other aliphatic amines [52]. Nonetheless, SMCC-based N-terminal targeting would almost certainly result in a less homogenous and well-defined product than the chemoenzymatic method employed here.

A major focus of this work was investigating the suitability of sortase reactions for the reliable modification of proteins using multi-step protocols (i.e. electrophile installation followed by crosslinking). While highly specific for both the site of electrophile incorporation and the subsequent crosslinking position, the method is challenged by the multiple purification steps required, as well as the poor catalytic efficiency of the sortase. Despite significant effort, optimization presented in this work was not able to recover as much electrophilic protein

from the sortase reaction as was desired for large-scale downstream applications for reasons mentioned above. Spectrophotometric measurements at 280nm before and after the IMAC sortase purification step consistently revealed that ~40% of the nanobody material was lost due to handling, on top of the less-than-full conversion by the sortase to the electrophilic product. Therefore, the results of this investigation raise questions as to the suitability of such multi-step methods for large-scale (i.e. industrial) applications. Nonetheless, the usefulness of this method for academic/research applications is highly intriguing, particularly when gentle reaction conditions, subtle genetic engineering requirements and impeccable specificity are the goals. Future work could certainly be done to establish the method with other enzymes like the subtiligase, or other chemoselective functionalities targeting, for example, amine-containing nucleophilic side chains (i.e. NHS-esters) [144]. This could enable the N-to-N covalent linkage of multiple proteins, which might be advantageous for certain targets when attempting to avoid interfering with other domains. However, the higher abundance of reactive amines in proteins might reduce the homogeneity of resulting products.

Generating more efficient sortase substrate nanobodies would likely be a fruitful area of investigation for improving the overall method yields. As discussed in section 3.1.2.3, the N-terminal methionine must be cleaved from a bacterially expressed protein by endogenous MAP activity before a recombinant glycine can be accessible to the sortase. Later in the method development it was recognized that MAP activity is significantly reduced in the periplasm, which in this case resulted in significantly impaired cleavage of the terminal methionine for PelB-containing nanobodies as observed by MS, rendering the polyglycine motif inaccessible to the sortase (Figure 63A). As a result, transpeptidation efficiency was extremely low. The next generation of glycine-containing nanobodies with protease recognition sequences were pursued at the time of investigation without fully understanding the poor labelling efficiency of this first generation by cloning into a new vector lacking the PelB sequence. This resulted in nanobodies with good labelling efficiency but poor solubility due to the absence of the PelB sequence (Figure 63B) as well as experiencing the multimerization effect when modified with electrophiles as discussed in section 3.1.6. Future investigations should re-introduce the PelB sequence for better soluble expression while maintaining the protease treatment to reliably generate an accessible glycine (Figure 63C).

This strategy would likely also prevent multimerization by preventing the disulfide oxidation observed in the cytosol through periplasmic export during expression. Recombinant introduction of additional glycine residues may also improve sortase labelling efficiency and thereby improve overall crosslinking yields.

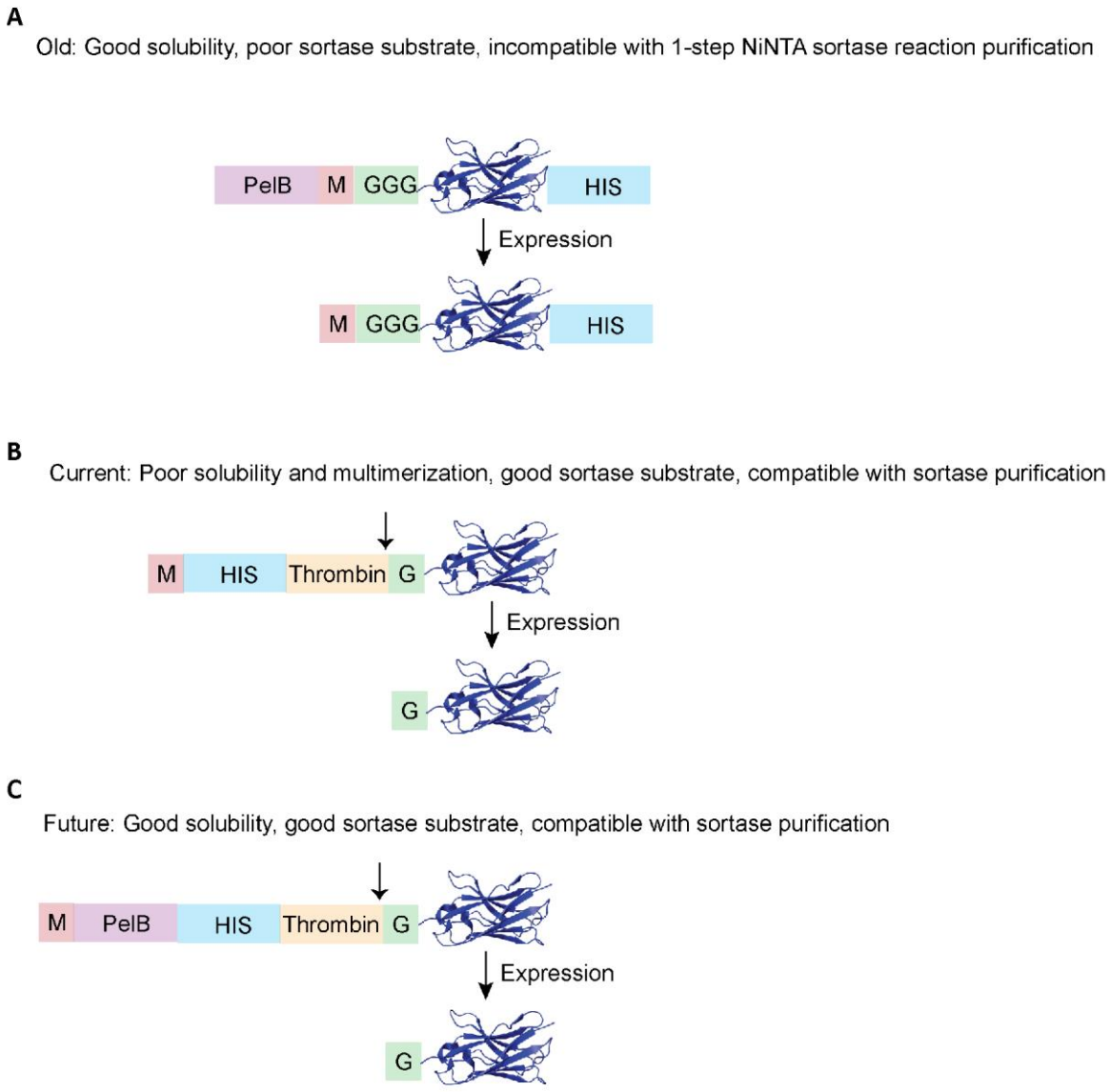


Figure 63: Different strategies pursued and envisioned for generating soluble nanobodies compatible with N-terminal sortase labeling.

4.2 Objective 2: Generation of azide-containing nanobodies and subsequent functionalization for caging and photo-deprotection

This project accomplished the site-specific installation of a photo-labile caging group into the substrate binding site of a nanobody using the Staudinger-phosphite reaction. First, an azide could be successfully incorporated at the nanobody R35 position within the substrate

recognition pocket using amber suppression with an azido-phenylalanine derivative. Next, the Staudinger-phosphite reaction performed on the nanobody azide resulted the steric envelopment of the recognition pocket by a light-cleavable PEG oligomer. Finally, UV-irradiation could remove the PEG oligomer to generate an aniline-containing nanobody with binding activity partially restored to that of the WT. Although not the first example of light-controlled antigen binding proteins (see work from the group of Henning Mootz [131]) this was a novel combination of both site-specific azide incorporation using amber suppression and subsequent chemoselective caging of the substrate recognition site with a light cleavable PEG group. To our knowledge it was also the first application of the Staudinger-phosphite reaction on a nanobody.

The nanobody selected for this investigation proved to harbor both advantages and disadvantages as a target. Although para-azido-phenylalanine could be successfully incorporated at the R35 position with a significant reduction in substrate binding affinity, the aniline resulting from decaging regrettably proved only partially successful in restoring substrate binding. This is probably not a fault of an incomplete decaging reaction, since ESI-MS showed complete generation of the aniline after UV irradiation. One likely explanation for this observation is that the arginine side change plays an essential role in substrate recognition that cannot be completely replicated by the aniline-containing tyrosine derivative resulting from decaging. Despite both containing nitrogenous functionalities, the aniline amine is attached to an electron-withdrawing aromatic functional group that bears little structural resemblance to the alkyl chain-like arginine side chain while also drastically lowering its pKa (~4.6) relative to the arginine side chain (>10) [145].

The GFP enhancer nanobody was especially convenient as a substrate given its ability to enhance GFP fluorescence upon binding. This proved essential for the workflow of both identifying essential active site residues, as well as for gauging the effectiveness of caging and decaging reactions. The fact that binding could be monitored with high sensitivity using tabletop fluorimeters instead of complicated methods like MST streamlined investigations significantly and allowed for more flexible iteration and testing [143].

Nonetheless, future applications of this method would benefit from a target that fully restores binding upon decaging, unlike the R35 position of the nanobody investigated herein. Moreover, the modularity of this strategy could be further exploited by enhancing the steric

bulk of the employed caging group (via larger PEG chains, for example). It would also be exciting to test the photo-activatable nanobodies in a functional assay. One such experiment could be based on fluorescence microscopy, in which the nanobody construct is functionalized with a fluorophore and then introduced to GFP-expressing cells (using, for example, cell-penetrating peptides to allow the construct to travel through the cell membrane) [146]. Co-localization of both fluorophores (perhaps detected by a FRET readout) should then only occur upon irradiation with UV light.

5. Materials and Methods

5.1 General Methods

5.1.1 High resolution mass spectrometry (HRMS)

Intact protein masses were measured via high resolution mass spectra recorded on a Xevo G2-XS Q-tof mass spectrometer (Waters). Samples were handled and separated on an Acquity UPLC system using water and acetonitrile as the mobile phases, both containing 0.01% formic acid. Spectra were collected ana analyzed using the MassLynx software (version 4.1, Waters) and deconvoluted using the MaxEnt 1 plugin.

5.1.2 Size exclusion chromatography (SEC)

Protein purifications involving an SEC step were done on an AKTA purification system (GE Healthcare).

5.1.3 SDS-PAGE

Protein samples were diluted in water and reducing Lamelli buffer (Bio-Rad) prior to boiling at 95°C in a heat block for 2 minutes. Proteins were loaded onto 10, 12 or 15% SDS-PAGE gels or 4-20% gradient SDS-PAGE gels (Bio-Rad) and run in a Mini-Protean cassette system (Bio-Rad). Following removal from the cassette, gels were stained with Coomassie staining and imaged on a ChemiDoc XRS+ gel scanner (Bio-Rad). Contrast adjustments and cropping were done using the Image Lab software (Version 5.1, Bio-Rad).

5.1.4 Analytical HPLC-MS

Determination of peptide identities performed via analytical HPLC were performed on a Waters H-class chromatography machine with an Acquity UPLC-BEH C18 2.1x50mm reverse-phase column connected to a Quaternary Solvent manager, a Waters autosampler and a UV detector with a Waters TUV detector set to 220nm. The method for the peptide analysis was set to: A = H2O + 0.1% TFA, B = MeCN + 0.1% TFA, gradient set to 5-95% B 0-5 min, flow rate 0.6 mL/min.

5.1.5 Software-based data analysis

Word processing and data tabulation was done with Microsoft Word and Excel, respectively. Figures were created with BioRender.com and/or Adobe Illustrator. Data analysis and

statistics were done with GraphPad Prism 8. Microscopy images and multichannel analysis were processed with ImageJ accessed via the FIJI package.

5.1.6 Protein concentration determination

Molar extinction coefficients for proteins were determined by inputting the amino acid sequence to the online tool ProtParam (Swiss Institute of Bioinformatics, https://web.expasy.org/protparam/). Spectrophotometric measurements were performed at 280nm, and concentrations were calculated using the Beer-Lambert law.

5.2 Buffers

5.2.1 Lysis buffer

Lysis buffer for protein expression in E. coli was 50mM Tris, 300mM NaCl and 0.1mM $MgCl_2$ supplemented with DNAse1 and a Complete Ultra EDTA-free protease inhibitor tablet (Roche). 10ml aliquots were prepared fresh before each lysis.

5.2.2 Sortase buffer

Sortase reactions were performed in 50mM Tris 150mM NaCl 10mM $CaCl_2$ (pH 7.5).

5.2.3 Crosslinking/conjugation buffer

Crosslinking reactions were performed in 50mM Tris 500mM NaCl (pH 8.5).

5.2.4 Chitin-column binding buffer

CBD-tagged proteins were immobilized on Chitin resin in 20mM HEPES buffer pH 8.5 with 500mM sodium chloride.

5.2.5 CBD-tagged Lysis buffer

CBD-tagged proteins were lysed in chitin-column buffer supplemented with an EDTA-free protease inhibitor cocktail (Roche) and DNAse I.

5.3 Peptide Synthesis

All peptides were synthesized with help from Ines Kretzschmar. Phosphonamidate reagents were kindly contributed by several colleagues in the Hackenberger group, including Christian Stieger and Sarah Hansen.

5.3.1 LPETGG/LPETGGHHHHHH

SORTag peptide (LPETGG-$CONH_2$ and LPETGGHHHHHH-$CONH_2$) was synthesized using FMOC-SPPS on an automated Activotec peptide synthesizer with Rink amide resin (0.05mmol scale, 0.23mmol/g) by Ines Kretzschmar. Subsequent functionalization steps (addition of fluorophores/electrophiles) were done by hand. Amino acid couplings were done using 8 equivalents of amino acid in the presence of an HOBt/HBTU/DIPEA solution. FMOC cleavage prior to each coupling and following the final coupling was done using 4ml of 20% piperidine in DMF. Peptide was cleaved from the resin using a cleavage cocktail consisting of Phenol: H2O:EDT:Methylphenylsulfide:TFA in ratios of 3:2:1:2:40 and purified via reverse-phase HPLC or left on the resin for subsequent functionalization.

5.3.2 TAMRA-LPETGG

TAMRA was coupled off-resin to the leucine residue of the LPETGG-$CONH_2$ peptide. Coupling was done in the presence of TAMRA (1.2eq), HATU (1.2eq) and DIPEA (2.4eq) in DMF via an overnight incubation at 25°C with 800RPM shaking. The peptide was precipitated in cold diethyl ether, dried, and then dissolved in 10% AcN, 0.1% TFA followed by purification on a reverse-phase preparatory C18 HPLC column (Gilson) using a 10% to 70% AcN gradient in 30 minutes. The peptide was isolated in 35% yield.

5.3.3 PEG_2-Alkyne-P(V)-LPETGG

NHS-PEG_2-Alkyne-P(V) (1.8eq) was coupled off-resin to the leucine residue of the LPETGG-$CONH_2$ peptide in the presence of DIPEA (3.6eq) in DMF via 2-hour incubation at room temperature with 800RPM shaking. The peptide was precipitated in cold diethyl ether, dried, and then dissolved in 10% AcN, 0.1% TFA followed by purification on a reverse-phase preparatory C18 HPLC column (Gilson) using a 10% to 70% AcN gradient in 30 minutes. The pure peptide was isolated in 24% yield.

5.3.4 Maleimide-LPETGG

SMCC (1.4eq) was coupled off-resin to the leucine residue of the LPETGG-$CONH_2$ peptide in the presence of DIPEA (1.7eq) in DMF via 1 hour incubation at room temperature with 800RPM shaking. The peptide was precipitated in cold diethyl ether, dried, and then dissolved in 10% AcN, 0.1% TFA followed by purification on a reverse-phase preparatory C18 HPLC column (Gilson) using a 10% to 70% AcN gradient in 30 minutes. The pure peptide was isolated in 39% yield.

5.3.5 Maleimide-LPETGGHHHHHH

SMCC (1.8eq) was coupled on-resin to the leucine residue of the LPETGGHHHHHH peptide in the presence of DIPEA (10eq) in DMF via overnight incubation at room temperature with 400RPM shaking. Cleavage was performed in the presence of TFA:TIS:H_2O in ratios of 38:1:1 for 1 hour at room temperature. The peptide was precipitated in cold diethyl ether, dried, and then dissolved in 5% AcN, 0.1% TFA followed by purification on a reverse-phase preparatory C18 HPLC column (Gilson) using a 5% to 95% AcN gradient in 35 minutes. The pure peptide was isolated in 11% yield.

5.3.6 TAMRA-LPETGGHHHHHH

TAMRA (4.4eq) was coupled on-resin to the leucine residue of the LPETGGHHHHHH peptide in the presence of HATU (4.1eq) and DIPEA (10eq) in DMF via overnight incubation at room temperature with 400RPM shaking. Cleavage was performed in the presence of TFA:TIS:H_2O in ratios of 38:1:1 for 1 hour at room temperature. The peptide was precipitated in cold diethyl ether, dried, and then dissolved in 5% AcN, 0.1% TFA followed by purification on a reverse-phase preparatory C18 HPLC column (Gilson) using a 5% to 95% AcN gradient in 35 minutes. The pure peptide was isolated in 11% yield.

5.3.7 PEG_2-Alkyne-P(V)-LPETGGHHHHHH

NHS-PEG_2-Alkyne-P(V) (1.1eq) was coupled on-resin to the leucine residue of the LPETGGHHHHHH peptide in the presence of DIPEA (10eq) in DMF via overnight incubation at room temperature with 400RPM shaking. Cleavage was performed in the presence of TFA:TIS:H_2O in ratios of 38:1:1 for 1 hour at room temperature. The peptide was precipitated in cold diethyl ether, dried, and then dissolved in 5% AcN, 0.1% TFA followed by purification on a reverse-phase preparatory C18 HPLC column (Gilson) using a 5% to 95% AcN gradient in 35 minutes. The pure peptide was isolated in 1.6% yield.

5.4 Cloning

5.4.1 Cloning *N*-terminal polyglycine tag onto GBP1

The first attempt to convert a nanobody into a sortase substrate was done by recombinantly engineering an *N*-terminal polyglycine tag using side-directed mutagenesis with the Q5 kit NEB (New England Biolabs, USA) while retaining the PelB leader sequence. PelB-GBP1-H6

(plasmid p47) was amplified using primers Alec89 and Alec107 and then treated with the kit as described in its protocol to generate a full-length plasmid for transformation.

5.4.2 Cloning GBP1 into pET28a

To generate an anti-GFP nanobody with a sortase-accessible *N*-terminal glycine residue, GBP1 was cloned into pET28a. pET28a (plasmid 34) was first digested with NdeI and HindIII. GBP1 in pET22b (plasmid p2) was then amplified using primers Alec112 and Alec113 and then ligated into pET28a using standard Gibson assembly with the HiFi DNA assembly kit (New England Biolabs, USA) before transformation into bacteria.

5.4.3 Cloning GBP1-AAAC

PelB-GBP1-H6 in an amber suppression vector (plasmid 47) was modified using the q5 site-directed mutagenesis kit to contain a three-alanine spacer followed by a C-terminal engineered cysteine using primers Alec105 and Alec106.

5.4.4 Cloning GBP1 active site mutations

Residue screening to identify essential residues, as well as the introduction of the TAG amber stop codon for amber suppression into the active site of GBP1 was accomplished using the q5 site directed mutagenesis kit (New England Biolabs, USA). Primers for each mutagenesis reaction can be identified in the primers table (Table 6) Plasmids used can be identified in the plasmids table (Table 7).

5.4.5 Transformation

Unless otherwise noted, transformation of recombinant genetic material into bacteria was performed via heat shock into chemically competent cells. 5-50ng of purified plasmid was added to an aliquot on ice and incubated for 30 minutes prior to a 30 second heat shock incubation in a hot water bath for 30 seconds. After 2 minutes on ice for recovery, super-optimal content media was added, and the cells were incubated at 37C 180RPM for 60 minutes before plating on LB plates with appropriate antibiotics for colony selection.

5.5 Protein expression

All proteins, unless otherwise indicated, were expressed from bacterial vectors using a T7 promoter expression system utilizing bacterial strains with the T7 RNA polymerase under control of a Lac operon.

5.5.1 Lac protein expression

BL21(DE3) is a chemically competent strain intended for T7 expressions. The DE3 marker indicates that it carries a copy of the phage DNA with the DE3 lysogen with the gene for T7 RNA polymerase under control of the lacUV5 promoter inserted into the bacterial chromosome. This means that the polymerase can be induced using IPTG, and therefore the protein of interest under the T7 promoter can be induced using IPTG. The induced T7 polymerase is highly specific for the T7 promoter. At low concentrations IPTG and lactose require lactose permease, a product of the lac operon, to be taken up into cells. At high IPTG concentrations used for induction, uptake can happen independently of the permease. IPTG induction therefore induces expression in two ways: first by removing the repressor on the LacUV5 promoter upstream of T7 RNA polymerase in the E coli genome to allow T7 Polymerase expression, and then by removing the repressor on T7 promoter upstream of the GOI to allow GOI expression. The IPTG induction induces expression of the T7 polymerase from the DE3 lysogen encoded in the Bl21(de3) genome. This expression is conducted by the native E. coli polymerase. The advantage of combining the LacI induction system with the t7 expression system is that it avoids unwanted basal protein expression.

5.5.2 araBAD protein expression

The arabinose operon is composed of three genes (AraB, AraA and AraD, collectively known as araBAD) that get transcribed as a single transcript by the AraC and the catabolite activator protein (CAP)-cAMP complex. The CAP-cAMP regulation is part of a global transcriptional regulation system involving catabolic genes. Low levels of glucose, for example, increase cAMP expression, which is sensed by CAP and induces catabolite gene expression (including araBAD expression). This regulation system is in parallel to the AraC regulation system. AraC is both a transcriptional repressor in the absence of arabinose, and a transcriptional activator in the presence of arabinose. AraC is a homodimer with a dimerization domain, a DNA binding domain, and an arabinose binding site. In the absence of arabinose, AraC binds to both the operator and a distant DNA site. Dimerization between these two monomers leads to formation of a DNA loop that blocks RNA polymerase from binding. In the presence of arabinose, AraC and CAP-cAMP work together to activate araBAD expression. Arabinose binding induces a conformational change in AraC, causing one AraC monomer to fall off the araBAD operon, thereby breaking the DNA loop. An AraC monomer also binds near the

araBAD promoter in this configuration, which helps recruit RNA polymerase to the promoter. In the absence of glucose, cAMP levels are increased, forming more CAP-cAMP complexes which bind to the araBAD operon, opening the DNA loop further to help AraC bind and recruit RNA polymerase.

