Development of spontaneous isopeptide bond formation for ligation of peptide tags

Jacob Orry Fierer, BSc, MSc

Department of Biochemistry
New College
University of Oxford

Trinity 2013

Thesis for a Doctorate in Philosophy

Abstract

Peptide tags are ubiquitous in the life sciences, with roles including purification and selective labeling of proteins. Because peptide tags are small they have a limited surface area for binding and hence usually form low affinity protein interactions. These weak interactions limit the uses of peptide tags in cases that require resistance to forces generated with macromolecular architectures or protein motors. Hence a way to create a covalent interaction with a peptide tag would be useful. It was found possible to create a covalent bond-forming peptide tag using the spontaneous isopeptide chemistry of the CnaB2 domain from the Gram-positive bacterium *Streptococcus pyogenes.* In the CnaB2 domain a reactive Lysine forms an isopeptide bond with an Aspartic acid, catalyzed by a Glutamic acid, creating an internal covalent linkage. Subsequently it was shown that the CnaB2 domain could be split into two parts, a domain with the Lysine and Glutamic acid called SpyCatcher and a peptide with the Aspartic acid called SpyTag, such that the isopeptide covalent linkage can be formed when SpyCatcher/SpyTag are mixed together. SpyCatcher/SpyTag was applied in this thesis and showed functionality in a wide array of scenarios. SpyCatcher/SpyTag covalently linked within the cytosol of *E. coli*, on surface membrane proteins of HeLa cells, and regardless of whether SpyTag was located on the N- or C-terminus or an internal site. Crystal structures of SpyCatcher/SpyTag were then obtained and it was found possible to shrink the SpyCatcher by 32 residues to a core domain of 83 residues. To create an even smaller covalent linkage system, SpyCatcher was split further to generate a protein (SpyLigase) ligating two peptide tags. The β-sheet with the reactive Lysine was removed from SpyCatcher and called KTag. SpyLigase could covalently link SpyTag and KTag. SpyLigase-induced ligation was independent of the location of SpyTag/KTag on the target proteins and was applied to create affibody polymers, which were shown to improve magnetic isolation of cells with low tumor antigen expression. Through this work protein-protein covalent linkage systems were refined and generated that have future applications for the creation of unique macromolecular structures, cellular labeling, and protein cyclization.

Authorship Declaration

I declare that all work performed in this thesis is my own, with the exceptions noted below. I have not submitted this thesis for any other degree.

Sections performed by other persons:

- Section 3.1 Precursor development of SpyCatcher/SpyTag, showing that split CnaB2 is capable of covalent reconstitution, performed by Dr. Bijan Zakeri.
- Section 3.5.1 Visualization of SpyCatcher-555 Reconstitution with a SpyTag-Membrane Protein Fusion on the Surface of HeLa Cells. Microscopy performed by Miss Emily Chittock.
- Section 3.6 Thermodynamics and Dissociation Constant Determination of SpyCatcher Reconstitution with Reactive and Inactive SpyTag(DA)-MBP. ITC experiment, analysis, and graph creation by Dr. Ulrich Schwarz-Linek.
- Section 4.3.1 Crystal Structure of SpyTag Reconstituted with SpyCatcherΔN1. Crystallization of SpyCatcherΔN1 reconstituted with SpyTag, three dimensional structure, and table of values in Section A.13 were generated by Dr. Long Li.
- Section 5.8.5 Application of KTag-AffiEGFR-SpyTag Polymers in Improving the Capture of a Cancer Cell Line with Low Expression of EGFR and Section A.10 Polymerization of KTag-AffiEGFR-SpyTag on Magnetic Beads performed by Mr. Gianluca Veggiani.

Publications submitted during this thesis:

Zakeri, B.*, Fierer, J. O*., Celik, E., Chittock, E. C., Schwarz-Linek, U., Moy, V. T., & Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), E690–7. doi:10.1073/pnas.1115485109 *Joint first authors

Li, L., Fierer, J. O., Rapoport, T. a., & Howarth, M. (2013). Structural Analysis and Optimization of the Covalent Association between SpyCatcher and a Peptide Tag. *Journal of molecular biology*. doi:10.1016/j.jmb.2013.10.021

Fierer, J. O.*, Veggiani, G.*, & Howarth, M. SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture. *(Under Submission)*. *Joint first authors

Acknowledgements

I would first like to thank my supervisor Dr. Mark Howarth, for accepting me into his lab, for his wisdom, guidance, and most importantly patience. You have helped to develop my mind and taught me a lot about leadership. I would like to thank the Clarendon fund and New College for their financial support, as it has allowed me to be the first person in my family to actually achieve the opportunity to study at a doctoral level, and at such a world renowned and prestigious institution.

I would to thank the previous members of the Howarth group Dr. Bijan Zakeri, Dr. Claire Chivers, Dr. Jayati Jain, Dr. Denis Krndija, Mr. Chetan Shingadia, Miss Eve Weatherill Miss Jenny Hayes, Miss Emily Chittock, and Mr. Dejan Mesner for their interesting conversations and contributions. However, I would like to thank in particular Dr. Michael Fairhead for his advice and insight, which always impressed me with the subtlety of approach but also the capacity to challenge my perspective both inside and outside the lab to long held views. And obviously I would like to like to thank Mr. Christopher Schoene and Mr. Gianluca Veggiani. You have both made my final year in the lab fun in unimaginable ways and helped me to survive till this point. It was always a pleasure to walk into the lab and be able to laugh.

Next I would like to thank my friends from all the College communities and especially the people whom I had the pleasure to serve on the 2012-2013 New College MCR committee. I truly have unforgettable memories from organizing events and working the bar with you guys. However I would like to thank a few people directly: Rob Daly thanks for listening to all my inanities; it helped me a lot, Tessa Baker for listening to my monologue ramblings, Catherine Chisholm for being an interesting housemate, Evgenia Ivanova for our eclectic chats, Alex ‘FSA’ Graham for making sure I had a ferociously fun time and got home in one piece, Shari Levine for our always unique conversations, Jon Storey for all the wonderful music and stories, Hannah Buckley for hosting interesting house parties, Peter Clark for Kiwi hospitality, Ian Cameron for Minnesotan kindness, Anca Farcas for helping to anchor me to reality, Gi Gi for life counseling, and Aileen Frost for being a confidante to chat about the general social status of the MCR. I would like to especially thank Alex *Elva* Lucas who taught me when a beautiful and strange woman makes you a strange offer you should take it, because it leads to wonderful things you never even imagined. And lastly but definitely not least, I would like to thank Miles Taylor for not only being a lot of fun to bartend with and have distinctive chats, but also for allowing me to be a Ranch hand at Freedom Ranch.

I would also like to thank four people for their close friendships: Sharon Ruane who is a good friend and dance partner and had the patience for when I put my foot in my mouth both on and off the dance floor, Razvan Florin Albu for introducing me to great comedy and being a good Bro during my Masters, Joanna Bagniewska for being there to chat over a cup of tea, and Ivan Mandadzhiev who was and still is a good friend when I needed it the most.

Finally I would like to thank my family. I would to thank my Sister Tali Fierer for teaching me how not to put up with anything and how to get things done. And to my Father Meir Fierer who has had unlimited patience with me and who has taught me that you don’t need a degree to have the knowledge of a Professor and that the most important lessons in life don’t come from books or lectures but from living and experiencing the real world.

Dedication

I would like to dedicate this thesis and work to my Mother, Hana Puterman Fierer (March 19th 1947-December 11th, 2006 /27th Adar, 5707-20th Kislev, 5767). I want to thank my Mother for her always being patient even when I was difficult, for always being compassionate when I was depressed, and for never accepting anything but my best work for when I was in school. I want to thank my Mother for showing me what a strong role model is truly. That you can be born in a displaced persons camp in Deggendorf Germany after World War II and the Holocaust, still become a graduate of UCLA with a Bachelors degree in Economics at the age of 18, and then create and manage your own business. That success is a matter of hard work and most importantly the will to achieve your goals. That the greatest accolades are not achieving the title of Doctor or winning the Nobel prize, but raising and caring for your family and being the best Mother you can to your children. And that the most important thing you can ever do, is just to be there, even when you don't have all the answers.

I miss you and love you a lot Mom.

Table of Contents

Abbreviations

AFM	Atomic force microscopy
BSA	bovine serum albumin
CD	Circular Dichroism
CTCs	Circulating tumor cells
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNL	'Dock-and-Lock'
DTT	Dithiothreitol
EGFR	Epidermal growth factor receptor
ER	Endoplasmic Reticulum
ESI-MS	Electrospray ionization mass spectrometry
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	green fluorescent protein
HA	human influenza hemagglutinin
HRP	horseradish peroxidase
ICAM-1	intercellular adhesion molecule-1
IMAC	immobilized metal-affinity chromatography
IPTG	isopropyl β-D-thiogalactoside
ITC	Isothermal titration calorimetry
K_d	dissociation constant
LB	Luria-Bertani
MBP	maltose binding protein
NCL	Native chemical ligation
NHS	*N*-hydroxysuccinimide
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
PMSF	phenylmethylsulfonyl fluoride
SAF	Self-assembling fiber
SDS	sodium dodecyl sulphate
TEV	Tobacco Etch Virus
TfR	Transferrin receptor
TMAO	Trimethylamine N-oxide

List of Figures and Tables

Chapter 1: Introduction

Chapter 2: Methods

Chapter 3: Characterization of the SpyCatcher-SpyTag Interaction

Chapter 4: The Use of Crystal Structures and Mutational Analysis to Understand and Improve the SpyCatcher-SpyTag Interaction

Chapter 5: SpyLigase: Creating a Tripartite Split Protein for Peptide-Peptide Ligation

Chapter 6: General Conclusions

Appendix

Chapter 1: Introduction

Chapter 1: Introduction

1.1 Peptide Fusion Tags:

Peptide fusion tags are a ubiquitous tool in the life sciences, which have the advantages of acting as an isolation and interaction point on a recombinant protein. Peptide tags are generally non-interfering stretches of 5-20 residues included onto a recombinant protein (Brizzard, 2008; Jarvik and Telmer, 1998; Terpe, 2003). Consequently peptide tags are simple to insert into a gene, only requiring established molecular biology techniques, avoiding chemical modifications following expression of the protein (Jarvik and Telmer, 1998; Malhotra, 2009). Furthermore peptide tags allow for the specific interaction between the tag and a binding agent which can range from an antibody, corresponding binding protein, or metal chelate (Brizzard, 2008; Fritze and Anderson, 2000; Jarvik and Telmer, 1998; Malhotra, 2009; Terpe, 2003). The specificity of the binding interaction can be increased further by sequential recognition with multiple peptide tags, which is feasible due to the aforementioned small size (Lee et al., 2007; Malhotra, 2009).

While peptide tags have these advantages they still have a few flaws limiting their applications. The first flaw is due to a peptide tag's small size which limits the binding area and the non-covalent nature of the interaction with a binding domain, leads to an overall weak interaction (Houk et al., 2003; Jarvik and Telmer, 1998; Stanfield and Wilson, 1995). This weak binding prevents their uses in many applications where forces are generated (Jain et al., 2013; McCloskey et al., 2003; Morris et al., 2001). The second flaw is that peptide tags are limited in their capacity for use inside live cells, regardless of whether it is prokaryotes or eukaryotes, because binding proteins are generally large and may not cross membranes or express well *in situ* (Terpe, 2003; Woof and Burton, 2004). The other consideration is whether the binding domain can withstand the

conditions within a cell; for example antibodies may not fold well in reducing compartments (Cumming et al., 2004; Sevier and Kaiser, 2002; Woof and Burton, 2004).

1.1.1 Applications of Peptide Tags:

Even though peptide tags have their limitations they are used for a variety of applications such as protein purification, isolation, labeling, and immobilization

The use of peptide tags for protein purification is best exemplified with the hexahistidine or His tag. The His tag is a simple peptide tag of typically six Histidines fused onto a protein, and selectively bound for purification by immobilized metal affinity chromatography (IMAC). The His tagged proteins are eluted in gentle conditions by use of the competitor imidazole to displace the Histidines interactions with the metal ions (Brizzard, 2008; Terpe, 2003). Peptide tags are also used for smaller scale isolations such as determining interaction partners by co-isolation (Lee et al., 2007; Terpe, 2003).

Peptide tags can be used for labeling of a tagged protein such as for localization within a cell (Hackbarth et al., 2004). Furthermore peptide tags are commonly used in immunoblotting to determine the presence of a desired protein (Jarvik and Telmer, 1998; Kurien and Scofield, 2003, 2006; Lee et al., 2007; MacPhee, 2010).

Peptide tags are also used for the immobilization of targeted proteins such as for Surface Plasmon Resonance (SPR) chips (Giese et al., 2012; Hodneland et al., 2002; Homola et al., 1999; van der Merwe, 2000), in lithography for 3D printing (Bhagawati et al., 2010), or even for single molecule measurements (Huang et al., 2009a).

Thus peptide tag capabilities demonstrate a wide array of functions with a simple modification. Yet what could improve upon peptide tag applicability would be the capacity to create a stronger interaction.

1.1.2 Protein-Peptide Interactions:

The factors that determine the strength of interaction between a protein and a peptide can be roughly divided into two sections. The first is the surface area where the larger the surface area of interaction typically the greater the binding (Houk et al., 2003). The reason for increased affinity has to do with the enthalpic and entropic considerations that due to the loss of freedom of movement of the peptide tag by the binding event, the increased amount of enthalpic interactions helps to compensate for the loss of entropy (Houk et al., 2003). The second is the types of interactions occurring between the binding domain and the ligand. The most frequent interactions for antibody-peptide binding are hydrophobic (Houk et al., 2003) though there are also hydrogen bonds and charged interactions (Stanfield and Wilson, 1995). However, the analysis of Houk et al. only considers antibodies as the binding domain for peptides, not other protein domains. It should be noted that there is a narrow selection of non-antibody peptide tag protein binders available, limiting the generality of these conclusions. The peptide binding domains with K_d information available are in Table 1.1.

Table 1.1: List of dissociation constants of peptide tags and non-antibody binding domains.

Peptide Tag	Binding Domain	Protein Domain of Origin	Dissocation Constant (K_d)	Reference
Strep-Tag II	Streptavidin	Streptavidin	~1 µM	(Korndörfer and Skerra, 2002)
S-Tag	S-protein	RNase A	~100 nM	(Terpe, 2003)
C-Tag	ePDZ1b (Affinity Clamp)	PDZ and FN3	~0.5 nM	(Huang et al., 2009a, 2009b)
Dsf	FimGt	*E. coli* type 1 pili	<1 pM, not directly determined	(Giese et al., 2012)

The Strep-Tag II and S-Tag bind to their respective domains with at least part of the tag exposed to the environment, limiting the surface area being interacted with (Hackbarth et al., 2004; Korndörfer and Skerra, 2002; Listgarten et al., 1995; Terpe, 2003). The C-Tag/ePDZ1b (affinity clamp) system involved using two engineered

domains (PDZ and FN3) fused together such that the domains dock with the C-Tag and then ‘clamp’ the tag, enveloping the peptide (Huang et al., 2009a, 2009b). Dsf/FimGt is able to achieve its strong binding via β-sheet complementation with multiple main-chain hydrogen bonds (Giese et al., 2012).

Yet even the best of these protein-peptide systems are limited because the interactions are non-covalent. Therefore these interactions are unlikely to withstand denaturing conditions, large forces from molecular motors (Morris et al., 2001), or forces from the cell surface (Jain et al., 2013; McCloskey et al., 2003). Thus the capacity to create a covalent interaction would be a valuable tool in terms of protein interactions.

1.2 Protein-Protein Ligation Systems:

1.2.1 Split Intein-Mediated Ligation:

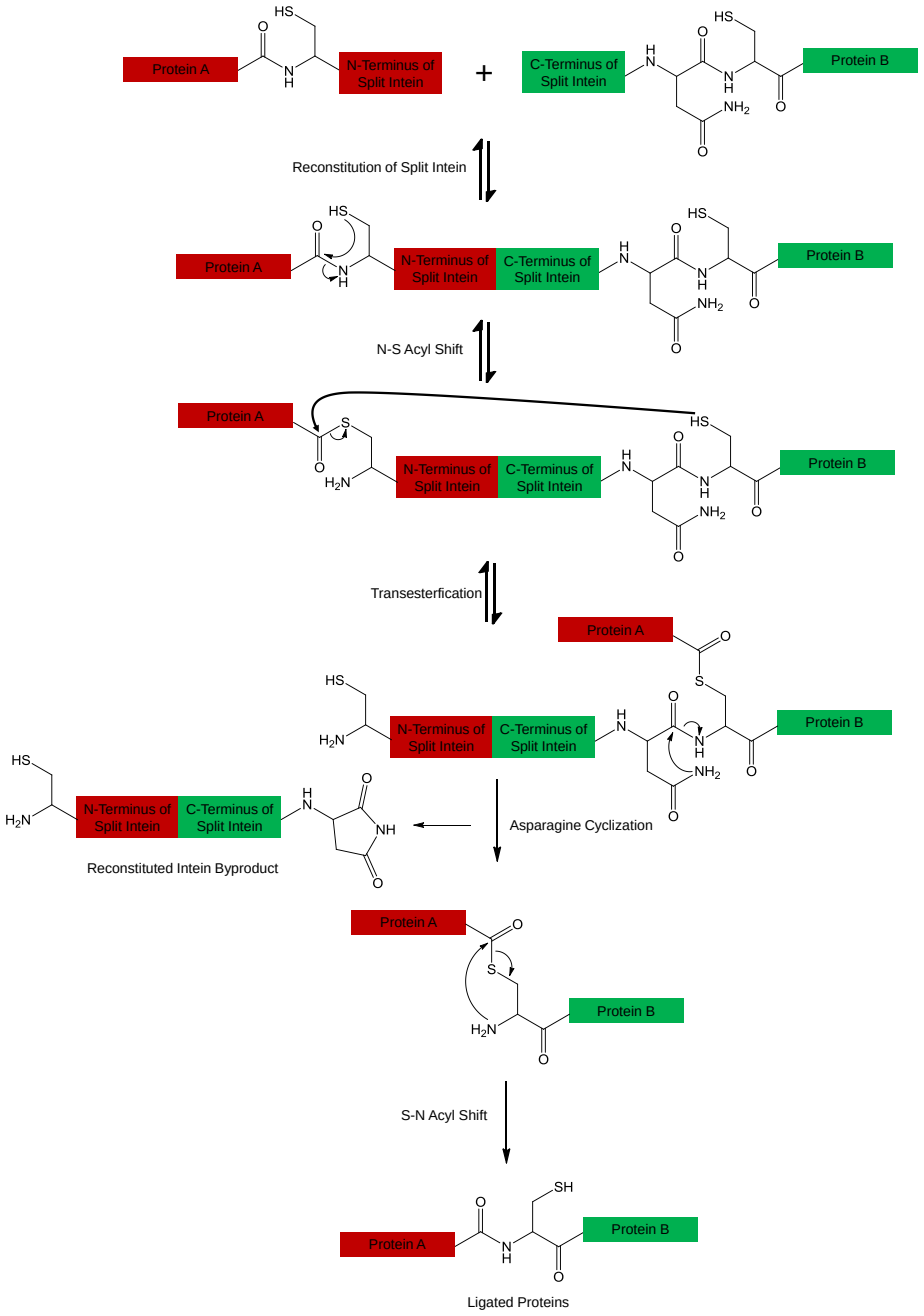


Figure 1.1: Mechanism of split intein-mediated ligation. Split inteins function such that initially the N- (Protein A) and C-terminal (Protein B) sections of an intein reconstitute, which allows for an N-S acyl shift creating a thioester intermediate. Then a transesterfication occurs by a nucleophilic attack by the C-terminal split intein nucleophile transferring the N-terminal protein (Protein A) to the C-terminal nucleophile creating an alternate branched intermediate. Then Asparagine cyclization occurs releasing the reconstituted split intein and allowing for the resolution of the branched intermediate via an S-N acyl shift. This figure was adapted from (Shah and Muir, 2011).

Inteins are self-cleaving protein elements found in many genes (Anraku et al., 2005). While generally considered parasitic genetic elements, their discovery has opened up a new approach for protein-protein ligation technology (Anraku et al., 2005; Lockless and Muir, 2009). The value of inteins is that they are able to ligate two proteins together in a traceless manner. Split inteins are an extension of inteins, in that they retain said traceless ligation capacity, however the intein is split into two parts, which can then be fused onto two target proteins. Subsequently the split intein sections can reconstitute and perform the protein-protein ligation (Lockless and Muir, 2009). Inteins accomplish this ligation via an auto-catalyzed reaction (Figure 1.1) which first requires an N-S acyl shift where the peptide bond of the N-terminal protein (Protein A, or commonly referred to as the N-extein) is transferred to the N-terminal nucleophile of the intein. Then the N-terminal protein (Protein A) is transferred to the intein's C-terminal nucleophile via transesterfication creating a branched intermediate. The branched intermediate is resolved by the cyclization of Asparagine on the C-terminal of the intein, releasing the reconstituted intein as a byproduct, and allowing for an S-N acyl shift creating the ligated product (Noren et al., 2000; Shah and Muir, 2011). Also not all inteins use Cysteines as the nucleophile, as both Threonine and Serine have been found within inteins (Noren et al., 2000; Xu and Evans, 2001).

The advantages of split inteins are that they use a chemistry based on the common amino acids and are usable both *in vitro* and *in vivo* (Noren et al., 2000). Certain split inteins have also been shown to be fast, going to completion within minutes (Shah et al., 2012; Zettler et al., 2009). The reaction can lead to a traceless ligation of the proteins (Lockless and Muir, 2009). On the other hand many inteins are sensitive to the redox environment (Noren et al., 2000; Xu and Evans, 2001). Inteins have also been shown to undergo side reactions (Siebold and Erni, 2002). Furthermore inteins must be located at the N- or C-termini of proteins (Lockless and Muir, 2009; Noren et al., 2000;

Shah and Muir, 2011), limiting their capacity for forming branched protein architectures. Although there have been attempts to use 3-part split inteins to control reconstitution, ligation yields were low (Sun et al., 2004). There have also been attempts to modulate the intein splicing by engineering an allosteric control with a small molecule (thyroid hormone), but this led to a slower reaction of ~1 h (Skretas and Wood, 2005).

1.2.2 Sortagging:

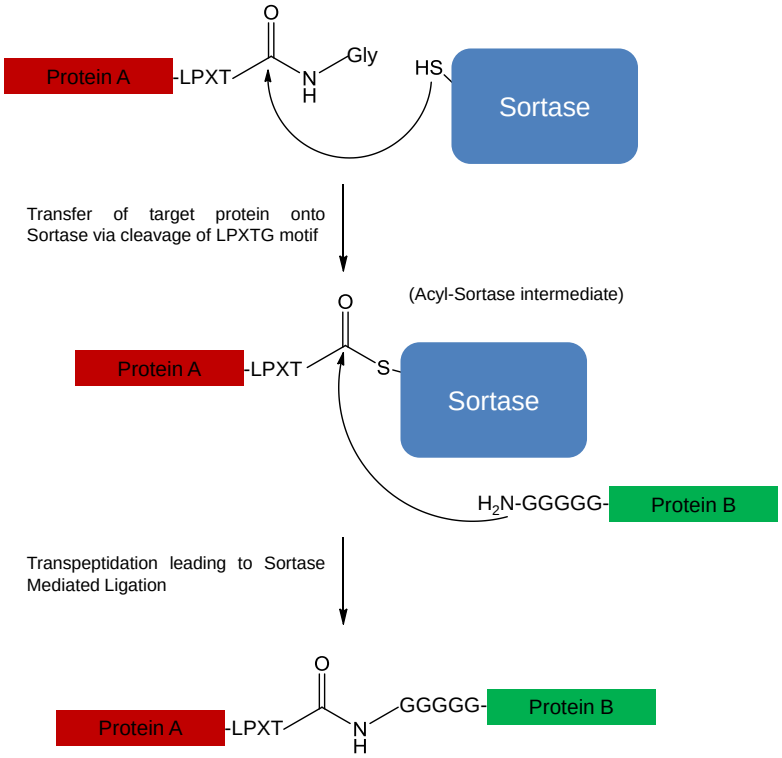


Figure 1.2: Mechanism of sortase-mediated ligation. The sortase enzyme uses a catalytic Cysteine to perform a cleavage of the LPXTG motif between the Threonine-Glycine at the Threonine carbonyl group. Then an acyl-sortase intermediate is formed containing a thioester. Finally the reaction is resolved by the transpeptidation via a nucleophilic attack by the N-terminus of a protein with a penta-Glycine tag. Figure adapted from (Tsukiji and Nagamune, 2009).

Sortases are a class of enzymes found in Gram-positive bacteria that covalently ligate proteins to cell wall components. Sortase targets a five residue motif included in the protein localization sequence (Hendrickx et al., 2011; Marraffini et al., 2006; Paterson and Mitchell, 2004; Ton-That and Schneewind, 2004). The most commonly applied of the Sortases is Sortase A from *Staphylococcus aureus* (Mazmanian et al.,

1999; Paterson and Mitchell, 2004; Tsukiji and Nagamune, 2009). Sortase A has the common target motif of LPXTG (X being any residue) found on the C-terminus of the protein and transfers to the N-terminus of the receptor sequence of a penta-Glycine which is found on peptidoglycan strands in the cell wall of *S. aureus* (Mazmanian et al., 1999; Tsukiji and Nagamune, 2009).

The mechanism of sortase-mediated ligation is shown in Figure 1.2, where initially the sortase performs a nucleophilic attack using a Cysteine on the carbonyl group of Threonine (LPXTG), cleaving the target motif. The nucleophilic attack creates an acyl-sortase intermediate with a thioester connecting to the LPXT motif. Finally the α-amino group of the penta-Glycine tagged protein (Protein B) performs a nucleophilic attack on the Threonine carbonyl group (LPXT-Sortase) of the target motif, completing the transpeptidation (Figure 1.2) (Hendrickx et al., 2011; Marraffini et al., 2006; Paterson and Mitchell, 2004; Ton-That and Schneewind, 2004; Tsukiji and Nagamune, 2009).

Sortase has a great many advantages as a protein ligation system. Sortase ligation is easy to modulate, with no reaction until the inclusion of the enzyme. Furthermore the only modification required is two small penta-peptides on the respective target proteins (Tsukiji and Nagamune, 2009). Sortases have been shown to reach completion within 30 min under neutral conditions (Ling et al., 2012). The system has also been used in a variety of cases for ligating proteins, cell surface labeling, and attaching synthetic chemical groups (Popp et al., 2007; Williamson et al., 2012). Sortases have their limitations in that they require specific reaction conditions; Sortase A requires milimolar amounts of Ca^{2+} (Popp et al., 2007; Williamson et al., 2012), which limits the use of the Sortase A *in vivo*. Other problems are that Sortase reactions are reversible and technically any N-terminally located Glycine can react with the acyl-Sortase intermediate, creating possible promiscuity (Williamson et al., 2012). Furthermore there are side-reactions, with water causing hydrolysis of the acyl-Sortase

intermediate (Ling et al., 2012). Finally Sortases are not traceless and require the tags to be at specific locations, with the LPXTG target motif at the C-terminus and the penta-Glycine receptor motif at the N-terminus (Popp et al., 2007).

1.2.3 Native Chemical Ligation:

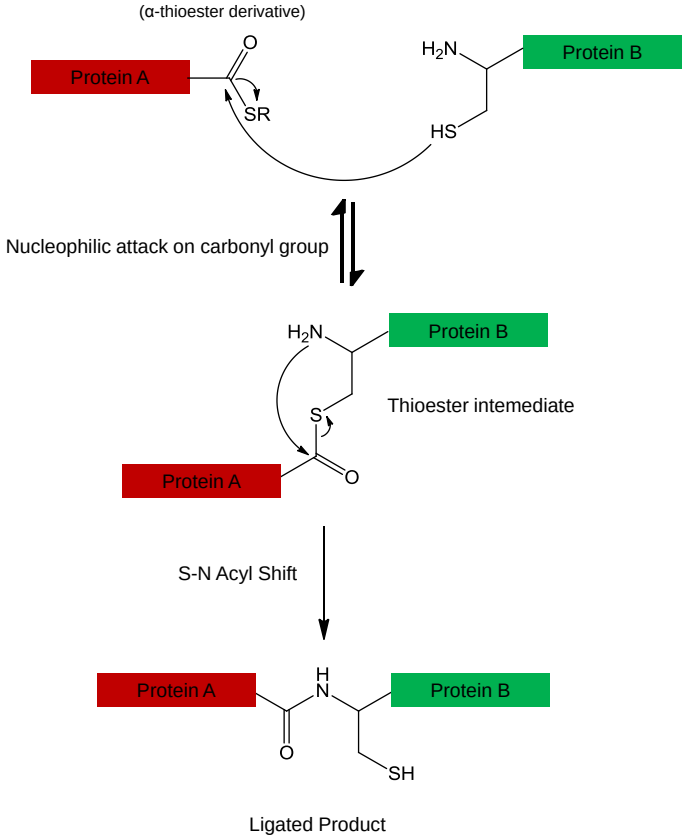


Figure 1.3: Mechanism of native chemical ligation. The process involves that the C-Terminal of a peptide/protein be functionalized with an α-thioester derivative. The thioester then allows for a nucleophilic attack on the carbonyl group by an N-terminally located Cysteine. Then a thioester intermediate is formed on the Cysteine sulfhydryl group, which allows for an S-N acyl shift, creating a peptide bond. Figure adapted from (Kent, 2009; Muralidharan and Muir, 2006).

Native Chemical Ligation (NCL) is a means to synthesize a whole protein without the use of molecular biology techniques or protecting groups, yet with chemo-selectivity (Kent, 2009). The mechanism involves three steps (Figure 1.3). The first is the generation of a peptide/protein with a thioester linkage on the 'C-terminus', which can be synthetically created (Kent, 2009). The thioester can also be semi-synthetically generated via expressed protein ligation, in which a protein has an intein expressed on

its C-terminus, then following the N-S acyl shift, thiols are used to perform a transthioesterification, creating an active substrate (Muralidharan and Muir, 2006). Then a protein/peptide with an N-terminally located Cysteine performs a thioester transfer onto the sulfhydryl group. This then leads to an internal nucleophilic attack causing an S-N acyl shift, generating the ligated product (Kent, 2009).

The NCL reaction has been shown to be functional at gentle conditions of pH ~7 and in aqueous buffers, with reactions taking anywhere from a few hours to 24 h (Johnson and Kent, 2006; Kent, 2009). NCL is also advantageous due to the fact that it is nearly traceless, only leaving a single Cysteine residue at the ligation site. The uses of NCL range from cyclization of proteins (Clark and Craik, 2010; Conlan et al., 2010; Muralidharan and Muir, 2006; White and Yudin, 2011), to simplifying N-terminal labeling using in conjunction a protease to specifically cleave before Cysteines (Busch et al., 2008). Due to the synthetic nature of NCL, the ligation of peptides or chemical groups is possible that would normally be difficult to perform through recombinant methods, such as fusing histone tails with chemical modifications to a core histone domain (Kumar et al., 2011; Voigt and Reinberg, 2011). However, the NCL reaction requires the use of thioesters, thus limiting the system in reducing conditions. Also the reactive groups must be located at the C- and N-termini of the respective proteins (Kent, 2009).

1.2.4 Disulfides and Dock-and-Lock System:

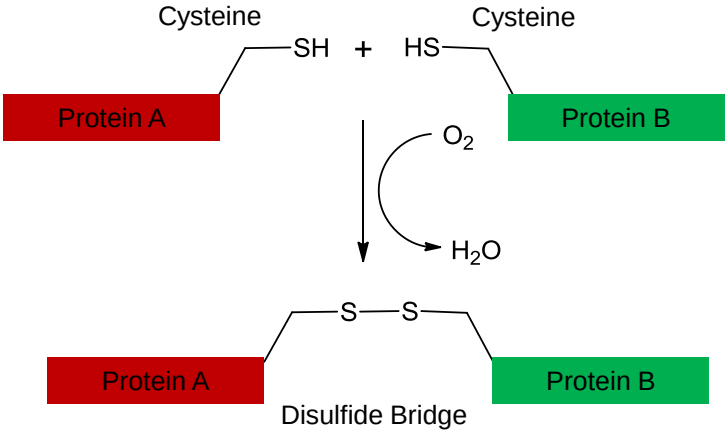


Figure 1.4: Mechanism of Cystine generation. Two Cysteine residues are oxidized leading to the formation of a disulfide bridge.

Disulfide bridges are a common example of a covalent linkage in proteins. Disulfide bridges are used as intramolecular crosslinks to stabilize the three-dimensional structure of a protein (Sevier and Kaiser, 2002) and assemble macromolecular structures such as antibodies (Woof and Burton, 2004). Though Cysteines are valuable in nature, would be simple to engineer, and are flexible in terms of location, they are not an ideal crosslink since they are not formed under reducing conditions and there can be promiscuous reactions between sulfhydryl groups.

However, disulfides have been used in conjunction with a non-covalent binding system to increase the robustness of the interaction in the Dock-and-Lock (DNL) system, which is based on the interaction between the Protein Kinase A (PKA) and the A-Kinase Anchoring Protein (AKAP) (Chang et al., 2007). The two proteins have α-helical binding domains that allow for a specific recognition between PKA and AKAP. Therefore Chang et al. fused the respective α-helices to different antibodies and was able to create bispecific antibodies. They mutated specific residues in the α-helices of PKA and AKAP binding domains to Cysteines such that after the initial binding, disulfide bridges would form specifically (Chang et al., 2007). While a clever method to improve the binding robustness, the DNL system still suffers as the PKA binding helix naturally dimerizes and

only then binds to the AKAP domain following that dimerization, restricting the system to where multivalency is desirable. Finally as the system is using disulfides, in reducing conditions the interactions would be based solely on non-covalent binding.

1.2.5 Other Covalent Ligation Systems:

The covalent ligation systems discussed above cover the most commonly used methods of ligating proteins to peptides/proteins. Here I will briefly summarize some of the other currently available covalent ligation systems.

Artificial amino acids represent a useful tool since by expanding the genetic code they allow for the inclusion of chemistries not normally found in cells. The artificial amino acids are most often included via the expression of a specific pair of tRNA synthetase and tRNA that target the amber codon. The orthogonal tRNA loading systems combined with feeding the cells the desired artificial amino acid allow for a simple *in vivo* method of introducing artificial residues (Chin et al., 2002; Hancock et al., 2010; Lang et al., 2012a, 2012b; Lee et al., 2009; Suchanek et al., 2005). Examples include protein labeling using amino acids with a fluorophore for an R-group (Lang et al., 2012a, 2012b) or reactive chemistries included in the side chain as a way to covalently lock a protein to a binding partner (Chin et al., 2002; Hancock et al., 2010; Lee et al., 2009; Suchanek et al., 2005). However, photocrosslinkings suffer from slow reaction times, low yields, and in many cases require either UV irradiation or toxic metals to activate them (Chin et al., 2002; Hancock et al., 2010; Lang et al., 2012a, 2012b; Lee et al., 2009; Suchanek et al., 2005).

Fluorescein Arsenical Helix Binder (FlAsH), can at low K_d (~pM range) react with the CCXXCC fusion tag (Griffin et al., 1998). FlAsH permeates cell membranes and is fluorescent only after interacting with the CCXXCC tag. FlAsH has also been used as a purification system by selectively isolating CCXXCC-tagged kinesin (Griffin et al., 1998; Thorn et al., 2000). However, the limitations of the FlAsH system are the redox

considerations with the CCXXCC tag (Griffin et al., 1998) and it is not suitable for protein-protein ligation.

HaloTag is a protein domain that can covalently attach to alkyl-halides and is a mutant of the DhaA (dehalogenase) enzyme that normally cleaves halides from alkyl-halide chains (Los et al., 2008). HaloTag uses the DhaA dehalogenation chemistry, which forms a covalent intermediate with the alkyl-halide by nucleophilic replacement of the halide and then is usually resolved by hydrolysis. However, HaloTag is mutated such that hydrolysis is blocked and the covalent intermediate remains (Los et al., 2008). The value of HaloTag is that synthetically produced alkyl-halides linked to other chemical groups can be selectively and quickly conjugated to HaloTag. Thus HaloTag can be used as a cell surface labeling system and as a purification system by coating resin with alkyl-halides (Los et al., 2008). Although the chemo-selectivity is useful, HaloTag is quite large at 33 kDa (Los et al., 2008), making fusions often disruptive.

SNAP-tag is a mutated O^6-alkylguanine-DNA-alkyltransferase (AGT) DNA repair enzyme. The system functions by using the SNAP-tag (mutated AGT) with nucleotide analogues modified with fluorophores that react irreversibly, leading to specific fluorophore labeling. The system allows for an opportune way of *in situ* labeling proteins in live cells by the use of SNAP-tag fusions and then permeating into said cells dyes conjugated to the nucleotide analogues (Gautier et al., 2008; Keppler et al., 2003). However, SNAP-tag is 20 kDa (Gautier et al., 2008) and was designed for fluorescence labeling of proteins *in vivo* rather than as a general system for protein ligation.

Finally there are transglutaminases, a class of enzymes that form isopeptide bonds between primary amines of Lysine with carboxyamide groups of Glutamine (Bailey, 1991; Greenberg et al., 1991; Griffin et al., 2002). However, the reaction is non-specific and will ligate wide ranges of proteins together (Griffin et al., 2002; Lai et al., 2002). Thus transglutaminases have limited use in specific protein ligation, but are ideal

for food processing where the enzyme is used as a 'glue' to create blocks of gelled meat (Motoki and Seguro, 1998).

Thus the systems discussed here in Section 1.2.5 are important tools for protein engineering and chemical biology, but tend to fill very specific niches. Therefore it would be important to have molecular tools for protein ligation that are easy to express recombinantly and flexible in terms of use.

1.2.6 Summary and Comparison of Protein-Protein and Other Ligation Systems:

Table 1.2: Summary and Comparison of Protein-Protein Ligation Systems Comparable to SpyCatcher-SpyTag and SpyLigase in Sections 1.2.1-1.2.4.

Ligation Systems	Advantages	Disadvantages	Reference(s)
Split Inteins	- Based on common amino acids - Recombinantly included - Functional *in vivo* and *in vitro* - Fast reaction; ~min timescale - Can be traceless	- Many inteins dependent on redox environment - Reaction difficult to modulate as immediately react when split inteins in vicinity - Side reactions - Must be located on the N- or C-termini	(Noren et al., 2000; Shah et al., 2012; Siebold and Erni, 2002; Xu and Evans, 2001; Zettler et al., 2009)
Sortagging	- Reaction easy to modulate as dependent on Sortase inclusion - Minimal recombinant modification required: only two small penta-peptide tags - Fast reaction >30 min in neutral conditions - Ligates *in vitro* and on mammalian cell surface - Can ligate synthetic chemical groups	- Require specific reaction conditions; Sortase requires mM Ca^{2+} (Limiting *in vivo*) - Reaction is reversible - Promiscuity as any N-terminal Gly can be a target - Side reactions - Not traceless - Tags must be located at the N- and C-termini	(Ling et al., 2012; Popp et al., 2007; Tsukiji and Nagamune, 2009; Williamson et al., 2012)
Native Chemical Ligation	- Function aqueous and gentle conditions (~pH 7) - Reaction time ≥24 h - Nearly traceless only leaving a single Cys - Synthetic nature increases flexibility compared to recombinant methods	- Nearly traceless leaving behind a Cysteine mutation - Requires thioesters for catalysis limiting use in reducing conditions - Synthetic nature limits applications to *in vitro* - Reactive groups must be located at N- and C-termini	(Clark and Craik, 2010; Conlan et al., 2010; Johnson and Kent, 2006; Kent, 2009; Muralidharan and Muir, 2006; White and Yudin, 2011)
Disulfides (Dock-and-Lock System)	- Easy to include recombinantly - Location flexible - Can increase binding robustness (DNL) - Create bispecific antibodies (DNL)	- Promiscuous with sulfhydryl groups - Non-functional in reducing conditions - DNL only creates multivalency	(Chang et al., 2007; Sevier and Kaiser, 2002; Woof and Burton, 2004)

Table 1.3: Summary and Comparison of Niche Ligation Systems in Section 1.2.5.

Ligation Systems	Advantages	Disadvantages	Reference(s)
Artificial Amino Acids	- Orthogonal chemistries can be introduced - Can be done *in vivo* - Flexible in terms of mutation location and control of chemistry	- Slow reaction times - Low yields - Can require UV irradiation or toxic metals for activation	(Chin et al., 2002; Hancock et al., 2010; Lang et al., 2012a, 2012b; Lee et al., 2009; Suchanek et al., 2005).
FlAsH	- High affinity interaction (K_d in ~pM range) - Small recombinantly encoded peptide tag - Membrane permeable (functional *in situ*)	- Sensitive to reducing conditions - Only suitable for protein labeling or isolation - Not feasible for protein-protein ligation	(Griffin et al., 1998; Thorn et al., 2000)
HaloTag	- Selectively and quickly conjugate to alkyl-halides - Synthetic nature of alkyl-halides allows for flexibility in application - Can be used *in vitro* and cell surface	- Synthetic nature of substrates limits *in vivo* applications - HaloTag is large at 33 kDa making fusions disruptive	(Los et al., 2008)
SNAP-tag	- Specific fluorophore labeling *in situ*	- SNAP-tag is 20 kDa - Designed for fluorescence labeling not protein ligation	(Gautier et al., 2008; Keppler et al., 2003)
Transglutaminases	- Can form an isopeptide between a Lys and Gln	- Low specificity	(Bailey, 1991; Greenberg et al., 1991; Griffin et al., 2002; Lai et al., 2002; Motoki and Seguro, 1998)

1.3 Non-Covalent Protein Linking Systems:

Non-covalent protein-peptide interactions were described in Section 1.1.2 but here I will describe one of the most important non-covalent protein-small molecule interactions.

1.3.1 Streptavidin-Biotin:

Streptavidin and Biotin are a hallmark affinity system. The reason is that Streptavidin and Biotin have one of the strongest non-covalent affinities in nature with a K_d in the femtomolar range (10^{-14} M) (Green, 1990). The Streptavidin-biotin interaction is robust to harsh conditions (Chivers et al., 2010; Diamandis and Christopoulos, 1991).

Therefore it is unsurprising that Streptavidin-biotin has been used in a multitude of applications. The advantages of Streptavidin-biotin extend to the fact that conjugating biotin to a target protein is simple, either by chemical means (Luna, 2001) or enzymatically using the BirA biotin ligase which can biotinylate the Acceptor Peptide (AP) tag (Howarth and Ting, 2008; Stephanopoulos and Francis, 2011).

Even though Streptavidin-biotin is simple to use and has high affinity, the system has its limitations. Streptavidin is a 56 kDa homotetramer, consequently each Streptavidin has four biotin binding sites (Howarth et al., 2006; Nordlund et al., 2005). While there has been the generation of a monovalent Streptavidin which only interacts with a single biotin, the procedure itself is quite work intensive and does not resolve the size problem of the tetramer (Howarth et al., 2006). The other limitation is that even with the extreme affinity of Streptavidin-biotin, the interaction is non-covalent and can be broken by force from molecular motors (Morris et al., 2001). Although there has been the generation of an improved Streptavidin variant, Traptavidin which has slower dissociation rate and better mechanostability, DNA motor protein FtsK was still able to displace the interaction (Chivers et al., 2010; Crozat et al., 2010).

Therefore while the Streptavidin-biotin interaction is an invaluable tool, it demonstrates that even one of the strongest non-covalent affinity interactions is not strong enough in certain conditions and there is a need for covalent systems.

1.4 Split Proteins: Non-Covalent Reconstitution:

The principle of protein splitting is that a target protein with a specific function is split into two or more fragments such that each individual component is inactive. The affinity of each fragment for one another depends on the situation; in the case of determining protein-protein interactions there should be minimal affinity between fragments, as they should not induce the interaction. In the case where one section acts as a tag and the other as a binding domain, a strong affinity is desirable. However, in

both cases, once the fragments are in close proximity this should lead to restoration of the protein function by reconstitution (Shekhawat and Ghosh, 2011). The process of protein splitting has been applied to numerous proteins for varying effects.

One of the most commonly used split proteins is GFP and its variants. In the work of Cabantous et al. they had used GFP akin to a peptide tag-protein domain binding system. They accomplished the unique split of GFP by creating multiple mutant variants and found one that could be split such that only a single β-strand could be removed to inactivate fluorescence but once reconstituted returned activity (Cabantous et al., 2005). The GFP β-stand was used as a tag on protein targets expressed inside live *E. coli.* Then the expression of the rest of the GFP was induced and used to determine a specific protein's localization and/or solubility *in situ* (Cabantous et al., 2005).

The use of split proteins has also been applied to enzymes. The value of split enzymes is that when determining protein-protein interactions it would amplify the signal by multiple turnovers in the case of rare or transient binding events. Split β-lactamase has been applied for binding events occurring within cells using a variety of membrane-permeable colorimetric/fluorogenic substrates (Galarneau et al., 2002). The process of splitting enzymes has also been done with Luciferase fused to split antibodies: antibody binding events were observable by the reconstitution of Luciferase (Stains et al., 2010).

While the examples above are all bipartite splits, the approach has also been shown for proteins split into more than two components. One example is with GFP where it has been split into three parts, the procedure involved removing two β-strands and leaving a core GFP domain that is non-fluorescent. The β-strands were then fused to RNA binding domains, which allowed for the detection of specific RNA sequences *in vitro* as when the two RNA binding domains are adjacent to one another due to the

sequence of the RNA the core domain can reconstitute and report the presence of that RNA (Kellermann et al., 2013). The tripartite splitting of proteins has also been applied to Cytochrome C (Juillerat et al., 1980) and RNase A (Lin et al., 1970) which both showed a restoration of activity following reconstitution.

Therefore split protein engineering is a powerful tool that allows for the spatial and temporal control of a protein activity only after reconstitution of inert fragments. Split proteins are simple to apply using general molecular biology techniques. Unlike protein ligation systems, split protein assembly tends to be functional both *in vitro* and *in vivo*. However, many of these described systems are large and have limited stability as fusion proteins.

1.5 Split Proteins: Covalent Reconstitution:

The generation of covalent bond forming split proteins is a new area developed by the Howarth group. The basis of this technology is the Gram-positive bacterial cell surface adhesin and pilin proteins, which have an unusual chemistry that allows for a spontaneous isopeptide bond formation. Therefore I shall discuss the isopeptide-forming chemistry and then how the covalent bond forming peptide tags were generated.

1.5.1 Chemistry and Possible Reaction Mechanism of Spontaneous Isopeptide Bond Formation:

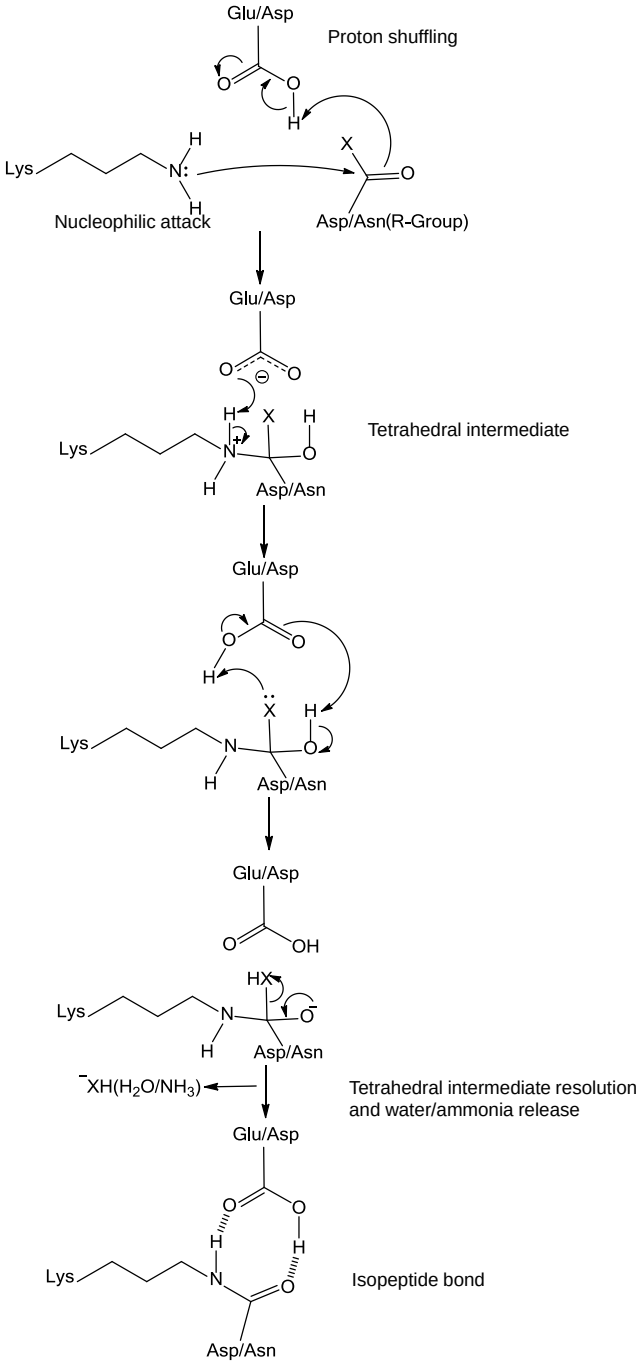


Figure 1.5: Proposed mechanism of spontaneous isopeptide bond formation in adhesin and pilin protein of Gram-positive bacteria. The mechanism follows as such that a Glutamate/Aspartate acts as an acid/base catalyst performing proton shuffling, which allows for the Lysine to perform a nucleophilic attack on the carbonyl group of a reactive Aspartate/Asparagine. Following the nucleophilic attack a tetrahedral intermediate is formed and then collapses into an amide bond between the Lysine and Aspartate/Asparagine, releasing one water or ammonia molecule in the process. This figure was adapted from (Kang and Baker, 2011).

The formation of a spontaneous isopeptide bond where the R-groups of two amino acids have an amide linkage was originally discovered in HK97 bacteriophage capsid (Wikoff et al., 2000) and then by Kang et al., 2007 in the pilin protein Spy0128 from *Streptococcus pyogenes* (Kang et al., 2007). Following the discovery of isopeptide bonds in *S. pyogenes*, the same chemistry was found in other cell surface adhesin and pilin proteins of Gram-positive bacteria (Kline et al., 2010; Ton-That and Schneewind, 2004; Vengadesan and Narayana, 2011).

What is the value of isopeptides for Gram-positive bacteria (Kang and Baker, 2011; Vengadesan and Narayana, 2011)? Isopeptide bonds increase the mechanical robustness of the domain, demonstrated with Spy0128 which withstood >800 pN of pulling force from an AFM cantilever (Alegre-Cebollada et al., 2010) compared to most proteins which unfold at ~100 pN (Crampton and Brockwell, 2010). Isopeptide crosslinked domains have also been shown to withstand other extreme conditions, such as pH 2.0, boiling at 100°C, and protease exposure (Hagan et al., 2010; Kang and Baker, 2009). Thus isopeptides may help the bacterial adhesion withstand high forces, which they are likely encounter as the bacteria attach to a target cell (Alegre-Cebollada et al., 2010).

Isopeptides are found in specific folds; CnaA- or CnaB-type domains which are named in regard to their similarity to either the A or B region of the Cna (collagen binding domain) protein of *Staphylcoccus aureus*. CnaA domains are composed of ~10 β-sheets and have some other secondary structural elements such as an α-helix between two of the β-sheets, while CnaB domains are composed of only β-sheets similar to an immunoglobulin fold and are composed of ~7 β-sheets (Deivanayagam et al., 2000; Kang and Baker, 2011; El Mortaji et al., 2010; Ponnuraj and Narayana, 2007; Ponnuraj et al., 2002; Vengadesan and Narayana, 2011). The isopeptides found in CnaA/B

domains are almost always composed of a Lysine forming an amide bond with an Asparagine, and then catalyzed by either a Glutamate or Aspartate situated between them (Kang and Baker, 2011). Many of the adhesin and pilin proteins have multiple isopeptide bonds (Kang et al., 2009). However, there are two exceptions to the Lysine-Asparagine type of isopeptide bonds, they are the CnaB2 domain of the fibronectin binding protein (FbaB) (Hagan et al., 2010; Oke et al., 2010) and the middle domain of the Spy0125 protein (Pointon et al., 2010) of *S. pyogenes*, which have a Lysine-Aspartate amide linkage and are catalyzed by a Glutamate (Hagan et al., 2010; Kang and Baker, 2011; Oke et al., 2010).

The proposed reaction mechanism of the isopeptide bond formation (Figure 1.5) has the catalytic Glutamate/Aspartate situated between the reactive residues acting as an acid/base catalyst to shift protons such that the Lysine can perform a nucleophilic attack on the carbonyl group of the Aspartate/Asparagine. A tetrahedral intermediate is then formed between the Lysine and Aspartate/Asparagine and collapses with water/ammonia acting as a leaving group (Hagan et al., 2010; Kang and Baker, 2011).

There are also other cases of enzyme-catalyzed isopeptide bond formation, such as SAMPyling in Archaea (Darwin and Hofmann, 2010) and post-translational modifications of Ubiquitin and SUMO in eukaryotes (Hershko and Ciechanover, 1998; Melchior, 2000).

1.5.2 Preceding Work of Isopeptide Forming Peptide Tags:

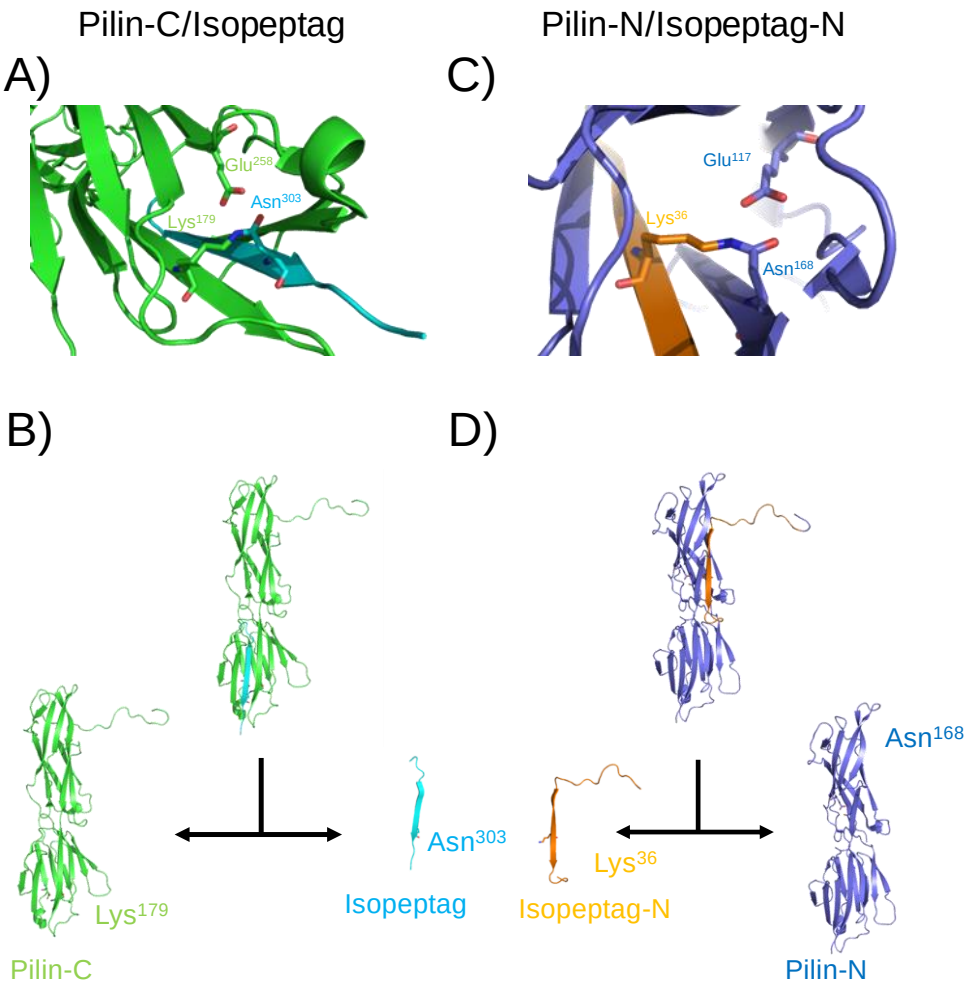


Figure 1.6: Isopeptide triads and different splittings of the Spy0128 protein. A and C show isopeptide triads used to make Pilin-C/Isopeptag and Pilin-N/Isopeptag-N, respectively. B and D show the splitting of Pilin-C/Isopeptag and Pilin-N/Isopeptag-N, respectively. Figure made in PyMOL using PDB 3B2M. Figure adapted from (Zakeri and Howarth, 2010).