5.5.3 Sortase

Sortase Δ59 from S. aureus lacking the membrane-bound domain of the wild-type variant was acquired in the pET28a vector as described previously [95]. The sortase 5M pentamutant was acquired from Addgene (#75144) [88]. BL21(DE3) E. coli cells were transformed via heat shock at 42°C with the appropriate plasmid and subjected to antibiotic selection on an LB+agar plate containing kanamycin. A single colony was selected from the resulting agar plate and grown overnight in 7ml LB medium supplemented with 40 µg/mL kanamycin. The next day, 1L of room-temperature LB medium was supplemented with 5ml of overnight culture and grown at 37°C until OD 0.5-0.8. Induction was performed with 0.4mM IPTG before incubation at 30°C for 3 hours 180RPM. Cells were harvested via centrifugation at 4000xG for 15minutes and the pellet lysed in lysis buffer via sonication (Branson Sonifier) for 2 minutes at 30% amplitude (10s bursts, 30 second intervals, max temperature 14°C). Lysate debris was pelleted at 25,000xG for 30 minutes and filtered with a 0.45µM filter. Lysate was purified first by hand via nickel-NTA chromatography using HisPur resin (ThermoFischer Scientific) equilibrated with 1X PBS containing 5mM imidazole (pH 7.5). After 80ml of washing with the equilibration buffer, elution was done with 6ml 1X PBS containing 500mM imidazole (pH 7.5) and the eluate concentrated in a VivaSpin 6 5kDa MWCO spin filter. Concentrated eluate was further purified and changed into sortase buffer via SEC with a Superdex 75 10/300GL column with a 0.8ml/min flow rate.

5.5.4 Nanobody expression

Sortase-substrate nanobodies were either expressed natively with glycine at their *N*-termini following methionine cleavage by bacterial methionine aminopeptidase (Figure 13), or were generated by protease cleavage at the *N*-terminus to expose an accessible glycine (Figure 14). Thiol-containing nanobody crosslinking partners were expressed as intein-CBD genetic fusions and then cleaved using cysteine to generate a *C*-terminal cysteine residue linked via a peptide bond (Figure 19).

5.5.4.1 antiEGFR nanobody, sortase substrate

WT Nanobody 7d12 was acquired from Addgene (#125268) in the pTrcHis vector to enable TEV cleavage following purification, thereby rendering the *N*-terminus accessible to sortase transpeptidation (Figure 13). Overnight expression at 30°C was done in Bl21(DE3) E. coli in 1L of LB supplemented with carbenicillin following inoculation with 5ml of overnight culture and induction at OD0.64 with 1mM IPTG. Cells harvested by centrifugation were resuspended in 50ml lysis buffer and lysed on a microfluidizer at 15000PSI before clarification at 25,000xG for 20minutes and filtration with a 0.45µm filter. Purification was done first by hand using HisPur NiNTA resin (ThermoFischer Scientific) equilibrated with 1X PBS containing 20mM imidazole (pH 7.5) with extensive washing followed by elution in 6ml 1X PBS containing 500mM imidazole (pH 7.5). Eluate was concentrated in a Vivaspin 6 5kDa MWCO spin filter and then subjected to overnight incubation with TEV protease (1:20 TEV:Protein by mass) in a dialysis membrane against 1X PBS without imidazole. Following TEV cleavage, the entire mixture was purified via reverse-NiNTA chromatography, again by hand, to remove uncleaved nanobody and the TEV in a single step. Purified nanobody was finally subjected to a final purification and buffer exchange step into sortase buffer via SEC with a Superdex75 10/300GL column and 0.8ml/min flow rate.

5.5.4.2 antiEGFR-Cys nanobody (thiol-specific crosslinking partner)

AntiEGFR-Intein-CBD (plasmid p57, cloned by Anselm Schneider into pTXB1) was the nanobody-intein fusion used to generate a nanobody variant with a *C*-terminal cysteine as a crosslinking handle via expressed protein ligation with cysteine as the ligation partner (Figure 19A). The plasmid was transformed into T7 Express *E. coli* cells as described above and plated on LB+ carbenicillin. Single colonies were selected, and overnight cultures made in LB+ carbenicillin. Following overnight incubation, 3.5ml ONC was diluted in 1L LB + carbenicillin and grown until OD 0.6 at 37°C, 180RPM prior to induction with 0.4mM IPTG. After expression for 2 hours and 40minutes at 30°C, cells were harvested by centrifugation at 4000G for 20 minutes. Cell pellets were resuspended in 10ml CBD lysis buffer and sonnicated with 1 second pulses separated by 1 second pauses for 4 minutes total at 30% amplitude (Branson Sonifier) on ice. Filtered lysate was passed over a 5ml bed-volume chitin-agarose resin column via gravity flow and washed with chitin column buffer (100ml). Fresh EPL cleavage buffer was prepared by supplementing column buffer with 100mM cysteine and 10mM MESNA (re-

adjusted to pH 8.5). The column was flushed with two bed volumes of cleavage buffer and then incubated with 5ml cleavage buffer overnight at 8°C in the cold room. The next day the eluate was analyzed by SDS-PAGE and product containing fractions were pooled and purified via SEC into conjugation buffer prior to aliquoting and storage at minus 80°C (Figure 19B).

5.5.4.3 GBP1 nanobody (active site mutants and sortase substrate)

GBP1 has been previously described and was simply cloned into the pET28a vector to add a H_6 purification tag that could be cleaved following purification using thrombin [12] (Figure 14). Following transformation, BL21(DE3) cells prepared from a single colony into an overnight culture were added to 1L LB containing kanamycin and grown until OD0.74, at which point they were induced with 1mM IPTG and grown overnight at 30°C. Cells were harvested by centrifugation and resuspended in lysis buffer before being lysed by sonication (2minutes, 30% amplitude, 10 second bursts, 30 second interval, max temp 14°C). Lysate was clarified by centrifugation and then purified by hand using HisPur NiNTA resin (ThermoFischer Scientific) with 1X PBS 5mM imidazole wash/equilibration buffer and 500mM imidazole elution buffer. Fractions were analyzed and product-containing fractions were pooled and dialyzed into PBS without imidazole.

To generate the nanobody sortase substrate, Rev-NiNTA product-containing fractions were concentrated and then subjected to overnight Thrombin cleavage (20U/ml) in a dialysis membrane against sortase buffer. The resulting mixture was subjected to reverse-NiNTA chromatography with sortase buffer as wash/equilibration buffer to remove the protease and exchange the buffer into sortase buffer.

5.5.4.4 GBP1 with C-terminal engineered cysteine (thiol-specific crosslinking partner) and N-terminal polyglycine motif

Following transformation, BL21(DE3) cells prepared from a single colony into an overnight culture were added to 1L LB containing tetracycline and grown until OD0.6, at which point they were induced with 2.0G arabinose and grown overnight at 30°C. Cells were harvested by centrifugation and resuspended in imidazole lysis buffer before being lysed by sonication (2minutes, 30% amplitude, 10s bursts, 30s interval, max temp 14°C). Lysate was clarified by centrifugation and then purified by hand using HisPur NiNTA resin (ThermoFischer Scientific) with 1X PBS 5mM imidazole wash/equilibration buffer and 500mM imidazole elution buffer.

Fractions were analyzed and product-containing fractions were pooled and dialyzed into conjugation buffer.

5.6 Sortase reaction and purification

Sortase reactions were performed on scales ranging from 200ul to 700ul depending on the application. Minimum volume requirements were determined by the need for sufficient product recovery for downstream applications following purification of the sortase. Maximum volumes were only set by the amount of resin used in purification and the availability of sortase/peptide reagents.

Prior to beginning the reaction, a Ni-NTA column was prepared with fresh resin and equilibrated to conjugation buffer. The reaction components were then thawed to room temperature and combined to reach the final reaction concentrations indicated in the tables below, with the sortase being added last to start the reaction. Reaction times varied depending on the application. Initial experiments using the wild-type sortase (SrtWT) were carried out for durations ranging from 1hour to overnight (Table 1), while the more-active pentamutant variant requiring only 5 minutes at room temperature with 400rpm shaking (Table 2). After incubation, the entire reaction was immediately added to the pre-equilibrated Ni-NTA column for removal of the sortase (and peptide if using the H6-containing variant). The reaction was eluted immediately using gravity flow and the entire elution, along with 8-10ml washes of conjugation buffer which were collected in a VivaSpin20 5kDa MWCO spin filter (Sigma, USA). The pool was concentrated to between 100 and 500ul and then used for downstream applications.

	[Stock], µM	[Final], µM	Eq.
LPETGG Peptide	1000	300	30
Glycine protein	18.57	10	1
SrtWT	457.06	40	4

Table 1: Sortase A Δ59 (SrtWT) reaction

	[Stock], µM	[Final], µM	Eq.
LPETGG-H6 peptide	10000	1000	20
Glycine protein	757.3	50	1

Srt5M	971.3	55	1.1

Table 2: Sortase A Δ59 Pentamutant (Srt5M) reaction

5.7 Sortase-enabled chemoselective crosslinking reactions

5.7.1 Protein-protein conjugation with P(V)

The thiol-containing conjugation partner was first reduced for 30 minutes with 2eq TCEP for 15minutes at room temperature, 400RPM shaking in conjugation buffer. Following reduction, the TCEP mixture was added directly with the purified, P(V)-containing conjugation partner and reacted for two nights in conjugation buffer. After crosslinking, the product was purified via SEC with a Superdex75 column into PBS (Cytiva, USA) with a 0.6ml/min flow rate (Table 3).

	[Stock], µM	[Final], µM	Eq.
Thiol-containing protein	60	22	0.89
P(V)-containing protein	55	24	1
TCEP	120	43.2	1.8

Table 3: Protein-protein conjugation with P(V)

5.7.2 Protein-protein conjugation with maleimide

The thiol-containing conjugation partner was first reduced for 30 minutes with 2eq TCEP for 15minutes at room temperature, 400RPM shaking. Following reduction, the TCEP mixture was added directly with the purified, maleimide-containing conjugation partner and reacted for between 1 and 12 hours in conjugation buffer. After crosslinking, the product was purified via SEC with a Superdex75 column into PBS (Cytiva, USA) with a 0.6ml/min flow rate (Table 4).

	[Stock], µM	[Final], µM	Eq.
Thiol-containing protein	60	22	0.85
Maleimide-containing protein	55	25	1
TCEP	120	43	1.7

Table 4: Protein-protein conjugation with maleimide

5.7.3 IgG-GFP crosslinking

Prior to crosslinking, the IgG antibody was reduced with 16eq. TCEP at 37C for 1 hour at 200RPM shaking to render the disulfide thiols accessible for targeting. The reduced antibody

mixture was then directly mixed with the maleimide-containing, purified GFP sortase product in conjugation buffer supplemented with 1mM EDTA and incubated overnight at 14°C. After crosslinking, the product was purified via SEC with a Superdex75 column into PBS (Cytiva, USA) with a 0.6ml/min flow rate.

	[Stock], µM	[Final], µM	Eq.
Trastuzumab, reduced	27	12	0.12
Maleimide-GFP	260	99	1.0
TCEP	430	190	1.9

Table 5: IgG-Protein crosslinking

5.7.4 IgG-GFP-(Cys) crosslinking

Prior to crosslinking, the IgG antibody was reduced with 16eq. TCEP at 37C for 1 hour at 200RPM shaking to render the disulfide thiols accessible for targeting. The reduced antibody mixture was then directly mixed with the maleimide-containing, purified cysteine-containing GFP (GFP-(Cys)) sortase product in conjugation buffer supplemented with 1mM EDTA and incubated overnight at 14°C. Purification of the sortase product was done at pH 7.5 to avoid undesired crosslinking, before rebuffering to pH 8.5 prior to IgG conjugation (Figure 64). Crosslinking was performed with 5eq of purified GFP sortase product. After crosslinking, the product was purified via SEC with a Superdex75 column into PBS (Cytiva, USA) with a 0.6ml/min flow rate.

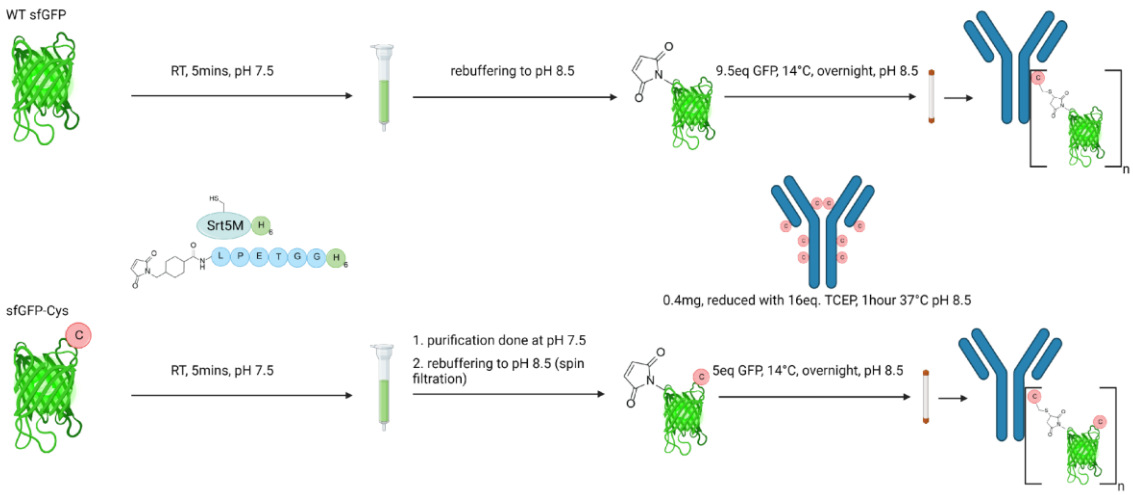


Figure 64: Comparison between WT and cysteine-containing GFP to IgG conjugation protocols.

5.8 PEG Phosphite synthesis

5.8.1 AM002: Pentaethylene glycol toluene-p-sulfonate

1eq (1.0 gram) pentaethylene glycol was mixed with 0.038g methylamine hydrochloride and 1.1ml triethyl amine in 10ml acetonitrile on ice and under an argon atmosphere in a round bottom flask. 2eq (1.53g) toluene sulfonyl chloride was slowly added and then left overnight with stirring. The next day, a turbid orange solution was evident. A small amount of water was added to reduce turbidity, and four rounds of phase separation in a separator funnel was applied to separate the product into ethyl acetate. The ethyl acetate phase was then separated from the aqueous phase and the solvent removed via roto-evaporation. TLC in pure ethyl acetate showed two resolvable spots, indicating that a silica column with pure ethyl acetate would be able to separate the two compounds. The product was therefore purified with a silica column using ethyl acetate as the mobile phase, with the product being eluted using up to 3% methanol. Roto-evaporation removed the ethyl acetate and produced a clean oil weighing 1.3 grams. 1-H NMR successfully confirmed the identity of the desired tosylated pentaethylene glycol (Spectrum 16).

5.8.2 AM003: Pentaethylene glycol ortho-nitrobenzyl alcohol

1eq (0.18g) of ortho-nitrobenzyl alcohol ett13 (kindly provided by Anett Hauser) was added to ~2.4eq (0.33g) potassium carbonate in 3.3ml acetonitrile to form a turbid yellow mixture. 1.4eq (0.50g) AM002 was then added dropwise using 1ml acetonitrile and an immediate color change to dark red was observed. The reaction was then refluxed at 80°C for 3 hours. Following reflux the potassium carbonate salt was filtered off and washed with ethyl acetate, then the solvent was evaporated. The resulting crude oil was stored overnight on the bench. The next day the crude AM003 was purified on a silica column to separate excess AM002. TLC confirmed that both species were mobile in methanol but had similar retention factors to silica in that mobile phase. Therefore, a silica column was performed with ethyl acetate as the first mobile phase, followed by a gradient of increasing methanol concentration to separate both species. 200ml of ethyl acetate was used to start the column, and then 150ml aliquots of Methanol in ethyl acetate were added ranging from 1% to 5% methanol. Sixty fractions were collected and analyzed by TLC, and product fractions were pooled and dried via roto-evaporation before 1H NMR analysis (Spectrum 17).

5.8.3 AM004: Pentaethylene glycol ortho-nitrobenzyl phosphite

152mg of AM003 alcohol in a round bottom flask was transferred to a Schlenk test tube containing a small stir bar using a small amount of THF. The tube was then connected to a Schlenk line for 5 minutes to evacuate the atmosphere and reduce the amount of THF while the stir bar was stirred. While the Schlenk tube was evacuated, a falcon tube for the phosphorous trichloride was prepared. Oxidation is always a fear when working with the phosphorous trichloride, but also hydrolysis from water in the air. Therefore, the falcon tube was opened and an argon stream with a needle was added to the top of the tube to force out the air and replace it with inert argon for 5 minutes. Into this falcon tube a ~0.5M phosphorous trichloride solution was prepared with 10.208µl PCL3 and 234µl THF. The alcohol was then dissolved in 0.468ml of THF for a ~0.75M solution and a sample was taken as a starting material for later TLC analysis. The tube was then lowered into an ice bucket and cooled, before 53.62µl triethylamine was added directly. The solution with the base was allowed to cool on ice for 5 minutes. To this cold solution the 0.5M solution of PCL3 was slowly added dropwise. Vapor was seen and a salt formed (the salt is the product of chloride ions being removed by the base to form ethyl ammonium salts). After stirring on ice for 10 minutes, the solution was removed from the ice bath, wrapped in foil, and an argon balloon was added. The solution was then stirred at room temperature overnight. TLCs of the overnight reaction and starting material were run in 100% Ethyl Acetate, 99% EE and 1% Methanol, and 98% EE with 2% methanol, all with 1% Triethyl Amine. The TEA was used to prevent hydrolysis of the phosphite by the silica gel itself. EE and methanol were shown to be able to separate the unreacted alcohol from the phosphite. A small silica column was prepared with ~250ml silica in Ethyl Acetate. The column was then flushed with 50ml of EE with 1% TEA. The column was loaded with the reaction mixture directly (i.e. no solvent removal or cleanup before loading onto the column) and the column was run in EE with 1% TEA and up to 5% Methanol in 75ml batches. After loading the reaction mixture, 50ml of EE with 1% TEA was added to start the column and collected in an Erlenmeyer flask. After the Erlenmeyer was ~half filled, fraction collection was started and continued as the 1% Methanol 1% TEA, 2% methanol 1% TEA, 3% Methanol 1% TEA, 4% methanol 1% TEA and 5% methanol 1% TEA batches were run. The unreacted alcohol eluted first, followed by the alcohol and phosphite together, and finally the pure phosphite (maybe) eluted last. Three fractions were collected (F1, F2 and F3). F1 contained tubes 4-12, F2 tubes 13-22, and F3 tubes 23-33. These

fractions were evaporated on the roto-evaporator, and most of the fraction was added to NMR tubes in deuterated acetonitrile. NMR analysis was done at the 600mHz NMR with phosphorous NNR, Phosphorous HMBC, and proton NMR for integration (3 runs per sample). The remaining volume of analyzed fractions not added to the NMR tube was washed into pre-weighed 10ml round bottom flasks using a small volume of EE, capped, and stored at -4C overnight.

5.9 GBP1-Mutant GFP binding characterization

5.9.1 Staudinger-phosphite reaction

GBP1 containing an azido-phenylalanine at position R35 in the substrate binding pocket (provided by Prof. Marina Rubini) at a concentration of 14.1uM was mixed with 100eq of phosphite in a black opaque Eppendorf tube and incubated at 26°C for 24 hours. The reaction was monitored by SDS-PAGE and intact protein MS. Unreacted phosphite was removed by gel filtration after 24hours using a 2ml ZEBA spin column. Protein concentration was then determined via BCA assay (Thermo, USA).

5.9.2 UV irradiation

UV irradiation was conducted in a custom-made apparatus composed of a metal cooling block tube holder atop a height-adjustable stand within a Styrofoam box containing a UV lamp on the top. A 3cm separation was created between the UV lamp and the samples contained in Eppendorf tubes. Irradiation was performed at 365nm for 60minutes, and the de-caging was monitored by intact protein MS.

5.9.3 GFP binding assay

The fluorimeter was turned on and allowed to settle for 1 hour prior to measurement, and the temperature controller was set to 25°C (JASCO, USA). Serial dilutions of nanobody were created in PBS. Dilutions were allowed to come to room temperature for 1 hour while kept in foil-wrapped tubes. Dilutions (500ul) were aliquoted into plastic cuvettes (Sarstedt, ref. 67.754), and fluorescence intensity was measured using the fixed wavelength program (ex. 480nm, em. 510nm, 3nm excitation bandwidth, 5nm emission bandwidth, 350V PMT) immediately after the addition of eGFP to the nanobody dilution (final volume 1ml). Measurements were performed in batches, with a standard sample of eGFP alone measured

at the beginning and end of each batch as a baseline to account for photobleaching over the course of each batch.