The previous work on the creation and development of spontaneous isopeptide bond forming peptide tags was done in the Howarth lab by Dr. Bijan Zakeri, with the purpose of combining the technologies of peptide tags, protein ligation, and split protein engineering all into one system. Dr. Zakeri split the Spy0128 pilin protein into two sections at the C-terminal domain (Figure 1.6B). The C-terminal β-sheet with the reactive Asparagine303 became the peptide tag Isopeptag, and the rest of the Spy0128 protein with the reactive Lysine179 and catalytic Glutamate258 was subsequently

named Pilin-C (Figure 1.6B) (Zakeri and Howarth, 2010). Isopeptag fused to maltose binding protein (MBP, to help with peptide solubility) and Pilin-C were mixed and were shown to form a covalent product (Zakeri and Howarth, 2010). The Pilin-C/Isopeptag system showed itself to be functional and flexible *in vitro* with the capacity to work in buffers ranging from pH 6-8, temperatures of 4, 25, and 37°C, and a wide range of buffers and detergents (Zakeri and Howarth, 2010). Pilin-C/Isopeptag also was functional *in vivo*, forming a covalent adduct in the cytosol of *E. coli* and forming covalent bonds when acting as a cell surface labeling system on an Isopeptag fused to a surface membrane protein (Zakeri and Howarth, 2010). The principle of splitting both the isopeptide chemistry and a protein domain simultaneously was shown to be general by the splitting of a second isopeptide bond located in the N-terminal domain of the Spy0128 pilin protein (Figure 1.6C). In the case of the N-terminal domain isopeptide, the N-terminal β-sheet with the reactive Lysine36 was made into a peptide tag (termed Isopeptag-N), while the catalytic Glutamate117 and reactive Asparagine168 remained on the truncated Spy0128 protein (termed Pilin-N) (Figure 1.6D) (Zakeri and Howarth, 2010). The Pilin-N/Isopeptag-N were then mixed and again showed the capacity to form a spontaneous covalent bond; however in this case a secondary product was also generated, believed to be due to a misformation of the isopeptide bond during Pilin-N expression (Zakeri and Howarth, 2010). The Pilin-C/Isopeptag system has also been applied by the Kamiya group in fusing fluorescent proteins for FRET analysis and as a system for immobilizing proteins to a surface (Abe et al., 2013). Even though split Spy0128 protein was a useful system, it still had its share of problems. Pilin-C/N was ~35 kDa and to achieve 50% covalent reaction took 24 h. Lastly there were side-products (Abe et al., 2013; Zakeri and Howarth, 2010).

The deficiencies of the split Spy0128 pilin protein led to a search for an alternative domain that would prove more robust: the SpyCatcher/SpyTag pair is explored in Chapter 3.

1.6 Thesis Objectives:

The first objective of this thesis was the application and characterization of the SpyCatcher/SpyTag system to achieve rapid and robust covalent protein/peptide interaction. The focus was then to improve the SpyCatcher/SpyTag interaction using insight from crystal structures of the reconstituted complex.

The second main area was the creation of a new type of peptide-peptide ligation technology termed SpyLigase, which was based upon the chemistry of the CnaB2 domain. Finally SpyLigase was applied to create a covalently-fused affibody protein polymers with the goal of enhancing the sensitivity of isolation of tumor cells for cancer diagnosis and prognosis.

Chapter 2: Methods

Some work in this chapter has previously been published in the manuscript (Zakeri et al., 2012), provided in the appendix.

Chapter 2: Methods

2.1 Cloning:

2.1.1 Cloning Conditions for Insert Amplification and Ligation:

PCR conditions for insert amplification were: 5% DMSO (Sigma), 1.5 mM $MgSO_4$ (Novagen), 1x of 10x Buffer for KOD Hot Start DNA polymerase (Novagen), 0.1 ng/µl template DNA, 2 µM sense primer (Invitrogen/Life Technologies), 2 µM antisense primer (Invitrogen/Life Technologies), 0.2 mM dNTPs (Novagen) 0.02 U/µl of KOD Hot Start DNA polymerase (Novagen), and MilliQ water to a total volume of 50 µl. The PCR cycling conditions were: initial denaturation for 5 min at 95°C, then 25 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C, elongation of 1 min for genes ≤1 kb with an additional 30s per kb at 68°C, with a final elongation of 7 min at 68°C before samples were cooled to 12°C. Annealing temperatures and times were optimized when needed. All PCR reactions were performed in the C1000™ Thermal Cycler (Bio-Rad). Analysis of PCR reactions was performed on 0.7-1% Agarose (MELFORD) gels stained with ethidium bromide at 130 V for 45 min. Gels were imaged with a ChemiDoc™ XRS+ Imager (Bio-Rad) and analyzed via Image Lab Software 3.0 (Bio-Rad).

Digestions of vectors and inserts were done with the corresponding New England Biolabs (NEB) restriction enzymes, under manufacturer-recommended conditions. Digested vectors and inserts were agarose gel-purified using a QIAquick Gel Extraction Kit (Qiagen).

Ligations were done with a 5:1 or 1:1 ratio of insert-to-vector, 1x of 10x T4 DNA Ligase buffer (NEB), 20 U/µl T4 DNA Ligase (NEB), and then the ligation reactions were topped up to 10 µl with MilliQ water. The ligation reactions were incubated for 1 h at 25°C. Ligation reactions were transformed by heat shock by mixing 3 µl of the ligation reaction with 50 µl DH5α *E. coli* (Invitrogen) heat-shock competent cells. The heat shock

procedure involved incubating the competent cell-ligated DNA mixture for 20 min on ice, incubating for 45 s at 42°C, incubating for 2 min on ice, adding 300 µl of Super Optimal Broth (SOB) media, incubating cells for 1 h at 37°C with 225 rpm shaking, and then plating cells on an LB agar plate with the appropriate antibiotic. Constructs were verified by DNA Sequencing by Source BioScience.

2.1.2 Overlap Extension PCR:

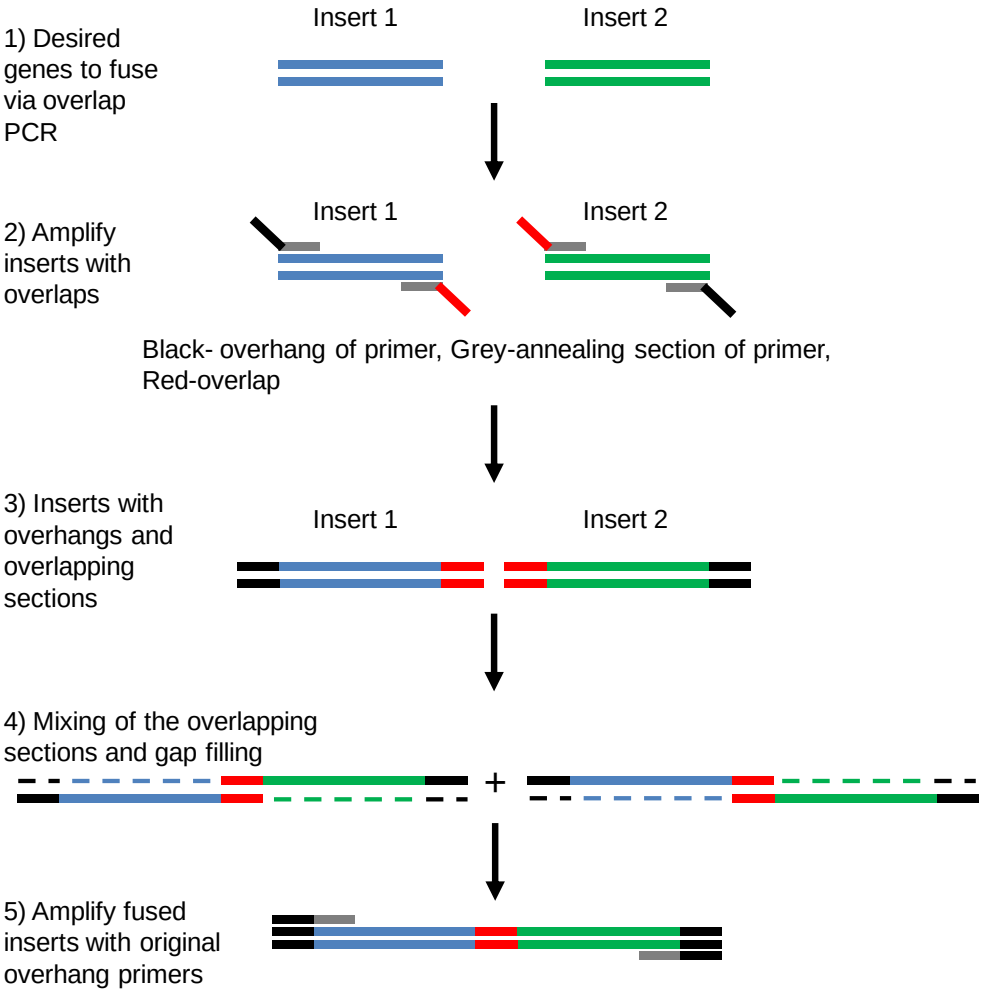


Figure 2.1: Diagram of overlap extension PCR

Overlap extension PCR process is shown in Figure 2.1. The process involved the regular amplification of individual desired segments containing corresponding overlap sequences (Figure 2.1 (2)). Then the amplified sections (Figure 2.1 (3)) would be

agarose gel-purified. Two desired segments would then be mixed in a 1:1 ratio and the combined amount of DNA included into the PCR would be a final concentration of ~0.1 ng/µl (Figure 2.1 (4)). Following the gel purification a two part cycling program would be performed. The program used to fill in the single-stranded regions was initial denaturation for 5 min at 95°C, then 10 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C, elongation of minimum time 1 min for genes ≤1 kb with an additional 30s per kb at 68°C, and then samples were cooled to 12°C (Figure 2.1 (4)). Then only the 5' end sense primer and the 3' end antisense primer of the total insert segment are added to the PCR reaction at the final concentrations of 2 µM each. Then the amplification program was 25 cycles of initial denaturation for 5 min at 95°C, denaturation for 30 s at 95°C, annealing for 30 s at 55°C, elongation of minimum time 1 min for genes ≤1 kb with an additional 30 s per kb at 68°C, final elongation for 7 min at 68°C, and then samples were cooled to 12°C to create the 'overlapped' insert (Figure 2.1 (5)). The insert was then analyzed by agarose gel and purified by agarose gel extraction. The new fusion insert can be used again for another round of overlap extension PCR or digested and ligated into the respective vector.

2.1.3 Cloning Conditions for QuikChange, Inverse PCR, and SLIM:

Site-directed mutagenesis, insertions, and deletions on whole plasmids via ligation-independent methods were performed using QuikChange (Stratagene), Inverse PCR (Kaluz et al., 1999), or Site-directed, ligase-independent mutagenesis (SLIM) (Chiu et al., 2004). Final concentrations for the PCR reactions were: 5% DMSO (Sigma), 2.25 mM $MgSO_4$ (Novagen), 1x of 10x Buffer for KOD Hot Start DNA polymerase (Novagen), 0.2 ng/µl template DNA, 0.4 µM sense primer (Invitrogen/Life Technologies), 0.4 µM antisense primer (Invitrogen/Life Technologies), 0.4 mM dNTPs (Novagen) 0.02 U/µl of KOD Hot Start DNA polymerase (Novagen), and MilliQ water to a final volume of 25 µl. In SLIM there are two extra primers both at 0.4 µM final concentration. The PCR cycling

conditions were: initial denaturation for 5 min at 95°C, then 20 cycles of denaturation for 20 s at 95°C, annealing for 20 s at 55°C, elongation for 4 min at 68°C, and then samples were cooled to 12°C. Annealing temperatures and times were optimized when needed. Following confirmation of PCR product size, reactions were mixed with 1 µl DpnI (NEB) (in a 20 µl PCR volume) and incubated at 37°C for 1 h. The template DNA digestion process was then repeated with the reaction volume transferred to a fresh tube and another 1 µl of DpnI added and incubated at 37°C for 1 h again, to stop any contamination of residual template DNA in the original reaction tube. Reactions were then transformed by heat shock as above.

2.2 Primers Used to Generate Constructs:

2.2.1 Bicistronic Construct:

The cloning of the bicistronic SpyTag-MBP and SpyCatcher/pET28a was based on the SpyTag-MBP/pET28a and SpyCatcher/pDEST14 constructs from Dr. Bijan Zakeri (Zakeri et al., 2012). SpyCatcher from the pDEST14 vector was amplified using 5'-AGGAAAGCTTTTAACTTTAAGAAGGAGATATACATATGTCGTACTACCATC and 5'-TTTACTCGAGTTATTAAATATGAGCGTCACCTTTAGTTGCTTTG. The insert was digested with HindIII (on the sense primer) and XhoI (on the antisense primer) and then ligated into SpyTag-MBP-pET28a vector using HindIII-XhoI sites.

2.2.2 SpyTag Fusion Constructs:

Cloning of the N-1GSpyTag-MBP and N-2GSpyTag-MBP to include 1 or 2 Glycines prior to the N-terminally located SpyTag in N-SpyTag-MBP used the N-SpyTag-MBP/pET28a construct provided by Dr. Bijan Zakeri with sense primer 5'-AGGACCATGGGAGCCCACATCGTGATGGTGGAC (used to add 1 Gly) or 5'-AGGACCATGGGTGGAGCCCACATCGTGATGGTGGAC (used to add 2 Gly), and the antisense primer which annealed to the 3' end of the MBP domain 5'-TTTAAAGCTTCTATTACGAGCTCGAATTAGTCTGCGCGTC (including a second stop

codon). The N-1GSpyTag-MBP and N-2GSpyTag-MBP constructs were digested with NcoI and HindIII and ligated into pET28a.

The mutants of SpyTag, AHIVMVDAYKTPK, where both Lysines were mutated to Arginine, Histidine, or Glutamine, were generated by QuikChange on SpyTag-MBP/pET28a. 2xR-SpyTag-MBP was generated with sense primer 5'-GGTGGACGCCTACAGGCCGACGAGAGGTAGTGGTGAAAGTGG and antisense primer 5'-CCACTTTCACCACTACCTCTCGTCGGCCTGTAGGCGTCCACC. 2xH-SpyTag-MBP mutant was generated with sense primer 5'-GGTGGACGCCTACCATCCGACGCACGGTAGTGGTGAAAGTGG and antisense primer 5'- CCACTTTCACCACTACCGTGCGTCGGATGGTAGGCGTCCACC. 2xQ-SpyTag-MBP was generated with sense primer 5'-GGTGGACGCCTACCAACCGACGCAGGGTAGTGGTGAAAGTGG and antisense primer 5'-CCACTTTCACCACTACCCTGCGTCGGTTGGTAGGCGTCCACC.

Circularly-permuted SpyTag-MBP, Perm1SpyTag-MBP and Perm2SpyTag-MBP were constructed by creating a larger primer where the minimal reactive peptide sequence of AHIVMVDA was followed by either a GS linker (Perm1) or GSG linker (Perm2) and then the N-terminal β-sheet sequence was fused downstream, to create the insert to encode the amino acid sequences AHIVMVDAGSGAMVDT (Perm1) or AHIVMVDAGSGGAMVDT (Perm2). The cloning used MBP/pMAL vector (NEB) as a template. The sense primers were 5'–GGGGCATATGGGAGCCCACATCGTGATGGTGGACGCCGGTAGTGGCGCCATGGTTGATACCGGTAGTGGTGAAAGTGGTAAAATCGAAGAAGGTAAA (Perm1SpyTag-MBP) and 5'–GGGGCATATGGGAGCCCACATCGTGATGGTGGACGCCGGTAGTGGTGGCGCCATGGTTGATACCGGTAGTGGTGAAAGTGGTAAAATCGAAGAAGGTAAA (Perm2SpyTag-MBP). The antisense primer was 5'-

GGGGAAGCTTTTACGAGCTCGAATTAGTCTG. The Perm1SpyTag-MBP and Perm2SpyTag-MBP constructs were digested with NcoI and HindIII and ligated into pET28a.

Mutations of SpyTag based on the crystal structures were generated from SpyTag-MBP/pET28a. Y9R-SpyTag-MBP mutation was by SLIM, while K10N-SpyTag-MBP and P11G-SpyTag-MBP mutation was by QuikChange. The Y9R-SpyTag-MBP (AHIVMVDAYKPTK→ AHIVMVDARKPTK) cloning used Reverse Tail primer 5'-GCCGCGCGGCAGCCATATGGGAGCCCACATCGTGATGGTGGACGCCCGCAAGCC GACGAAG, Reverse Short primer 5'-TCCCATATGGCTGCCGCGCGGC, Forward Tail primer 5'-GCCCACATCGTGATGGTGGACGCCCGCAAGCCGACGAAGGGTAGTGGTGAAAGTG GTAAAATCGAAGAAGGTAAACTGGTAATCTG, and Forward short primer 5'-CAGATTACCAGTTTACCTTCTTCGATTTTACCACTTTCACCACTACC. K10N-SpyTag-MBP (AHIVMVDAYKPTK→ AHIVMVDAYNPTK) cloning used sense primer 5'-GATGGTGGACGCCTACAATCCGACGAAGGGTAGTGGTG and antisense primer 5'-CACCACTACCCTTCGTCGGATTGTAGGCGTCCACCATC. P11G-SpyTag-MBP (AHIVMVDAYKPTK→AHIVMVDAYKGTK) cloning used sense primer 5'-GATGGTGGACGCCTACAAGGGGACGAAGGGTAGTGGTGAAAG and antisense primer 5'- CTTTCACCACTACCCTTCGTCCCCTTGTAGGCGTCCACCATC.

SUMO-SpyTag/pET28a was generated by SLIM based on SUMO-KTag/pET28a by replacing the C-terminal KTag with SpyTag using 5'-GCATCGCGAACAGATCGGTGGTGGCGCCCACATCGTGATGGTGGACGCCTACAAG CCGACGAAG, 5'-GCCACCACCGATCTGTTCGCGATGC, 5'-GCCCACATCGTGATGGTGGACGCCTACAAGCCGACGAAGGGTTATTAAGCTTGCG GCCGCACTCGAG, and 5'-CTCGAGTGCGGCCGCAAGCTTAATAACC.

2.2.3 SpyCatcher Constructs:

All numbering of residues of SpyCatcher was based on the CnaB2 PDB structure 2X5P (Oke et al., 2010). Cloning of TfR-sfGFP-SpyCatcher started with cloning out the transferrin receptor (TfR) transmembrane domain from TfR-YPet-SpyCatcher/pcDNA3. The initial cloning was to mutate the TfR transmembrane domain with the Y20C-F23A double point mutation. The process involved a site-directed mutagenesis of the TfR-YPet-SpyCatcher/pcDNA3 construct using sense primer 5'-CCATTGTCATGTACCCGGGCCAGCCTGG and antisense primer 5'-CCAGGCTGGCCCGGGTACATGACAATGG. Following the creation of the TfR-Y20C-F23A-YPet-SpyCatcher/pcDNA3 construct, the TfR-Y20C-F23A domain was amplified via sense primer 5'-GAAGGAATTCACCATGATGGATCAAGCTAGATC and antisense primer 5'-CCCGGATCCTCCGCTTCCATAGCCCAAGTAGCCAATC (including a GS linker to overlap with sfGFP domain). Then superfolder GFP (sfGFP) from sfGFP/pNIC28-Bsa4 (Nathani et al., 2013) was amplified with sense primer 5'-GGAAGCGGAGGATCCGGGATGCGTAAAGGCGAAGAGCTGTTCACTG and antisense primer 5'-GGATCCACCCGAGCCACCGGAACCCAGATCCTCTTCTGAGATGAGTTTTTGTTCTT TGTATAGTTCATCCATACCATGCGTGATGC. SpyCatcher from SpyCatcher/pDEST14 was then amplified with primers 5'-CCGGTGGCTCGGGTGGATCCGGCGCCATGGTTGATACCTTATCAGG and antisense primer 5'-CTGGGCGGCCGCTTAAATATGAGCGTCACC. The three parts of the insert were joined by overlap extension PCR and with the desired PINCO vector was digested with EcoRI and NotI and ligated.

SpyCatcherΔN1, SpyCatcherΔN2, and SpyCatcherΔN3 were generated from SpyCatcher/pDEST14 by SLIM. SpyCatcherΔN1 was generated by deleting residues 25-46 of SpyCatcher and replacing with a GSG linker via 5'-

CCAACGACCGAAAACCTGTATTTTCAGGGCTCCGGTGATAGTGCTACC, 5'-CTGTATTTTCAGGGCTCCGGTGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAGGAC, 5'-GTCCTCATCACGTTTTGAGAATTTAATATGGGTAGCACTATC, and 5'-GCCCTGAAAATACAGGTTTTCGGTCGTTGG. SpyCatcherΔN2 was generated by inverse PCR, deleting residues 25-34, with 5'-GCCCTGAAAATACAGGTTTTCGGTCGTTGGGATGTCGTAATCGTG and 5'-CAGGGCTCAAGTGAGCAAGGTCAGTCCGGTGATATGACAATTG. SpyCatcherΔN3 was generated by SLIM, deleting residues 11-46 and replacing with a GSG linker using 5'-CATCACCATCACTCCGGTGATAGTGCTACCCATATTAAATTCTCAAAACGTGATGAG 5'-CTCATCACGTTTTGAGAATTTAATATGGGTAGCACTATC, 5'-GTGATGGTGATGGTGATGGTAGTACGACATATG, and 5'-CATATGTCGTACTACCATCACCATCACCATCACTCCGGTGATAGTGCTACC.

The generation of SpyCatcher mutants M0T/K, L4T/K, I58T/K, V63T, V76T/D, and subsequent permutations were generated via site-directed mutagenesis of SpyCatcher/pDEST14 using QuikChange. M0T was generated with 5'-CTGTATTTTCAGGGCGCCAAGGTTGATACCTTATCAGG and 5'-CCTGATAAGGTATCAACCTTGGCGCCCTGAAAATACAG. M0K was generated with 5'-CTGTATTTTCAGGGCGCCACGGTTGATACCTTATCAG and 5'-CTGATAAGGTATCAACCGTGGCGCCCTGAAAATACAG. L4T was generated with 5'-GCGCCATGGTTGATACCACATCAGGTTTATCAAGTGAG and 5'-CTCACTTGATAAACCTGATGTGGTATCAACCATGGCGC. L4K was generated with 5'-GCGCCATGGTTGATACCAAATCAGGTTTATCAAGTGAG and 5'-CTCACTTGATAAACCTGATTTGGTATCAACCATGGCGC. I58T was generated with 5'-CATCTGGTAAAACTATTAGTACATGGACTTCAGATGGACAAGTGAAAGATTTCTAC and 5'-

GTAGAAATCTTTCACTTGTCCATCTGAAGTCCATGTACTAATAGTTTTACCAGATG

I58K was generated with 5'-CATCTGGTAAAACTATTAGTACATGGAAATCAGATGGACAAGTGAAAGATTTCTAC and 5'-GTAGAAATCTTTCACTTGTCCATCTGATTTCCATGTACTAATAGTTTTACCAGATG. V63T was generated with 5'-GTACATGGATTTCAGATGGACAAACGAAAGATTTCTACCTGTATCC and 5'-GGATACAGGTAGAAATCTTTCGTTTGTCCATCTGAAATCCATGTAC. V76D was generated with 5'-CCAGGAAAATATACATTTGACGAAACCGCAGCACCAG and 5'-CTGGTGCTGCGGTTTCGTCAAATGTATATTTTCCTGG. V76T was generated with 5'-CCAGGAAAATATACATTTACCGAAACCGCAGCACCAG and 5'-CTGGTGCTGCGGTTTCGGTAAATGTATATTTTCCTGG. The generation of the M0T-L4T-SpyCatcher mutant used primers 5'-CCTGTATTTTCAGGGCGCCAAGGTTGATACCAAATCAGGTTTATCAAGTGAG and 5'-CTCACTTGATAAACCTGATTTGGTATCAACCTTGGCGCCCTGAAAATACAGG. The I58T-V63T-V76T-SpyCatcher mutant was generated using V76T-SpyCatcher/pDEST14 and 5'-CATCTGGTAAAACTATTAGTACATGGACTTCAGATGGACAAACGAAAGATTTCTACC TGTATCCAG and 5'-CTGGATACAGGTAGAAATCTTTCGTTTGTCCATCTGAAGTCCATGTACTAATAGTTTT ACCAGATG. The M0T-L4T-I58T-V63T-V76T-SpyCatcher mutant used I58T-V63T-V76T-SpyCatcher as a template and then used the M0T-L4T primers to make the complete set of mutations. The M0K-L4K mutant was generated with 5'-CCTGTATTTTCAGGGCGCCAAGGTTGATACCAAATCAGGTTTATCAAGTGAG and 5'-CTCACTTGATAAACCTGATTTGGTATCAACCTTGGCGCCCTGAAAATACAGG. The I58K-V63K-V76D-SpyCatcher mutant was generated from V76D-SpyCatcher/pDEST14

and 5'-CATCTGGTAAAACTATTAGTACATGGAAATCAGATGGACAAAAGAAAGATTTCTACCTGTATCCAG and 5'-CTGGATACAGGTAGAAATCTTTCTTTTGTCCATCTGATTTCCATGTACTAATAGTTTTACCAGATG. The M0K-L4K-I58K-V63K-V76D-SpyCatcher mutant used I58K-V63K-V76D-SpyCatcher as a template and then used the M0K-L4K primers to make the complete set of mutations.

I113A-SpyCatcher, I113V-SpyCatcher, and I113G-SpyCatcher were based on the SpyCatcher/pDEST14 vector and generated by QuikChange. I113A-SpyCatcher used primers 5'-CAACTAAAGGTGACGCTCATGCTTAATAGGTTGATGCTTGAGGATCCG and 5'-CGGATCCTCAAGCATCAACCTATTAAGCATGAGCGTCACCTTTAGTTG. I113V-SpyCatcher used primers 5'-CAACTAAAGGTGACGCTCATGTTTAATAGGTTGATGCTTGAGGATCCG and 5'-CGGATCCTCAAGCATCAACCTATTAAACATGAGCGTCACCTTTAGTTG. I113G-SpyCatcher used primers 5'-CAACTAAAGGTGACGCTCATGGTTAATAGGTTGATGCTTGAGGATCCG and 5'-CGGATCCTCAAGCATCAACCTATTAACCATGAGCGTCACCTTTAGTTG.

SpyCatcherΔN1ΔC1 was generated by deleting the last 4 residues of SpyCatcherΔN1 using SLIM via 5'-GCTCGAATTCGGATCCTCAAGCATCAACCATTTA, 5'-TAAATGGTTGATGCTTGAGGATCCGAATTCGAGC, 5'-ACCTTTAGTTGCTTTGCCATTTACAGTAACCTGACCTTG, 5'-CAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTTAAATGGTTGATGCTTGAGGATC. SpyCatcherΔN1ΔC2 was based on the SpyCatcherΔN1 construct and generated by deleting the last 9 residues from SpyCatcherΔN1 using SLIM with 5'-GCTCGAATTCGGATCCTCAAGCATCAACCATTTA, 5'-

TAAATGGTTGATGCTTGAGGATCCGAATTCGAGC, 5'-GCCATTTACAGTAACCTGACCTTGCTCATTAACTGTAAAGGTAATAGC, 5'-GCTATTACCTTTACAGTTAATGAGCAAGGTCAGGTTACTGTAAATGGCTAAATGGTTGATGCTTGAGGATC.

Generation of SpyCatcherΔN1 T107 point mutants based on the crystal structure of Dr. Long Li were done via QuikChange. T107V-SpyCatcherΔN1 was generated using 5'-GGTTACTGTAAATGGCAAAGCAGTTAAAGGTGACGCTCATATTTAAAT, 5'-ATTTAAATATGAGCGTCACCTTTAACTGCTTTGCCATTTACAGTAACC. T107L-SpyCatcherΔN1 was generated using 5'-GGTTACTGTAAATGGCAAAGCACTTAAAGGTGACGCTCATATTTAAAT, 5'-ATTTAAATATGAGCGTCACCTTTAAGTGCTTTGCCATTTACAGTAACC.

2.2.4 SpyLigase Constructs:

All numbering of residues of SpyLigase and variants was based on the CnaB2 structure PDB 2X5P (Oke et al., 2010). The SpyLigase construct was produced in three consecutive steps. The initial cloning was based on the SpyCatcher/pDEST14 construct produced by Dr. Bijan Zakeri (Zakeri et al., 2012). The initial deletion removed the 'KTag' β-strand (residues 41-60) using primers, 5'-GAGCAAGGTCAGTCCGGTGACGGCAAAGAGTTAGC and 5'-GCTAACTCTTTGCCGTCACCGGACTGACCTTGCTC, which was called SpyLigase1/pDEST14.

SpyLigase1 then was truncated via SLIM PCR (Residues 14-38, premised on the SpyLigase primary structure) based on the following primers 5'-GTCGTACTACCATCACCATCACCATCACGATTACGACGGTCAGTCCGGTGACGGC, 5'-GTCGTAATCGTGATGGTGATGGTGATGGTAGTACGAC, 5'-CACCATCACGATTACGACGGTCAGTCCGGTGACGGCAAAGAGTTAGCTG, and 5'-

CAGCTAACTCTTTGCCGTCACCGGACTGACC. This truncation also generated the SpyLigase variant ‘SpyLigase2’.

The principal “SpyLigase” construct, also known as SpyLigase2C, was generated from the SpyLigase2/pDEST14 vector, replacing the last 4 residues and inserting the following residues, GSGGSGGSGEDSATHI. The SLIM cloning used primers 5’-GCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGGGAGTGGTGGCAGCGGAGGTAGTGGCGAGGACAGCGCTACCCATATT, 5’-ACCTTTAGTTGCTTTGCCATTTACAGTAACCTGACCTTGC, 5’-GGGAGTGGTGGCAGCGGAGGTAGTGGCGAGGACAGCGCTACCCATATTTAAATGGTTGATGCTTGAGGATCCGAATTCGAGC, and 5’-GCTCGAATTCGGATCCTCAAGCATCAACCATTTA.

SpyLigase variant SpyLigase2B was generated using the template SpyLigase2/pDEST14, replacing the last 4 C-terminal residues of the SpyLigase2 and inserting the following residues, GSGGSGGSGEDSATHIFK. The SLIM cloning used the following primers 5’-GCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTGGGAGTGGTGGCAGCGGAGGTAGTGGCGAGGACAGCGCTACCCATATTAAATTC, 5’-ACCTTTAGTTGCTTTGCCATTTACAGTAACCTGACCTTGC, 5’-GGGAGTGGTGGCAGCGGAGGTAGTGGCGAGGACAGCGCTACCCATATTAAATTCTAAATGGTTGATGCTTGAGGATCCGAATTCGAGC, and 5’-GCTCGAATTCGGATCCTCAAGCATCAACCATTTA

SpyLigase3 was generated by SLIM from SpyLigase1/pDEST14, inserting RDI immediately after position 17, using 5’-ACCGGACTGACCTTGCTCACTTGATAAACCTGATAAG, 5’-CTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTCGTGACATCGACGGCAAA,

5’- CTCCATAGTTGCACCAGCTAACTCTTTGCCGTC, and 5’-GTCAGTCCGGTCGTGACATCGACGGCAAAGAGTTAGCTGGTGCAACTATGGAG.

SpyLigase4 was generated by SLIM and deletes residues 18-20 from SpyLigase1/pDEST14 with 5’-ACCTTTAGTTGCTTTGCCATTTACAGTAACCTGACCTTGC, 5’-GCAAGGTCAGGTTACTGTAAATGGCAAAGCAACTAAAGGTTAAATGGTTGATGCTTGAGGATCC, 5’- GCTCGAATTCGGATCCTCAAGCATCAACCATTTA, and 5’-TAAATGGCAAAGCAACTAAAGGTTAAATGGTTGATGCTTGAGGATCCGAATTCGAGC.

SpyLigase5 was generated by SLIM and deletes residues 18-23 using SpyLigase1/pDEST14 with 5’-ACCGGACTGACCTTGCTCACTTGATAAACCTGATAAG, 5’-CTTATCAGGTTTATCAAGTGAGCAAGGTCAGTCCGGTGAGTTAGCTGGTGCAACTATGGAG, 5’- GAATCACGCAACTCCATAGTTGCACCAGCTAACTC, and 5’-GTGAGCAAGGTCAGTCCGGTGAGTTAGCTGGTGCAACTATGGAGTTGCGTGATTC.

SpyLigase disulfide mutants were generated by QuikChange. The L4C-H112C mutant was generated from SpyLigase1/pDEST14 using 5’-CGCCATGGTTGATACCTGTTCAGGTTTATCAAGTG and 5’-CACTTGATAAACCTGAACAGGTATCAACCATGGCG to generate L4C and then 5’-GCAACTAAAGGTGACGCTTGTATTTAAATGGTTGATGCT and 5’-AGCATCAACCATTTAAATACAAGCGTCACCTTTAGTTGC to generate H112C. The A42C-S59C mutant was based on SpyLigase2/pDEST using primers 5’-GGCAAAGAGTTAGCTGGTTGTACTATGGAGTTGCGTG and 5’-CACGCAACTCCATAGTACAACCAGCTAACTCTTTGCC to generate the A42C mutation and then 5’-GTAAAACTATTAGTACATGGATTTGCGATGGACAAGTGAAAGATTTCTAC and 5’-

GTAGAAATCTTTCACTTGTCCATCGCAAATCCATGTACTAATAGTTTTAC to generate S59C. The T43C-A79C mutant was generated from SpyLigase2/pDEST14 using 5'-CAAAGAGTTAGCTGGTGCATGTATGGAGTTGCGTGATTCATC and GATGAATCACGCAACTCCATACATGCACCAGCTAACTCTTTG for T43C and 5'-CATTTGTCGAAACCGCATGTCCAGACGGTTATGAGG and 5'-CCTCATAACCGTCTGGACATGCGGTTTCGACAAATG for A79C. Disulfides were selected by Disulfide by Design Version 1.2 software (Wayne State University, Dr. Alan Dombkowski).

Cloning of SUMO-SpyLigase fusions was based on overlap PCR (Figure 2.1). The first part was the amplification of the two sections to be fused, the SUMO insert and SpyLigase1 and SpyLigase2 inserts. The SUMO insert was amplified four times, the reason being is that the SUMO insert section needed to be overlapped fused with two variants of SpyLigase1 and SpyLigase2 each with different linker lengths (short was for SUMO-SpyLigase1,2 and long was SUMO-GS-SpyLigase1,2). Therefore using the SUMO/pOPIN construct as the template, the sense primer used in all SUMO amplifications was 5'-AGGACCATGGGTAGCAGCCATCACCATCATC. The antisense primer used for overlapping with SpyLigase1 with a short linker to generate SUMO-SpyLigase1 was 5'-GCTCACTTGATAAACCTGATAAGGTATCGCCACCACCGATCTGTTCGCGATGCGC, the antisense primer used for overlapping with SpyLigase1 with a long linker to generate SUMO-GS-SpyLigase1 was 5'-CCTGATAAGGTATCTCCGCTACCACTGCCACCACCGATCTGTTCGCGATGCGC, the antisense primer used for overlapping with SpyLigase2 with a short linker to generate SUMO-SpyLigase2 was 5'-GCCGTCACCGGACTGACCGCCACCACCGATCTGTTCGCGATGCGC and the antisense primer used for overlapping with SpyLigase2 with a long linker to generate

SUMO-GS-SpyLigase2 was 5'-GGACTGACCTCCGCTACCACTGCCACCACCGATCTGTTCGCGATGCGC.

The cloning of the second section of the SUMO-SpyLigase fusions required the amplifications of SpyLigase1 and SpyLigase2 inserts. As there are two versions of SpyLigase insert, one with a short linker (SUMO-SpyLigase1,2) and one with a long linker (SUMO-GS-SpyLigase1,2) it requires that there be two sense primers per SpyLigase insert, and therefore four different amplifications of SpyLigase. The sense primer to amplify SpyLigase1 with the short linker to overlap and generate SUMO-SpyLigase1 using template SpyLigase1/pDEST14 was 5'-CGAACAGATCGGTGGTGGCGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAG, the sense primer to amplify SpyLigase1 with the long linker to overlap and generate SUMO-GS-SpyLigase1 using template SpyLigase1/pDEST14 was and 5'-GTGGTGGCAGTGGTAGCGGAGATACCTTATCAGGTTTATCAAGTGAGCAAGGTCAG (to generate SUMO-GS-SpyLigase1). The sense primer to amplify SpyLigase2 with the short linker to overlap and generate SUMO-SpyLigase2 using template SpyLigase2/pDEST14 was 5'-CGAACAGATCGGTGGTGGCGGTCAGTCCGGTGACGGCAAAGAGTTAG, and the sense primer to amplify SpyLigase2 with the long linker to overlap and generate SUMO-GS-SpyLigase2 using template SpyLigase2/pDEST14 was 5'-GGTGGCAGTGGTAGCGGAGGTCAGTCCGGTGACGGCAAAGAGTTAG. As the 3' end of SpyLigase1 and SpyLigase2 are the same, one antisense primer could be used for the 3' amplification of all the aforementioned variants of SpyLigase regardless of the linker version. The antisense primer used to amplify short and long linker SpyLigase1,2 was 5'-TTTAAAGCTTTTAAATATGAGCGTCACCTTTAGTTGCTTTGCC (which included a HindIII site).

The SUMO-SpyLigase inserts were mixed with the corresponding linker section (Figure 2.1 (4)). As an example the SUMO amplified with the short SpyLigase1 linker would be mixed with the SpyLigase1 insert amplified with the sense primer with the short SpyLigase1 linker. Once the gap filling had been completed as Figure 2.1 (4), the SUMO-SpyLigase fusions were amplified with the sense primer used to amplify SUMO (5'-AGGACCATGGGTAGCAGCCATCACCATCATC) and the antisense primer used to amplify all the SpyLigase inserts (5'-TTTAAAGCTTTTAAATATGAGCGTCACCTTTAGTTGCTTTGCC). After this final amplification and gel purification all four of the inserts SUMO-SpyLigase1, SUMO-GS-SpyLigase1, SUMO-SpyLigase2, and SUMO-GS-SpyLigase2 were digested with NcoI/HindIII and ligated into pET28a.

2.2.5 KTag Fusion Constructs:

SUMO-KTag/pET28a was cloned by Miss Jenny Hayes from SUMO/pOPIN with 5'-TATACCATGGGTAGCAGCCATC and 5'-GTGAAGCTTAATAACCATCACGTTTTGAGAATTTAATATGGGTAGCGCCACCACCGATCTGTTCGCG, followed by digestion with NcoI and HindIII and ligation into pET28a.

KTag-MBP/pET28a was cloned by amplifying MBP of pMAL using 5'-GGGGCATATGGGAGCTACCCATATTAAATTCTCAAAACGTGATGGTAGTGGTGAAAGTGGTAAAATCGAAGAAGGTAAA and 5'-GGGGAAGCTTTTACGAGCTCGAATTAGTCTG. The amplified insert was then digested with NdeI and HindIII and ligated into pET28a.

2.2.6 KCatcher Construct:

KCatcher/pET28a cloning involved the deletion of the KTag β-strand (residues 41-60) from CnaB2/pDEST14 by inverse PCR using 5'-GAGCAAGGTCAGTCCGGTGACGGCAAAGAGTTAGC and 5'-GCTAACTCTTTGCCGTCACCGGACTGACCTTGCTC.

2.2.7 Affibody Fusion Construct:

KTag-AffiEGFR-SpyTag/pET28a was based on AffiEGFR/pET21b, received from Dr. Michael Fairhead in the Howarth group. The cloning involved amplification of the anti-EGFR affibody (Friedman et al., 2008) with 5'-AGGACCATGGGCGCTACCCATATTAAATTCTCAAAACGTGATGGTTCTGGAGCTAG CATGACTGGTGGACAGCAAATGG and 5'-TTTAAAGCTTTTATCACTTCGTCGGCTTGTAGGCGTCCACCATCACGATGTGGGCG CCGGATCCGTGGTGGTGGTGGTGGTGCTCGAG. The insert was then digested with NcoI-HindIII and then ligated into pET28a.

2.2.8 Quantification of DNA:

DNA was quantified via the ND-1000 Nanodrop (NanoDrop) from A_{260}.

2.3 Protein Expression, Purification, and Measurement of Concentration:

2.3.1 Protein Expression:

Expression of pDEST14 constructs (SpyCatcher, KCatcher, and SpyLigase variants) was done in BL21 DE3 pLysS cells (Stratagene), while pET28a constructs (SpyTag, KTag, and Affibody fusions) were expressed in BL21 DE3 RIPL (Stratagene) cells. Cultures were inoculated at 37°C in LB and 0.8% glucose with either kanamycin (0.5 mg/ml) for pET28a vectors or ampicillin (0.1 mg/ml) for pDEST14 vectors. The overnight culture was diluted 1:100 in 800 ml LB and grown until OD 0.5-0.6 at 37°C at 225 rpm, before inducing with 0.4 mM IPTG and growing at 30°C for 4 h. The cells were then spun down at 7,000 rpm, supernatant decanted and bacterial pellets stored at -80°C.

2.3.2 Ni-NTA Protein Purifications:

His_6-tag fusion proteins were purified using standard Ni-NTA methodology. The bacterial pellet was resuspended in 8 ml Ni-NTA binding buffer (50 mM Tris Base, 300 mM NaCl, pH 7.8), keeping chilled on ice throughout the procedure. The resuspension

included the protease inhibitors phenylmethylsulfonylfluoride (PMSF) at 10 mM final concentration and EDTA-free mixed protease inhibitors at 1x final concentration (Roche 11836170001, 3 tablets resuspended in 4.5 ml of MilliQ water was 100x stock). The resuspended pellet was then transferred into 15 ml Falcon tubes and sonicated (Ultrasonic Processor) four times at 40% duty cycle, set to '4' on micro-tip limit for 60 s. The samples were then transferred into 1.5 ml tubes and spun down at 17,000 g for 20 min at 4°C in a microcentrifuge (Heraeus Fresco 17-Thermo Scientific). 0.9 ml Ni-NTA resin (Qiagen) was washed three times with Ni-NTA buffer. The supernatant of the sonicated cell lysate was then transferred onto the washed Ni-NTA resin and rotated end-over-end (Stuart Rotator SB2) for 30 min at 4°C. The samples were then washed sequentially with Ni-NTA binding buffer, Ni-NTA binding buffer containing 10 mM imidazole, and Ni-NTA binding buffer containing 30 mM imidazole. The resin was then transferred to a Poly-prep column and eluted with Ni-NTA binding buffer containing 200 mM imidazole. 0.5 ml fractions were collected and protein concentration was measured via A_{280}. The fractions with the highest protein concentrations were then combined and dialyzed three times in a 1,000 fold excess of Phosphate buffered saline (PBS: 140 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \cdot H_2O$, 2 mM KH_2PO_4 pH 7.4), for 3 h for each round at 4°C.

2.3.3 Amylose Agarose Protein Purifications:

MBP-SpyTag-Zif-SpyTag, N-SpyTag-MBP, N-1GSpyTag-MBP, and N-2GSpyTag-MBP proteins were purified on Amylose agarose (NEB). The cell disruption and sonication was done as for the Ni-NTA purification (section 2.3.2), but pellets were resuspended in Amylose agarose buffer: 50 mM Tris, 300 mM NaCl, 1 mM EDTA, pH 7.8. The lysed cells following sonication and centrifugation were bound with 0.8 ml of Amylose agarose on an end-over-end rotator for 30 min at 4°C. The Amylose agarose was then washed twice with Amylose agarose buffer containing 600 mM NaCl, which

helped increase purity. The samples were then eluted with 10 mM maltose in Amylose agarose buffer. 0.5 ml fractions were collected and measured via A_{280} to determine the protein concentrations. The most suitable fractions were then pooled and dialyzed as above.

2.3.4 Quantification of Protein Concentration:

A_{280} measurements used the ND-1000 Nanodrop (NanoDrop), with the ProtParam online tool from the ExPASy server to determine the extinction coefficient based on the amino acid sequence.

The second way concentrations were determined, for the majority of proteins used in this thesis, was with the Microplate BCA Protein Assay Kit – Reducing Agent Compatible (Pierce/Thermo Scientific - 23252) following manufacturer-recommended protocols.

2.4 Alexa Fluor-555 Labeling of Amine Functional Groups:

Dye conjugation was performed with 10 µl of 1 M $NaHCO_3$, final concentration of 726 µM of SpyLigase or 480 µM of SpyLigase EQ, 10 µl of Alexa Fluor-555 succinimidyl ester (stock 10 mg/ml in DMSO, Life Technologies), and PBS to 100 µl. The samples were wrapped in aluminum foil and rotated end-over-end for 4 h at room temperature. The labeled proteins were then separated from the excess dye by gel filtration via 0.8 ml packed volume of Sephadex G-25 (Sigma-Aldrich) column. 0.2 ml fractions were collected, measured by A_{555} to determine which fractions had the highest amount of dye-labeled protein, and then those fractions were pooled and dialyzed three times in 1,000-fold excess PBS for 3 h at 4°C.

2.5 Polyacrylamide Gel Electrophoresis:

2.5.1 SDS-PAGE:

Reactions were stopped by the addition of 6x SDS Loading dye (PBS, 46 mM Tris HCl, 5 mM EDTA, 4.8% glycerol, 46 mM SDS, and 24 µM bromophenol blue). The

reactions were boiled for 5 min at 95°C in a C1000 Thermal Cycler and 5 µl of each sample was loaded per well (with the exception of *in vivo* reconstitution lysate section 2.6.4 where 18 µl was loaded). The samples were then analyzed via SDS-PAGE using 14% polyacrylamide gels for SpyCatcher-SpyTag reconstitution and 16% polyacrylamide gels for SpyLigase reactions. Gels were run at 200 V for either 67 min (14% SDS PAGE) or 75 min (16% SDS PAGE) in an XCell SureLock (Life Technologies) container. The gels were then stained with InstantBlue Coomassie stain (Triple Red Ltd.), de-stained in MiliQ water, imaged by a ChemiDoc™ XRS+ Imager (Bio-Rad), and band intensities quantified by Image Lab Software 3.0 (Bio-Rad).

2.5.2 Silver Stain:

Silver staining was performed using the Bio-Rad Silver Stain kit (161-0449) on a 14% SDS-PAGE exactly following the manufacturer's instructions.

2.5.3 Native PAGE:

The samples (5 µl of reaction with Native PAGE loading dye were added per well) were run on an 18% Native polyacrylamide gel (same composition as in section 2.5.1 for SDS polyacrylamide gels in terms of buffer composition, water, and acrylamide with the exception that all components were devoid of SDS), using 10% glycerol (final concentration) as Native PAGE loading dye, and run for 2 h 45 min at 200V on ice in an XCell SureLock container. The gel was then washed with MilliQ water and imaged in the FLA-3000 (Fujifilm) by the Image reader V2.02 (Fujifilm) software with the settings: Mode-Fluor.532nm, Filter O580, gradation 65536 (16bit), resolution 200, sensitivity F1000. The gel was stained in Instant Blue and imaged by the ChemiDoc XRS+ Imager as above.

2.6 SpyCatcher-SpyTag Reconstitution Reactions:

2.6.1 In Vitro Reconstitution of SpyCatcher Variants with SpyTag-MBP Variants on the N-terminus:

10 µM SpyCatcher variant was mixed with 10 µM SpyTag-MBP variant in 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 7.0. Reactions were incubated at 25°C for either 1 or 10 min. Reactions were stopped and analyzed as in section 2.5.1.

2.6.2 In Vitro Reconstitution of SpyCatcher with SpyTag-MBP Under Reducing Conditions:

SpyCatcher reconstitution with SpyTag-MBP performed under reducing conditions followed the same reaction conditions as in section 2.6.1 except that reactions labeled 'DTT' were performed with freshly dissolved Dithiothreitol (DTT) at a final concentration of 10 mM.

2.6.3 In Vitro Reconstitution of SpyCatcher with MBP-SpyTag-Zif-SpyTag:

The reconstitution of SpyCatcher with MBP-SpyTag-Zif-SpyTag used the standard buffer composition of 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 7.0. The reactions were incubated for 30 min at 25°C. MBP-SpyTag-Zif-SpyTag was used at 10 µM and SpyCatcher (EQ) was included at 10 µM except where labeled '3x', where 30 µM was used.

2.6.4 In Vivo Reconstitution of SpyCatcher with SpyTag-MBP in the Cytosol of E. Coli and Ni-NTA Pulldown of Reconstituted Adduct:

Following the generation of the bicistronic construct of SpyTag-MBP and SpyCatcher/pET28a (section 2.2.1), the plasmid was transformed into BL21 DE3 RIPL *E. coli* cells. The cells were grown to A_{600} of 0.5 in 800 ml LB, 0.8% glucose, kanamycin (0.5 mg/ml) at 37°C and then induced with 0.4 mM IPTG at 30°C for 4 h.

As a control for possible reconstitution reactions occurring outside the cells from leakage of SpyCatcher and SpyTag-MBP, cells expressing SpyCatcher alone and SpyTag-MBP alone were mixed together at 25°C for 1 h and then lysed.

As a control for reconstitution occurring during the lysis, cells expressing SpyCatcher alone and SpyTag-MBP alone were lysed by adding SDS loading dye and then mixed together at 25°C for 1 h, before boiling.

Pull-downs were performed via Ni-NTA as in section 2.3.2 and protein concentrations were determined via BCA assay (section 2.3.4).

2.6.5 Quantification of In Vitro Reconstitution of SpyCatcher Variants with SpyTag-MBP Variants:

Quantification of the variant SpyCatcher and SpyTag reactions was performed using Image Lab Software 3.0 (Bio-Rad). The band intensity of the reconstituted adduct was divided by the sum of the intensities of all the reactants and reconstituted adduct, and then multiplied by 100.

2.7 SpyLigase Ligation Reactions:

2.7.1 General Conditions of Ligation of SpyTag and KTag Fusion Proteins Directed by SpyLigase:

SpyTag and KTag fusion proteins each at 10 µM were mixed with SpyLigase at 40 µM, under the conditions of SpyLigase reaction buffer: 40 mM Na_2HPO_4 with 20 mM Citric Acid at pH 5.0 was mixed with 1.5 M Trimethylamine N-Oxide Dihydrate (TMAO) (Sigma Aldrich 92277-50G FLUKA), giving a final pH of 7.0. The reactions were incubated for 24 h at 4°C. Reactions were stopped and analyzed as described in section 2.5.1.

2.7.2 Testing Incubation Time, Temperature-dependence, and SpyLigase Concentration-Dependence:

Reactions testing the dependence on incubation time, temperature, and SpyLigase concentration were performed under the same conditions as the *in vitro* ligation in section 2.7.1 but with the single factor changed.

2.7.3 Ligation of SpyTag and KTag Fusions at the N- and/or C-termini of Varying Fusion Proteins:

SpyTag or KTag were fused on the N-terminus of MBP and the C-terminus of the SUMO protein. The reaction conditions were as in section 2.7.1.

2.7.4 Ligation of KTag-MBP and SUMO-KTag to Internally Located SpyTag on MBP-SpyTag-Zif-SpyTag:

Conditions were as in 2.7.1 except for protein concentrations: 40 µM SpyLigase (EQ), 10 µM MBP-SpyTag-Zif-SpyTag and 20 µM KTag-MBP or 20 µM SUMO-KTag.

2.7.5 Ligation of SUMO-KTag and SUMO-SpyTag for Mass Spectrometry Analysis:

40 µM SpyLigase was incubated with 10 µM SUMO-SpyTag and 10 µM SUMO-KTag as in 2.7.1.

2.7.6 Ligation of SUMO-KTag and SpyTag-MBP for Testing Covalent Adduct Release via Native PAGE:

Conditions were as in section 2.7.1, except that SpyLigase (EQ)-555 was in place of SpyLigase (EQ).

2.7.7 Ligation of KTag-MBP and SpyTag-MBP for Testing Covalent Adduct Release via Gel Filtration:

The ligation reaction for gel filtration (Section 2.10) was scaled up to improve detection, containing 40 µM SpyLigase (EQ)-555, 30 µM SpyTag-MBP, 30 µM KTag-MBP, and was incubated as in 2.7.1.

2.7.8 Ligation of SpyTag-MBP and SUMO-KTag using SpyLigase1:

The first ligation reaction was performed in 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 7.0 with 10 µM SpyLigase1, 20 µM SUMO-KTag, 20 µM SpyTag-MBP, incubating at 4°C for 24 h.

2.7.9 Ligation of SpyTag-MBP and SUMO-KTag using L4C-H112C-SpyLigase1, A42C-S59C-SpyLigase2, and T43C-A79C-SpyLigase2:

The ligation using L4C-H112C-SpyLigase1, A42C-S59C-SpyLigase2, and T43C-A79C-SpyLigase2 was performed in 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 5.0

with 10 µM SpyLigase variant, 10 µM SUMO-KTag, and 10 µM SpyTag-MBP, incubating at 4°C for 24 h.

2.7.10 Ligation of SpyTag-MBP and SUMO-KTag using SUMO-SpyLigase1, SUMO-GS-SpyLigase1, SUMO-SpyLigase2, and SUMO-GS-SpyLigase2:

Ligations with SUMO-KTag using SUMO-SpyLigase1, SUMO-GS-SpyLigase1, SUMO-SpyLigase2, and SUMO-GS-SpyLigase2 were performed in 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 5.0 with 10 µM SpyLigase variant, 10 µM SUMO-KTag, and 10 µM SpyTag-MBP, incubating at 4°C for 24 h.

2.7.11 Testing Additives with SpyLigase2:

The additives were tested in 40 mM Na_2HPO_4 with 20 mM Citric Acid at pH 5.0, using 10 µM of SpyLigase2, 10 µM SUMO-KTag, and 10 µM SpyTag-MBP, incubated at 4°C for 24 h. Additives included in Appendix Figure A.6A were at following final concentrations; 50 mM potassium acetate, 5 or 10 mM 2-(N-morpholino)ethanesulfonic acid (MES), 20% (w/v) ammonium sulfate, 125 or 250 mM Arginine, 0.5, 1, 2 M Urea, 50 or 100 mM Sorbitol, and 1.6 M Trehalose. All reactions in Appendix Figure A.6A were at final pH 5.0. Additives included in Appendix Figure A.6B were at the following final concentrations; 15% (w/v) polyethylene-glycol M_W 800 (PEG), 15% (w/v) PEG M_W 3350, 24% (w/v) sucrose, 10 mM MES, 100 mM Sorbitol, and 384 mM Trehalose. All the additives in Appendix Figure A.6B also included 2M TMAO, therefore the pH of all these reactions was 7.0.

2.7.12 Quantification of Ligations Directed by SpyLigase:

The percentage of ligated product was determined by taking the band intensity of the ligated covalent adduct, dividing it by the total band intensities of the SpyTag-fusion and the KTag-fusion and the ligated covalent adduct, then multiplying by 100. Determination of statistical significance was done using a two-tailed unpaired t-test via the online GraphPad Software tool.

2.8 Affibody Polymerization:

2.8.1 Standard Reaction for SpyLigase-Directed Polymerization of KTag-AffiEGFR-SpyTag:

Polymerization of KTag-AffiEGFR-SpyTag was run under the conditions in 2.7.1 except final protein concentrations were 100 µM SpyLigase (EQ) and 37.5 µM KTag-AffiEGFR-SpyTag. Reactions were stopped by the addition of SDS loading dye and boiling for 5 min at 95°C. The samples were then visualized via SDS-PAGE and Coomassie staining using 8 and 16% gels.

2.8.2 Time Course of Polymerization:

The reaction conditions for the time course of the polymerization of KTag-AffiEGFR-SpyTag were as in Section 2.7.1 except reactions were stopped at 0, 0.5, 1, 2, 4, 8 or 24 h (Figure 5.14).

2.8.3 Testing Influence of TMAO Concentration on KTag-AffiEGFR-SpyTag Polymerization and Testing Polymer Aggregates after 37°C Incubation:

The reaction conditions for the polymerization of KTag-AffiEGFR-SpyTag by SpyLigase were the same as in Section 2.7.1, except that TMAO concentrations were 0, 0.5, 1 or 1.5 M.

Reactions were performed in duplicate in the case of Figure 5.15. Following the 24 h incubation in Figure 5.15, reactions of the TMAO titration were treated differently. One set of reactions at 0, 0.5, 1 or 1.5 M TMAO had SDS loading dye added to stop the reaction and were boiled at 95°C for 5 min (lanes labeled ‘B’ to signify both the supernatant and pellet). The other set were spun down for 10 min at 17,000 g, 4°C. The supernatant was then transferred to another set of tubes and had SDS loading dye added and were boiled at 95°C for 5 min, ‘S’. The remaining pellet was resuspended with 30 µl PBS, had SDS loading dye added, and then was boiled at 95°C for 5 min, ‘P’.