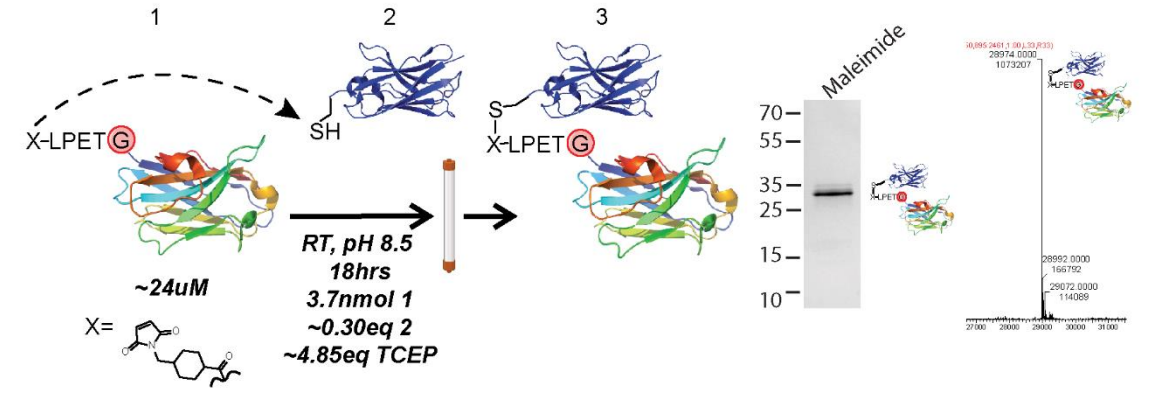
6. Appendices

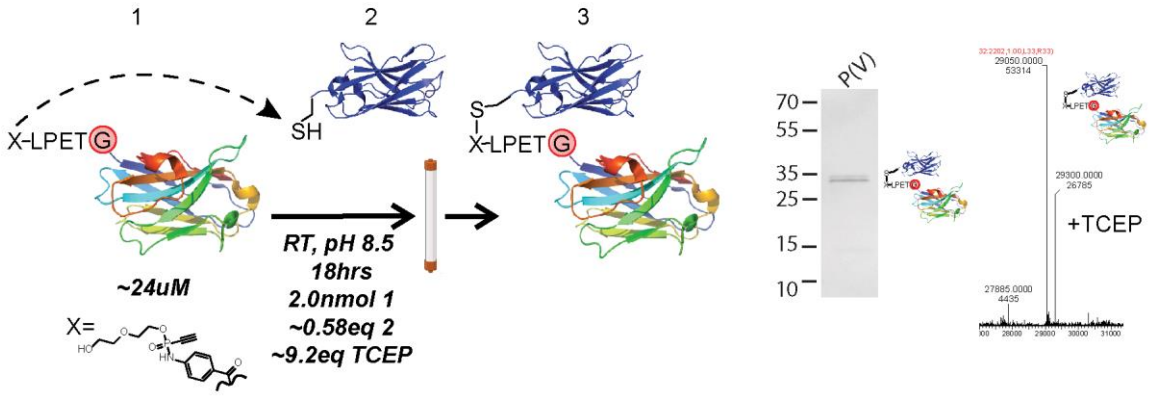
6.1 Abbreviations

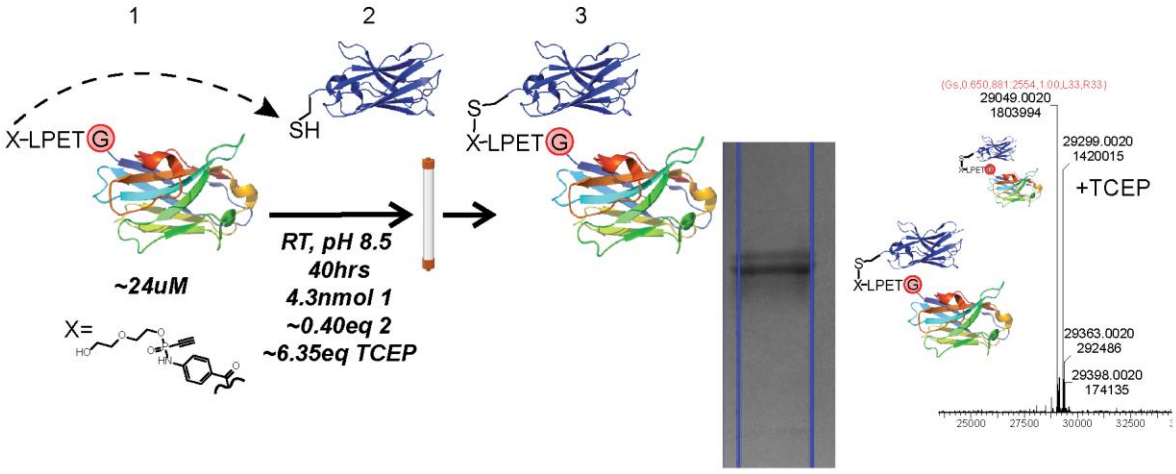
PBS	Phosphate-buffered saline
SEC	Size-exclusion chromatography
CV	Column volume
CBD	Chitin-binding domain
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
NB	Nanobody
ADC	Antibody-drug conjugate
EGFR	Epidermal growth factor receptor
GBP1	GFP Binding Protein 1 (GFP enhancer)
PEG	Poly(ethylene glycol)
Ni	Nickel
NTA	Nitrilotriacetic acid
GFP	Green fluorescent protein
P(V)	Phosphorus five
MS	Mass spectrometry
ESI-MS	Electrospray Ionization – Mass Spectrometry
ONB	Ortho-Nitrobenzyl
TEA	Triethyl amine
THF	Tetrahydrofuran
TLC	Thin layer chromatography
NMR	Nuclear magnetic resonance
HMBC	Heteronuclear multiple bond correlation
EE	Ethyl acetate
pAzF	Para-azido phenylalanine
WT	Wild type

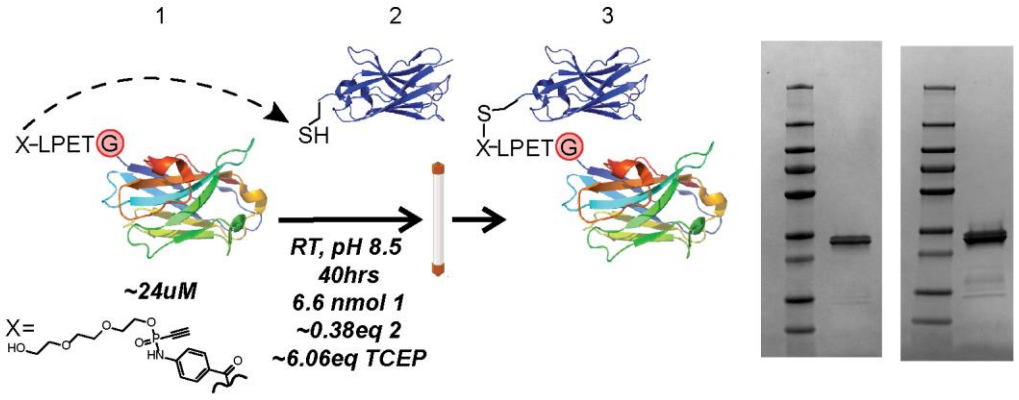
UDP	Uridine diphosphate
CuAAC	Copper-catalyzed azide-alkyne cycloaddition
NHS	*N*-hydroxysuccinimide
SPPS	Solid-phase peptide synthesis
SMCC	Succinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate
FRET	Förster resonance energy transfer
DBCO	Dibenzocyclooctyne
NCL	Native chemical ligation
LDS	lithium dodecyl sulfate
SDS	sodium dodecyl sulfate
IMAC	Immobilized metal affinity chromatography
TCEP	tris(2-carboxyethyl)phosphine

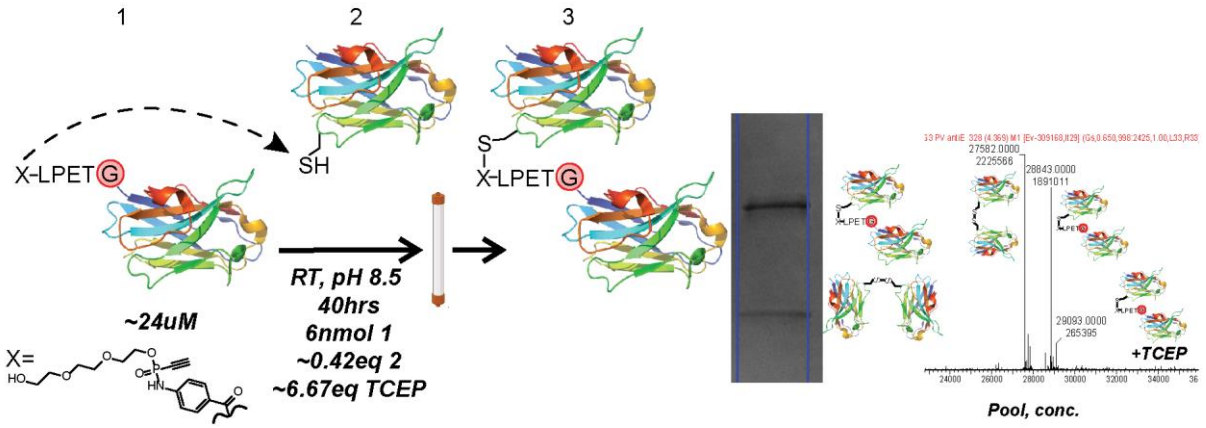
6.2 Crosslinking Optimization Reactions

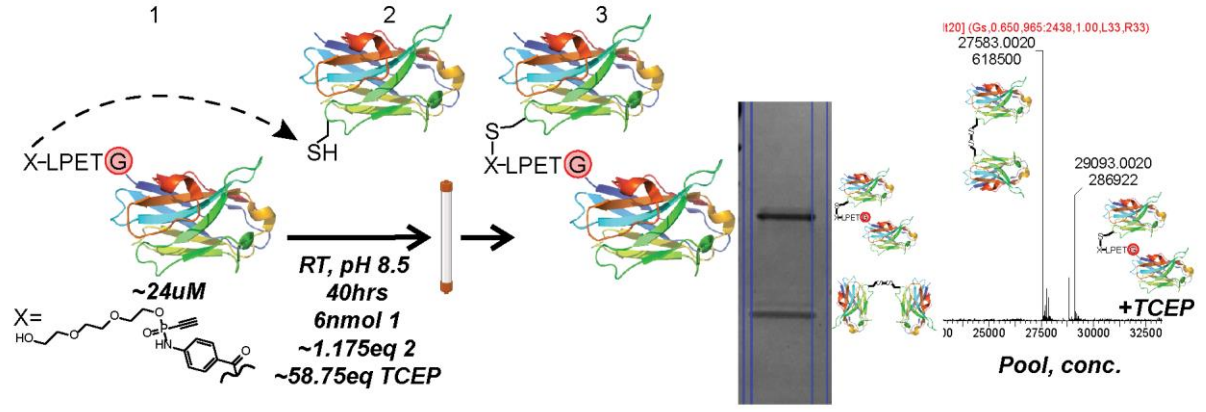


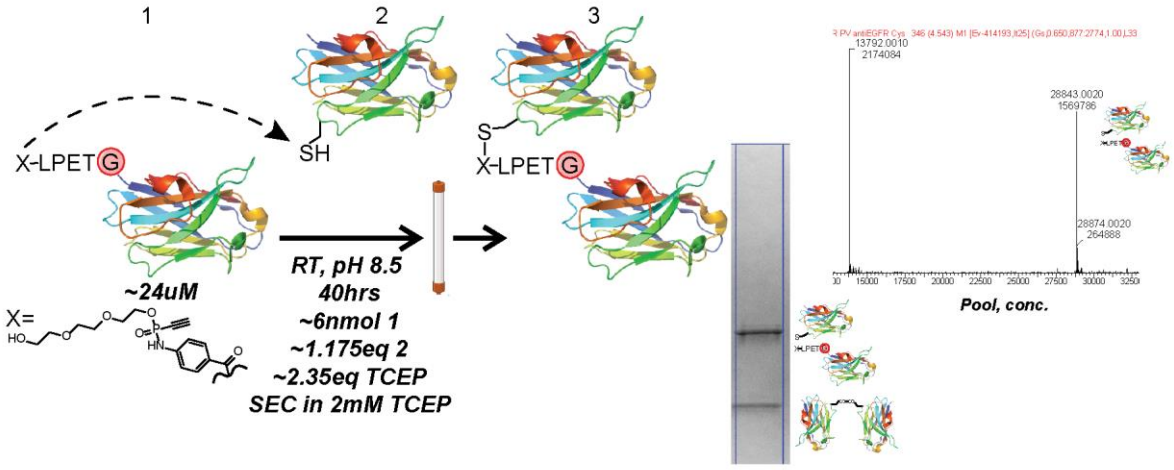


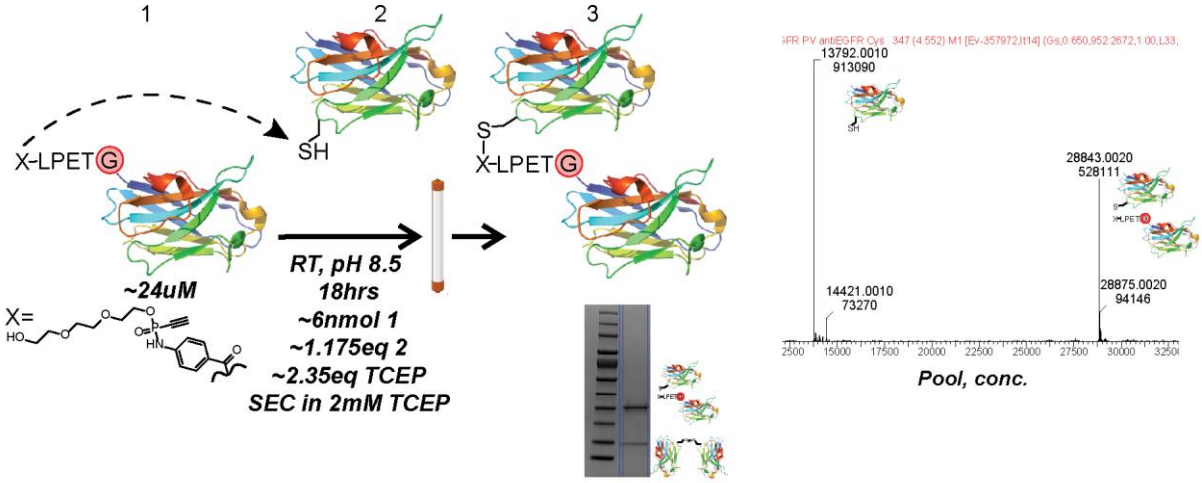


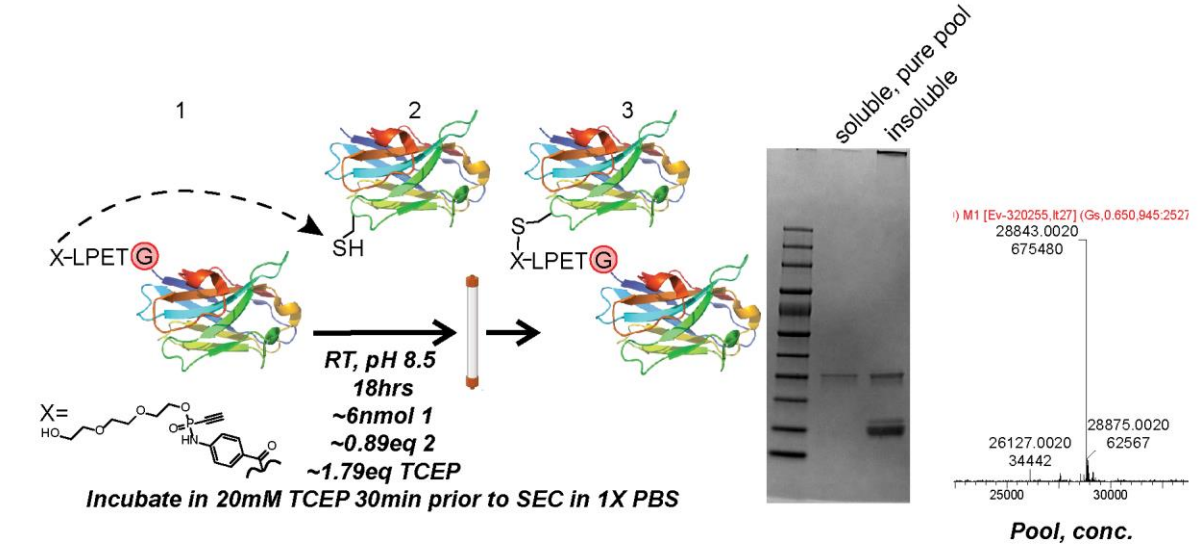

1
2
3
X-LPETG
SH
S
X-LPETG
~24uM
X=
HO
RT, pH 8.5
40hrs
6.6 nmol 1
~0.38eq 2
~6.06eq TCEP

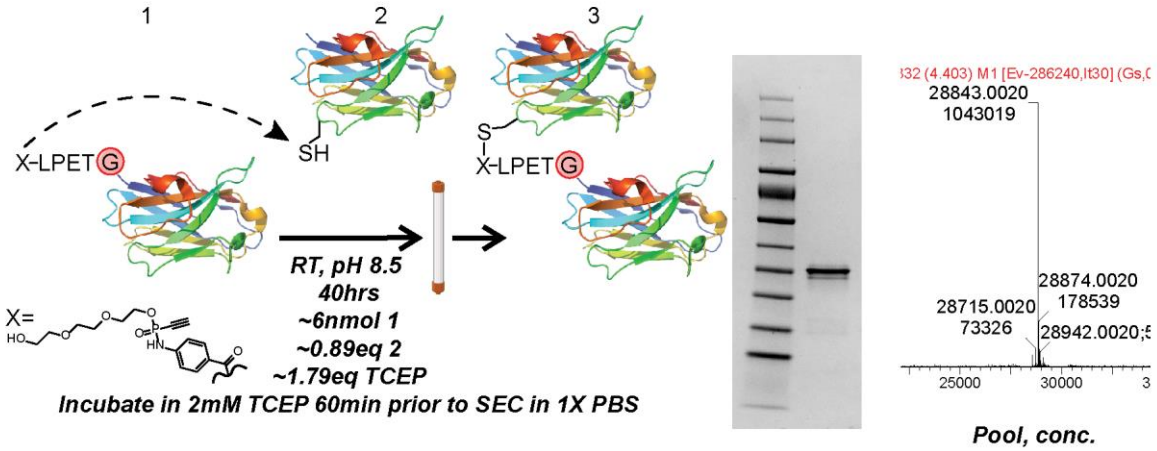

1
2
3
X-LPETG
SH
S
X-LPETG
~24uM
X=
HO
RT, pH 8.5
40hrs
6nmol 1
~0.42eq 2
~6.67eq TCEP
27582.0000
2225566
28843.0000
1891011
29093.0000
265395
+TCEP
Pool, conc.

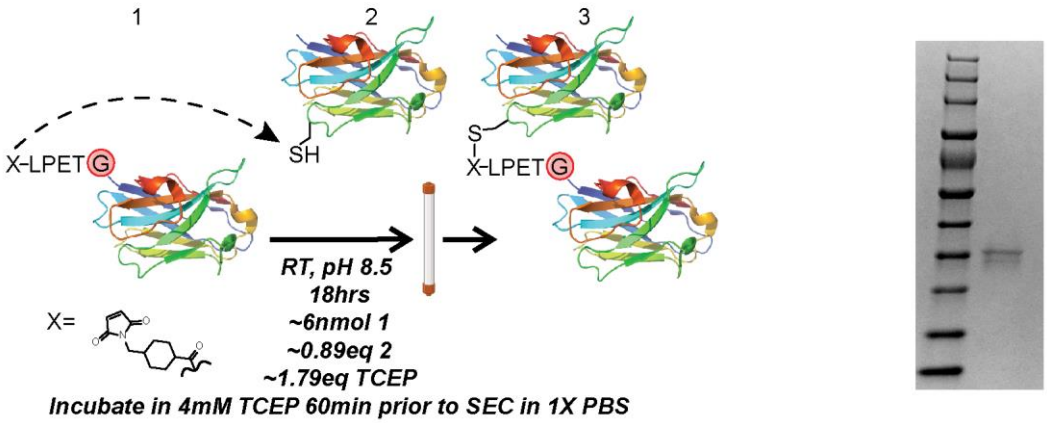

1
2
3
X-LPETG
SH
S
X-LPETG
~24uM
X=
HO
RT, pH 8.5
40hrs
6nmol 1
~1.175eq 2
~58.75eq TCEP
27583.0020
618500
29093.0020
286922
+TCEP
Pool, conc.


1
2
3
X-LPETG
SH
S
X-LPETG
~24uM
X=
HO
RT, pH 8.5
40hrs
~6nmol 1
~1.175eq 2
~2.35eq TCEP
SEC in 2mM TCEP
13792.0010
2174084
28843.0020
1569786
28874.0020
264888
Pool, conc.


1
2
3
X-LPETG
SH
S
X-LPETG
~24uM
X=
HO
RT, pH 8.5
18hrs
~6nmol 1
~1.175eq 2
~2.35eq TCEP
SEC in 2mM TCEP
13792.0010
913090
28843.0020
528111
14421.0010
73270
28875.0020
94146
Pool, conc.


soluble, pure pool
insoluble
1
2
3
X-LPETG
SH
S
X-LPETG
X=
HO
RT, pH 8.5
18hrs
~6nmol 1
~0.89eq 2
~1.79eq TCEP
Incubate in 20mM TCEP 30min prior to SEC in 1X PBS
28843.0020
675480
26127.0020
34442
28875.0020
62567
Pool, conc.





6.3 Primers and Plasmids

Alec1_trpD_Fwd	GCTGACATTCTGCTGCTCG
Alec2_GBP1_R35Amber_F	CCCCGTCAATCGCTATAGTATGTAGTGGTACCGCCAGGCTCCAGG
Alec3_GBP1_R35Amber_R	CCTGGAGCCTGGCGGTACCACTACATACTATAGCGATTGACGGGG
Alec4_GBP1_R35L_F	CCCCGTCAATCGCTATAGTATGCTGTGGTACCGCCAGGCTCCAGG
Alec5_GBP1_R35L_R	CCTGGAGCCTGGCGGTACCACAGCATACTATAGCGATTGACGGGG
Alec6_GBP1_R35I_F	CCCCGTCAATCGCTATAGTATGATCTGGTACCGCCAGGCTCCAGG
Alec7_GBP1_R35I_R	CCTGGAGCCTGGCGGTACCAGATCATACTATAGCGATTGACGGGG
Alec8_GBP1_R57Amber_F	GAGTAGTGCTGGTGATTAGTCAAGTTATGAAGACTCCGTGAAGGG
Alec9_GBP1_R57Amber_R	CCCTTCACGGAGTCTTCATAACTTGACTAATCACCAGCACTACTC

Alec10_GBP1_R57L_F	GAGTAGTGCTGGTGATCTTTCAAGTTATGAAGACTCCGTGAAGGG
Alec11_GBP1_R57L_R	CCCTTCACGGAGTCTTCATAACTTGAAAGATCACCAGCACTACTC
Alec12_GBP1_R57I_F	GAGTAGTGCTGGTGATATTTCAAGTTATGAAGACTCCGTGAAGGG
Alec13_GBP1_R57I_R	CCCTTCACGGAGTCTTCATAACTTGAAATATCACCAGCACTACTC
Alec14_GBP1_K43Amber_F	GGTACCGCCAGGCTCCAGGGTAGGAGCGCGAGTGGGTCGCGGG
Alec15_GBP1_K43Amber_R	CCCGCGACCCACTCGCGCTCCTACCCTGGAGCCTGGCGGTACC
Alec16_GBP1_K43L_F	GGTACCGCCAGGCTCCAGGGCTGGAGCGCGAGTGGGTCGCGGG
Alec17_GBP1_K43L_R	CCCGCGACCCACTCGCGCTCCAGCCCTGGAGCCTGGCGGTACC
Alec18_GBP1_K43I_F	GGTACCGCCAGGCTCCAGGGATCGAGCGCGAGTGGGTCGCGGG
Alec19_GBP1_K43I_R	CCCGCGACCCACTCGCGCTCGATCCCTGGAGCCTGGCGGTACC
Alec20_GBP1_K65Amber_F	GTTATGAAGACTCCGTGTAGGGCCGATTCACCATCTCCAGAGACG
Alec21_GBP1_K65Amber_R	CGTCTCTGGAGATGGTGAATCGGCCCTACACGGAGTCTTCATAAC
Alec22_GBP1_K65L_F	GTTATGAAGACTCCGTGCTGGGCCGATTCACCATCTCCAGAGACG
Alec23_GBP1_K65L_R	CGTCTCTGGAGATGGTGAATCGGCCCAGCACGGAGTCTTCATAAC
Alec24_GBP1_K65I_F	GTTATGAAGACTCCGTGATCGGCCGATTCACCATCTCCAGAGACG
Alec25_GBP1_K65I_R	CGTCTCTGGAGATGGTGAATCGGCCGATCACGGAGTCTTCATAAC
Alec26_GBP1_K87Amber_F	CTGCAAATGAACAGCCTGTAGCCTGAGGACACGGCCGTG
Alec27_GBP1_K87Amber_R	CACGGCCGTGTCCTCAGGCTACAGGCTGTTCATTTGCAG
Alec28_GBP1_K87L_F	CTGCAAATGAACAGCCTGCTACCTGAGGACACGGCCGTG
Alec29_GBP1_K87L_R	CACGGCCGTGTCCTCAGGTAGCAGGCTGTTCATTTGCAG
Alec30_GBP1_K87I_F	CTGCAAATGAACAGCCTGATACCTGAGGACACGGCCGTG
Alec31_GBP1_K87I_R	CACGGCCGTGTCCTCAGGTATCAGGCTGTTCATTTGCAG
Alec32_GBP1_R35K_F	CCCCGTCAATCGCTATAGTATGTAGTGGTACCGCCAGGCTCCAGG
Alec33_GBP1_R35K_R	CCTGGAGCCTGGCGGTACCACTACATACTATAGCGATTGACGGGG
Alec34_GBP1_R57K_F	GAGTAGTGCTGGTGATAAGTCAAGTTATGAAGACTCCGTGAAGGG
Alec35_GBP1_R57K_R	CCCTTCACGGAGTCTTCATAACTTGACTTATCACCAGCACTACTC
Alec36_pET22b_Seq	GGGACGCGCCCTGTAGCGGCG
Alec37_pET22b_Seq	AGTTCTGCTATGTGGCGCGG
Alec38_pET22b_Seq	AAGATCAAAGGATCTTCTTG
Alec39_pET22b_Seq	GGTATTTTCTCCTTACGCATCTG
Alec40_pET22b_Seq	CGCGTATCGGTGATTCATTCTG

Alec41_pET22b_Seq	GAACGATGCCCTCATTCAGCA
Alec42_pET22b_Seq	CGACGCTCTCCCTTATGCGA
Alec43_pMAL_Seq	CCGACACCATCGAATGGTGC
Alec44_pMAL_Seq	GCGATGCTGGTTGCCAACGA
Alec45_pMAL_Seq	CGCCAAAATCGAAGAAGGTA
Alec46_pMAL_Seq	ATTAACGCCGCCAGTCCGAACA
Alec47_pMAL_Seq	GCGGCATTTTGCCTTCCTGT
Alec48_pMAL_Seq	CGCTGAGATAGGTGCCTCACTGA
Alec49_pMAL_Seq	GAGTGAGCTGATACCGCTCGCC
Alec50_GBP1_Forward	GGGGCCATGGGATGTGCAGCTGGTGGAGTCTGGGGGAGCCTTGGTGCAGC
Alec51_GBP1_Reverse	CCCCGAATTCTTATGAGGAGACGGTGACCTGGGTCCCCTGGCCCCAGTAC
Alec52_GBP4_Forward	GGGGCCATGGGGCCGATGTGCAGCTGCAGGAGTCTGGGGGAGGC
Alec53_GBP4_Reverse	CCCCGAATTCTTATGAGGAGACGGTGACCTGGGTCCCTTTGCCCC
Alec54_PCNA_Forward	GGGGCCATGGGCTCAGGTGCAGCTGGTGGAGTCTGGGGG
Alec55_PCNA_Reverse	CCCCGAATTCTTATGAGGAGACGGTGACCTGGGTCCCC
Alec56_GBP1_Blunt_Forward	GATGTGCAGCTGGTGGAGTCTGGGGGAGCC
Alec57_GBP4_Blunt_Forward	GCCGATGTGCAGCTGCAGGAGTCTGG
Alec58_PCNA_Blunt_Forward	GCTCAGGTGCAGCTGGTGGAGTCTGGGGG
Alec59_Y37F_GBP1_Forward	CGCTATAGTATGAGGTGGTTCCGCCAGGCTCCAGGG
Alec60_Y37F_GBP1_Reverse	CCCTGGAGCCTGGCGGAACCACCTCATACTATAGCG
Alec61_Y116F_GBP4_Forward	CGTCGGACTTTCGTGTTTAGACTTCGTCATGGACTACTGGGGCAAAGG
Alec62_Y116F_GBP4_Reverse	CCTTTGCCCCAGTAGTCCATGACGAAGTCTAAACACGAAAGTCCGACG
Alec63_G15TAG_GBP1_Forward	GGGGGAGCCTTGGTGCAGCCGTAGGGGTCTCTGAGACTCTCCTGTG
Alec64_G15TAG_GBP1_Reverse	CACAGGAGAGTCTCAGAGACCCCTACGGCTGCACCAAGGCTCCCCC
Alec65_G15TAG_GBP4_Forward	GGGGGAGGCTCGGTGCAGGCTTAGGGGTCTCTGAGACTCTCATGTG
Alec66_G15TAG_GBP4_Reverse	CACATGAGAGTCTCAGAGACCCCTAAGCCTGCACCGAGCCTCCCCC
Alec67_GBP4_Blunt2_Forward	GATGTGCAGCTGCAGGAGTCTGGGGGAGGC

Alec68_Y37F_GBP1_Forward	CGCTATAGTATGAGGTGGTTCCGCCAGGCTCCAGGG
Alec69_Y37F_GBP1_Reverse	CCCTGGAGCCTGGCGGAACCACCTCATACTATAGCG
Alec70_Y116F_GBP4_Forward	CGTCGGACTTTCGTGTTTAGACTTCGTCATGGACTACTGGGGCAAAGG
Alec71_Y116F_GBP4_Reverse	CCTTTGCCCCAGTAGTCCATGACGAAGTCTAAACACGAAAGTCCGACG
Alec72_G15TAG_GBP1_Forward	GGGGGAGCCTTGGTGCAGCCGTAGGGGTCTCTGAGACTCTCCTGTG
Alec73_G15TAG_GBP1_Reverse	CACAGGAGAGTCTCAGAGACCCCTACGGCTGCACCAAGGCTCCCCC
Alec74_G15TAG_GBP4_Forward	GGGGGAGGCTCGGTGCAGGCTTAGGGGTCTCTGAGACTCTCATGTG
Alec75_G15TAG_GBP4_Reverse	CACATGAGAGTCTCAGAGACCCCTAAGCCTGCACCGAGCCTCCCCC
Alec76_GBP1_Y37F_NEB_Forward	ATGAGGTGGTTCCGCCAGGCT
Alec77_GBP1_Y37F_NEB_Rev	ACTATAGCGATTGACGGGG
Alec78_GBP4_Y118F_NEB_Forward	TGTTTAGACTTCGTCATGGACTACTGGGG
Alec79_GBP4_Y118F_NEB_Rev	CGAAAGTCCGACGGCGCC
Alec80_GBp1_R35K_NEB_F	CTATAGTATGAAGTGGTACCGCC
Alec81_GBP1_R35K_NEB_R	CGATTGACGGGGAATCCA
Alec82_GBP1_R35G_NEB_F	CTATAGTATGGGGTGGTACCGCC
Alec83_GBP1_E44G_NEB_F	TCCAGGGAAGGGGCGCGAGTGGG
Alec84_GBP1_E44G_NEB_R	GCCTGGCGGTACCACCTCATAC
Alec85_GBP1_S33G_NEB_F	CAATCGCTATGGGATGAGGTGGTACC
Alec86_GBP1_S33G_NEB_R	ACGGGGAATCCAGAGGCT
Alec87_GBP1_S58G_NEB_F	TGGTGATCGTGGGAGTTATGAAGACTCC
Alec88_GBP1_S58G_NEB_R	GCACTACTCATACCCGCG
Alec89_GBP1_PolyGly_NEB_F	CGGTGCCGATGTGCAGCTGGTG
Alec90_GBP1_PolyGly_NEB_R	CCACCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAG
Alec91_GBP1R35Y_NEB_F	CTATAGTATGTACTGGTACCGCCAGGCTCC
Alec92_GBP1R35Y_NEB_R	CGATTGACGGGGAATCCA

Alec93_Gibson1_Forward	CCCAGGTCACCGTCTCCTCACATCATCACCATCACCAT
Alec94_Gibson1_Reverse	GTCGGCAGCAGGTATTTCATGGTTAATTCCTCCTGTTAG
Alec95_Gibson2_Forward	GCTAACAGGAGGAATTAACCATGAAATACCTGCTGCC
Alec96_Gibson2_Reverse	CAATGGTGATGGTGATGATGTGAGGAGACGGT
Alec97_GBP1_R35TAG_NEB_F	CTATAGTATGTAGTGGTACCGCCAG
Alec98_GBP1_R35TAG_NEB_R	CGATTGACGGGGAATCCA
Alec99_GBP1_Y37TAG_NEB_F	TGAGGTGGTAGCGCCAGGCTC
Alec100_GBP1_Y37TAG_NEB_R	TACTATAGCGATTGACGGGGAATCC
Alec101_GNE/MNK_qPCR_F	TAACTATTTCACCCTTCATGC
Alec102_GNE/MNK_qPCR_R	GAGATGGTTCGAGTGATGC
Alec103_GAPDH_qPCR_F	CCACCCATGGCAAATTCCATGGCA
Alec104_GADPH_qPCR_R	TCTAGACGGCAGGTCAGGTCCACC
Alec105_GBP1_AAAC_F	GCATGCCATCATCACCATCACCATTGAG
Alec106_GBP1_AAAC_R	TGCTGCTGAGGAGACGGTGACCTG
Alec107_GBP1_PolyGly2_NEB_R	CCACCCATGGCCATCGCCGGCTG
Alec108_eGFPC70M_PolyG_F	GGCGGTGGCAGCAGCCATCATCAT
Alec109_eGFPC70M_PolyG_R	CATGGTATATCTCCTTCTTAAAGTTAAAC
Alec110_HiFi_P2_F	CTGGTGCCGCGCGGCAGCCAGATGTGCAGCTGGTGGAG
Alec111_HiFi_P2_R	GGTGCTCGAGTGCGGCCGCATGAGGAGACGGTGACCTG
Alec112_HiFi_P2_Fv2	CTGGTGCCGCGCGGCAGCCATGATGTGCAGCTGGTGGAG
Alec113_HiFi_P2_Rv2	GGTGCTCGAGTGCGGCCGCATTATGAGGAGACGGTGACCTG
Alec114_HiFi_P35_F	GCCTGGTGCCGCGCGGCAGCGATGTGCAGCTGCAGGAG
Alec115_HiFi_P35_R	GGTGCTCGAGTGCGGCCGCATTATGAGGAGACGGTGACCTG
Alec116_HiFi_p64_F	GCCTGGTGCCGCGCGGCAGCCAGGTGCAGCTCGTGGAG
Alec117_HiFi_p64_R	GGTGCTCGAGTGCGGCCGCATTATGAGGAGACGGTGACCTG
Alec118_HiFI_p65_F	GCCTGGTGCCGCGCGGCAGCCAGGTGCAGCTCGTGGAG
Alec119_HiFi_p65_R	GGTGCTCGAGTGCGGCCGCATTATGAGGAGACGGTGACCTG
Alec120_HiFI_p66_F	GCCTGGTGCCGCGCGGCAGCCAGGTGCAGCTCGTGGAG
Alec121_HiFi_p66_R	GGTGCTCGAGTGCGGCCGCATTATGAGGAGACCGTGACCTG
Alec122_q5SDM_p61_F	GCTGCTCCTCGCTGCCCAGCCGGCGATGGCCGGCAGCAGCCATCATCAT

Alec123_q5SDM_p61_R	AGACCAGCAGCAGCGGTCGGCAGCAGGTATTTCATGGTATATCTCCTTCTTAAA GTTAAAC

Table 6: Primers

Plasmid Number	Plasmid Name
1	pET22b (empty) 3/10/17
2	PelB-H10-enh-TRM-pET22b 3/10/17
3	H10-GFP-enh-TAG-TRM-pET22b 3/10/17 (3_0410)
4	GBP1-Lin-R10-pET22b 3/10/17
5	GBP4-Lin-TAT-pET22b 3/10/17
6	GBP1-Lin-TAT-pET22b 3/10/17
7	MmPylTS-Trp-pJZ (1) 3/10/17
8	MmPylTS-Trp-pJZ (2) 3/10/17

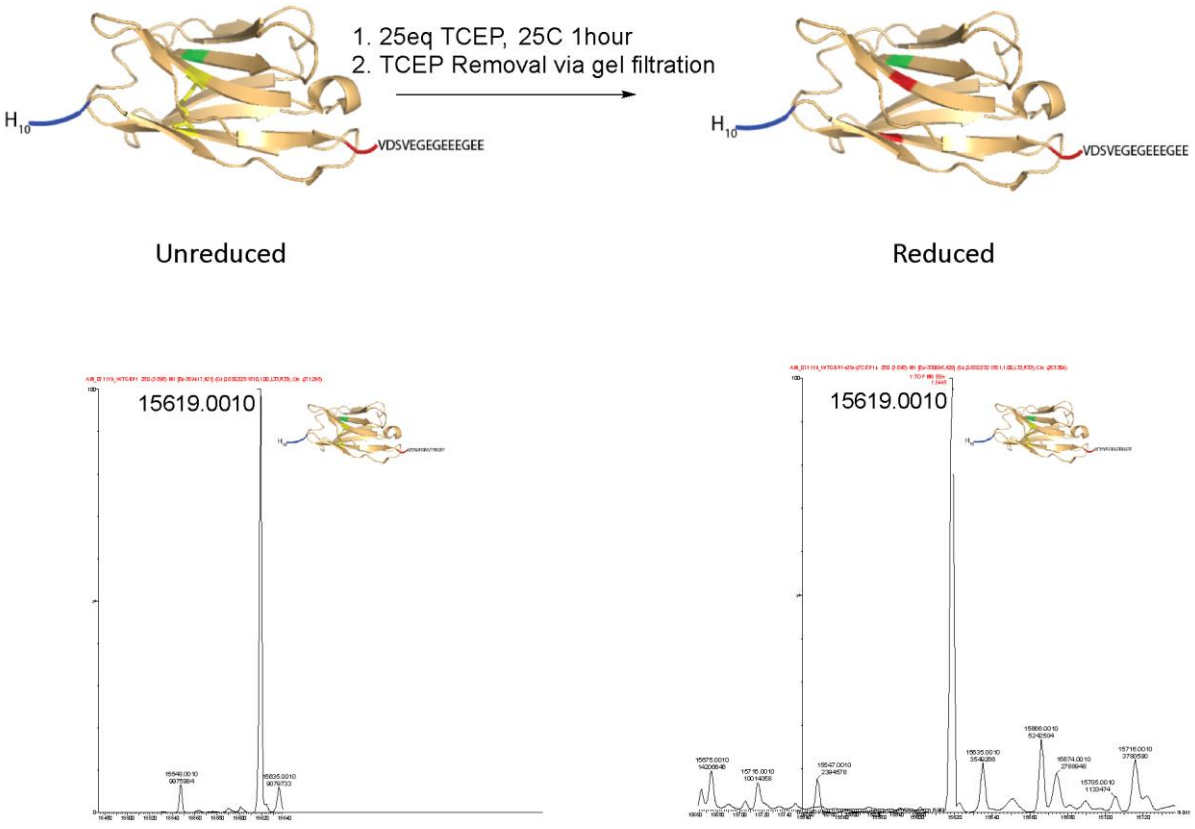
9	GFPmin-pHEN6 3/10/17
10	GFPmin-TAG-pHEN6 3/10/17
11	GFPmin-TAG-pET15b 3/10/17
12	GFPmin-TAG-pGEX4T1 3/10/17
13	H10-Enh-TRM-pET22b 3/10/17
14	GBP1-LinTAT-pET22b 3/10/17
15	GBP1-Lin-R10-pET22b 3/10/17
16	GBP4-LinTAT-pET22b 3/10/17
17	GBP1 R35I
18	GBP1 R57Amber
19	GBP1 R57I
20	GBP1 K43L
21	GBP1 K43I
22	GBP1 K65I

23	GBP1 K87TAG
24	GBP1 K87L
25	GBP1 K87I
26	GBP1 R35TAG
27	GBP1 K65TAG
28	PCNA_Binder-in-ptxb1
29	GBP1 in pMal-p5x, extra res
30	GBP1(TAG) in pMAL-p5x, extra res
31	WT eGFP
32	PelB-H10-GBP1 Y37F-TUB in pET22b
33	Pelb-h10-GBP4 Y118F-Tub in pHEN6
34	PET28a empty
35	PelB-GBP4-h10 in pET15b
36	PET22b empty
37	PelB-GBP4Y118TAG-H10
38	PelB-GBP4Y118F-H10
39	PelB-H10-GBP1 R35K-TUB
40	G3-GBP1-A3-Intein-CBD
41	Sortase from Addgene
42	MmPCKRS-mCherry-TAG-eGFP-HA
43	p4CMVE-U6-PylT
44	BKamp-PCKRS
45	PMyo4TAG PylT
46	PelB-H10-GBP1 R35Y
47	PelB-GBP1-H6 PylT
48	PelB-GBP1- R35TAG-H6 PylT
49	SUMO3-TTL1

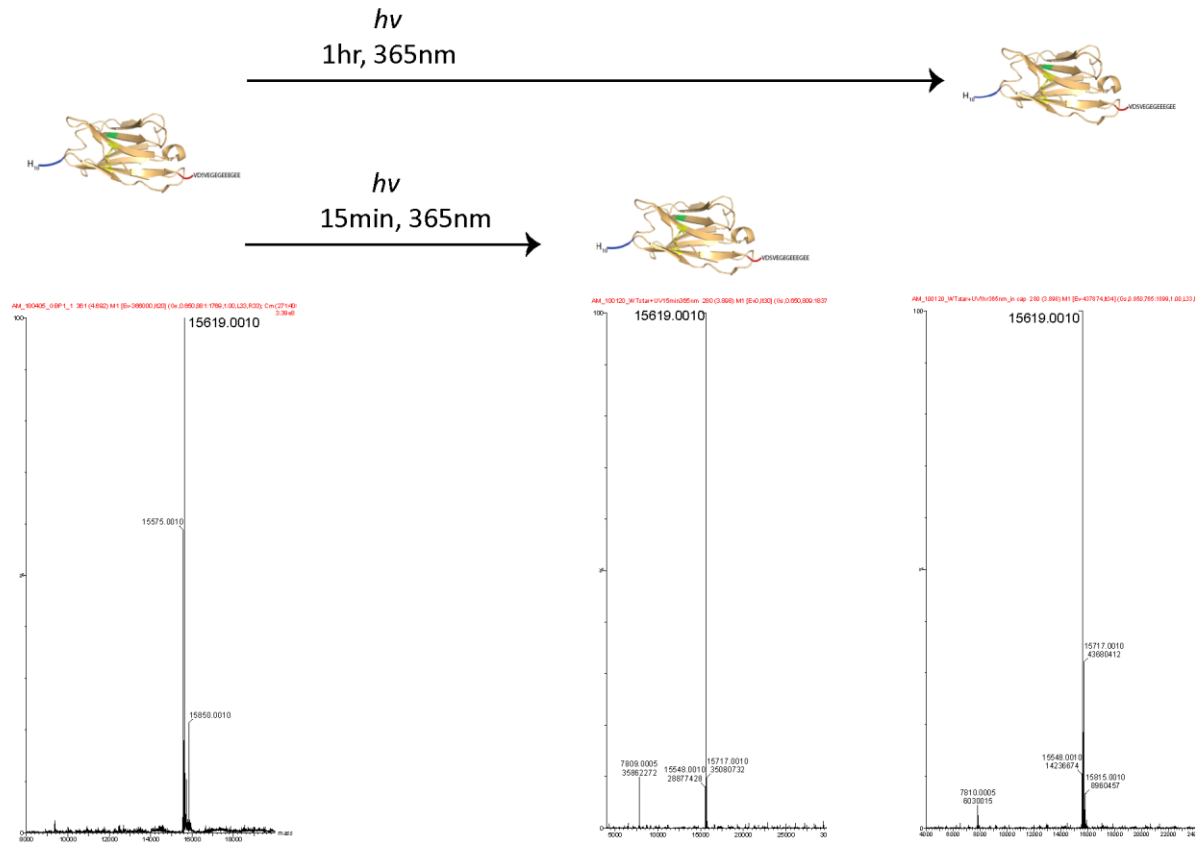
50	Tau WT
51	H6-Thrombin-MNK
52	pULTRA-CNF-synthetase
53	PelB-GBP1 Y37TAG - H6 PylT
54	PelB-GBP1_AAAC_H6 PylT
55	PelB-GBP1R35TAG_AAAC_H6 PylT
56	eGFP C70M
57	Anti_EGFR_Nanobody-AAA-Intein-CBD
58	PelB-G3-GBP1-H6 PylT
59	G3-H6-Thrombin-eGFPC70M
60	anti_EGFR_nanobody WT 7d12 in pTrcHis
61	GBP1 in pET28a H6-Thromb-GBP1
62	SortaseA 5M (eSrtA) in pET29
63	GBP4 in pET28a
64	SarsCov2 3186
65	Sars Cov2 3194
66	Sars Cov 2 3363