In Figure 5.16 the TMAO titration was repeated at 0, 0.5, 1, and 1.5 M TMAO in duplicate. After 24 h, one set of reactions were kept at 4°C for another 24 h and the

other set was transferred to 37°C for 24 h incubation. Following the 4 & 37°C incubations, each set of reactions were resuspended and then split in half, such that from a 30 µl volume, 15 µl was removed and each of these tubes were mixed immediately with SDS loading dye and boiled at 95°C for 5 min. These samples that had SDS loading dye directly added after the 4/37°C incubation are the samples loaded into the lanes labeled with 'B' for they should have both the composition of the supernatant and pellet. The other remaining half of the 4/37°C incubation volumes were then were centrifuged in a table top centrifuge for 10 min at 17,000 g, 4°C. Following centrifugation supernatants were transferred to new tubes, then loaded with SDS loading dye and boiled at 95°C for 5 min. The supernatant samples are run in the lanes labeled 'S'. The tubes containing the pellets were then resuspended in 15 µl of PBS then SDS loading dye was added, and then boiled at 95°C for 5 min. The pellets are run in the lanes labeled 'P'.

2.9 Application of KTag-AffiEGFR-SpyTag Polymers:

The KTag-AffiEGFR-SpyTag and SpyLigase (EQ) were supplied by myself. The rest of the reagent generation and experiments in Section 2.9 were done by Mr. Gianluca Veggiani in the Howarth laboratory.

2.9.1 Polymerization of KTag-AffiEGFR-SpyTag on Magnetic Beads:

Cys-SpyCatcher (generated by Dr. Bijan Zakeri), a Tyr→Cys point mutant on the N-terminal region of SpyCatcher, was first anchored on to magnetic beads. Primary amine-coated superparamagnetic polystyrene beads (Dynabeads® M-270 Amine. 2.8 µm diameter at concentration 2×10^9 beads/ml, Life Technologies) were washed three times with coupling buffer (PBS with 0.1 M $NaHCO_3$ pH 8.0). To functionalize the surface with a sulfhydryl group, 2.5 mM sulfosuccinimidyl 6-(3'-[2-pyridyldithio]-propionamido)hexanoate (sulfo-LC-SPDP, Thermo Scientific) was incubated at 25°C for 1 h, shaking at 1,000 rpm in a thermomixer. Excess sulfo-LC-SPDP was removed by

placing the beads in a magnetic rack (MagRack 6, GE Healthcare) and washing three times with PBS. Then the magnetic beads at 2×10^9 beads/ml were resuspended with 200 µM Cys-SpyCatcher at 25°C for 16 h, shaking at 1,000 rpm in a thermomixer, to create disulfide bridges between the Cys-SpyCatcher and the beads. Excess Cys-SpyCatcher was removed by washing three times with 0.5% BSA, 0.1% Triton-X-100 in PBS. The magnetic beads were then resuspended in 3% BSA, 0.05% NaN_3 in PBS at 25°C for 1 h, shaking at 1,000 rpm in a thermomixer. The magnetic beads were then washed three more times with PBS, followed by resuspension in 0.5% BSA, 0.05% NaN_3 in PBS and storage at 4°C.

KTag-AffiEGFR-SpyTag polymers were formed directly on the beads. The process involved washing the beads three times with PBS then the beads were incubated at 25°C for 2 h and shaken at 1,200 rpm with 8.3 µM KTag-AffiEGFR-SpyTag. The beads were washed once with PBS and resuspended in 1 M TMAO, 40 mM Na_2HPO_4 & 20 mM Citric Acid at pH 5.0, and 76 µM of KTag-AffiEGFR-SpyTag. In the tube where polymerization was directed, SpyLigase at 200 µM final concentration was added. In the negative control, SpyLigase EQ was added to 200 µM. The magnetic beads were then shaken at 1,200 rpm at 4°C for 72 h (Appendix Figure A.9).

2.9.2 Capture Assay on Magnetic Beads:

The BT474 and LBL 771.221 cells were resuspended in DMEM with 1% FCS, 50 U/ml penicillin, and 50 µg/ml streptomycin (D1). Each cell line was used at a concentration of 2.5×10^6 cells/ml, with 100 µl of cells used per experiment. Then 12.5 µl of polymeric and monomeric beads (from Section 2.9.1) were washed three times with D1. The magnetic beads were then mixed with the cells and the volume raised to 500 µl with D1 and at 25°C for 20 min mixed end-over-end. An aliquot of 100 µl of beads and cell mix were taken for counting. The tubes were then put onto the magnetic rack; media with the unbound cells were taken for counting. The cells still bound to magnetic beads

were washed with 500 µl D1 and incubated at 25°C for 30 s at 1,100 rpm. The magnetic beads were then resuspended into 100 µl D1.

Cell counting was done with a Coulter Counter (CasyR Model TT, Innovatis) using a 150 µm measuring capillary, with sample volume of 400 µl and evaluation cursor of 7.5-50 µm. The percentage of recovered cells was quantified by taking the number of recovered cells, dividing by the number of cells in the original bead-and-cell mixture, and then multiplying by 100 (Figure 5.17C).

2.9.3 EGFR Expression Level Determination by Flow Cytometry:

The BT474 and 771.221 cells were collected and resuspended at a concentration of 2.5×10^6 cells/ml. In each sample, 125,000 cells were mixed with FACS A buffer (PBS with 1 % BSA and 0.1% NaN_3) with and without 10 µg/ml of anti-human EGFR antibody (Millipore, Ab-1 clone 528) at 25°C for 10 min. The cell lines were then washed three times with 100 µl of FACS A buffer. 100 µl of FACS A buffer with 1:100 goat anti-mouse IgG-phycoerythrin (Life Technologies) was added to each sample and incubated on ice for 10 min. The cells were washed three times with 100 µl FACS A buffer. The cells were then resuspended in 300 µl FACS A buffer, loaded into the FACScalibur flow cytometer and analyzed with CellQuest Pro version 5.2.1 (Becton Dickinson).

2.10 Gel Filtration for Determination of Covalent Adduct Release by SpyLigase:

Gel Filtration was performed on the ÄKTA™ Purifier (GE Healthcare) with a 24 ml volume Superdex™ 200 10/30 column (old Amersham notation of 10 *mm* by 30 *cm*) (GE Healthcare). The running buffer was 1x PBS and total volume of sample loaded per injection was 100 µl, at a flow rate of 0.5 ml/min, and total flow of 1.5 column volumes. Gel filtration was performed at 4°C and absorbance was measured at 280 nm to measure total protein concentration and at 555 nm to measure SpyLigase(EQ) labeled with Alexa Fluor-555 dye. The condition of the ligation reactions are reported in section

2.7.7. The protein concentrations of 'alone' control samples were 40 μM SpyLigase (EQ)-555, 30 μM SpyTag-MBP, and 30 μM KTag-MBP.

2.11 Isothermal Titration Calorimetry (ITC) of SpyCatcher-SpyTag Interaction:

Proteins and buffer were supplied by myself and all other experimental work was performed by Dr. Uli Schwarz-Linek at the University of St. Andrews.

"ITC measuremens were performed on a VP-ITC micro-calorimeter (MicroCal). SpyCatcher, SpyTag-MBP, and SpyTag DA-MBP were dialyzed in 10 mM potassium phosphate and 50 mM NaCl at pH 7.0 (phosphate buffer), at 4˚C for 12 h. Concentrations were determined by A_{280} measurements and extinction coefficients calculated from the amino acid sequence using the ProtParam online tool (ExPASy). Solutions were degassed and ITC was carried out at 25 ˚C. The cell volume of 1.43 ml had 50 μM SpyCatcher; the syringe contained ~290 μl of 750 μM SpyTag-MBP or 650 μM SpyTag DA-MBP. Titrations were performed by one preliminary injection of 2 μl and then followed by 30 injections of 5 μl (SpyTag-MBP) or 6 μl (SpyTag DA-MBP). The cell was stirred at 307 rpm and the delay time between the injections was adjusted to take into account the time for completion of reaction with reactive SpyTag-MBP or fixed to 4 min with the inactive SpyTag DA-MBP. The averaged values of the endothermic heats of dilution at the end of the experiments were subtracted from the experiment values and the corrected data was fitted to a one-site model using MicroCal Origin software" (Zakeri et al., 2012).

2.12 Mass Spectrometry of SpyLigase-Directed Covalent Adduct:

Ligation reactions in Section 2.7.5 were spin-filtered to enrich the SUMO-KTag:SUMO-SpyTag adduct using a Vivaspin 30,000 molecular weight cut-off column (Sartorius) and microdialyzed against MilliQ water. Samples were then injected into the Micromass LCT time-of-flight electrospray ionization mass spectrometer (Micromass UK). The results were analyzed by MassLynx V4.00.0 (Waters). Masses were predicted using the ExPASy ProtParam online tool based on amino acid sequences.

2.13 Crystallization of SpyCatcherΔN1 Reconstituted with SpyTag:

Plasmids were supplied by myself; all other work and analysis was performed by Dr. Long Li at Harvard University.

"The SpyCatcher-SpyTag protein was concentrated to 110 mg/ml in a Tris buffer (Tris 20mM, NaCl 50mM, pH8.0). Protein crystallization was performed in 24-well plates by hanging drop vapor diffusion. 1 μl of the protein solution was mixed with 1 μl of the crystallization solution (sodium acetate 0. pH4.5, PEG3350 30%). The crystallization drop was then streak seeded. The thin plate crystals appeared in one day and grew to full sizes in three days. The crystals were transferred to a drop of the cryo-solution containing the crystallization solution plus 5% glycerol before flash freezing in liquid

nitrogen. The SpyCatcherΔN1-Spytag protein was crystallized similarly without streak seeding. The protein was concentrated to 34 mg/ml and the crystallization solution contained sodium acetate 0.1 M pH 4.5, sodium chloride 200 mM, PEG3350 30%, and glycerol 5%. The crystals were flash frozen directly without using additional cryo-solution. The X-ray diffraction data were collected on beam ID-C of NE-CAT at Advanced Photon Source in Chicago. The data were processed with HKL2000. The structure were solved by molecular replacement with the program Phaser and using the CnaB2 structure (PDB access number: 2X5P) as the search model. The refinement was performed with the program Phenix." (Personal Communication with Dr. Long Li)

2.14 Western Blotting:

Samples were run on 8% SDS-PAGE and transferred to a PVDF (polyvinylidene fluoride) membrane (Immobilon-P, 0.45 µm pore size, Millipore) in the case of section 2.14.1 or to a nitrocellulose trans-blot 0.2 µm membrane (Bio-Rad) for section 2.14.2. Transfer buffer was 12 mM Tris base, 96 mM Glycine and 20% Methanol and transfer used an XCell SureLock container with the XCell II™ Blot Module (Life Technologies) on ice at 230 mA for 2h. The membranes were then blocked with 12.5% fat-free milk in PBS overnight at 4°C.

2.14.1 Western Blotting of Pull-Down of In Vivo Reconstitution of SpyCatcher with SpyTag-MBP:

The samples used for the Western blotting were prepared as in section 2.6.4, except an 8% SDS-PAGE was used as it helped in the transfer.

The PVDF membrane with the Ni-NTA Pull-Down proteins from the *in vivo* reconstitution (section 2.6.4) were immunoblotted. PBS with 0.05% Tween 20 (PBST), 2.5% fat free milk, and mouse 1:500 anti-penta-His (Qiagen) was incubated at room temperature for 1 h with the membrane. The membrane was then washed four times with PBST for 1 min each round at room temperature while on a rocker. Then the membrane was incubated with PBST mixed with 1:5,000 goat anti-mouse IgG-horseradish peroxidase (Sigma) for 1 h at room temperature while rocking. The membrane was then washed again four times with PBST for 5 min each round on a rocker. The membrane was incubated with SuperSignal West Pico (Pierce), following manufacturer instructions. The membrane was imaged on the ChemiDoc™ XRS+

Imager with the development observed over 20 min and analyzed via Image Lab Software 3.0.

2.14.2 Western Blotting of SpyCatcher Reconstitution with HA-SpyTag-ICAM-EGFP on the Surface of HeLa Cells:

HeLa or HeLa HA-SpyTag-ICAM-EGFP were grown in 75 cm^2 flasks to full confluency, washed with 5 mM EDTA in PBS, and released by incubation with 5 mM EDTA in PBS at 37°C for 5 min. Cells were washed twice with PBS and then incubated for 15 min at 25°C with 1% BSA, 5 mM $MgCl_2$ and mixed with either 5 µM SpyCatcher (EQ) or PBS. Cells were pelleted and washed twice with ice cold PBS, before being lysed with SDS loading dye. The samples were then pre-heated to 70°C for 10 min and then boiled at 90°C for 10 min. Cells were sonicated in an Ultrawave Ltd sonicating water bath at 25°C for 5 min to reduce DNA viscosity. Cell lysates were then boiled with 10 mM DTT at 95°C for 5 min to reduce any disulfide bridges.

The nitrocellulose membrane was immunoblotted with PBST, 2.5% fat free milk, and the appropriate antibody: mouse 1:500 anti-penta-His antibody (Qiagen) to target SpyCatcher (EQ), rabbit 1:300 anti-HA (Rockland) to detect HA-SpyTag-ICAM-EGFP, or mouse 1:1,000 anti-Glyceraldehyde-3-phosphate dehydrogenase antibody (clone GAPDH-71.1, Sigma) as a loading control. The membranes were incubated at room temperature for 1 h and then washed four times with PBST for 1 min each at room temperature while rocking. Then the membranes were incubated with PBST mixed with 1:5,000 goat anti-mouse IgG-horseradish peroxidase for anti-His and anti-GAPDH or 1:5,000 goat anti-rabbit IgG-horseradish peroxidase for anti-HA, for 1 h at room temperature while rocking. The membranes were then washed again four times with PBST for 5 min each round. The membranes were then imaged as above.

2.15 Cell Culture:

2.15.1 Culture of HeLa, BT474, and 771.221 Cells:

HeLa cells were grown in DMEM (Dulbecco's Modified Eagle's Medium) with 10% fetal calf serum (PAA labs), streptomycin 50 µg/ml, and penicillin 50 U/ml. Cells were passaged for less than 6 months.

The cell culture work in Chapter 5 was performed by Mr. Gianluca Veggiani in the Howarth laboratory. The cell lines BT474 and LBL 771.221 were grown in DMEM with 10% fetal calf serum (PAA labs), streptomycin 50 µg/ml, and penicillin 50 U/ml. 5 µg/ml of insulin was included for the BT474 cell line.

2.15.2 Transfection of HeLa Cells:

The cells used in Chapter 3 were transfected with the PINCO vector, which keeps the plasmid episomally retained (Grignani et al., 1998). Transfection of HA-SpyTag-ICAM-EGFP/PINCO into HeLa and then freezing at -80°C of the cell line was performed by Dr. Mark Howarth. I performed the transfections and maintenance of TfR-sfGFP-SpyCatcher (EQ)/PINCO.

Transfections were performed using Turbofect (Thermo Scientific) following manufacturer-recommended instructions.

2.15.3 Cell Surface Labeling of HeLa Cells with HA-SpyTag-ICAM-EGFP or TfR-sfGFP-SpyCatcher (EQ):

The labeling of HeLa cells expressing HA-SpyTag-ICAM-EGFP was performed by Miss Emily Chittock. Cells were washed with PBS with 5 mM $MgCl_2$ (PBS-Mg) and then incubated with 5 µM SpyCatcher (EQ) labeled with Alexa Fluor-555 (Section 2.4), 1% BSA, in PBS-Mg at 25°C for 15 min. The cells were then washed twice with PBS-Mg and imaged live.

HeLa expressing TfR-sfGFP-SpyCatcher were washed twice with PBS-Mg, and then incubated in PBS-Mg with 10 µM SpyTag-MBP for 15 min at room temperature. Cells were washed twice with PBS-Mg. Then cells were then incubated with 100 µl 1:100

anti-penta-His antibody on ice for 30 min. The cells were washed twice with PBS-Mg. The cells were then incubated with 100 µl 1:100 goat anti-mouse Alexa Fluor-555 on ice for 30 min. Cells were washed twice with PBS-Mg and then fixed by 3.7% formaldehyde (Sigma) in PBS for 20 min at room temperature.

2.15.4 Microscopy:

Microscopy in Figure 3.8 for HA-SpyTag-ICAM-EGFP was performed by Miss Emily Chittock. Microscopy experiments were performed using the DeltaVision Elite core wide-field deconvolution systems (Applied Precision) at 40x magnification with an oil immersion lens. The filters used to image sfGFP/EGFP were excitation 475/28, emission 525/48, and dichroic "Quad" API 52-852111-001. For Alexa Fluor-555, filters were excitation 575/25, emission 625/45, and dichroic "Quad" API 52-852111-001. Images were background-corrected and exposed for 0.1-1s. Images were analyzed using the softWoRx 5.0 software (Applied Precision) or ImageJ (National Institutes of Health).

2.16 Circular Dichroism:

Circular Dichroism (CD) was performed on a Jasco J-720 Spectropolarimeter. Absorption was measured from 195-260 nm. The proteins were in PBS at 0.1-0.5 mg/ml. PBS alone was measured for background.

2.17 Images of Protein Structures:

Images of protein structures, alignments and polar contacts were generated in PyMOL (DeLano Scientific LLC).

Chapter 3: Characterization of the SpyCatcher-SpyTag Interaction

Some work in this chapter has previously been published in the manuscript (Zakeri et al., 2012), provided in the appendix.

Chapter 3: Characterization of the SpyCatcher-SpyTag Interaction

The discovery that a protein with an intramolecular isopeptide bond can be split into two parts, a protein domain and peptide tag, and then those two parts can be mixed to re-form the isopeptide bond created a new class of peptide tag. The development of a covalent bond-forming class of peptide tags was an evolutionary process. Covalent bond-forming peptide tags started with a protein domain and peptide tag based on the Spy0128 pilin protein of *Streptococcus pyogenes*. The Spy0128 protein domain was split into the Pilin-C (domain) and a 16-residue Isopeptag (peptide tag). Though the system was functioning, Pilin-C/Isopeptag was slow in that it took 24 h to reach 50% covalent reconstitution (Zakeri and Howarth, 2010). It was an obvious target to improve on this system by developing a faster reacting split protein system

Thus work began on developing an alternative system, based on another isopeptide-containing domain from *S. pyogenes*. The chosen domain was the small 16 kDa Collagen Adhesion Binding Domain (CnaB2) from the Fibronectin binding protein (FbaB) (Hagan et al., 2010; Oke et al., 2010). The CnaB2 domain was again split into two parts; a domain called SpyCatcher and a small 13-residue peptide tag called SpyTag (name coming from the 'S' and 'py' of *S. pyogenes)*. The basis of the spontaneous isopeptide formation is CnaB2's 'isopeptide triad', in which a reactive Lysine and Aspartate are positioned adjacent to a Glutamate which acts as an acid-base catalyst (Figure 3.1B). CnaB2 was split such that the reactive Lysine and catalytic Glutamate reside in the SpyCatcher domain and the reactive Aspartate is located on SpyTag (Figure 3.1A) (Hagan et al., 2010; Zakeri et al., 2012). The initial splitting, engineering, and partial characterization of the CnaB2 domain was done by Dr. Bijan Zakeri.

We then wanted to further characterize and apply the split protein system under a variety of conditions, to better understand its limits. The weak interactions of peptides (Houk et al., 2003; Stanfield and Wilson, 1995) limit their uses in a variety situations, especially inside and on the surface of cells, where the varying conditions and forces would make a covalent interaction more valuable. Thus SpyCatcher-SpyTag was tested to determine whether the reaction could work in assorted scenarios. The scenarios included *in vitro* testing of SpyCatcher-SpyTag reconstitution with and without reducing agent or in the cytosol of *E. coli*. Another scenario was on the surface of mammalian cells, as an alternative to antibody labeling. Finally thermodynamic characterization of SpyCatcher-SpyTag reconstitution was performed.

3.1 Precursor Development of SpyCatcher-SpyTag Interaction by Dr. Bijan Zakeri:

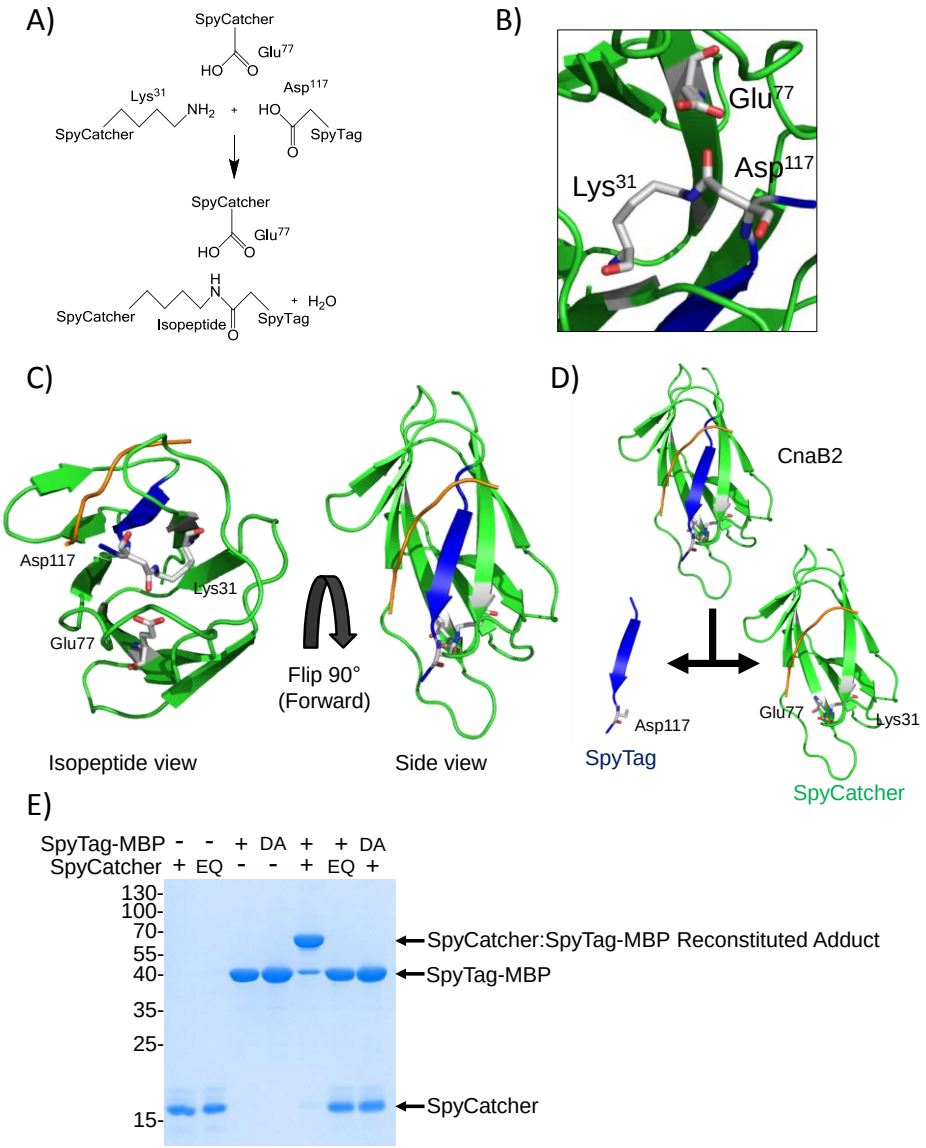


Figure 3.1: Covalent interaction of SpyCatcher-SpyTag. A) Chemistry of isopeptide bond formation of reactive Lysine with the reactive Aspartate catalyzed by Glutamate. Made in ChemBioDraw 12.0 B) Structure of Isopeptide bond from the CnaB2 structure (PDB 2X5P). C) CnaB2 structure overview, the blue colored β-strand represents SpyTag, green domain represents SpyCatcher, and orange is a small section of the N-terminal that is visible in the 2X5P PDB entry. D) Displaying how SpyTag (blue) with reactive Asp117 was split from the CnaB2 domain, SpyCatcher (green) with the reactive Lys31 and catalytic Glu77 E) Reconstitution of SpyCatcher with SpyTag-labeled MBP. The inactive SpyTag D→A mutant (SpyTagDA) and the catalytic Glutamate with the inactive E→Q (SpyCatcher EQ) mutant act as negative controls. Experimental work done by Dr. Bijan Zakeri and this figure is published in (Zakeri et al., 2012).

In Figure 3.1, the product of work done by Dr. Bijan Zakeri can be seen by the splitting of the CnaB2 domain (Figure 3.1D) and rational design which led to the

development of SpyCatcher and SpyTag. The gel shows (Figure 3.1E) that when MBP labeled with SpyTag (45 kDa) is mixed with SpyCatcher (15.5 kDa), a third band appears in the SDS-PAGE: a ~60 kDa product is formed by the reconstitution of SpyCatcher with SpyTag-MBP. Thus the system is shown to be functioning and capable of forming an interaction capable of withstanding boiling at 95°C with SDS. The interaction was later shown to be covalent with mass spectrometry (Zakeri et al., 2012). Figure 3.1 also shows that the reactive Aspartate located on the SpyTag and catalytic Glutamate on the SpyCatcher are crucial for the reconstitution (Figure 3.1C) (Zakeri et al., 2012).

3.2 SpyCatcher-SpyTag Interaction under Reducing Conditions:

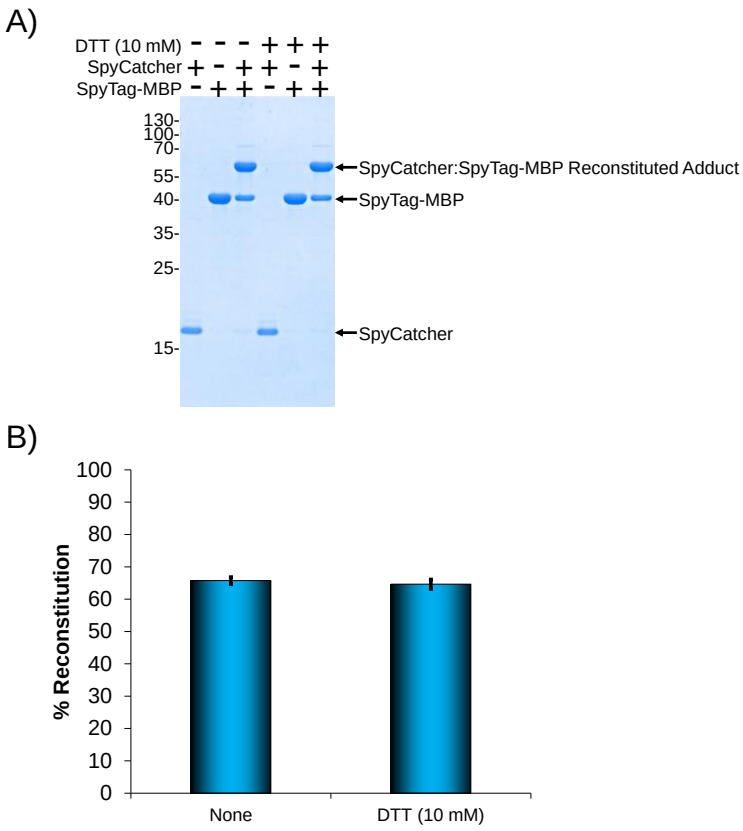


Figure 3.2: SpyCatcher-SpyTag-MBP reconstitution with and without 10 mM DTT. A) Illustrative gel of SpyCatcher-SpyTag-MBP reconstitution examining whether DTT influences the reaction. B) Quantification of the SpyCatcher-SpyTag-MBP reconstitution after 10 min incubation without DTT (None) and with 10 mM DTT (DTT). Values are an average of reactions in triplicate, ±1 standard deviation, and showing a non-significant P-value=0.35 in a two-tailed unpaired t-test. Note this figure is published in (Zakeri et al., 2012).

The CnaB2 domain does not have any Cysteine residues (Oke et al., 2010; Zakeri et al., 2012) but it was considered worth confirming whether a reducing agent would influence the reconstitution capacity between SpyCatcher and SpyTag. Figure 3.2A compares the reconstitution of SpyCatcher with SpyTag-MBP in 10 mM DTT, where there is no significant difference between the amount of reconstituted adduct formed with and without the reducing agent (Figure 3.2B).

3.3 Determining Whether SpyCatcher-SpyTag Reconstitution is Affected Based on SpyTag Location:

This section examined location-dependence of SpyTag interaction with SpyCatcher: whether SpyCatcher could reconstitute with SpyTag located directly on the N-terminus (following the fMet residue), C-terminus, or located on a linker between two protein domains.