Table 7: Plasmids

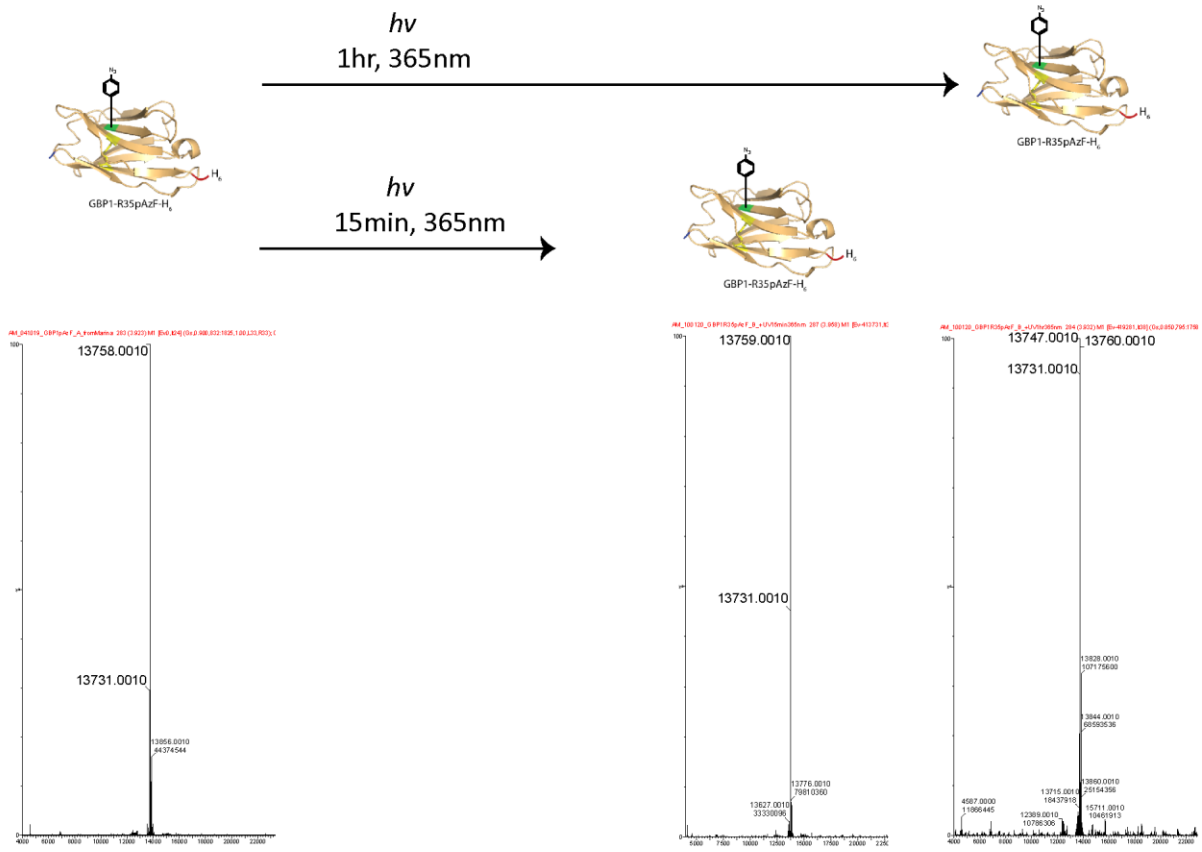
6.4 ESI-MS Spectra



Spectrum 1: ESI-MS of WT GBP1 before and after TCEP reduction. Control for azide reduction GFP binding assay.

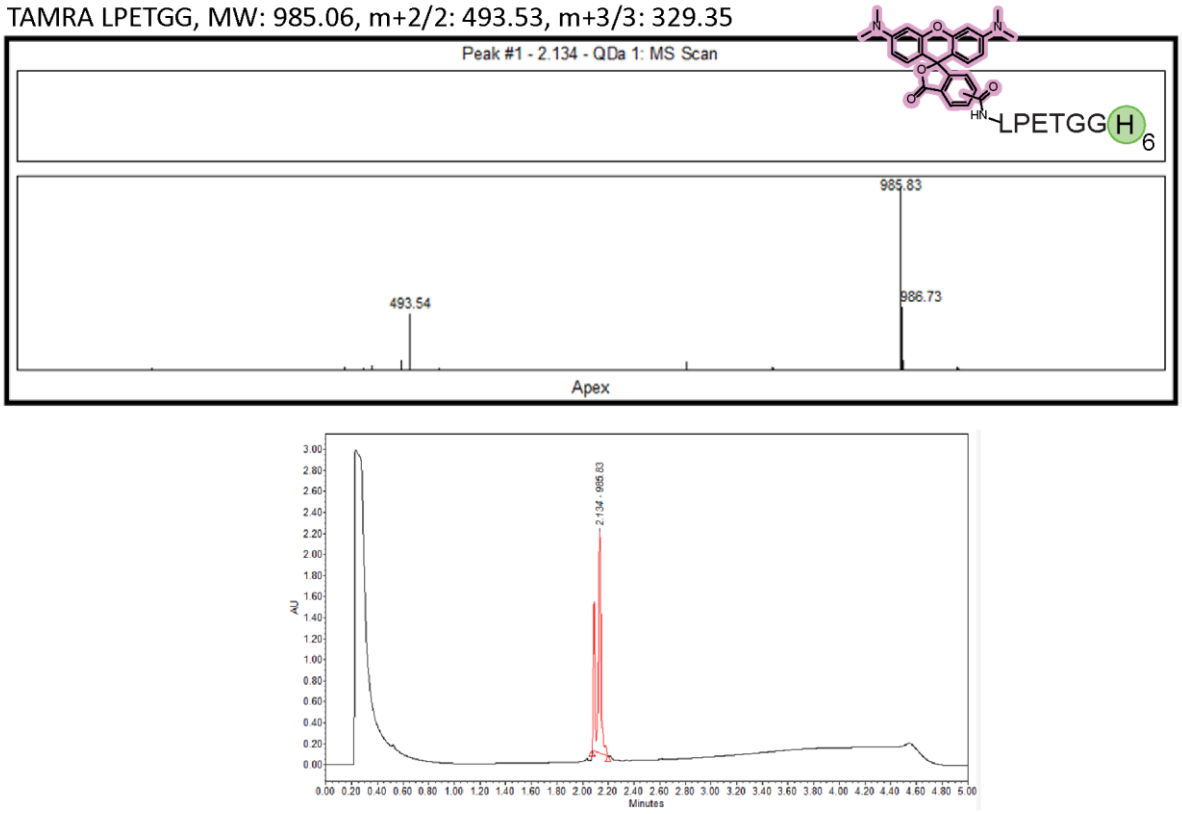


Spectrum 2: ESI-MS of WT GBP1 before and after UV irradiation. Control for GFP binding assay.



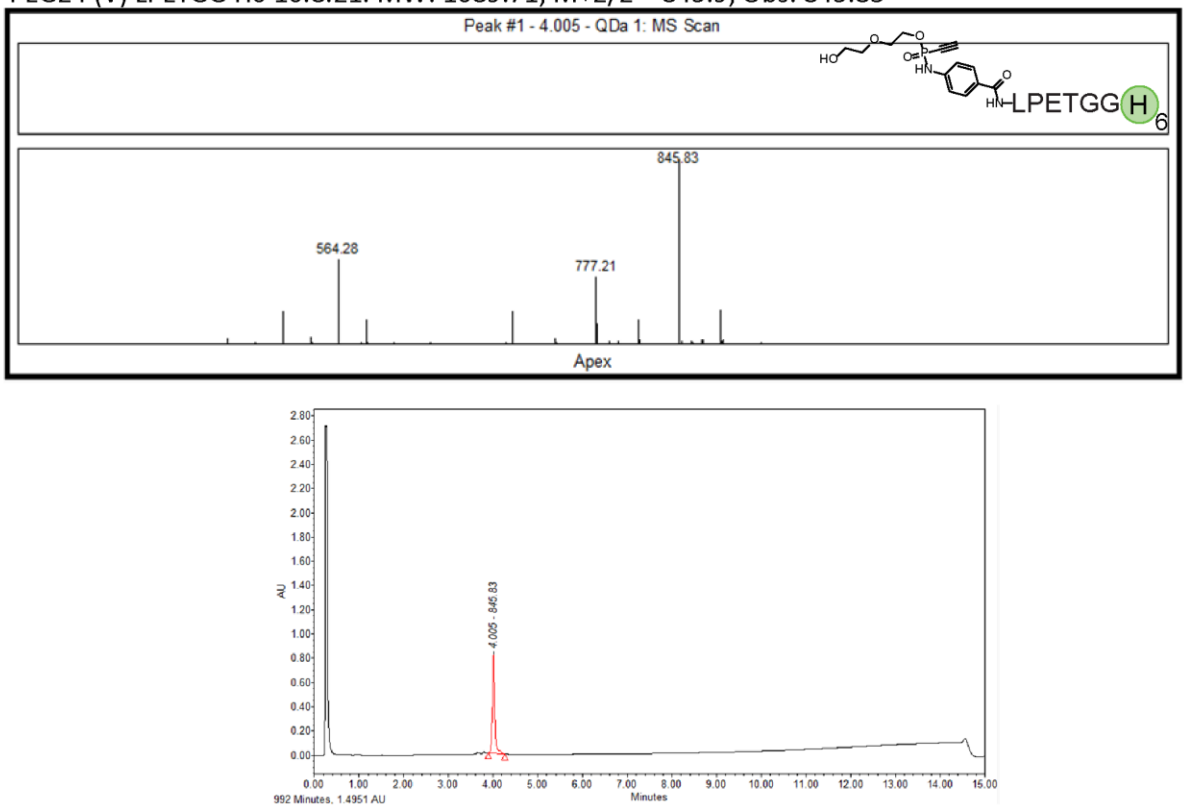
Spectrum 3 ESI-MS of GBP1R35pAzF before and after UV irradiation. Control for GFP binding assay.

6.5 HPLC-MS Spectra



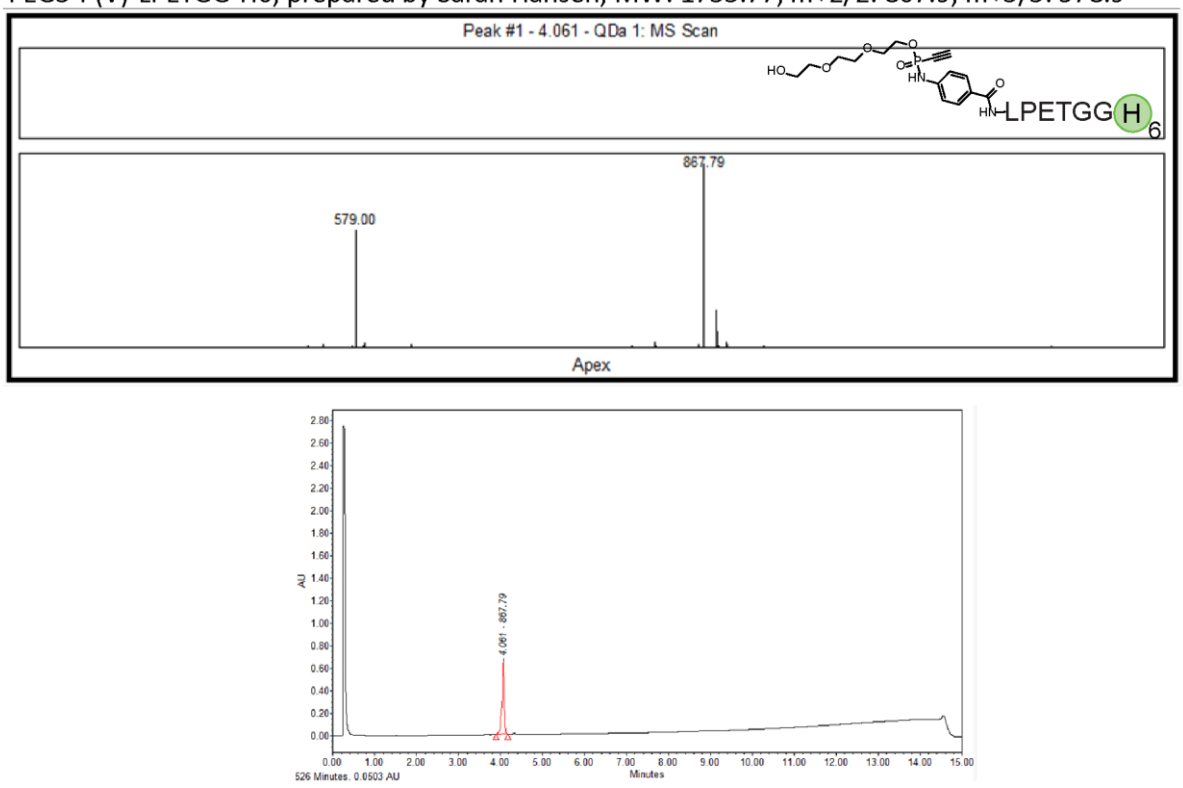
Spectrum 4: HPLC-MS

PEG2 P(V) LPETGG H6 16.8.21. MW: 1689.71, M+2/2 = 845.9, Obs: 845.83



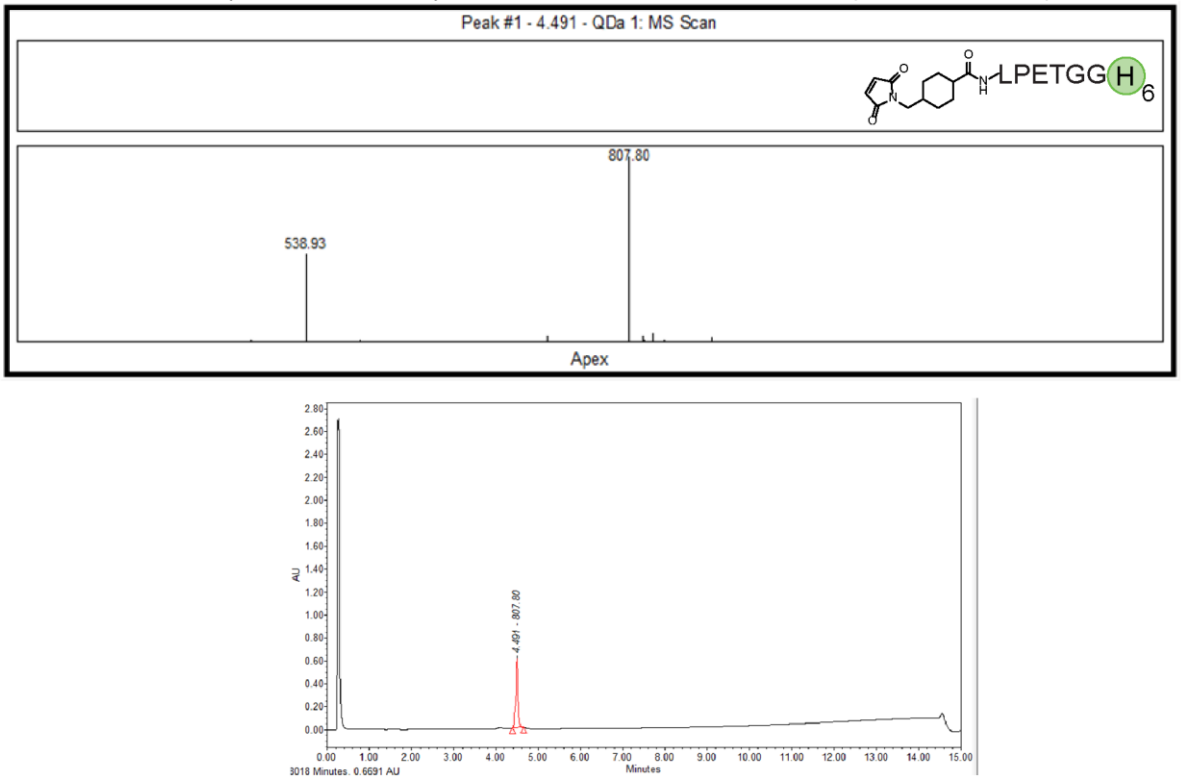
Spectrum 5: HPLC-MS

PEG3-P(V)-LPETGG-H6, prepared by Sarah Hansen, MW: 1733.77, m+2/2: 867.9, m+3/3: 578.9



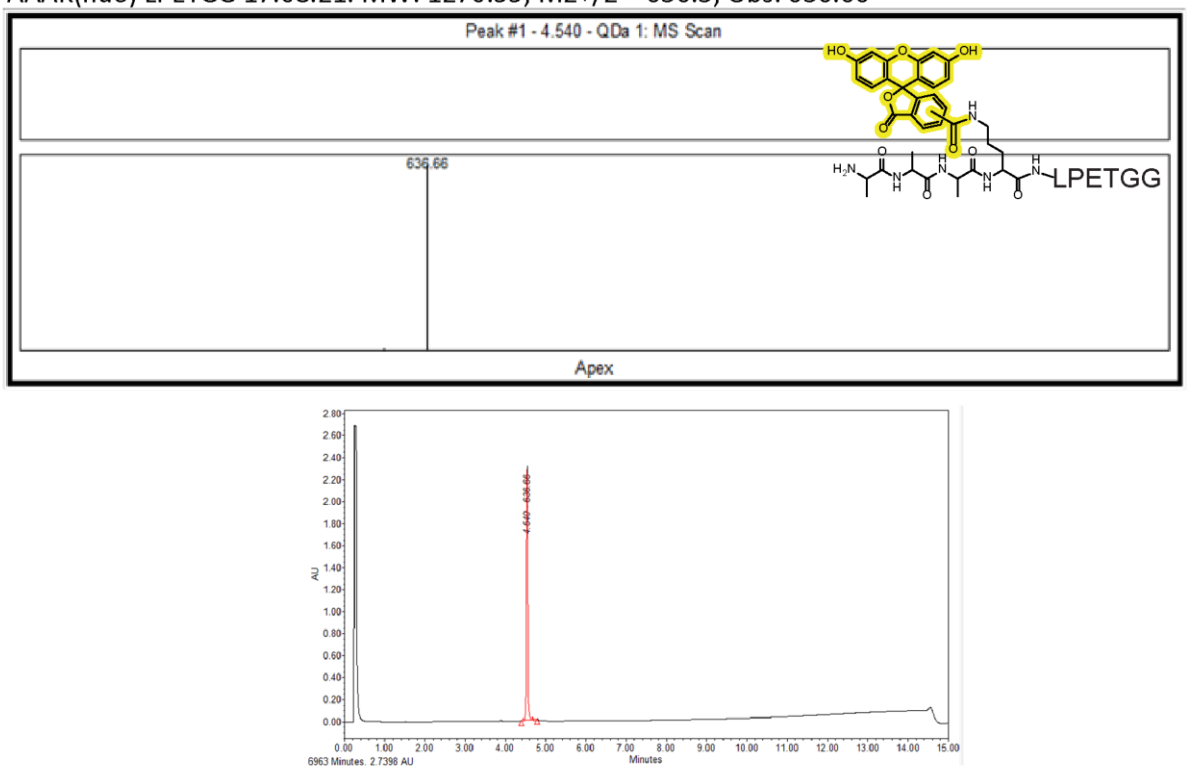
Spectrum 6: HPLC-MS

SMCC-LPETGG-H6 (from 01.05.2021) measured 19.08.21. MW: 1613.72, M2+2/2: 807.9, Obs 807.8



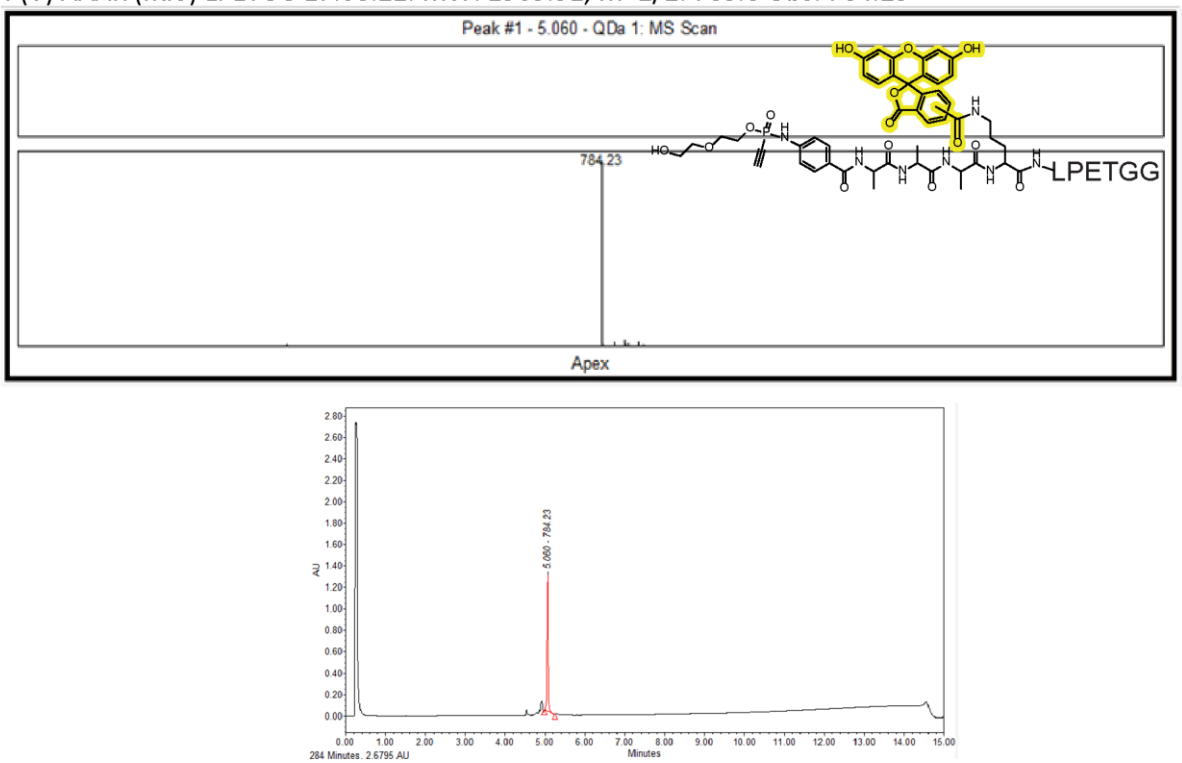
Spectrum 7: HPLC-MS

AAAK(fluo) LPETGG 17.08.21. MW: 1270.55, M2+/2 = 636.3, Obs: 636.66



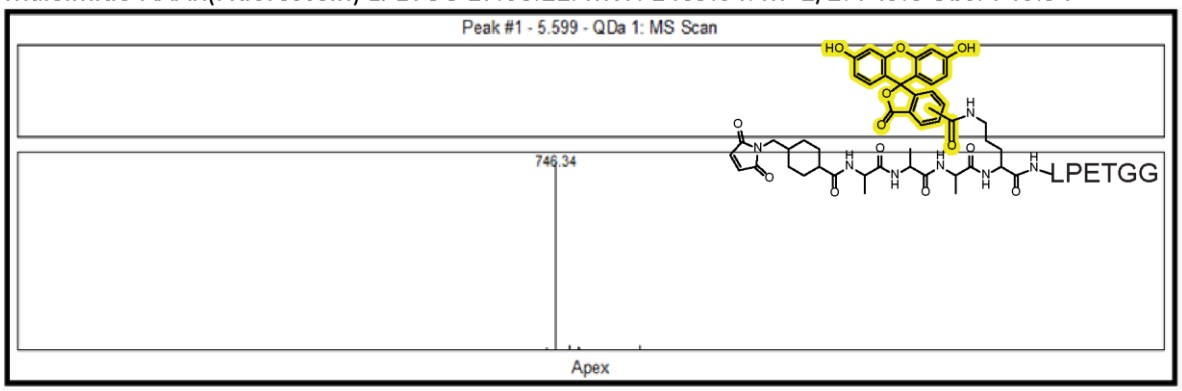
Spectrum 8: HPLC-MS

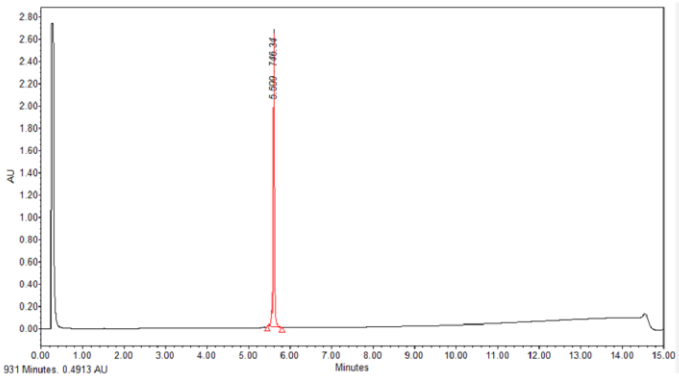
P(V) AAAK (fluo) LPETGG 17.08.21. MW: 1565.51, M+2/2: 783.8 Obs: 784.23



Spectrum 9: HPLC-MS

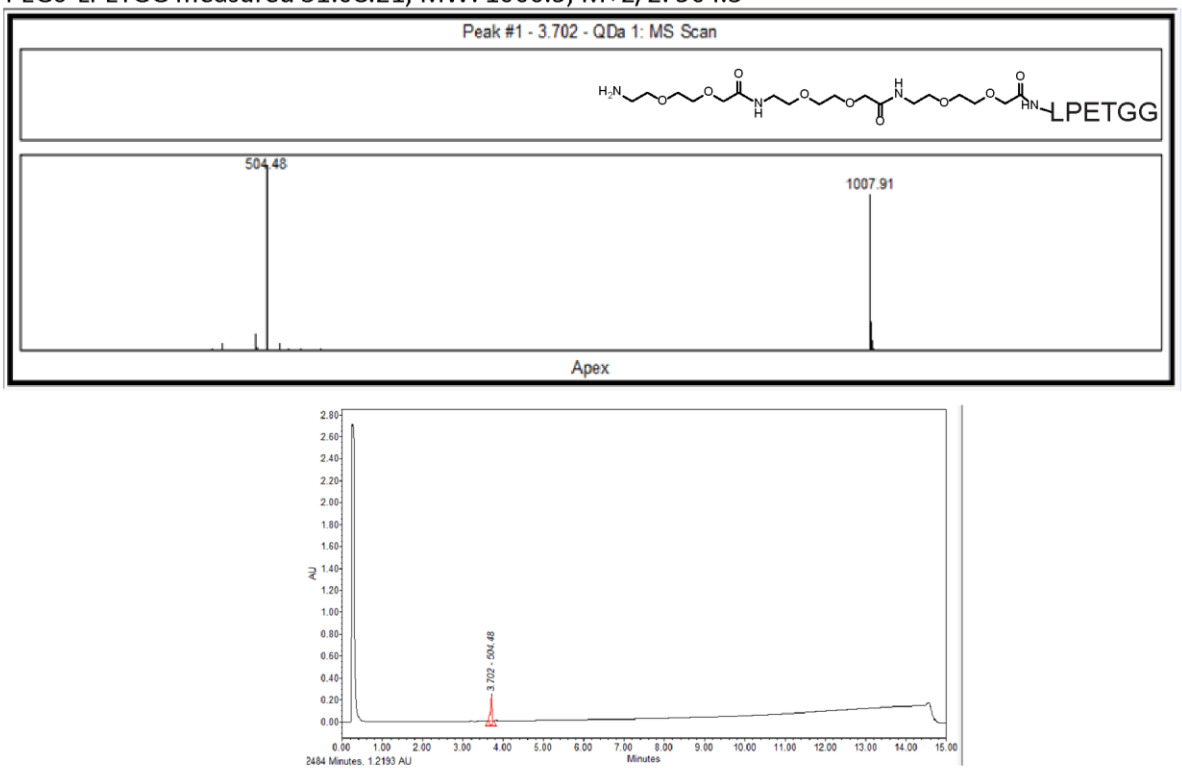
Maleimide-AAAK(Fluorescein) LPETGG 17.08.21. MW: 1489.64. M+2/2: 745.8 Obs: 746.34





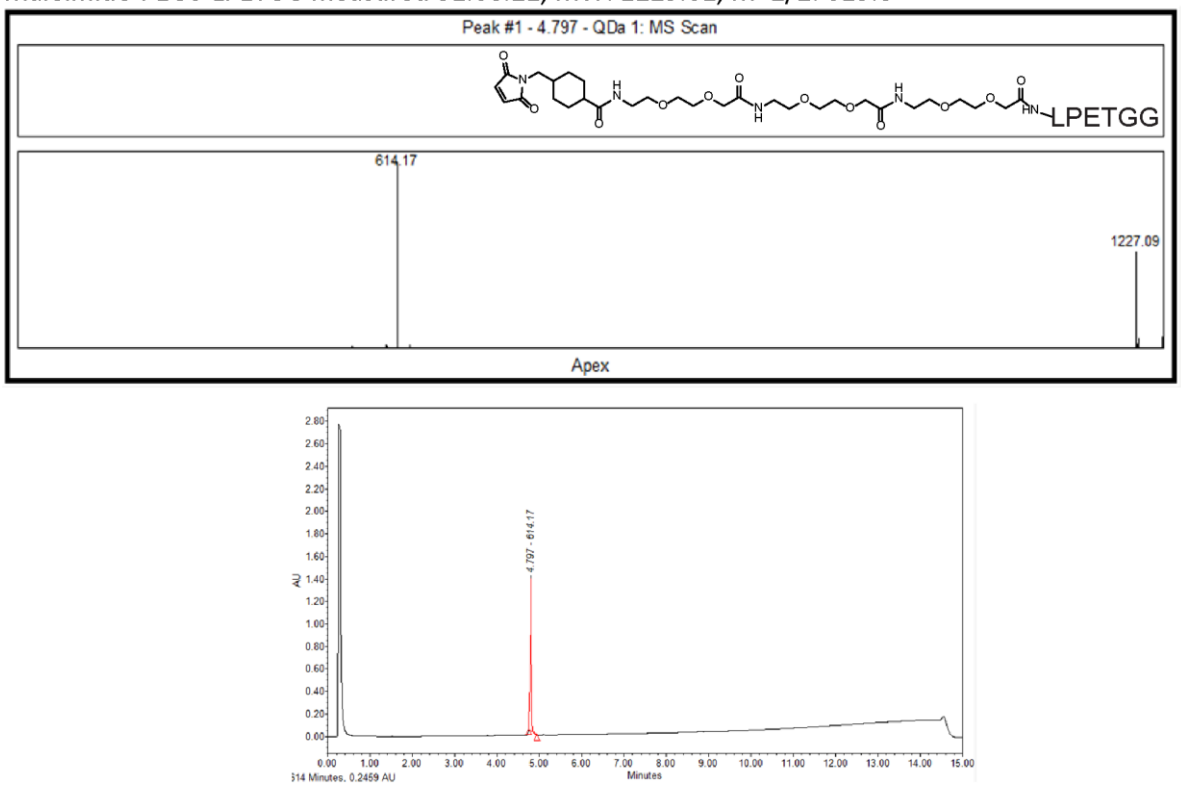
Spectrum 10: HPLC-MS

PEG6-LPETGG measured 31.08.21, MW: 1006.5, M+2/2: 504.3



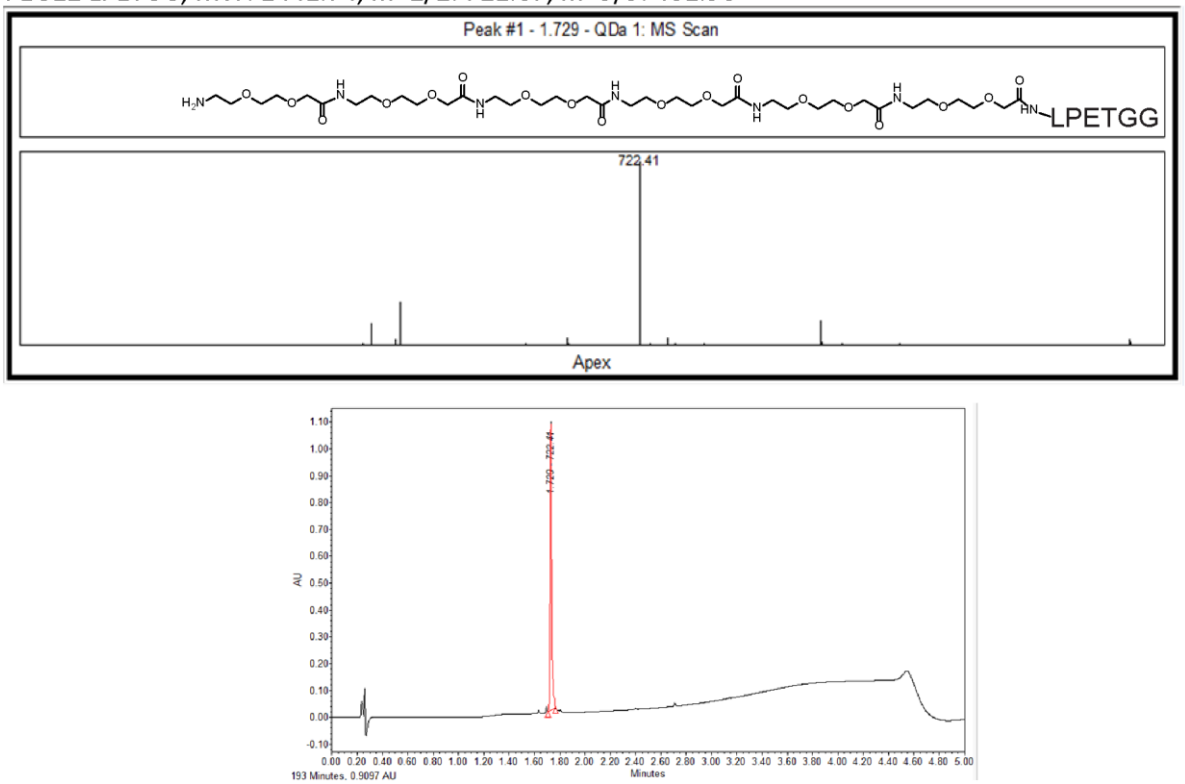
Spectrum 11: HPLC-MS

Maleimide-PEG6-LPETGG measured 31.08.21, MW: 1225.61, M+2/2: 613.8



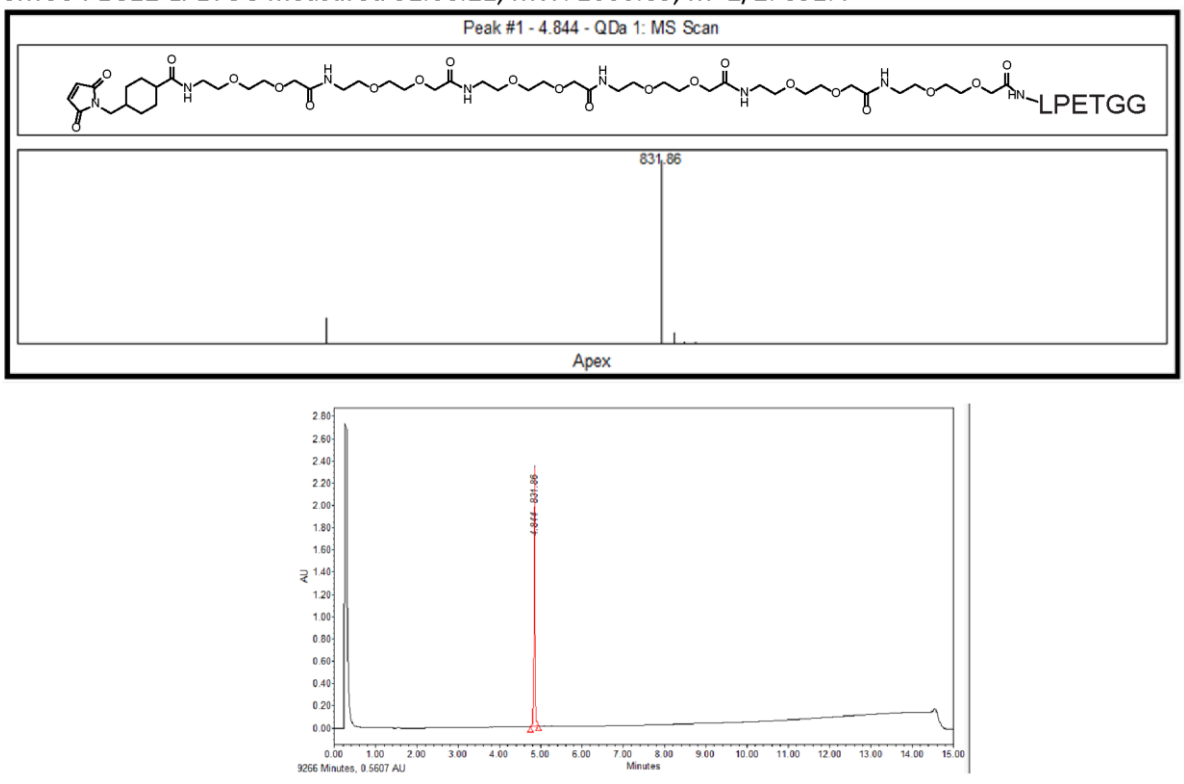
Spectrum 12: HPLC-MS

PEG12 LPETGG, MW: 1441.74, m+2/2: 721.87, m+3/3: 481.58



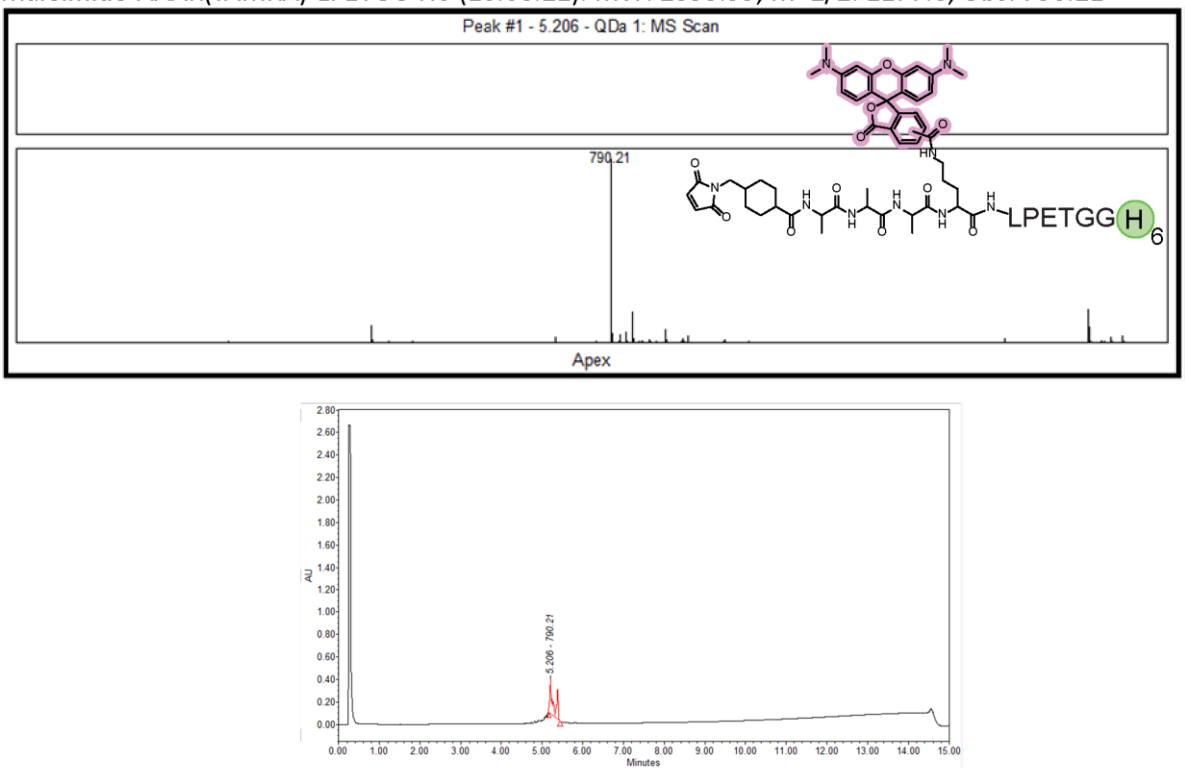
Spectrum 13: HPLC-MS

SMCC-PEG12-LPETGG measured 31.08.21, MW: 1660.83, M+2/2: 831.4



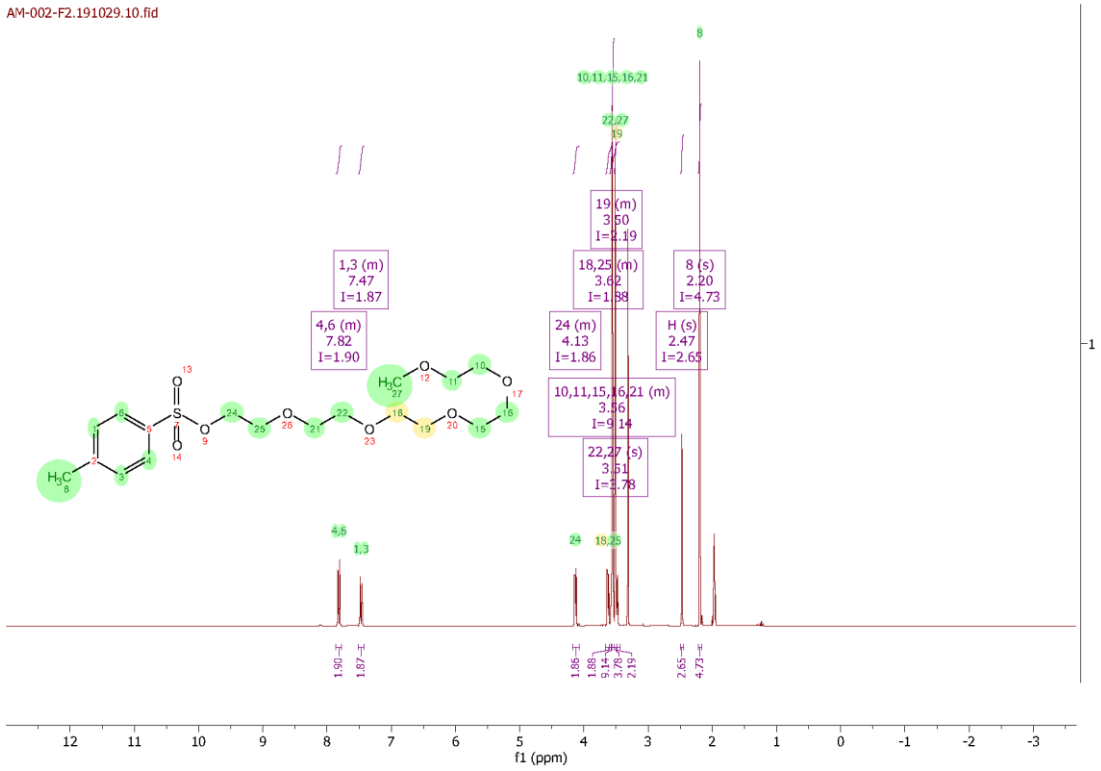
Spectrum 14: HPLC-MS

Maleimide-AAAK(TAMRA)-LPETGG-H6 (16.08.21). MW: 2353.55, M+2/2: 1177.8, Obs: 790.21