3.3.1 Activity of SpyCatcher with SpyTag Located on the N-Terminus of a Protein:

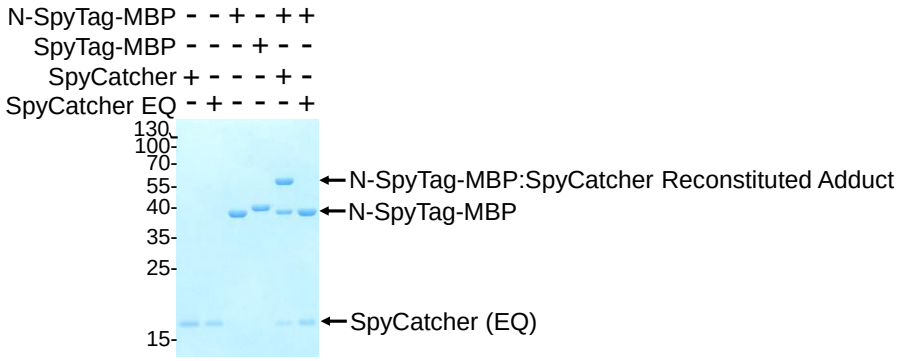


Figure 3.3: Testing SpyCatcher reconstitution with an MBP protein labeled with SpyTag directly on the N-terminus (N-SpyTag-MBP), i.e. the residue following the fMet. Note this figure is published in (Zakeri et al., 2012).

One consideration in the original SpyTag-MBP protein is that the SpyTag itself was located downstream from the N-terminus of the protein, following the His-Tag and thrombin site available from the pET28a vector (Novagen) (Figure 3.4A). Hence, the N-SpyTag-MBP protein was created such that the SpyTag is located directly following the fMet residue. Thus following the post-translational cleavage of fMet (Hirel et al., 1989),

the Alanine of SpyTag represents the initial residue of the N-Terminus. Figure 3.3 shows that SpyCatcher can interact and efficiently form a covalently reconstituted adduct with a SpyTag located on the N-terminus.

3.3.2 Activity of SpyCatcher with SpyTag at Differing Locations on the N-Terminus:

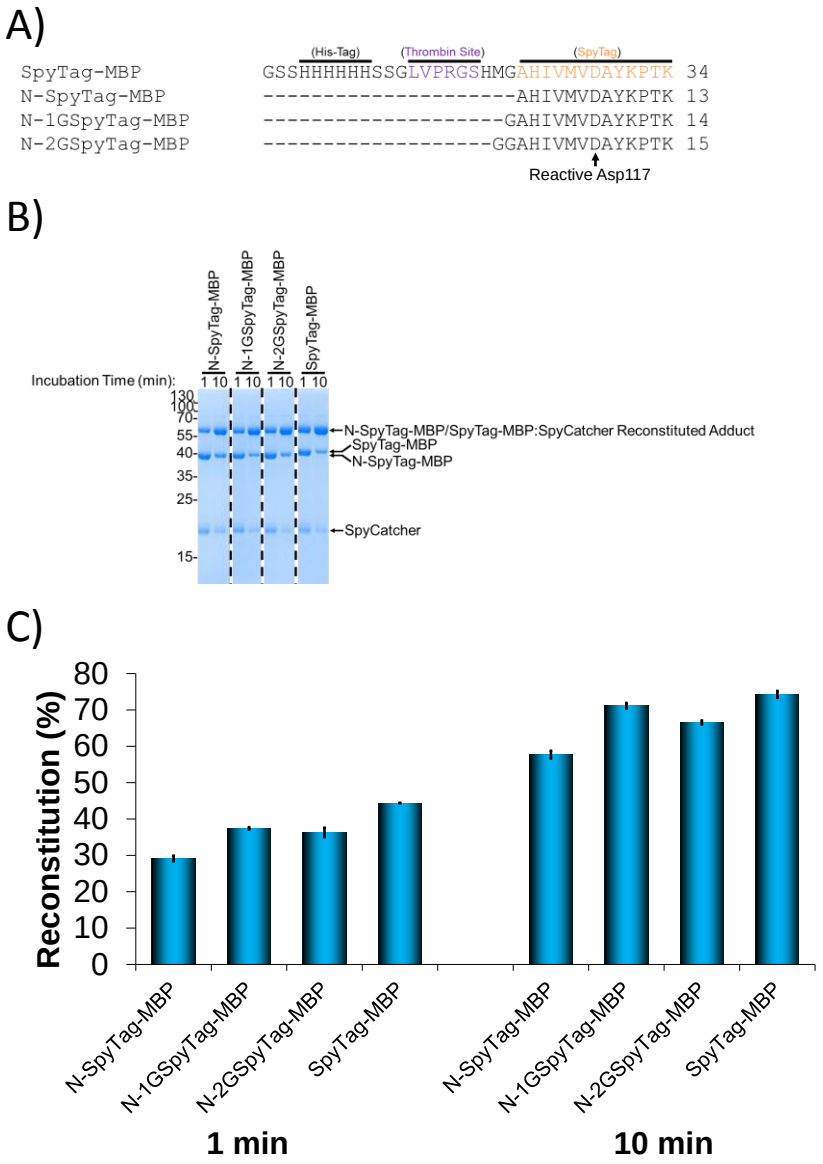


Figure 3.4: Examining the importance of location of SpyTag on the N-terminus. A) Sequence alignments of the N-termini of the four SpyTag variants tested. B) Demonstrative gel of the reconstitution done by incubation of SpyCatcher with each of the four N-Terminal SpyTag variants for 1 and 10 min. C) Quantification of the reconstitution of SpyCatcher with the N-terminal-SpyTag-MBP variants of (B), done under standard SpyCatcher-SpyTag reconstitution conditions. Values are averages of triplicate reactions, ±1 standard deviation.

Multiple types of N-terminal SpyTag-MBP variants were generated (Figure 3.4A), with the SpyTag directly on the N-terminus (N-SpyTag-MBP), a single Glycine preceding the SpyTag (N-1GSpyTag-MBP), two Glycines preceding the SpyTag on the N-terminus (N-2GSpyTag-MBP), and finally the standard version with the N-terminal leader section from the pET28a vector (SpyTag-MBP). The N-terminal variants were all shown functional (Figure 3.4B), though when the N-terminal variant rates were compared to the standard SpyTag (SpyTag-MBP), the N-terminal SpyTags were slower at both 1 and 10 min (Figure 3.4C). The significance of the differences in reaction rate between the different SpyTag locations is minor when there is at least one residue in front of SpyTag. The amount of N-terminal leader sequence or 'space' before SpyTag was shown not to be crucial, but did help improve activity when compared to N-SpyTag-MBP which had a ~10% drop in activity.

3.3.3 Activity of SpyCatcher with SpyTag in a Linker between Protein Domains and on the C-Terminus:

Figure 3.5: SpyCatcher reconstitution with MBP-SpyTag-Zif-SpyTag. A) The SpyTag fusion protein in this case contains two SpyTags, one in a linker between the MBP domain and Zif (Zinc finger) domain, and the second following the Zif domain on the C-terminus. B) The reactions were performed under standard reconstitution conditions except where denoted by '3x' which meant that 30 μM of SpyCatcher or SpyCatcher EQ was used rather than the standard 10 μM. Note this figure is published in (Zakeri et al., 2012).

Though SpyCatcher-SpyTag reconstitution can occur at the N-terminus, it was also important to characterize whether the system was more flexible allowing for SpyTag to be located at the C-terminus and also if possible in a linker between protein domains. The issue is that many tags tend to have preferences or simply for one reason or another must be located at a specific terminus (Brizzard, 2008; Fritze and Anderson, 2000; Lichty et al., 2005; Terpe, 2003). In Figure 3.5 a double SpyTag-labeled fusion protein was generated to test for SpyCatcher activity in different locations. In this case as the fusion name suggests, MBP-SpyTag-Zif-SpyTag, has one tag located on the C-terminal directly after the Zinc finger domain (Zif) (Christy and Nathans, 1989), and another located in a linker between MBP and the Zif domains (Fig 3.5A). The reconstitution reaction was done under two conditions (Fig 3.5B): 10 µM MBP-SpyTag-Zif-SpyTag, with either 10 or 30 µM of SpyCatcher. The 30 µM reaction of SpyCatcher saturated the SpyTag sites of MBP-SpyTag-Zif-SpyTag showing that both sites were accessible and could be reconstituted. The 10 µM SpyCatcher reaction shows an intermediate that is singly reacted with while the other SpyTag site is left vacant.

3.4 SpyCatcher-SpyTag Reconstitution within the Cytosol of *E. coli*:

The purpose of this section was to examine the robustness of the SpyCatcher-SpyTag system inside the cytosol of what is one of the most widely used model organisms, *E. coli*.

3.4.1 Examining SpyCatcher-SpyTag Reconstitution in the Cytosol of E. coli:

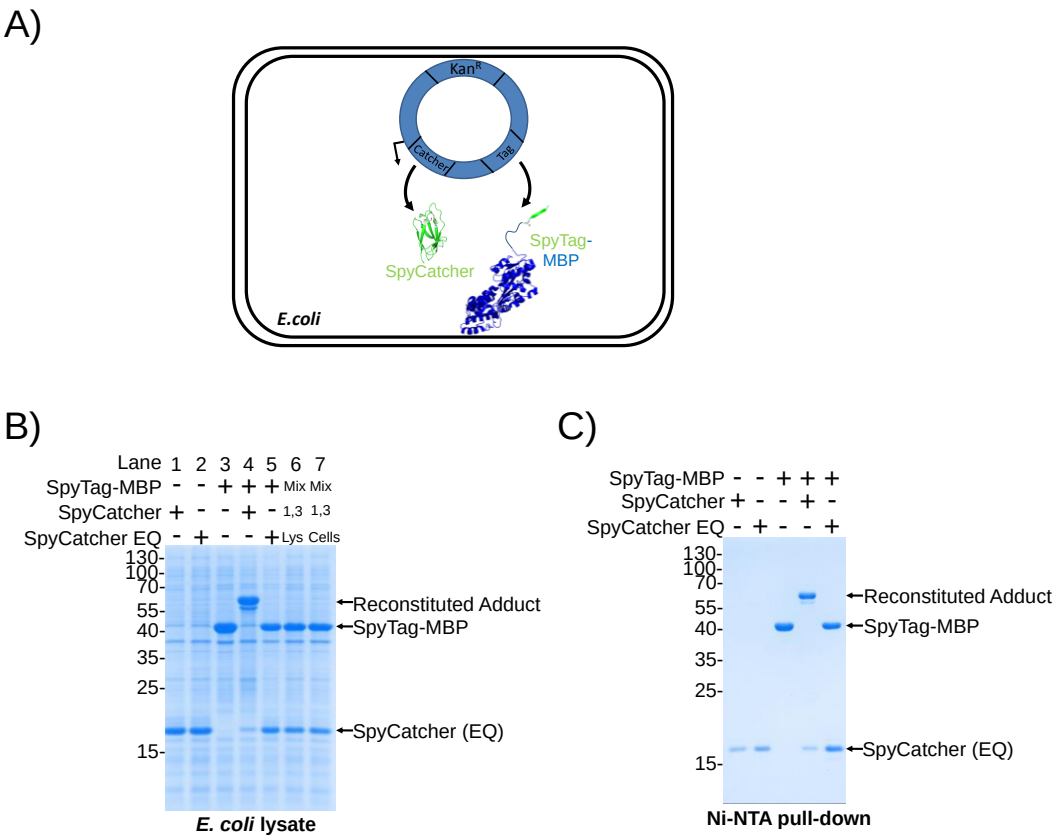


Figure 3.6: *In vivo* reconstitution of SpyCatcher-SpyTag-MBP in the cytosol of *E. coli*. A) Demonstrative image of bicistronic construct of SpyTag-MBP and SpyCatcher co-expressed within the cytosol of *E. coli* B) A bicistronic plasmid containing SpyCatcher and SpyTag-MBP was co-expressed inside BL21 DE3 RIPL cells, the cells were then lysed, boiled with SDS loading dye, and run on SDS-PAGE, with SpyCatcher EQ acting as the negative control. Further controls of lane 6 and 7 were performed to show that the reconstitution occurred inside the *E. coli* cytosol. In lane 6 lysate of SpyCatcher expressed alone and SpyTag-MBP expressed on its own (lanes 1 and 3, respectively) were boiled with SDS loading dye and mixed to ensure that the reaction did not occur following lysis of the cells. Lane 7 had the total cells of Lane 1 (SpyCatcher alone induction) and Lane 3 (SpyTag-MBP induction), mixed post over-expression, then SDS loading dye was added, and the cells were boiled. This control was to ensure that the proteins were not somehow exported or leaking out of the *E. coli* such that the reaction was occurring in the LB media surrounding the cells. C) The Ni-NTA purification of the co-overexpression of the SpyCatcher and SpyTag-MBP bicistronic plasmid. Note this figure is published in (Zakeri et al., 2012).

Checking whether SpyCatcher has any activity *in vivo*, or more explicitly within the cytosol of *E. coli*, was tested via a simple induction experiment. A bicistronic plasmid was constructed by modifying the original SpyTag-MBP/pET28a construct by inserting

SpyCatcher (inclusive of the ribosomal binding site) downstream of the SpyTag-MBP gene (Figure 3.6A). The bicistronic construct ensured that the transcription and translation of the two proteins were at an approximate 1:1 ratio, which allowed for noting whether there was any inhibition of reconstitution rather than just expression or copy number issues with two separate plasmids. *In vivo* reconstitution can be seen in Figure 3.6B: after induction with IPTG and overexpression for 4 h at 30°C, the SpyCatcher and SpyTag-MBP had reconstituted efficiently, with only minimal amount of either reactant being unreacted in the cytosol (lane 4). Reconstitution is shown more clearly by comparing the bicistronic expression of SpyCatcher with SpyCatcher EQ: the depletion of the unreacted SpyCatcher and SpyTag-MBP can be seen (Figure 3.6B).

Following the overexpression analysis of the bicistronic constructs, His-tag containing proteins were purified via Ni-NTA pull-down. The experiment (Figure 3.6C) served the function of determining reaction specificity, to see if SpyCatcher (EQ) or SpyTag-MBP had any promiscuous activity. At least from Coomassie staining (Figure 3.6C) the proteins were only seen to interact with their cognate partner. It should also be noted that the reaction could also still be occurring during the Ni-NTA purification, however that will require future experimentation to confirm SpyCatcher-SpyTag reconstitution in Ni-NTA buffer.

3.4.2 Silver and Immunostaining of SpyCatcher-SpyTag-MBP Reconstitution In Vivo for Enhanced Resolution of Promiscuous Activity:

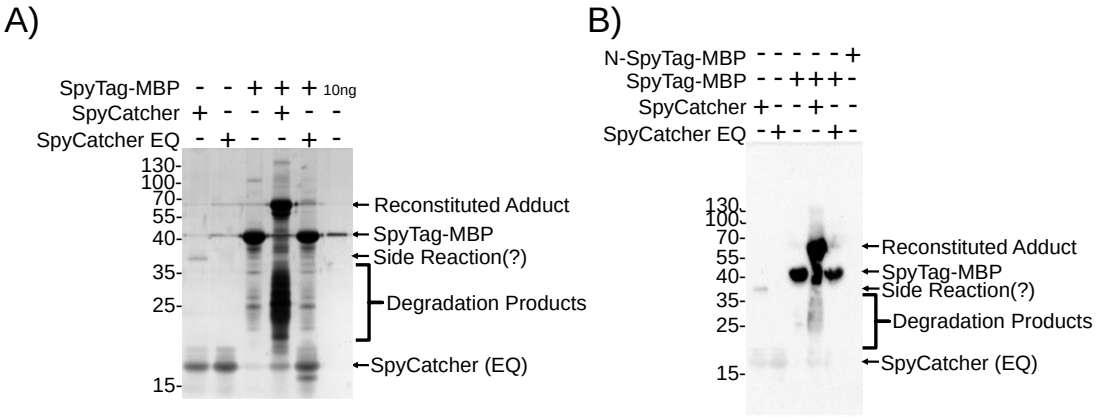


Figure 3.7: SpyCatcher-SpyTag-MBP reconstitution *in vivo* using a bicistronic construct and Ni-NTA purification, with protein analysis by silver staining and Western blot. A) Silver staining of the pulled-down *in vivo* SpyCatcher-SpyTag-MBP reconstitution. B) Anti-His western blot of the Ni-NTA pulled down SpyCatcher-SpyTag-MBP *in vivo* reconstitutions. Note this figure is published in (Zakeri et al., 2012).

Though from Figure 3.6B,C it could be taken that the SpyCatcher reconstitution with SpyTag-MBP within the cytosol of *E. coli* had no side reactions and that the proteins showed no stability issues, it was important to confirm whether this was indeed the case via two methods that have higher sensitivity for low protein concentrations, silver staining and immunostaining. The value of silver staining is that the stain can have 100-fold greater sensitivity compared to Coomassie stains, being able to easily detect ~10 ng of protein (Merril et al., 1988).

Figure 3.7A is the silver staining of the Ni-NTA purified *in vivo* reconstitution between SpyCatcher and SpyTag-MBP. The 10 ng control lane of SpyTag-MBP was a control to help gauge the sensitivity of the system. It should be noted that silver stain binding to proteins is variable, thus intensity cannot be taken as a linear estimation of concentration (Merril, 1990; Merril et al., 1988). It can be seen that in the control SpyCatcher-only lane, there is another band at ~37.5 kDa, possibly signifying a side reaction primarily due to the fact that the band does not appear in the SpyCatcher EQ only lane. Another aspect revealed by the silver staining is the large smear where the

SpyCatcher-SpyTag-MBP *in vivo* reconstitution occurred. Between the SpyTag-MBP and SpyCatcher bands there is a large smear of proteins, indicating degradation of the reconstituted adduct.

To confirm whether the band and smear were either coincidental co-purifications from Ni-NTA pull-down or actual side reactions and degradation products, Western blotting of the Ni-NTA pull-down was performed with anti-His antibody. It is seen that there is indeed a side reaction with SpyCatcher and degradation of the reconstituted adduct (Figure 3.7B). In the SpyCatcher-only lane the anti-His antibody is detecting the band at the same molecular weight (~37.5 kDa) as in the silver staining. The detection of this promiscuity is due to the covalent interaction between the SpyCatcher and the unknown protein. Other side reactions could be occurring but they are neither detectable via silver stain nor Western blotting. In the case of the degradation of the reconstituted adduct the anti-His antibody is detecting the smear visible in the silver staining, indicating that indeed the reconstituted adduct is becoming degraded. However, the majority of the covalently locked proteins are still un-degraded, as shown by the intensity of the reconstituted adduct band (Figure 3.7B).

3.5 SpyCatcher-SpyTag Reconstitution with Membrane Proteins on the Surface of HeLa Cells:

For the characterization of SpyCatcher-SpyTag as a molecular tool, it is important to examine its functionality in other model systems. In this case SpyCatcher and SpyTag were examined as tools for mammalian cell surface labeling.

3.5.1 Visualization of SpyCatcher-555 Reconstitution with a SpyTag-Membrane Protein Fusion on the Surface of HeLa Cells:

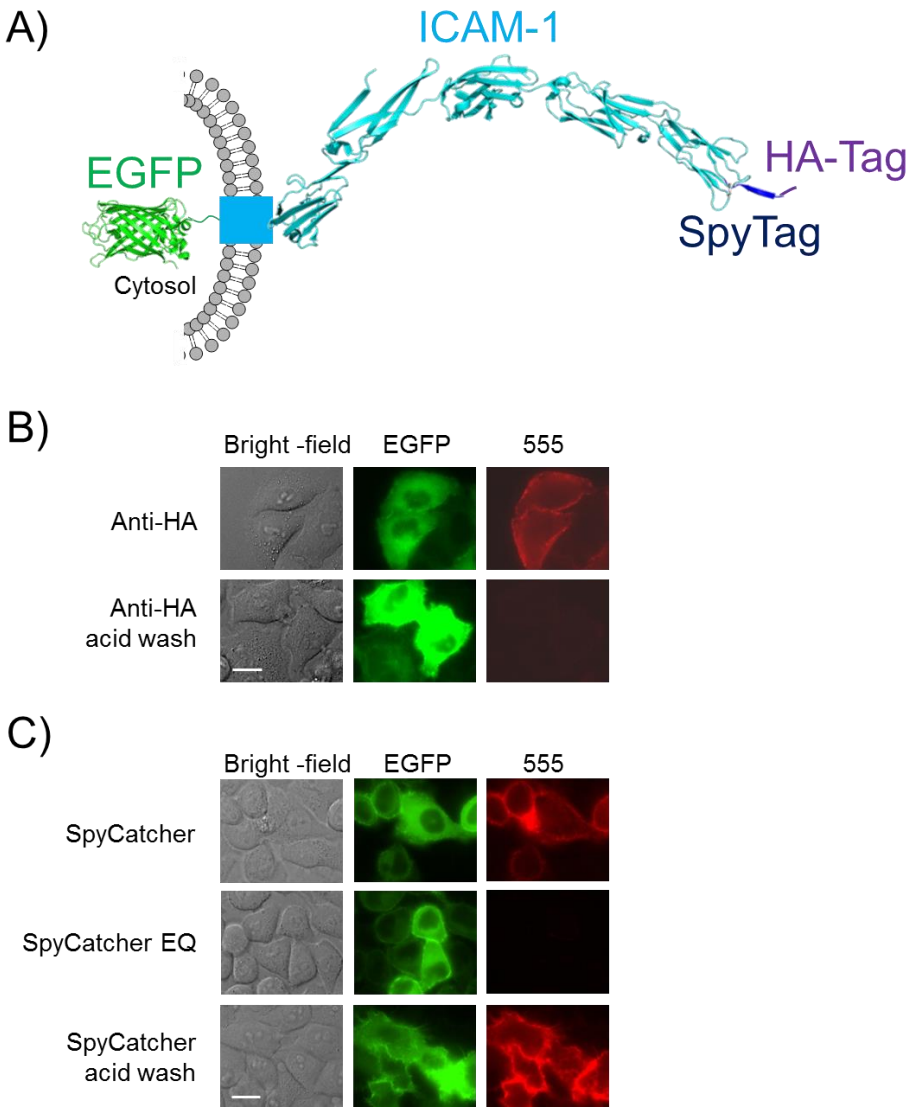


Figure 3.8: SpyCatcher reconstitution on the surface of HeLa cells. A) Cartoon of the fusion ICAM-1 cell membrane protein used in the experimental work contained both HA-Tag and SpyTag on the N-Terminus and EGFP fused to the C-terminus. The HA-SpyTag-ICAM1 is exposed extracellularly, while the EGFP is in the cytosol. Cartoon adapted from (Zakeri et al., 2012). B) Detecting expression and localization of the HA-SpyTag-ICAM1-EGFP fusion on the cell surface. Cells were incubated with anti-HA antibody and visualized with secondary anti-rabbit Alexa Fluor-568 conjugated antibody, following cells were incubated in an acidic buffer. C) SpyCatcher (EQ) conjugated with Alexa Fluor-555 was incubated with the HA-SpyTag-ICAM1-EGFP expressing HeLa cells. The cells incubated with the active SpyCatcher-555 were also acid-washed while cells labeled with SpyCatcher EQ were simply washed with PBS. Scale bar 20 µM. Note: Experimental cell culture and microscopy work done by Miss Emily Chittock. ICAM-1 fusion protein cloned by Dr. Bijan Zakeri. This figure is published in (Zakeri et al., 2012). Demonstrative model (A) composed in PyMOL using PDB structures 1IC1 & 1P53 (ICAM-1), 2Y0G (EGFP), and 2X5P (SpyTag).

Further examining the robustness of SpyCatcher-SpyTag was to test whether the system could function in the complex environment of DMEM media and the surface of a mammalian cell, in this case HeLa. A construct was made based on the human intercellular adhesion molecule-1 (ICAM-1) (Yang et al., 2004). ICAM-1 was fused with EGFP to visualize the protein and also labeled with HA-tag (a section of the human influenza hemagglutinin glycoprotein (Field et al., 1988) and SpyTag (Figure 3.8A). The ICAM-1 fusion was inserted into the PINCO vector, to allow for episomal expression in HeLa cells (Grignani et al., 1998).

The cells were initially stained with anti-HA antibody to determine that the cells were indeed expressing the HA-SpyTag-ICAM1-EGFP fusion protein at the cell-surface. The cells were then incubated with and without an acid stripping buffer (Figure 3.8B), shown previously to break a wide-range of non-covalent interactions on the cell surface (Sugawara et al., 1987). The acid stripping buffer was shown to be effective at removing the high affinity interaction between antibody and HA tag. The experiment was then repeated using SpyCatcher (EQ)-555 (SpyCatcher NHS conjugated with Alexa-555 dye) (Figure 3.8C). 5 µM SpyCatcher (EQ)-555 was incubated 15 min with the cells and initial microscopy determined that the cells were indeed specifically labeled by only the active SpyCatcher-555 after a PBS washing (Figure 3.8C, bottom right and middle right panels). It was important to determine whether the interaction was actually the covalent reconstitution and not just a very strong non-covalent interaction between the ICAM-1 fusion and the SpyCatcher and so the cells were acid-washed again (bottom panel of Figure 3.8C). Co-localization between HA-SpyTag-ICAM1-EGFP and SpyCatcher-555 is maintained, consistent with a covalent interaction.

3.5.2 Immunostaining Confirmation of Covalent SpyCatcher Reconstitution with HA-SpyTag-ICAM-EGFP:

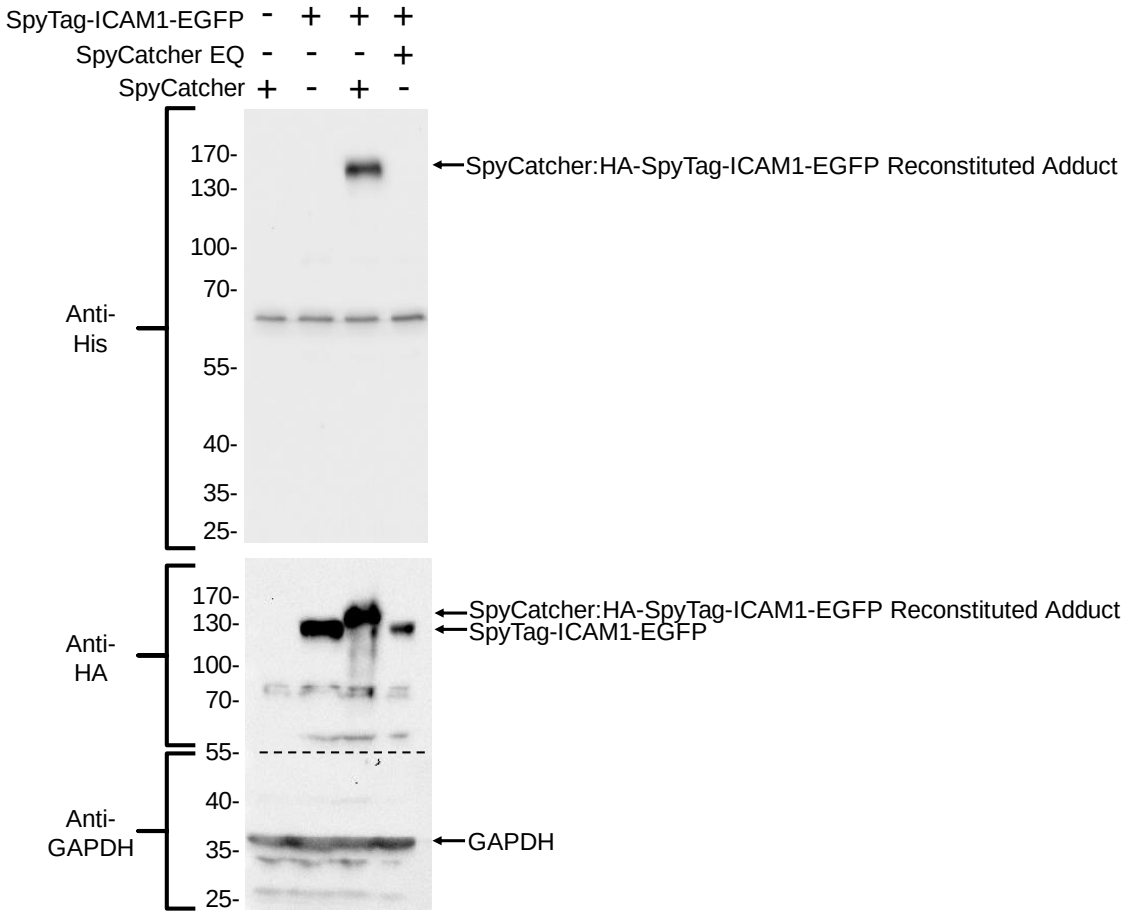


Figure 3.9: Determining SpyCatcher covalent reconstitution on the surface of HeLa cells via Western blot. HeLa cells expressing HA-SpyTag-ICAM1-EGFP were reacted with SpyCatcher (EQ) (last lanes on the right) and then the cells were boiled with SDS loading dye to stop the reaction. As controls, HeLa cells transfected with the HA-SpyTag-ICAM1-EGFP were incubated with just PBS (no SpyCatcher, second lane from the left) and non-transfected regular HeLa cells were incubated with SpyCatcher to test for promiscuity (left lane). The top panel shows the anti-His immunostaining, to detect SpyCatcher (EQ) covalent interactions. The lower set of bands seen in every lane is an endogenous protein recognized by the anti-His antibody (Salichs et al., 2009). In the middle panel, an anti-HA antibody was used to detect the HA-SpyTag-ICAM1-EGFP fusion. Bottom panel anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) staining was performed as a loading control. Note this figure is published in (Zakeri et al., 2012).