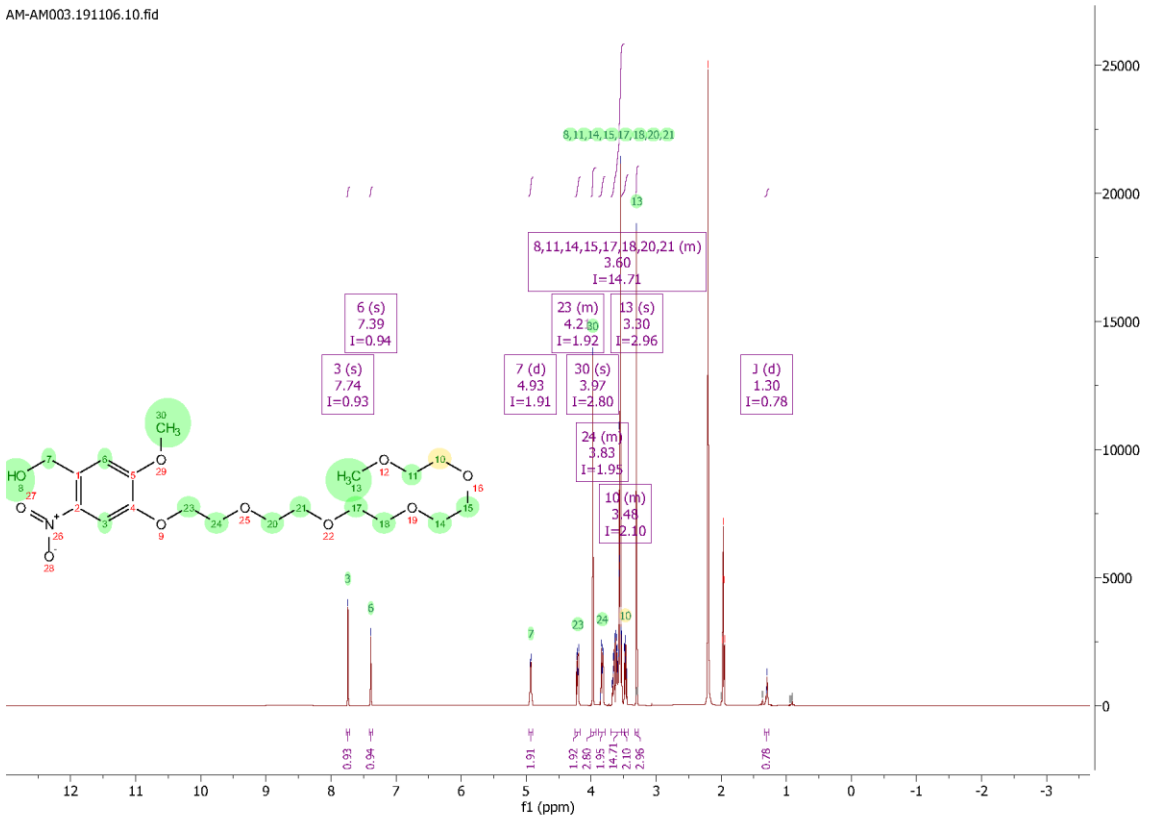
Spectrum 15: HPLC-MS

6.6 NMR Spectra



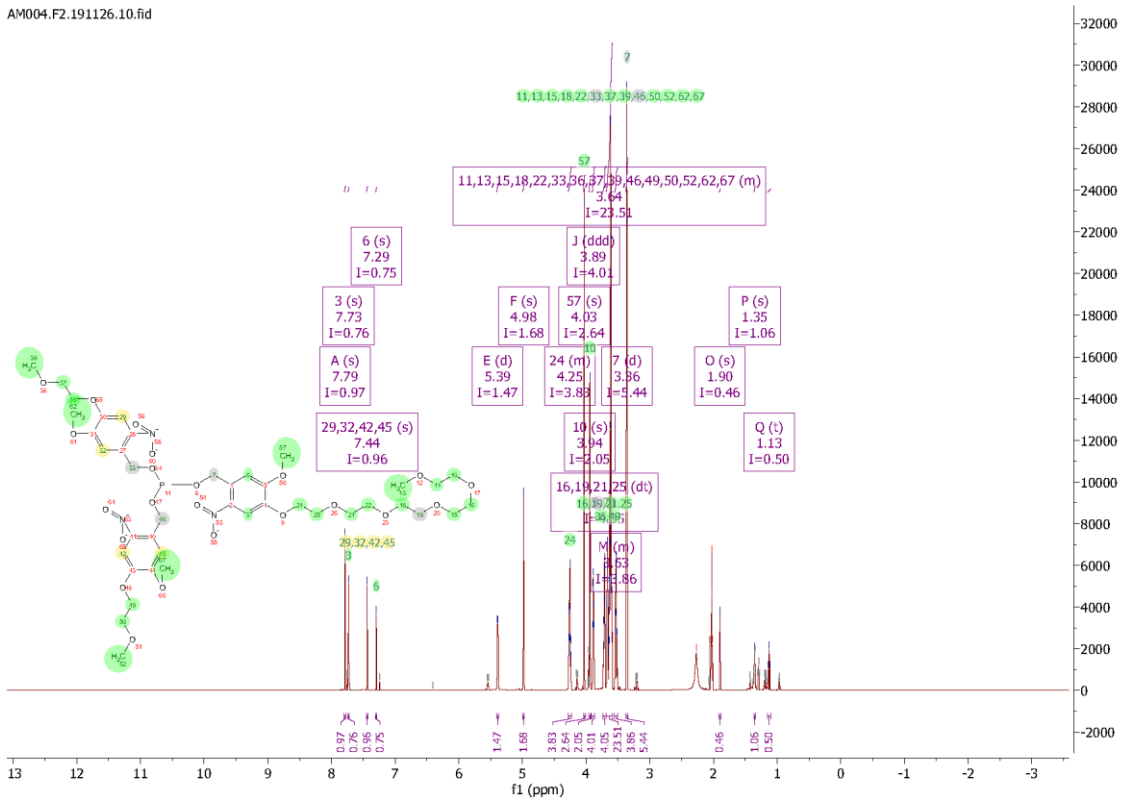
Spectrum 16: AM002 1H NMR.

AM002: ^{1}H NMR (300 MHz, CD_3CN) δ 7.86 – 7.76 (m, 2H), 7.52 – 7.42 (m, 2H), 4.18 – 4.07 (m, 2H), 3.67 – 3.59 (m, 2H), 3.62 – 3.52 (m, 1H), 3.56 (s, 4H), 3.58 – 3.43 (m, 10H), 2.47 (s, 3H), 2.20 (s, 5H).



Spectrum 17: AM003 1H NMR.

AM003: ^{1}H NMR (300 MHz, CD_3CN) δ 7.74 (s, 1H), 7.39 (s, 1H), 4.93 (d, *J* = 3.8 Hz, 2H), 4.25 – 4.17 (m, 2H), 3.97 (s, 3H), 3.89 – 3.79 (m, 2H), 3.70 – 3.50 (m, 15H), 3.53 – 3.43 (m, 2H), 3.30 (s, 3H), 1.30 (d, *J* = 4.5 Hz, 1H).



Spectrum 18: AM004 1H NMR.

AM004: 1H NMR (600 MHz, CD3CN) δ 7.79 (s, 1H), 7.73 (s, 1H), 7.44 (s, 1H), 7.29 (s, 1H), 5.39 (d, J = 7.5 Hz, 1H), 4.98 (s, 2H), 4.28 – 4.23 (m, 4H), 4.03 (s, 3H), 3.94 (s, 2H), 3.89 (ddd, J = 4.4, 3.6, 1.6 Hz, 4H), 3.71 (dt, J = 5.4, 3.1 Hz, 4H), 3.69 – 3.58 (m, 24H), 3.55 – 3.50 (m, 4H), 3.36 (d, J = 3.9 Hz, 5H), 1.90 (s, 0H), 1.35 (s, 1H), 1.13 (t, J = 7.2 Hz, 0H).

Spectrum 19: AM004 phosphorous NMR.

7. References

1. Pál, C., B. Papp, and M.J. Lercher, *An integrated view of protein evolution.* Nature Reviews Genetics, 2006. **7**(5): p. 337-348.
2. Jamison, J., *Screening protein libraries for new activities.* Nature Biotechnology, 2000. **18**(5): p. 480-480.
3. Packer, M.S. and D.R. Liu, *Methods for the directed evolution of proteins.* Nature Reviews Genetics, 2015. **16**(7): p. 379-394.
4. Mosier, N.S.L.M.R., *Modern Biotechnology Connecting Innovations in Microbiology and Biochemistry to Engineering Fundamentals.* 2011.
5. Brannigan, J.A. and A.J. Wilkinson, *Protein engineering 20 years on.* Nature Reviews Molecular Cell Biology, 2002. **3**(12): p. 964-970.
6. Walsh, S.J., J.D. Bargh, F.M. Dannheim, A.R. Hanby, H. Seki, A.J. Counsell, X. Ou, E. Fowler, N. Ashman, Y. Takada, A. Isidro-Llobet, J.S. Parker, J.S. Carroll, and D.R. Spring, *Site-selective modification strategies in antibody–drug conjugates.* Chemical Society Reviews, 2021. **50**(2): p. 1305-1353.
7. Shakib, F. and D.R. Stanworth, *Human IgG subclasses in health and disease (A review)Part I.* Ricerca in clinica e in laboratorio, 1980. **10**(3): p. 463.
8. Liu, H., G. Ponniah, H.-M. Zhang, C. Nowak, A. Neill, N. Gonzalez-Lopez, R. Patel, G. Cheng, A.Z. Kita, and B. Andrien, *In vitro and in vivo modifications of recombinant and human IgG antibodies.* mAbs, 2014. **6**(5): p. 1145-1154.
9. Szijj, P. and V. Chudasama, *The renaissance of chemically generated bispecific antibodies.* Nature Reviews Chemistry, 2021. **5**(2): p. 78-92.
10. Schumacher, D., J. Helma, A.F.L. Schneider, H. Leonhardt, and C.P.R. Hackenberger, *Nanobodies: Chemical Functionalization Strategies and Intracellular Applications.* Angew Chem Int Ed Engl, 2018. **57**(9): p. 2314-2333.
11. Harmsen, M.M. and H.J. De Haard, *Properties, production, and applications of camelid single-domain antibody fragments.* Applied Microbiology and Biotechnology, 2007. **77**(1): p. 13-22.
12. Kirchhofer, A., J. Helma, K. Schmidthals, C. Frauer, S. Cui, A. Karcher, M. Pellis, S. Muyldermans, C.S. Casas-Delucchi, M.C. Cardoso, H. Leonhardt, K.P. Hopfner, and U. Rothbauer, *Modulation of protein properties in living cells using nanobodies.* Nat Struct Mol Biol, 2010. **17**(1): p. 133-8.
13. Schneider, F., C. Eggeling, and E. Sezgin, *Influence of nanobody binding on fluorescence emission, mobility and organization of GFP-tagged proteins*. 2020, Cold Spring Harbor Laboratory.
14. Singh, P., L. Sharma, S.R. Kulothungan, B.V. Adkar, R.S. Prajapati, P.S.S. Ali, B. Krishnan, and R. Varadarajan, *Effect of Signal Peptide on Stability and Folding of Escherichia coli Thioredoxin.* PLoS ONE, 2013. **8**(5): p. e63442.
15. Lagassé, H.A.D., A. Alexaki, V.L. Simhadri, N.H. Katagiri, W. Jankowski, Z.E. Sauna, and C. Kimchi-Sarfaty, *Recent advances in (therapeutic protein) drug development.* F1000Research, 2017. **6**: p. 113.
16. Knowles, J.R., *Tinkering with enzymes: what are we learning?* Science, 1987. **236**(4806): p. 1252-8.
17. Schuster, J., A. Koulov, H.-C. Mahler, P. Detampel, J. Huwyler, S. Singh, and R. Mathaes, *In Vivo Stability of Therapeutic Proteins.* Pharmaceutical Research, 2020. **37**(2).
18. De Groot, A.S. and D.W. Scott, *Immunogenicity of protein therapeutics.* Trends in Immunology, 2007. **28**(11): p. 482-490.
19. Frokjaer, S. and D.E. Otzen, *Protein drug stability: a formulation challenge.* Nature Reviews Drug Discovery, 2005. **4**(4): p. 298-306.
20. Dhalluin, C., A. Ross, L.A. Leuthold, S. Foser, B. Gsell, F. Müller, and H. Senn, *Structural and biophysical characterization of the 40 kDa PEG-interferon-alpha2a and its individual positional isomers.* Bioconjug Chem, 2005. **16**(3): p. 504-17.

21. Hoyt, E.A., P.M.S.D. Cal, B.L. Oliveira, and G.J.L. Bernardes, *Contemporary approaches to site-selective protein modification.* Nature Reviews Chemistry, 2019. **3**(3): p. 147-171.
22. Devaraj, N.K., *The Future of Bioorthogonal Chemistry.* ACS Central Science, 2018. **4**(8): p. 952-959.
23. Rabuka, D., *Chemoenzymatic methods for site-specific protein modification.* Current Opinion in Chemical Biology, 2010. **14**(6): p. 790-796.
24. Rashidian, M., J.K. Dozier, and M.D. Distefano, *Enzymatic Labeling of Proteins: Techniques and Approaches.* Bioconjugate Chemistry, 2013. **24**(8): p. 1277-1294.
25. Effertz, K., S. Hinderlich, and W. Reutter, *Selective loss of either the epimerase or kinase activity of UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase due to site-directed mutagenesis based on sequence alignments.* J Biol Chem, 1999. **274**(40): p. 28771-8.
26. Winkler, K., A. Kramer, G. Küttner, M. Seifert, C. Scholz, H. Wessner, J. Schneider-Mergener, and W. Höhne, *Changing the Antigen Binding Specificity by Single Point Mutations of an Anti-p24 (HIV-1) Antibody.* The Journal of Immunology, 2000. **165**(8): p. 4505-4514.
27. Fekner, T. and M.K. Chan, *The pyrrolysine translational machinery as a genetic-code expansion tool.* Current Opinion in Chemical Biology, 2011. **15**(3): p. 387-391.
28. Brabham, R. and M.A. Fascione, *Pyrrolysine Amber Stop-Codon Suppression: Development and Applications.* ChemBioChem, 2017. **18**(20): p. 1973-1983.
29. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool.* Biochim Biophys Acta, 2014. **1844**(6): p. 1059-70.
30. Kiga, D., K. Sakamoto, K. Kodama, T. Kigawa, T. Matsuda, T. Yabuki, M. Shirouzu, Y. Harada, H. Nakayama, K. Takio, Y. Hasegawa, Y. Endo, I. Hirao, and S. Yokoyama, *An engineered Escherichia coli tyrosyl synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system.* Proceedings of the National Academy of Sciences, 2002. **99**(15): p. 9715-9720.
31. Spicer, C.D. and B.G. Davis, *Selective chemical protein modification.* Nature Communications, 2014. **5**(1): p. 4740.
32. Strable, E., D.E. Prasuhn, A.K. Udit, S. Brown, A.J. Link, J.T. Ngo, G. Lander, J. Quispe, C.S. Potter, B. Carragher, D.A. Tirrell, and M.G. Finn, *Unnatural Amino Acid Incorporation into Virus-Like Particles.* Bioconjugate Chemistry, 2008. **19**(4): p. 866-875.
33. Hackenberger, C.P.R. and D. Schwarzer, *Chemoselective Ligation and Modification Strategies for Peptides and Proteins.* Angewandte Chemie-International Edition, 2008. **47**(52): p. 10030-10074.
34. Kleineweischede, R. and C.P. Hackenberger, *Chemoselective peptide cyclization by traceless Staudinger ligation.* Angew Chem Int Ed Engl, 2008. **47**(32): p. 5984-8.
35. Chin, J.W., S.W. Santoro, A.B. Martin, D.S. King, L. Wang, and P.G. Schultz, *Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli.* J Am Chem Soc, 2002. **124**(31): p. 9026-7.
36. Saxon, E. and C.R. Bertozzi, *Cell surface engineering by a modified Staudinger reaction.* Science, 2000. **287**(5460): p. 2007-2010.
37. Agard, N.J., J.M. Baskin, J.A. Prescher, A. Lo, and C.R. Bertozzi, *A Comparative Study of Bioorthogonal Reactions with Azides.* ACS Chemical Biology, 2006. **1**(10): p. 644-648.
38. Presolski, S.I., V.P. Hong, and M.G. Finn, *Copper-Catalyzed Azide–Alkyne Click Chemistry for Bioconjugation.* Current Protocols in Chemical Biology, 2011. **3**(4): p. 153-162.
39. Li, S., H. Cai, J. He, H. Chen, S. Lam, T. Cai, Z. Zhu, S.J. Bark, and C. Cai, *Extent of the Oxidative Side Reactions to Peptides and Proteins During the CuAAC Reaction.* Bioconjugate Chemistry, 2016. **27**(10): p. 2315-2322.
40. Nikić, I., J.H. Kang, G.E. Girona, I.V. Aramburu, and E.A. Lemke, *Labeling proteins on live mammalian cells using click chemistry.* Nature Protocols, 2015. **10**(5): p. 780-791.
41. Kim, E. and H. Koo, *Biomedical applications of copper-free click chemistry: in vitro, in vivo, and ex vivo.* Chemical Science, 2019. **10**(34): p. 7835-7851.

42. van Geel, R., G.J.M. Pruijn, F.L. van Delft, and W.C. Boelens, *Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide–Alkyne Cycloaddition.* Bioconjugate Chemistry, 2012. **23**(3): p. 392-398.
43. Lin, F.L., H.M. Hoyt, H. van Halbeek, R.G. Bergman, and C.R. Bertozzi, *Mechanistic Investigation of the Staudinger Ligation.* Journal of the American Chemical Society, 2005. **127**(8): p. 2686-2695.
44. Heiss, T.K., R.S. Dorn, and J.A. Prescher, *Bioorthogonal Reactions of Triarylphosphines and Related Analogues.* Chem Rev, 2021. **121**(12): p. 6802-6849.
45. Serwa, R., I. Wilkening, G. Del Signore, M. Muhlberg, I. Claussnitzer, C. Weise, M. Gerrits, and C.P. Hackenberger, *Chemoselective Staudinger-phosphite reaction of azides for the phosphorylation of proteins.* Angew Chem Int Ed Engl, 2009. **48**(44): p. 8234-9.
46. Bohrsch, V., T. Mathew, M. Zieringer, M.R. Vallee, L.M. Artner, J. Dernedde, R. Haag, and C.P. Hackenberger, *Chemoselective Staudinger-phosphite reaction of symmetrical glycosyl-phosphites with azido-peptides and polygycerols.* Org Biomol Chem, 2012. **10**(30): p. 6211-6.
47. Serwa, R., P. Majkut, B. Horstmann, J.M. Swiecicki, M. Gerrits, E. Krause, and C.P.R. Hackenberger, *Site-specific PEGylation of proteins by a Staudinger-phosphite reaction.* Chemical Science, 2010. **1**(5): p. 596-602.
48. Nischan, N., A. Chakrabarti, R.A. Serwa, P.H. Bovee-Geurts, R. Brock, and C.P. Hackenberger, *Stabilization of peptides for intracellular applications by phosphoramidate-linked polyethylene glycol chains.* Angew Chem Int Ed Engl, 2013. **52**(45): p. 11920-4.
49. Vallee, M.R., P. Majkut, I. Wilkening, C. Weise, G. Muller, and C.P. Hackenberger, *Staudinger-phosphonite reactions for the chemoselective transformation of azido-containing peptides and proteins.* Org Lett, 2011. **13**(20): p. 5440-3.
50. Vallee, M.R., L.M. Artner, J. Dernedde, and C.P. Hackenberger, *Alkyne phosphonites for sequential azide-azide couplings.* Angew Chem Int Ed Engl, 2013. **52**(36): p. 9504-8.
51. Weerapana, E., C. Wang, G.M. Simon, F. Richter, S. Khare, M.B. Dillon, D.A. Bachovchin, K. Mowen, D. Baker, and B.F. Cravatt, *Quantitative reactivity profiling predicts functional cysteines in proteomes.* Nature, 2010. **468**(7325): p. 790-5.
52. Rosen, C.B. and M.B. Francis, *Targeting the N terminus for site-selective protein modification.* Nature Chemical Biology, 2017. **13**(7): p. 697-705.
53. Paulsen, C.E. and K.S. Carroll, *Cysteine-Mediated Redox Signaling: Chemistry, Biology, and Tools for Discovery.* Chemical Reviews, 2013. **113**(7): p. 4633-4679.
54. Ochtrop, P. and C.P.R. Hackenberger, *Recent advances of thiol-selective bioconjugation reactions.* Current Opinion in Chemical Biology, 2020. **58**: p. 28-36.
55. Kasper, M.A., M. Glanz, A. Stengl, M. Penkert, S. Klenk, T. Sauer, D. Schumacher, J. Helma, E. Krause, M.C. Cardoso, H. Leonhardt, and C.P.R. Hackenberger, *Cysteine-Selective Phosphonamidate Electrophiles for Modular Protein Bioconjugations.* Angew Chem Int Ed Engl, 2019. **58**(34): p. 11625-11630.
56. Kasper, M.A., M. Glanz, A. Oder, P. Schmieder, J.P. von Kries, and C.P.R. Hackenberger, *Vinylphosphonites for Staudinger-induced chemoselective peptide cyclization and functionalization.* Chem Sci, 2019. **10**(25): p. 6322-6329.
57. Baumann, A.L., S. Schwagerus, K. Broi, K. Kemnitz-Hassanin, C.E. Stieger, N. Trieloff, P. Schmieder, and C.P.R. Hackenberger, *Chemically Induced Vinylphosphonothiolate Electrophiles for Thiol–Thiol Bioconjugations.* Journal of the American Chemical Society, 2020. **142**(20): p. 9544-9552.
58. Nair, D.P., M. Podgórski, S. Chatani, T. Gong, W. Xi, C.R. Fenoli, and C.N. Bowman, *The Thiol-Michael Addition Click Reaction: A Powerful and Widely Used Tool in Materials Chemistry.* Chemistry of Materials, 2014. **26**(1): p. 724-744.
59. Baldwin, A.D. and K.L. Kiick, *Reversible maleimide–thiol adducts yield glutathione-sensitive poly(ethylene glycol)–heparin hydrogels.* Polym. Chem., 2013. **4**(1): p. 133-143.

60. Jones, M.W., R.A. Strickland, F.F. Schumacher, S. Caddick, J.R. Baker, M.I. Gibson, and D.M. Haddleton, *Polymeric Dibromomaleimides As Extremely Efficient Disulfide Bridging Bioconjugation and Pegylation Agents.* Journal of the American Chemical Society, 2012. **134**(3): p. 1847-1852.
61. Lyon, R.P., J.R. Setter, T.D. Bovee, S.O. Doronina, J.H. Hunter, M.E. Anderson, C.L. Balasubramanian, S.M. Duniho, C.I. Leiske, F. Li, and P.D. Senter, *Self-hydrolyzing maleimides improve the stability and pharmacological properties of antibody-drug conjugates.* Nature Biotechnology, 2014. **32**(10): p. 1059-1062.
62. De Rosa, L., R. Di Stasi, A. Romanelli, and L.D. D'Andrea, *Exploiting Protein N-Terminus for Site-Specific Bioconjugation.* Molecules, 2021. **26**(12): p. 3521.
63. Dozier, J. and M. Distefano, *Site-Specific PEGylation of Therapeutic Proteins.* International Journal of Molecular Sciences, 2015. **16**(10): p. 25831-25864.
64. Kinstler, O., G. Molineux, M. Treuheit, D. Ladd, and C. Gegg, *Mono-N-terminal poly(ethylene glycol)–protein conjugates.* Advanced Drug Delivery Reviews, 2002. **54**(4): p. 477-485.
65. Chan, A.O.-Y., C.-M. Ho, H.-C. Chong, Y.-C. Leung, J.-S. Huang, M.-K. Wong, and C.-M. Che, *Modification of N-Terminal α-Amino Groups of Peptides and Proteins Using Ketenes.* Journal of the American Chemical Society, 2012. **134**(5): p. 2589-2598.
66. Zhang, Y., K.-Y. Park, K.F. Suazo, and M.D. Distefano, *Recent progress in enzymatic protein labelling techniques and their applications.* Chemical Society Reviews, 2018. **47**(24): p. 9106-9136.
67. Shadish, J.A. and C.A. Deforest, *Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials.* Matter, 2020. **2**(1): p. 50-77.
68. Peracchi, A., *The Limits of Enzyme Specificity and the Evolution of Metabolism.* Trends in Biochemical Sciences, 2018. **43**(12): p. 984-996.
69. Choi, J.-S. and S.H. Joo, *Recent Trends in Cyclic Peptides as Therapeutic Agents and Biochemical Tools.* Biomolecules & Therapeutics, 2020. **28**(1): p. 18-24.
70. Schumacher, D., O. Lemke, J. Helma, L. Gerszonowicz, V. Waller, T. Stoschek, P.M. Durkin, N. Budisa, H. Leonhardt, B.G. Keller, and C.P.R. Hackenberger, *Broad substrate tolerance of tubulin tyrosine ligase enables one-step site-specific enzymatic protein labeling.* Chem Sci, 2017. **8**(5): p. 3471-3478.
71. Pishesha, N., J.R. Ingram, and H.L. Ploegh, *Sortase A: A Model for Transpeptidation and Its Biological Applications.* Annual Review of Cell and Developmental Biology, 2018. **34**(1): p. 163-188.
72. Ton-That, H., G. Liu, S.K. Mazmanian, K.F. Faull, and O. Schneewind, *Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif.* Proceedings of the National Academy of Sciences, 1999. **96**(22): p. 12424-12429.
73. Ton-That, H. and O. Schneewind, *Anchor Structure of Staphylococcal Surface Proteins.* Journal of Biological Chemistry, 1999. **274**(34): p. 24316-24320.
74. Suree, N., S.W. Yi, W. Thieu, M. Marohn, R. Damoiseaux, A. Chan, M.E. Jung, and R.T. Clubb, *Discovery and structure–activity relationship analysis of Staphylococcus aureus sortase A inhibitors.* Bioorganic & Medicinal Chemistry, 2009. **17**(20): p. 7174-7185.
75. Ton-That, H., S.K. Mazmanian, K.F. Faull, and O. Schneewind, *Anchoring of Surface Proteins to the Cell Wall ofStaphylococcus aureus.* Journal of Biological Chemistry, 2000. **275**(13): p. 9876-9881.
76. Mao, H., S.A. Hart, A. Schink, and B.A. Pollok, *Sortase-Mediated Protein Ligation: A New Method for Protein Engineering.* Journal of the American Chemical Society, 2004. **126**(9): p. 2670-2671.
77. Huang, X., A. Aulabaugh, W. Ding, B. Kapoor, L. Alksne, K. Tabei, and G. Ellestad, *Kinetic Mechanism of Staphylococcus aureus Sortase SrtA.* Biochemistry, 2003. **42**(38): p. 11307-11315.

78. Frankel, B.A., R.G. Kruger, D.E. Robinson, N.L. Kelleher, and D.G. McCafferty, *Staphylococcus aureus Sortase Transpeptidase SrtA: Insight into the Kinetic Mechanism and Evidence for a Reverse Protonation Catalytic Mechanism.* Biochemistry, 2005. **44**(33): p. 11188-11200.
79. Aulabaugh, A., W. Ding, B. Kapoor, K. Tabei, L. Alksne, R. Dushin, T. Zatz, G. Ellestad, and X. Huang, *Development of an HPLC assay for Staphylococcus aureus sortase: Evidence for the formation of the kinetically competent acyl enzyme intermediate.* Analytical Biochemistry, 2007. **360**(1): p. 14-22.
80. Schoonen, L., J. Pille, A. Borrmann, R.J.M. Nolte, and J.C.M. van Hest, *Sortase A-Mediated N-Terminal Modification of Cowpea Chlorotic Mottle Virus for Highly Efficient Cargo Loading.* Bioconjugate Chemistry, 2015. **26**(12): p. 2429-2434.
81. Parthasarathy, R., S. Subramanian, and E.T. Boder, *Sortase A as a Novel Molecular "Stapler" for Sequence-Specific Protein Conjugation.* Bioconjugate Chemistry, 2007. **18**(2): p. 469-476.
82. Suree, N., C.K. Liew, V.A. Villareal, W. Thieu, E.A. Fadeev, J.J. Clemens, M.E. Jung, and R.T. Clubb, *The Structure of the Staphylococcus aureus Sortase-Substrate Complex Reveals How the Universally Conserved LPXTG Sorting Signal Is Recognized.* Journal of Biological Chemistry, 2009. **284**(36): p. 24465-24477.
83. Heck, T., P.-H. Pham, A. Yerlikaya, L. Thöny-Meyer, and M. Richter, *Sortase A catalyzed reaction pathways: a comparative study with six SrtA variants.* 2014. **4**(9): p. 2946.
84. Bentley, M.L., H. Gaweska, J.M. Kielec, and D.G. McCafferty, *Engineering the Substrate Specificity of Staphylococcus aureus Sortase A.* Journal of Biological Chemistry, 2007. **282**(9): p. 6571-6581.
85. Antos, J.M., M.C. Truttmann, and H.L. Ploegh, *Recent advances in sortase-catalyzed ligation methodology.* Current Opinion in Structural Biology, 2016. **38**: p. 111-118.
86. Bierlmeier, J., M. Alvaro-Benito, M. Scheffler, K. Sturm, L. Rehkopf, C. Freund, and D. Schwarzer, *Sortase mediated multi-fragment assemblies by ligation site switching.* Angewandte Chemie International Edition, 2021.
87. Hirakawa, H., S. Ishikawa, and T. Nagamune, *Design of Ca2+-independent Staphylococcus aureus sortase A mutants.* Biotechnol Bioeng, 2012. **109**(12): p. 2955-61.
88. Chen, I., B.M. Dorr, and D.R. Liu, *A general strategy for the evolution of bond-forming enzymes using yeast display.* Proceedings of the National Academy of Sciences, 2011. **108**(28): p. 11399-11404.
89. Podracky, C.J., C. An, A. Desousa, B.M. Dorr, D.M. Walsh, and D.R. Liu, *Laboratory evolution of a sortase enzyme that modifies amyloid-β protein.* Nature Chemical Biology, 2021.
90. Bellucci, J.J., J. Bhattacharyya, and A. Chilkoti, *A Noncanonical Function of Sortase Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins.* Angewandte Chemie International Edition, 2014: p. n/a-n/a.
91. Glasgow, J.E., M.L. Salit, and J.R. Cochran, *In Vivo Site-Specific Protein Tagging with Diverse Amines Using an Engineered Sortase Variant.* Journal of the American Chemical Society, 2016. **138**(24): p. 7496-7499.
92. Sarpong, K. and R. Bose, *Efficient sortase-mediated N-terminal labeling of TEV protease cleaved recombinant proteins.* Analytical Biochemistry, 2017. **521**: p. 55-58.
93. Schmohl, L. and D. Schwarzer, *Sortase-mediated ligations for the site-specific modification of proteins.* Current Opinion in Chemical Biology, 2014. **22**: p. 122-128.
94. Race, P.R., M.L. Bentley, J.A. Melvin, A. Crow, R.K. Hughes, W.D. Smith, R.B. Sessions, M.A. Kehoe, D.G. McCafferty, and M.J. Banfield, *Crystal Structure of Streptococcus pyogenes Sortase A.* Journal of Biological Chemistry, 2009. **284**(11): p. 6924-6933.
95. Guimaraes, C.P., M.D. Witte, C.S. Theile, G. Bozkurt, L. Kundrat, A.E.M. Blom, and H.L. Ploegh, *Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions.* Nature Protocols, 2013. **8**(9): p. 1787-1799.

96. Theile, C.S., M.D. Witte, A.E. Blom, L. Kundrat, H.L. Ploegh, and C.P. Guimaraes, *Site-specific N-terminal labeling of proteins using sortase-mediated reactions.* Nat Protoc, 2013. **8**(9): p. 1800-7.
97. Reed, S.A., D.A. Brzovic, S.S. Takasaki, K.V. Boyko, and J.M. Antos, *Efficient Sortase-Mediated Ligation Using a Common C-Terminal Fusion Tag.* Bioconjugate Chemistry, 2020. **31**(5): p. 1463-1473.
98. Qi, Y., A. Simakova, N.J. Ganson, X. Li, K.M. Luginbuhl, I. Ozer, W. Liu, M.S. Hershfield, K. Matyjaszewski, and A. Chilkoti, *A brush-polymer/exendin-4 conjugate reduces blood glucose levels for up to five days and eliminates poly(ethylene glycol) antigenicity.* Nature Biomedical Engineering, 2017. **1**(1): p. 0002.
99. Hou, Y., Y. Zhou, H. Wang, R. Wang, J. Yuan, Y. Hu, K. Sheng, J. Feng, S. Yang, and H. Lu, *Macrocyclization of Interferon–Poly(α-amino acid) Conjugates Significantly Improves the Tumor Retention, Penetration, and Antitumor Efficacy.* Journal of the American Chemical Society, 2018. **140**(3): p. 1170-1178.
100. Popp, M.W., S.K. Dougan, T.Y. Chuang, E. Spooner, and H.L. Ploegh, *Sortase-catalyzed transformations that improve the properties of cytokines.* Proceedings of the National Academy of Sciences, 2011. **108**(8): p. 3169-3174.
101. Stephanopoulos, N. and M.B. Francis, *Choosing an effective protein bioconjugation strategy.* Nature Chemical Biology, 2011. **7**(12): p. 876-884.
102. Tayri-Wilk, T., M. Slavin, J. Zamel, A. Blass, S. Cohen, A. Motzik, X. Sun, D.E. Shalev, O. Ram, and N. Kalisman, *Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins.* Nature Communications, 2020. **11**(1).
103. Mattson, G., E. Conklin, S. Desai, G. Nielander, M.D. Savage, and S. Morgensen, *A practical approach to crosslinking.* Molecular Biology Reports, 1993. **17**(3): p. 167-183.
104. Brinkley, M., *A brief survey of methods for preparing protein conjugates with dyes, haptens and crosslinking reagents.* Bioconjugate Chemistry, 1992. **3**(1): p. 2-13.
105. Jin Lee, Y., *Mass spectrometric analysis of cross-linking sites for the structure of proteins and protein complexes.* Molecular BioSystems, 2008. **4**(8): p. 816.
106. Levary, D.A., R. Parthasarathy, E.T. Boder, and M.E. Ackerman, *Protein-Protein Fusion Catalyzed by Sortase A.* PLoS ONE, 2011. **6**(4): p. e18342.
107. Witte, M.D., C.S. Theile, T. Wu, C.P. Guimaraes, A.E.M. Blom, and H.L. Ploegh, *Production of unnaturally linked chimeric proteins using a combination of sortase-catalyzed transpeptidation and click chemistry.* Nature Protocols, 2013. **8**(9): p. 1808-1819.
108. Lobba, M.J., C. Fellmann, A.M. Marmelstein, J.C. Maza, E.N. Kissman, S.A. Robinson, B.T. Staahl, C. Urnes, R.J. Lew, C.S. Mogilevsky, J.A. Doudna, and M.B. Francis, *Site-Specific Bioconjugation through Enzyme-Catalyzed Tyrosine–Cysteine Bond Formation.* ACS Central Science, 2020. **6**(9): p. 1564-1571.
109. Jain, N., S.W. Smith, S. Ghone, and B. Tomczuk, *Current ADC Linker Chemistry.* Pharmaceutical Research, 2015. **32**(11): p. 3526-3540.
110. Dai, J.-M., X.-Q. Zhang, J.-Y. Dai, X.-M. Yang, and Z.-N. Chen, *Modified Therapeutic Antibodies: Improving Efficacy.* Engineering, 2021. **7**(11): p. 1529-1540.
111. Lu, R.-M., Y.-C. Hwang, I.J. Liu, C.-C. Lee, H.-Z. Tsai, H.-J. Li, and H.-C. Wu, *Development of therapeutic antibodies for the treatment of diseases.* Journal of Biomedical Science, 2020. **27**(1).
112. Gébleux, R., M. Briendl, U. Grawunder, and R.R. Beerli, *Sortase A Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody-Drug Conjugates.* Methods Mol Biol, 2019. **2012**: p. 1-13.
113. Möhlmann, S., C. Mahlert, S. Greven, P. Scholz, and A. Harrenga, *In vitro sortagging of an antibody fab fragment: overcoming unproductive reactions of sortase with water and lysine side chains.* Chembiochem, 2011. **12**(11): p. 1774-80.

114. Beerli, R.R., T. Hell, A.S. Merkel, and U. Grawunder, *Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency.* PLOS ONE, 2015. **10**(7): p. e0131177.
115. Dickgiesser, S., M. Rieker, D. Mueller-Pompalla, C. Schröter, J. Tonillo, S. Warszawski, S. Raab-Westphal, S. Kühn, T. Knehans, D. Könning, J. Dotterweich, U.A.K. Betz, J. Anderl, S. Hecht, and N. Rasche, *Site-Specific Conjugation of Native Antibodies Using Engineered Microbial Transglutaminases.* Bioconjugate Chemistry, 2020. **31**(4): p. 1070-1076.
116. Yu, W., K.P. Gillespie, B. Chhay, A.-S. Svensson, P.-Å. Nygren, I.A. Blair, F. Yu, and A. Tsourkas, *Efficient Labeling of Native Human IgG by Proximity-Based Sortase-Mediated Isopeptide Ligation.* Bioconjugate Chemistry, 2021.
117. Kasper, M.A., A. Stengl, P. Ochtrop, M. Gerlach, T. Stoschek, D. Schumacher, J. Helma, M. Penkert, E. Krause, H. Leonhardt, and C.P.R. Hackenberger, *Ethynylphosphonamidates for the Rapid and Cysteine-Selective Generation of Efficacious Antibody-Drug Conjugates.* Angew Chem Int Ed Engl, 2019. **58**(34): p. 11631-11636.
118. Furter, R., *Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli.* Protein Science, 1998. **7**(2): p. 419-426.
119. Wals, K. and H. Ovaa, *Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins.* Frontiers in chemistry, 2014. **2**: p. 15-15.
120. Kiick, K.L., E. Saxon, D.A. Tirrell, and C.R. Bertozzi, *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(1): p. 19-24.
121. Yanagisawa, T., R. Ishii, R. Fukunaga, T. Kobayashi, K. Sakamoto, and S. Yokoyama, *Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode N(epsilon)-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification.* Chemistry & Biology, 2008. **15**(11): p. 1187-1197.
122. Yamaguchi, S., *Recent Advances in Protein Caging Tools for Protein Photoactivation.* Applied Sciences, 2022. **12**(8): p. 3750.
123. Pelliccioli, A.P. and J. Wirz, *Photoremovable protecting groups: reaction mechanisms and applications.* Photochemical & Photobiological Sciences, 2002. **1**(7): p. 441-458.
124. Turecek, P.L., M.J. Bossard, F. Schoetens, and I.A. Ivens, *PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs.* J Pharm Sci, 2016. **105**(2): p. 460-475.
125. Luo, J., Q. Liu, K. Morihiro, and A. Deiters, *Small-molecule control of protein function through Staudinger reduction.* Nat Chem, 2016. **8**(11): p. 1027-1034.
126. Il'ichev, Y.V., M.A. Schworer, and J. Wirz, *Photochemical reaction mechanisms of 2-nitrobenzyl compounds: methyl ethers and caged ATP.* J Am Chem Soc, 2004. **126**(14): p. 4581-95.
127. Klan, P., T. Solomek, C.G. Bochet, A. Blanc, R. Givens, M. Rubina, V. Popik, A. Kostikov, and J. Wirz, *Photoremovable protecting groups in chemistry and biology: reaction mechanisms and efficacy.* Chem Rev, 2013. **113**(1): p. 119-91.
128. Gautier, A., D.P. Nguyen, H. Lusic, W. An, A. Deiters, and J.W. Chin, *Genetically encoded photocontrol of protein localization in mammalian cells.* J Am Chem Soc, 2010. **132**(12): p. 4086-8.
129. Deiters, A., D. Groff, Y. Ryu, J. Xie, and P.G. Schultz, *A genetically encoded photocaged tyrosine.* Angew Chem Int Ed Engl, 2006. **45**(17): p. 2728-31.
130. Brieke, C., F. Rohrbach, A. Gottschalk, G. Mayer, and A. Heckel, *Light-Controlled Tools.* Angewandte Chemie-International Edition, 2012. **51**(34): p. 8446-8476.
131. Jedlitzke, B., Z. Yilmaz, W. Dorner, and H.D. Mootz, *Photobodies: Light-Activatable Single-Domain Antibody Fragments.* Angew Chem Int Ed Engl, 2020. **59**(4): p. 1506-1510.
132. Reimann, O., *Tag-free semisynthetic tau proteins and novel antibodies targeted against phospho-tau*. 2016, Freie Universität Berlin: Berlin, Germanz.

133. Jacobitz, A.W., M.D. Kattke, J. Wereszczynski, and R.T. Clubb, *Sortase Transpeptidases: Structural Biology and Catalytic Mechanism*. 2017, Elsevier. p. 223-264.
134. Sechi, S. and B.T. Chait, *Modification of cysteine residues by alkylation. A tool in peptide mapping and protein identification.* Anal Chem, 1998. **70**(24): p. 5150-8.
135. Popp, M.W.L., J.M. Antos, and H.L. Ploegh, *Site-Specific Protein Labeling via Sortase-Mediated Transpeptidation.* Current Protocols in Protein Science, 2009. **56**(1).
136. Theile, C.S., M.D. Witte, A.E.M. Blom, L. Kundrat, H.L. Ploegh, and C.P. Guimaraes, *Site-specific N-terminal labeling of proteins using sortase-mediated reactions.* Nature Protocols, 2013. **8**(9): p. 1800-1807.
137. Bechtel, T.J. and E. Weerapana, *From structure to redox: The diverse functional roles of disulfides and implications in disease.* Proteomics, 2017. **17**(6): p. 10.1002/pmic.201600391.
138. Nagy, P., *Kinetics and mechanisms of thiol-disulfide exchange covering direct substitution and thiol oxidation-mediated pathways.* Antioxidants & redox signaling, 2013. **18**(13): p. 1623-1641.
139. Gupta, V. and K.S. Carroll, *Sulfenic acid chemistry, detection and cellular lifetime.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2014. **1840**(2): p. 847-875.
140. Devarie Baez, N.O., J.A. Reisz, and C.M. Furdui, *Mass spectrometry in studies of protein thiol chemistry and signaling: Opportunities and caveats.* Free Radical Biology and Medicine, 2015. **80**: p. 191-211.
141. Kuo, Y.H., A.M. Konopko, N.B. Borotto, J.D. Majmudar, S.E. Haynes, and B.R. Martin, *Profiling Protein S-Sulfination with Maleimide-Linked Probes.* ChemBioChem, 2017. **18**(20): p. 2028-2032.
142. Li, L., I. Vorobyov, and T.W. Allen, *The different interactions of lysine and arginine side chains with lipid membranes.* The journal of physical chemistry. B, 2013. **117**(40): p. 11906-11920.
143. Wienken, C.J., P. Baaske, U. Rothbauer, D. Braun, and S. Duhr, *Protein-binding assays in biological liquids using microscale thermophoresis.* Nature Communications, 2010. **1**(1): p. 100.
144. Mädler, S., C. Bich, D. Touboul, and R. Zenobi, *Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities.* Journal of Mass Spectrometry, 2009. **44**(5): p. 694-706.
145. Fitch, C.A., G. Platzer, M. Okon, B. Garcia-Moreno E, and L.P. McIntosh, *Arginine: Its pKa value revisited.* Protein Science, 2015. **24**(5): p. 752-761.
146. Schneider, A.F.L., A.L.D. Wallabregue, L. Franz, and C.P.R. Hackenberger, *Targeted Subcellular Protein Delivery Using Cleavable Cyclic Cell-Penetrating Peptides.* Bioconjug Chem, 2019. **30**(2): p. 400-404.