To confirm the acid wash result of Section 3.5.1, the cells were incubated for 15 min with 5 μM of SpyCatcher (EQ) and it was observed whether the interaction could survive boiling in SDS. In Figure 3.9, immunostaining of the reaction was done from two

directions, the SpyCatcher-side and the ICAM-1 fusion-side, to confirm the covalent reconstitution. From the SpyCatcher-side, the anti-His antibody was used to detect the SpyCatcher. As can be seen from the top panel, only a solitary band exists when active SpyCatcher is incubated with the HeLa cells expressing the HA-SpyTag-ICAM1-EGFP fusion. As an extra verification, immunostaining was approached from the HA-SpyTag-ICAM1-EGFP-side, where the anti-HA antibody detected the ICAM-1 fusion protein. In the middle panel the addition of SpyCatcher has led to a band shift compared to the HA-SpyTag-ICAM1-EGFP fusion alone or incubation with SpyCatcher EQ. The GAPDH (glyceraldehyde-3-phosphate dehydrogenase) staining acted as a loading control.

3.5.3 Testing Membrane Protein Fused with SpyCatcher for Cell Labeling:

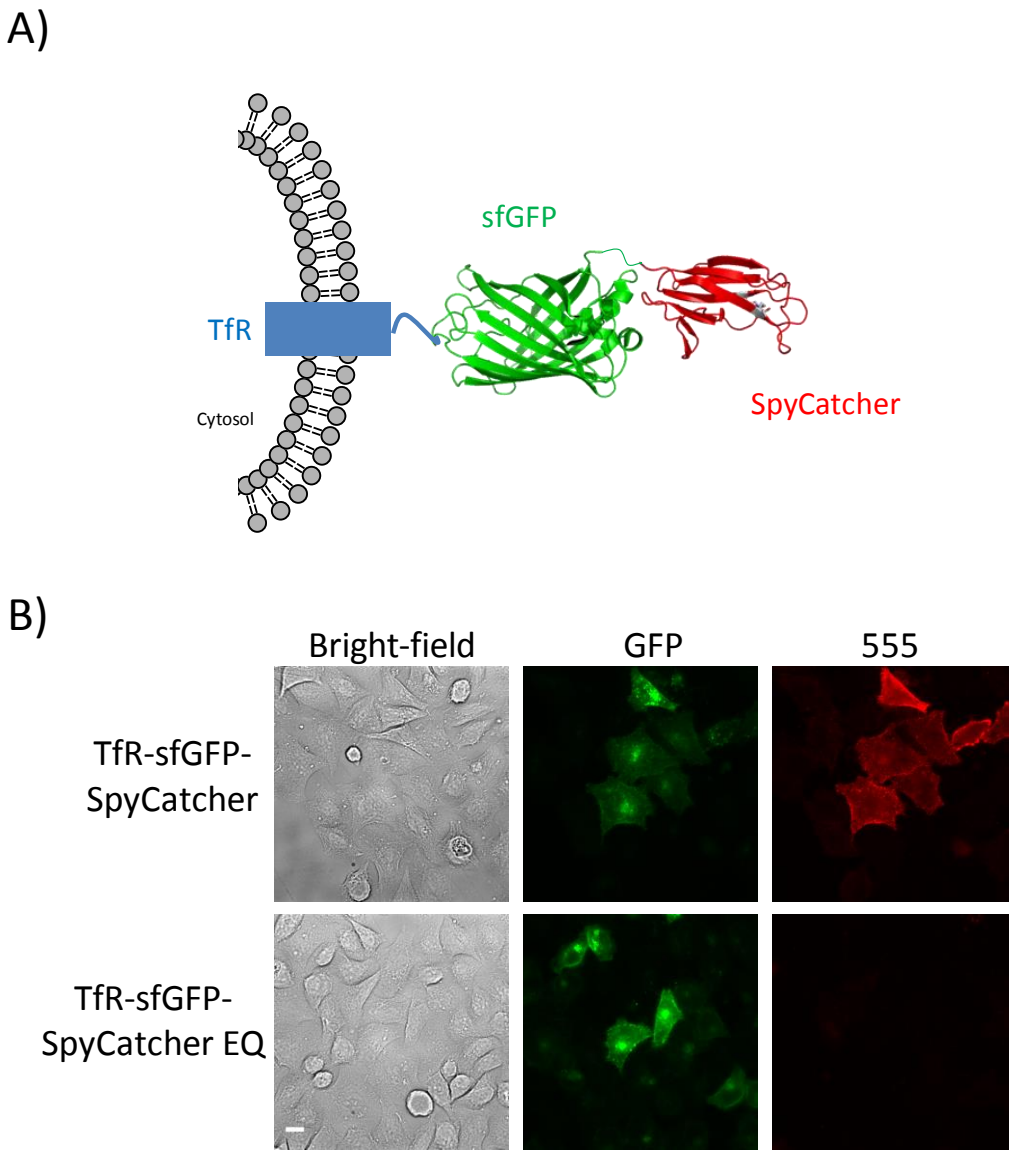


Figure 3.10: Testing that SpyCatcher fused to a membrane protein can covalently reconstitute with SpyTag-MBP at the surface of HeLa cells. A) Cartoon of SpyCatcher (EQ) transmembrane fusion protein, TfR-sfGFP-SpyCatcher. B) HeLa cells expressing TfR-sfGFP-SpyCatcher (EQ). The cells were incubated with SpyTag-MBP. Then the cells were washed twice with PBS and then incubated with Anti-His antibody followed by anti-mouse Alexa Fluor-555 conjugated antibody. Scale bar represents 100 µm. Structure in (A) composed of PDB files 2Q6P (super folder (sf) GFP) and 2X5P (SpyCatcher).

Figure 3.10 tests the reverse of Figure 3.8, in determining whether it is feasible to fuse SpyCatcher onto a membrane protein and use a SpyTagged protein for cell labeling. The SpyCatcher construct in this case was quite different from the HA-SpyTag-

ICAM1-EGFP, in that the SpyCatcher construct required a different scaffold. In this case the Transferrin receptor transmembrane domain (TfR) was chosen, due to two factors. The first is that the N-terminus is facing towards the cytosol and the C-terminus extracellularly (Rutledge et al., 1994), allowing for the linker to be fused to the N-terminus of the SpyCatcher leaving the reactive cleft at the C-terminus un-obstructed (Zakeri et al., 2012) (Figure 3.10A). The second factor is that a double mutation (Y20C and F23A) in TfR reduces endocytosis, increasing the fraction of the protein on the surface of a cell (Rutledge et al., 1994). The TfR was also fused with super-folder (sf) GFP, to aid visualization and to help extend SpyCatcher away from the cell surface allowing for more opportunities for interactions with a SpyTagged protein (Figure 3.10A).

The cells were then incubated with 10 μM of SpyTag-MBP for 15 min and immunostained with anti-His antibody (Figure 3.10B). There is only co-localization of Alexa Fluor-555 with the active TfR-sfGFP-SpyCatcher but not with TfR-sfGFP-SpyCatcher EQ, indicating specific recognition.

3.6 Thermodynamics and Dissociation Constant Determination of SpyCatcher Reconstitution with Reactive and Inactive SpyTag(DA)-MBP:

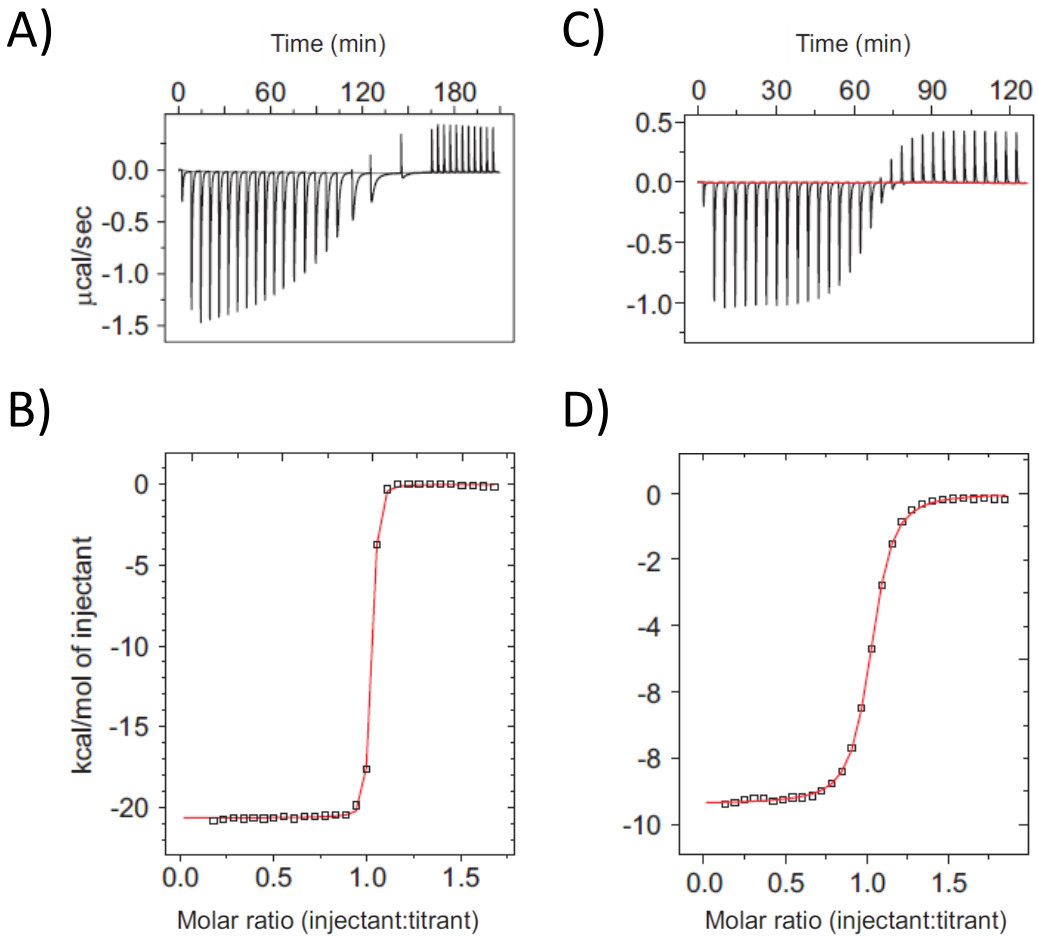


Figure 3.11: Isothermal titration calorimetry (ITC) of SpyCatcher reconstitution with SpyTag-MBP and inactive SpyTagDA-MBP. A) & C) ITC profile of the injection of reactive SpyTag-MBP (A) and inactive SpyTagDA-MBP (C) into SpyCatcher. B) & D) The red line represents the best fit of the incremental enthalpy changes (squares). The values obtained from fitting were (B) ΔH= -20.2 kcal/mol for active SpyTag-MBP with SpyCatcher, and (D) K_d= 0.2 µM, ΔH= -9.4 kcal/mol, ΔS= -0.9 cal/(mol K) for inactive SpyTagDA-MBP with SpyCatcher. The stoichiometry was measured to be near 1.0. Note: Protein expression, purification, and dialysis was performed myself; ITC experimentation, data analysis, and graph creation by Dr. Ulrich Schwarz-Linek. Part of this figure is published in (Zakeri et al., 2012).

Isothermal titration calorimetry (ITC) allowed for the determination of the thermodynamics and K_d between SpyCatcher and SpyTag(DA)-MBP. Due to the covalent nature of the interaction, the information from the ITC between SpyCatcher and

SpyTag-MBP is limited to the thermodynamics, due to the single turnover nature of the interaction. From fitting of the enthalpic changes SpyCatcher-SpyTag-MBP interaction was shown to have a ΔH= -20.2 kcal/mol (Figure 3.11B). However, this information becomes more interesting when comparing it to the non-covalent interaction between the inactive SpyTagDA-MBP and SpyCatcher which had ΔH= -9.4 kcal/mol (Figure 3.11D).

Furthermore the ITC measurement of inactive SpyTagDA-MBP interaction with SpyCatcher allowed for determination of the K_d which was 0.2 μM. The affinity represented here is in the same order of magnitude of another (non-covalent) split protein system, S-Protein/S-Tag with a K_d of 0.1 μM (Terpe, 2003) (Table 1.1).

3.7 Discussion

The initial splitting and reconstitution of a protein into two parts was an interesting development in terms of chemical biology (Zakeri et al., 2012). The basis of this development was the isopeptide forming 'isopeptide triad' of Lysine, Aspartate, and (catalytic) Glutamate (Figure 3.1A, B). Those three main residues could then be put onto separate components, Lysine and Glutamate on a protein domain and Aspartate on a peptide-tag. Then two components, protein and tag, could be mixed together and in short order reconstitute into a covalent complex (Figure 3.1C) (Zakeri et al., 2012).

The characterization here examined the limits of the SpyCatcher-SpyTag system in three areas. First *in vitro* work focused on the influence of the reaction conditions and location of SpyTag. The second area was examining the SpyCatcher-SpyTag interaction *in vivo*, by testing the system in the cytosol of *E. coli* and on the surface of HeLa cells. Third was the thermodynamic characterization of the interaction energies and dissociation constants of SpyCatcher interacting with SpyTag.

The initial aspect tested *in vitro* was to determine whether the SpyCatcher-SpyTag interaction could function in the presence of a reducing agent (Section 3.2).

Though it may be taken as a foregone conclusion because the CnaB2 structure does not have a single Cysteine residue (Oke et al., 2010), it was important to ensure for future uses of SpyCatcher. Thus SpyCatcher-SpyTag can be used both in oxidizing cellular compartments, such as in the endoplasmic reticulum in mammalian cells or the periplasmic space in Gram-negative bacteria (Sevier and Kaiser, 2002), or in reducing cytosolic/nuclear environments (Cumming et al., 2004). This flexibility contrasts with epitope-tags targeted by antibodies, where antibodies may not function or fold well under reducing conditions because of their multiple disulfide bonds (Woof and Burton, 2004).

The next aspect was observing the influence of SpyTag location on reconstitution (Section 3.3). SpyCatcher showed the capacity to reconstitute with SpyTag directly located on the N-terminus (following the fMet) of MBP (Figure 3.3), and with varying amounts of peptide preceding SpyTag in the N-terminal region (Figure 3.4A & B). It was also determined that the location does have an influence on the reconstitution rate of SpyCatcher with SpyTag in the varying parts of the N-terminus of MBP (Figure 3.4C). Reconstitution generally improved with a longer stretch of peptide preceding the SpyTag (Figure 3.4C) when comparing SpyTag directly on the N-terminus (N-SpyTag-MBP) and SpyTag located downstream of a His-tag and thrombin site (SpyTag-MBP) (Figure 3.4A). However, one should not generalize the results determined in Figure 3.4C for every protein available. These data signify that some refinement of the SpyTag location may be required for different proteins to achieve the best possible reconstitution rate.

I then checked whether SpyTag could function in between protein domains and at the C-terminus (Figure 3.5). SpyCatcher can interact with SpyTag at both locations, as when putting SpyCatcher in excess, the majority of the doubly-SpyTagged protein (MBP-SpyTag-Zif-SpyTag) was consumed and formed a double-covalent adduct. These results shows that the SpyTag system is flexible in terms of tag location, in contrast to some other epitopes which are rigid in terms of location requirements (Terpe, 2003).

The second area I investigated was *in vivo* labeling, where there are high numbers of alternative protein contacts to check for promiscuity. Reconstitution was checked inside the cytosol of *E. coli*. The reaction inside the cytosol showed that almost all of the SpyCatcher and SpyTag-MBP had reacted to form the reconstituted adducts (Figure 3.6B). This was followed up by a Ni-NTA pull-down (Figure 3.6C) of each induction. The purpose was to detect for any secondary reactions by observing for unexpected masses with either SpyCatcher or SpyTag-MBP. Figure 3.6C looked clean, though Coomassie staining has its limitations. Therefore the far more sensitive method of silver staining (Merril et al., 1988) was used to see any bands below the detection limit of Coomassie. Figure 3.7A reveals two parts that were not visible in Figure 3.6C. In the SpyCatcher-only lane a faint band is visible at ~37.5 kDa from silver staining, which is not seen in the inactive control (SpyCatcher EQ). In the lane where SpyCatcher and SpyTag-MBP were co-expressed and reconstituted together there is a large series of bands between the sizes of ~20-36 kDa. To confirm that these are indeed SpyCatcher or SpyTag-MBP the same samples were immunoblotted with anti-His antibody to detect the His-tags used to purify them (Figure 3.7B). The band at ~37.5 kDa is still visible in the SpyCatcher-only control lane, signifying that indeed there is at least one promiscuous interaction. The immunoblotting (Figure 3.7B) does also confirm that the series of bands between ~20-36 kDa are also His-tagged, most likely representing a series of degradation products. Though the promiscuous band and degradation smear may be considered a negative feature in both cases, the results themselves are interesting for a few reasons. In the case of the promiscuous band of the SpyCatcher control lane, the band is quite weak with either silver staining or immunoblotting and is the only side-reaction visible (Figure 3.7). The side-reaction is of interest because the promiscuous band can become a possible basis for evolution of new SpyCatcher targets. The degradation products are also interesting because this is one of the first cases of

creating an intermolecular isopeptide bond within *E. coli.* It is interesting to observe how *E. coli* deal with unfamiliar protein products and its slow degradation of a protein chemistry that it is believed not to be normally encountered.

The next test, in an environment ripe for promiscuity, was the use of SpyCatcher-SpyTag for cell-surface labeling. The surface of the cell is dynamic, with a large variety of different proteins (McDonald et al., 2009). Cell surface labeling is also a common technique and it would be interesting to use SpyTag to overcome the limitations of antibodies. Thus a SpyTagged membrane fusion protein was generated, HA-SpyTag-ICAM1-EGFP (Figure 3.8A). The SpyTag-ICAM1 fusion expressed and localized to the surface of HeLa cells and enabled specific and stable labeling (Figure 3.8B, C and Figure 3.9) (Zakeri et al., 2012). The significance is that it shows SpyCatcher-SpyTag could be an alternative system to antibodies and epitope tags. However, there are still limitations: primarily that our approach requires that the protein to be labeled is a recombinantly generated one, unlike for antibodies which are available for a variety of endogenous targets (Leenaars and Hendriksen, 2005).

The testing of SpyCatcher in the cell biological arena continued with what can be considered the reverse experiment of Section 3.5.1 with a SpyTagged protein, in that a membrane fusion protein with SpyCatcher was generated (Figure 3.10A). The SpyCatcher membrane fusion protein (TfR-sfGFP-SpyCatcher) was shown to be functional in that the TfR-sfGFP-SpyCatcher localized to the cell surface and was able to reconstitute with SpyTag-MBP (Figure 3.10B). The SpyCatcher fusion could traffic through the HeLa cell secretory pathway (Ellgaard, 1999) and was still functional on the cell surface (Figure 3.10B). The other value is that the system is flexible in that either component, SpyCatcher or SpyTag, can be used for labeling cells, depending on the needs of the researcher. It would also be of interest to use both SpyCatcher and SpyTag cell surface labels on two separate cell types to induce or lock cell interactions, to create

a new type of cell:cell synapse in contrast to the network of reversible interactions that usually enable the precise regulation of cell-cell interaction and signaling (van der Merwe and Dushek, 2011).

The final characterization was the determination of the interaction energies of SpyCatcher with SpyTag-MBP or the inactive SpyTagDA-MBP (Figure 3.11). The ITC analysis of this interaction is very unusual in the case of SpyCatcher with the active SpyTag-MBP, where the interaction is a single turn-over binding event, which obviously makes determination of a dissociation constant not feasible (Figure 3.11A, B). Thus it required the use of the inactive SpyTagDA-MBP, which showed that the dissociation constant is ~200 nM. This affinity is similar to other optimized peptide-protein interactions (Houk et al., 2003; Korndörfer and Skerra, 2002; Terpe, 2003). The reason this is interesting is that, unlike many other protein-peptide interactions, the domain and peptide were not evolved to develop this intermolecular interaction; this affinity simply results from an artificial process of splitting the CnaB2 domain into two components. What is also interesting is the importance of the isopeptide bond formation for the thermodynamic stability of the reconstitution, comparing the enthalpies of the SpyCatcher with SpyTag-MBP the ΔH= -20.2 kcal/mol (Figure 3.11B) while with SpyTagDA-MBP ΔH= -9.4 kcal/mol (Figure 3.11D). However, further experiments will be required to determine what fraction of this enthalpy difference derives from initial docking versus isopeptide bond formation.

Chapter 4: The Use of Crystal Structures and Mutational Analysis to Understand and Improve the SpyCatcher-SpyTag Interaction

Sections of this chapter have been published in (Li et al., 2013).

Chapter 4: The Use of Crystal Structures and Mutational Analysis to Understand and Improve the SpyCatcher-SpyTag Interaction

The process of understanding and improving the SpyCatcher-SpyTag interaction was initially based on the amino acid sequence and the crystal structure of CnaB2 from Oke et al., 2010, (PDB 2X5P). My work in this chapter started by trying to improve SpyCatcher by creating different N-terminal truncations; the N-terminal region initially truncated did not give clear electron density in the CnaB2 structure, so I tested how crucial was this N-terminal region for SpyCatcher (Figure 4.1A). Then a rational design campaign of SpyCatcher was pursued, modifying surface exposed hydrophobic residues on SpyCatcher and the C-terminal residue of SpyCatcher which faces the cleft where SpyTag docks (Figure 4.1B).

SpyTag was also mutated, based on the CnaB2 structure, using a two pronged strategy. First I examined the importance of Lysine residues in SpyTag (Figure 4.5A). Second I tested circular permutation (Yu and Lutz, 2010), where the first N-terminal β-strand visible in the CnaB2 structure next to the SpyTag β-strand was fused to the C-terminus of SpyTag (Oke et al., 2010) (Figure 4.5B). The intent was that the circular permutation of SpyTag would provide extra docking contacts with SpyCatcher.

In collaboration with the Rapoport group at Harvard, a crystal structure of the SpyCatcherΔN1 domain I generated was solved after reconstitution with SpyTag (PDB 4MLS). Based on these crystal structures I generated further mutants of SpyCatcherΔN1 and SpyTag to exploit these structural insights.

4.1 Mutations of SpyCatcher Based on the CnaB2 Structure:

4.1.1 Overview of the Varying SpyCatcher Mutants Based on the CnaB2 Primary and Tertiary Structure:

A)

(His-Tag) (Linker) (TEV Site)

```
CnaB2          MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDI 60
SpyCatcher     MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDE 60
SpyCatcherΔN1  MSYYHHHHHHDYDIPTTENLYFQG----------------------SGDSATHIKFSKRDE 39
SpyCatcherΔN2  MSYYHHHHHHDYDIPTTENLYFQG---------SSEQGQSGDMTIEEDSATHIKFSKRDE 51
SpyCatcherΔN3  MSYYHHHHHH-----------------------------------SGDSATHIKFSKRDE 25

CnaB2          DGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTV 120
SpyCatcher     DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 120
SpyCatcherΔN1  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 99
SpyCatcherΔN2  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 111
SpyCatcherΔN3  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 85

CnaB2          NEQGQVTVNGKATKGDAHIVMVDA 144
SpyCatcher     NEQGQVTVNGKATKGDAHI----- 139
SpyCatcherΔN1  NEQGQVTVNGKATKGDAHI----- 118
SpyCatcherΔN2  NEQGQVTVNGKATKGDAHI----- 130
SpyCatcherΔN3  NEQGQVTVNGKATKGDAHI----- 104
```

Underlined residues are visible in the CnaB2 crystal structure (PDB 2X5P)

B)

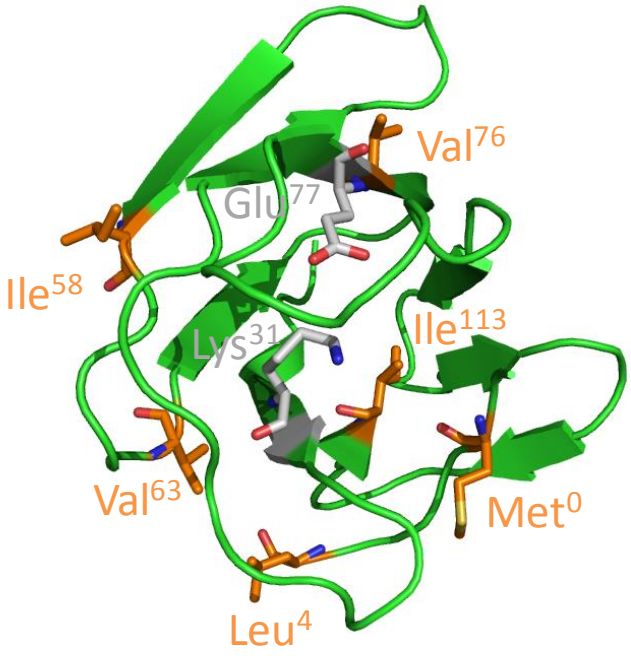


Figure 4.1: Mutants of SpyCatcher based upon the sequence of SpyCatcher and the crystal structure of CnaB2. A) Sequence alignment of CnaB2 and SpyCatcher with three N-terminal truncation variants. Residues underlined in the CnaB2 sequence are visible in the CnaB2 crystal structure (PDB 2X5P) B) Cartoon of SpyCatcher highlighting residues selected for mutation in orange, with reactive Lys31 and catalytic Glu77 in silver. Each of the mutants colored in orange, with the exception of Ile113, were exposed hydrophobics (based on PDB 2X5P).

N-terminal truncation mutants were first made to determine whether the regions recombinantly included from the plasmid (pDEST14), such as a linker region and TEV protease cleavage site, affected SpyTag reconstitution (Figure 4.1A). The second reason was that much of the N-terminal region is not visible in the crystal structure of CnaB2 (Hagan et al., 2010; Oke et al., 2010). The only part of the N-terminal region of CnaB2 visible in the crystal structure is the sequence GAMVDTL, which starts from the last residue of the TEV site. The N-terminal region of CnaB2 became an interest as to whether the inherent flexibility influences the structural stability of SpyCatcher.

The second set of mutants were designed by looking for exposed hydrophobic residues, with the goal of improving SpyCatcher solubility and cellular application. Thus each surface hydrophobic residue was mutated individually with a charged or polar residue and then multiple mutants of SpyCatcher were combined.

4.1.2 Rates of SpyTag Reaction of N-Terminal Truncated SpyCatcher Variants:

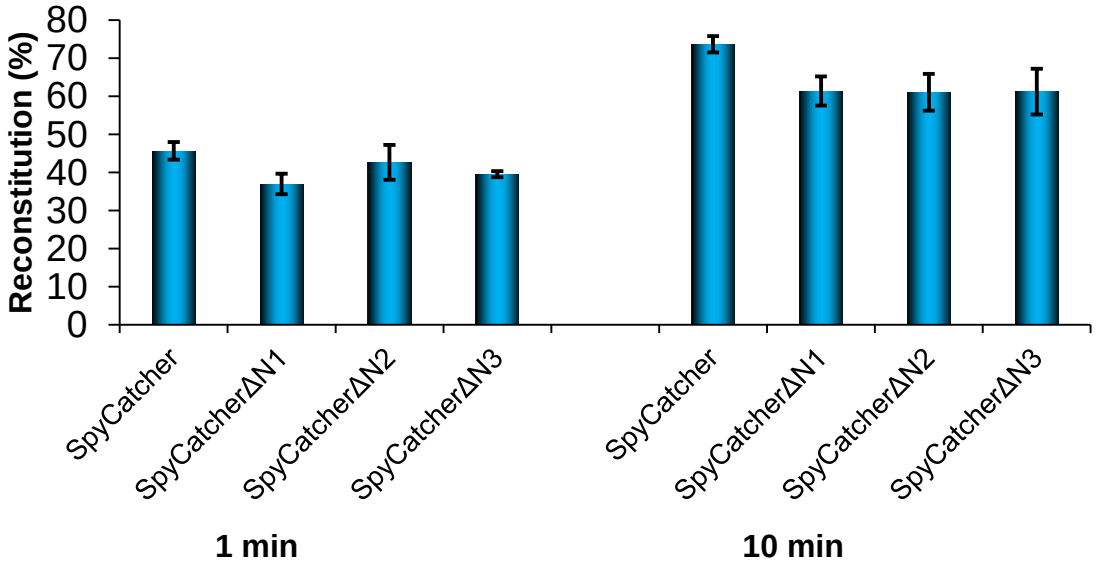


Figure 4.2: Determining the reconstitution rate of SpyCatcher with SpyTag-MBP compared to three N-Terminal SpyCatcher variants (SpyCatcherΔN1, 2, and 3 (Figure 4.1A)). Reactions were analyzed by SDS-PAGE with Coomassie staining. All reactions were performed in triplicate and error bars represent ± 1 standard deviation.

Comparison of the rates of the different N-Terminal truncation mutants of SpyCatcher (Figure 4.1A) can be seen in Figure 4.2, at 1 and 10 min. SpyCatcher was still superior relative to the N-terminal truncation mutants. Though the N-Terminal region of CnaB2/SpyCatcher is not seen in the crystal structure of CnaB2 (Oke et al., 2010), the N-terminal region is playing a role in helping the reconstitution between SpyCatcher and SpyTag.

4.1.3 Mutations at Ile113 in SpyCatcher to Determine Influence on the Reconstitution Rate with SpyTag:

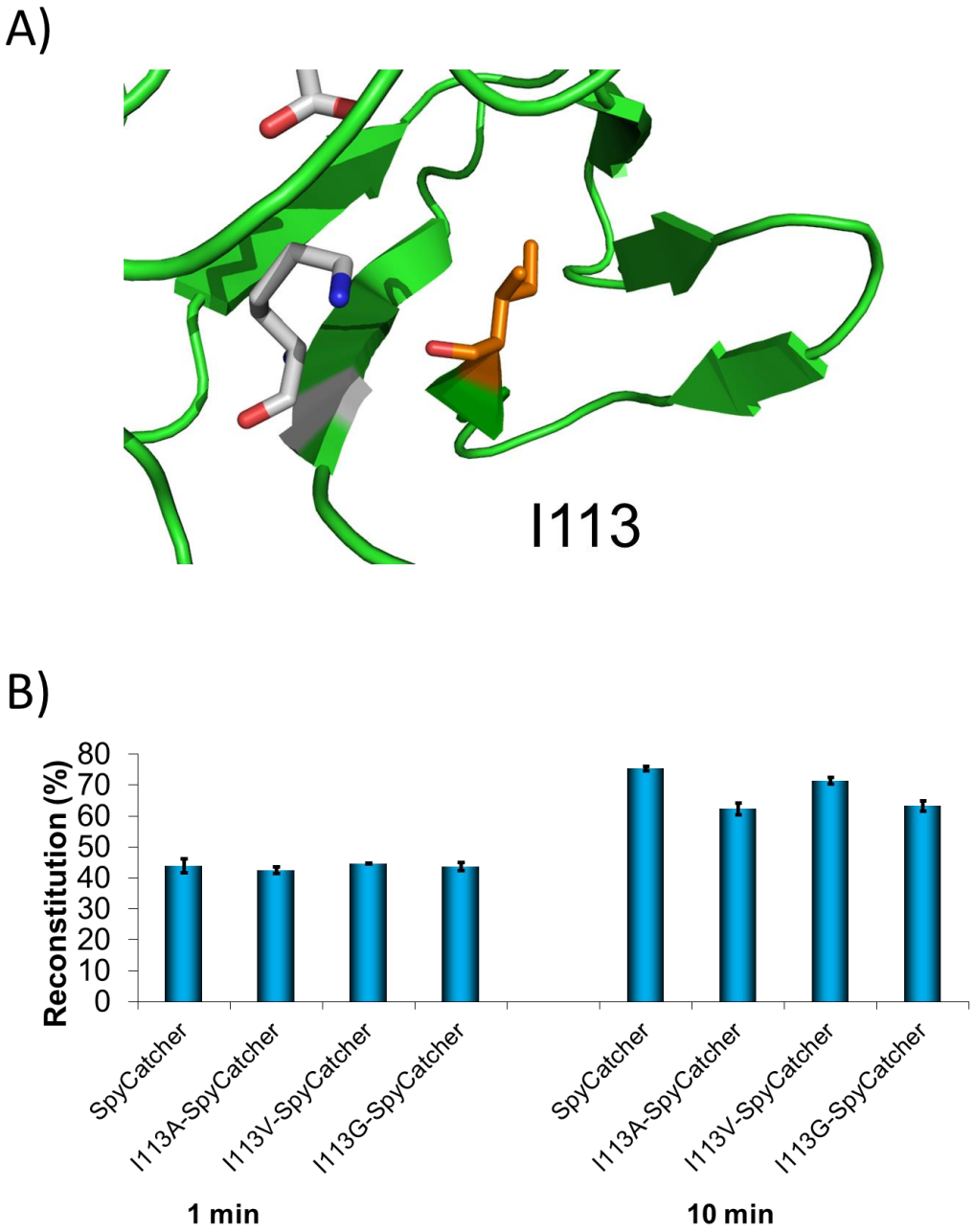


Figure 4.3: Influence on the reconstitution rate of SpyCatcher by mutating Ile113. A) I113 in cartoon of SpyCatcher (from PDB 2X5P). B) Rates of the different I113 mutants and SpyCatcher reacting with SpyTag-MBP. Reactions were analyzed by SDS-PAGE with Coomassie staining. All reactions were performed in triplicate showing mean ± 1 standard deviation.

Ile113 can be seen in a key position (Figure 4.1B), where the N-terminal part of SpyTag interacts with SpyCatcher after SpyTag docks into the cleft. Also Ile113 may form a strong interaction with the hydrophobic core of SpyCatcher, so slow displacement of this residue could hinder SpyTag docking. Thus SpyCatcher was mutated to Ala, Val, or Gly at position 113 (Figure 4.3A). The reconstitution rates of the mutants at 1 min were similar to the parent SpyCatcher. After 10 min incubation with SpyTag-MBP, SpyCatcher is superior compared to all the Ile113 mutants, suggesting that having a larger hydrophobic residue at that location helps in docking SpyTag.

4.1.4 Charged and Polar Mutations of SpyCatcher Based on the CnaB2 Structure:

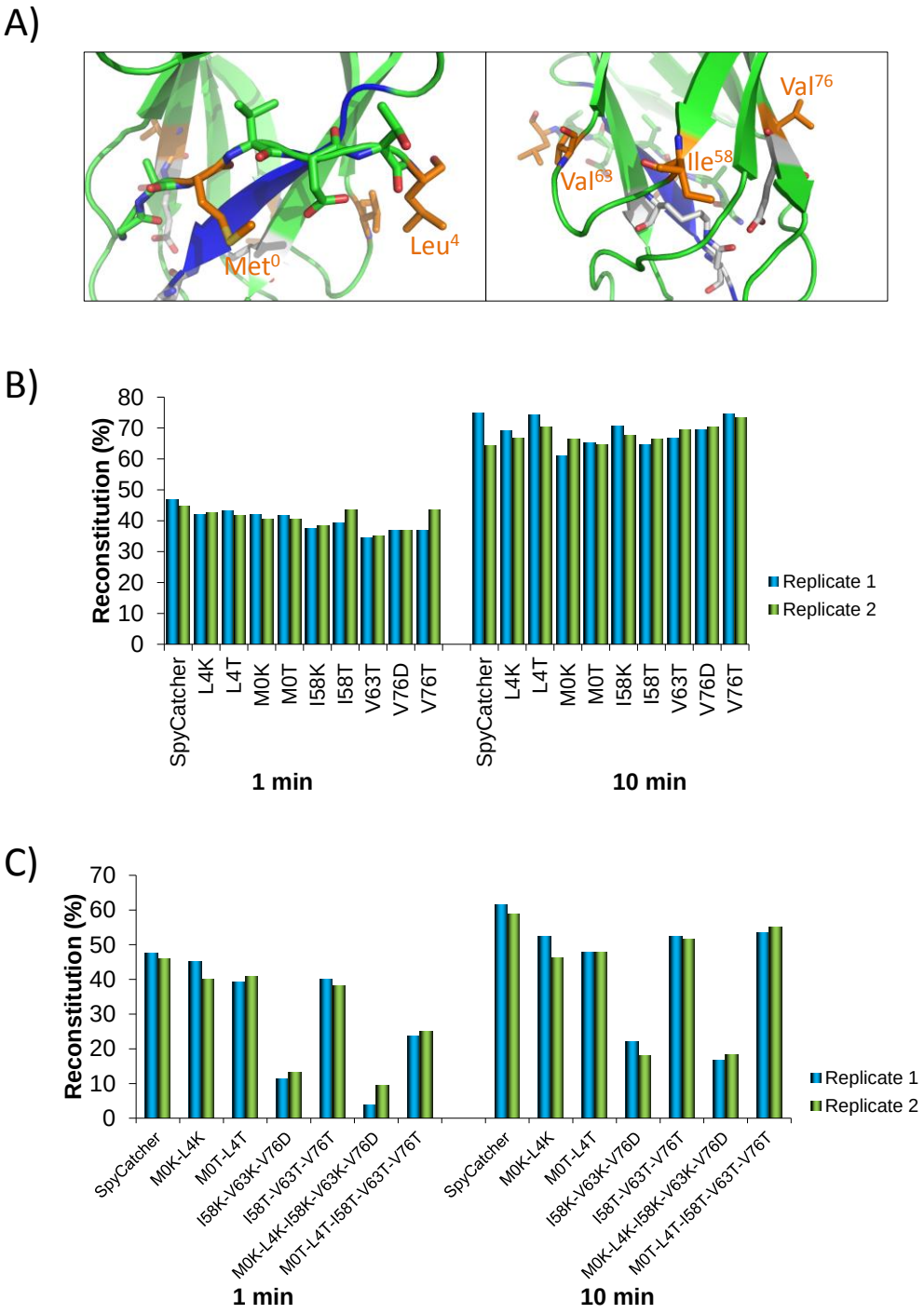


Figure 4.4: Determining the influence of polar and charged point mutations on the SpyCatcher reconstitution rate. A) Location of non-polar surface SpyCatcher residues mutated in this test: Met0, Leu4, Ile58, Val63, and Val76, from PDB 2X5P. B) Reconstitution rates of single polar/charged mutants of SpyCatcher with SpyTag-MBP. C) Reconstitution rates of combinations of mutants of SpyCatcher with SpyTag-MBP. Reactions were performed in duplicate and quantification was performed by SDS-PAGE with Coomassie staining

The sequence space of SpyCatcher was further explored by a series of mutations of exposed hydrophobic residues. Met0 and Leu4 are on the β-strand adjacent to the cleft for SpyTag binding (Figure 4.4A, left panel and Figure 4.1B). Hence if the cleft can be stabilized and kept unobstructed, this could lead to more efficient SpyTag docking. Ile58, Val63, and Val76 are located near the isopeptide triad, thus it could possibly stabilize the reactive residues which could improve the rate of amide bond formation (Figure 4.4A, right panel and Figure 4.1B).

The initial broad screening of single mutants gave only minimal changes: at 1 min the reconstitution was only slightly perturbed by the mutants in most cases, with V63T and V76D showing the largest drops in reconstitution rates (Figure 4.4B). At 10 min there are no substantial changes in reconstitution rate; thus the individual mutations had negligible effect on SpyCatcher activity.

The mutations were then combined in varying permutations (Figure 4.4C). The combination of charged mutants was more detrimental than the polar mutations (Figure 4.4C). I58K-V63K-V76D and M0K-L4K-I58K-V63K-V76D mutants have dramatically lowered the reconstitution rate of SpyCatcher. Mutation near the reactive Lysine and catalytic Glutamate (Figure 4.4A) seems to be the major cause of the loss of activity. In the case of the polar mutants, there is only a slight loss of activity, with the exception of the completely polar mutant of M0T-L4T-I58T-V63T-V76T, as at 1 min the reconstitution rate is greatly lowered. In contrast to the multi-charged mutant M0K-L4K-I58K-V63K-V76D, the multi-polar mutant M0T-L4T-I58T-V63T-V76T regains much of the activity after 10 min (Figure 4.4C). This relative reactivity may signify that the initial docking is hindered but once the protein has docked with SpyTag, there is no problem in catalysis of the isopeptide bond formation.

4.2 Mutations and Circular Permutations of SpyTag:

4.2.1 Overview of Mutations and Circular Permutations of SpyTag:

A)

```
                          (His-Tag)     (Thrombin Site)     (SpyTag)(Circular Permutation)
SpyTag          MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAYKPTK---- 35
2xR-SpyTag      MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAYRPTR---- 35
2xH-SpyTag      MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAYHPTH---- 35
2xQ-SpyTag      MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAYQPTQ---- 35
Perm1-SpyTag    MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAGS-GAMVDT 38
Perm2-SpyTag    MGSSHHHHHHSSGLVPRGSHMGAHIVMVDAGSGGAMVDT 39

CnaB2           MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQ
                SGDMTIEEDSATHIKFSKRDIDGKELAGATMELRDSSGK
                TISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAIT
                FTVNEQGQVTVNGKATKGDAHIVMVDA
```

B)

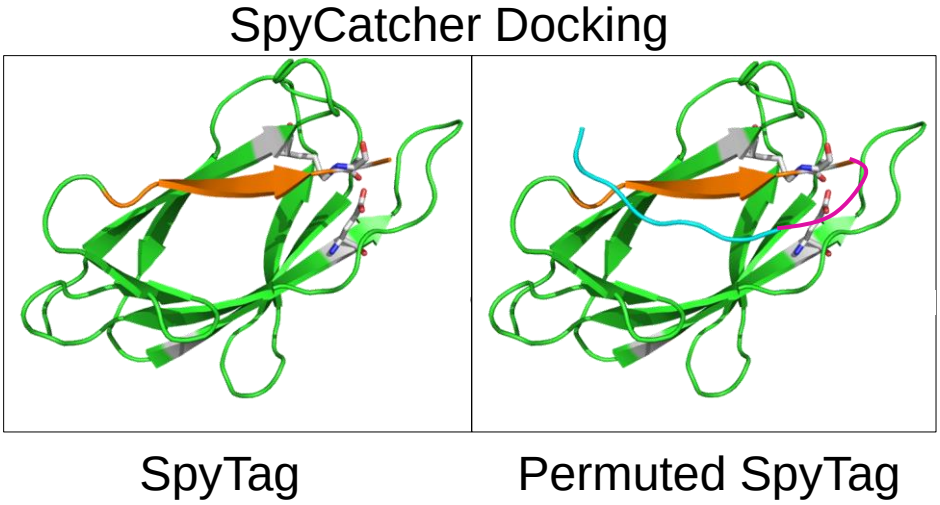


Figure 4.5: Mutants of SpyTag. A) Sequence alignment of the N-terminal region of SpyTag-MBP variants before the MBP sequence. The mutants fall into two categories: the first is the double mutants, where (in red) the Lysines within SpyTag (AHIVMVDAYKPTK) were both mutated to Arg, His, or Gln. The second category involves the two circularly permuted SpyTags, with CnaB2 sequence colored in coordinate to help display where the circular permutation comes from in the CnaB2 structure. The only difference between the two circular permutations is a single Glycine in the linker (pink). B) Cartoon of desired orientation of SpyTag (left panel, orange) and circularly permuted SpyTag (right panel, orange for SpyTag, pink for Gly-Ser linker which was drawn in and *not* included in the original PDB structure, and Cyan for circular permutation) docking with SpyCatcher, based on PDB 2X5P.

The SpyTag mutants generated were an attempt at improving the docking with SpyCatcher via two approaches. The first approach was double point mutations of SpyTag (AHIVMVDAYKPTK), replacing the Lysines with Arginines, Histidines, or

Glutamines (Figure 4.5A). The C-terminal stretch of SpyTag mutated (…YKPTK) was not present in the PDB 2X5P structure and was included from the primary sequence of FbaB (Zakeri et al., 2012). The C-terminal stretch of SpyTag was found in previous work to improve reconstitution with SpyCatcher (Zakeri et al., 2012). Mutating the Lysines was to facilitate applications requiring NHS-mediated labeling of SpyTag, so less nucleophilic positively-charged residues or another polar residue (Gln) were tested.

The second approach was circular permutation of SpyTag, removing the last residues of SpyTag (YKPTK) and replacing with the initial β-strand of the CnaB2 domain (Figure 4.5A, bottom two sequences, cyan representing the first β-strand of CnaB2). The circular permutation was designed first to stabilize the SpyTag conformation by the inclusion of an adjacent β-strand and second so that the extra β-strand would provide further interaction with SpyCatcher to stabilize SpyTag docking (Figure 4.5B).

4.2.2. Effect of Double Mutations on SpyTag Reconstitution Rate:

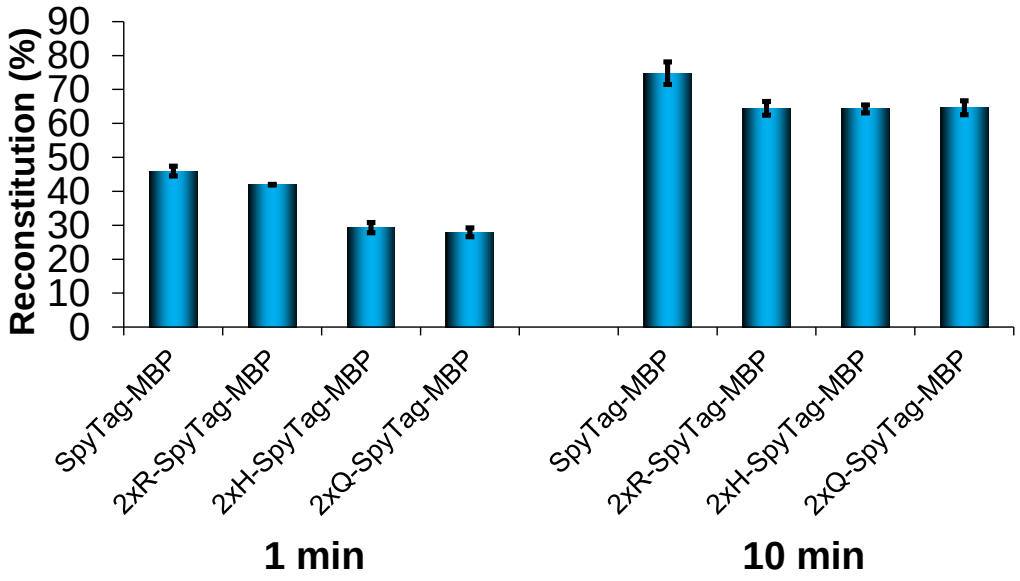


Figure 4.6: Testing reconstitution rate of 2xR, 2xH, 2xQ-SpyTag-MBP mutants with SpyCatcher. Reactions were analyzed by SDS-PAGE with Coomassie staining. Mean of triplicate ± 1 standard deviation.

The point mutants shown in Figure 4.5A were tested with SpyCatcher to determine their rate of reaction (Figure 4.6). All the mutations are detrimental. At 1 min

the 2xH and 2xQ-SpyTag-MBP constructs led to a substantial loss of activity (≥10%), while 2xR-SpyTag-MBP was only marginally worse than SpyTag-MBP at 1 min. The reconstitution rates agree with the previous data on SpyTag mutations, that even conservative changes to SpyTag hinder the reconstitution with SpyCatcher (Zakeri et al., 2012).

4.2.3 Testing Circularly Permuted SpyTag with SpyCatcher and N-Terminal Truncated SpyCatcher Variants:

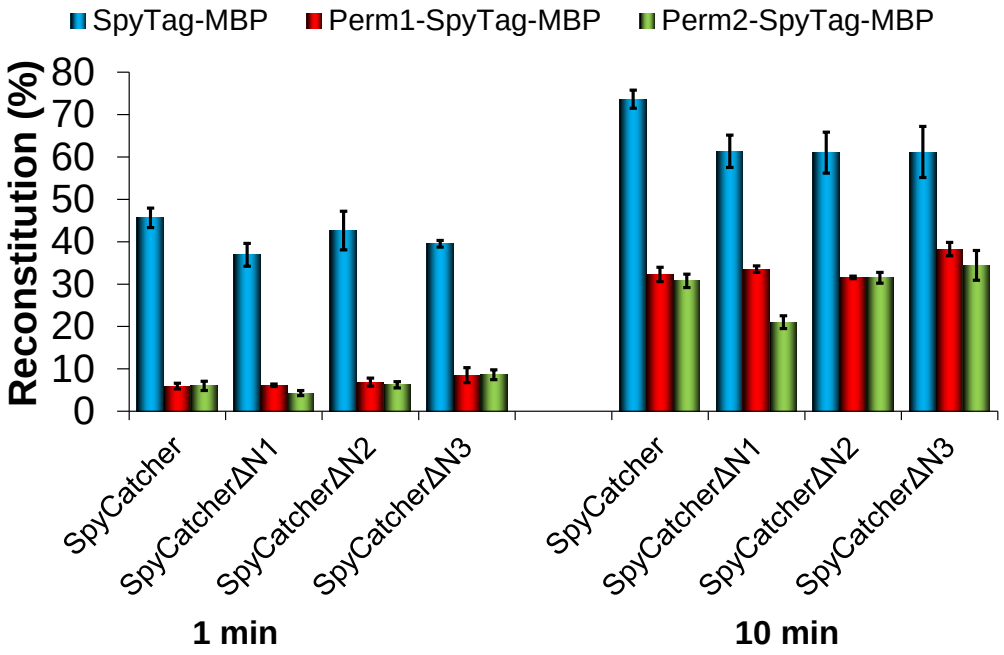


Figure 4.7: Determining the influence of circular permutation on SpyTag reconstitution with SpyCatcher, SpyCatcherΔN1, SpyCatcherΔN2, and SpyCatcherΔN3. Reactions were analyzed by SDS-PAGE with Coomassie staining. Mean of triplicate ± 1 standard deviation.

The circularly permuted SpyTag variants (Figure 4.5A, bottom two sequences), were tested with SpyCatcher and the N-terminal truncations of SpyCatcher (Figure 4.1A). The reason for testing with the N-terminal truncation mutants is that the original SpyCatcher includes the N-terminal section that was circularly permuted onto SpyTag to generate Perm1-SpyTag and Perm2-SpyTag (Figure 4.1A). Thus the N-terminal section of SpyCatcher could obstruct the docking of the Perm1-SpyTag and Perm2-SpyTag.

Though the hypothesis was that the N-terminal truncation mutants of SpyCatcher would be superior with the circularly permuted versions of SpyTag, the version of SpyCatcher did not matter. As seen in Figure 4.7, the circularly permuted SpyTags were dramatically worse than the original SpyTag. The circular permutation of SpyTag hampered the capacity to interact with SpyCatcher regardless of how much of the N-terminal region was present on SpyCatcher.

4.3 Structural Information of SpyTag Reconstituted with SpyCatcherΔN1:

Through collaboration with the Rapoport group, Dr. Long Li solved the structure of SpyCatcherΔN1 reconstituted with SpyTag (PDB 4MLS). Based on comparing this new SpyCatcherΔN1-SpyTag structure with the CnaB2 structure, a new series of mutants for both SpyCatcherΔN1 and SpyTag were generated.

4.3.1 Crystal Structure of SpyTag Reconstituted with SpyCatcherΔN1:

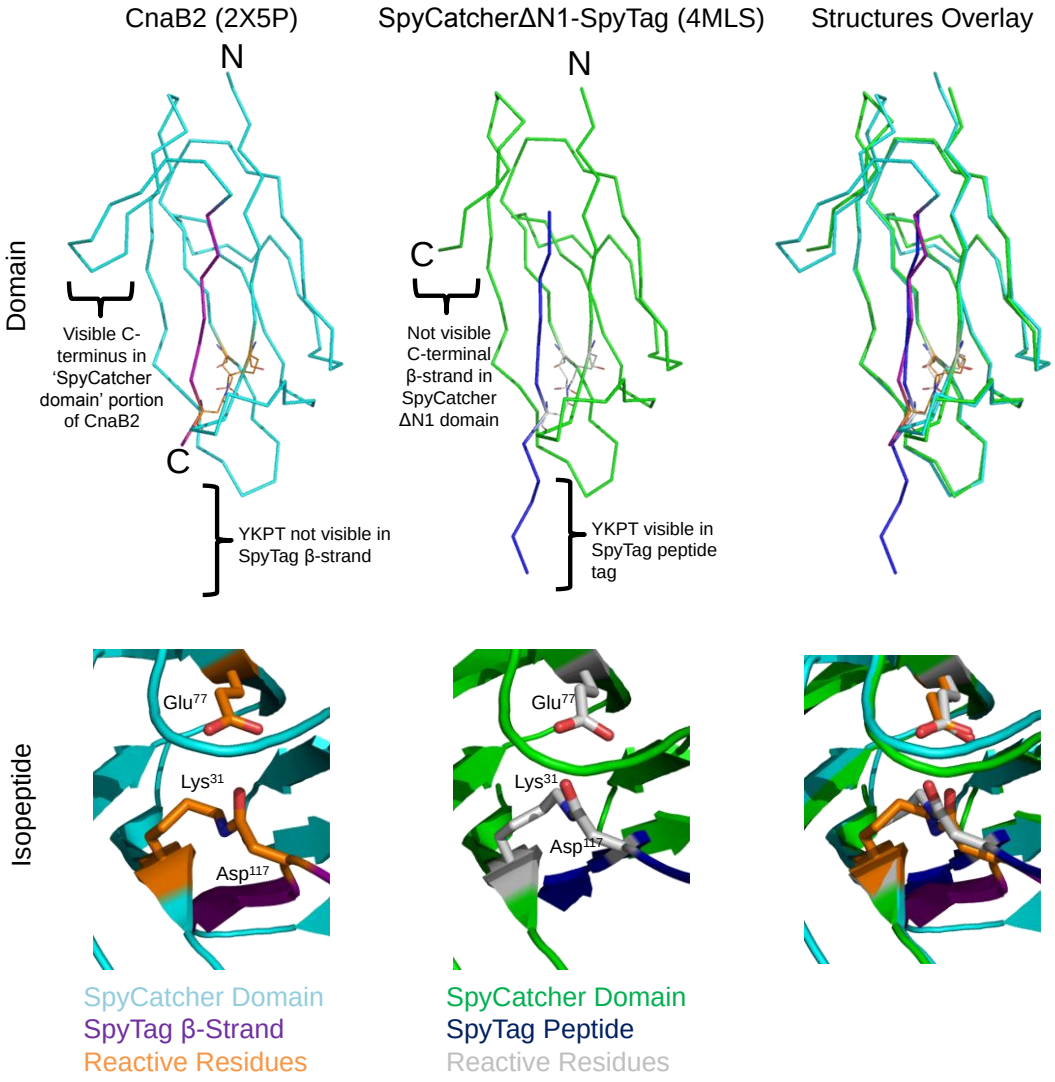


Figure 4.8: Comparison of the structures of CnaB2 (PDB 2X5P) and SpyCatcherΔN1-SpyTag (PDB 4MLS) using PyMOL.

SpyCatcherΔN1 reconstituted with SpyTag was crystallized and the structure solved by Dr. Long Li at a resolution of 2.0 Å (Appendix Table A.2) (PDB 4MLS). Comparison of the SpyCatcherΔN1-SpyTag with the CnaB2 structure shows, as expected, close overlay (Figure 4.8, right panel).

Two areas of interest arose from the SpyCatcherΔN1-SpyTag structure compared to CnaB2: the C-terminal β-strand of the SpyCatcherΔN1 domain is not visible in the crystal structure and so it is a question whether the disordered C-terminal β-strand

is playing a role in the interaction with SpyTag (Figure 4.8 top left and middle panels). The second area of interest was the newly visible residues of SpyTag (…YKPT) extending out of the SpyCatcherΔN1 domain (Figure 4.8, top-middle panel). Though these were included into SpyTag originally based on the FbaB primary sequence, they were not visible in CnaB2 (Figure 4.8, top-left panel) (Zakeri et al., 2012). Therefore a series of mutants were generated which focused on the differences between the SpyCatcherΔN1-SpyTag structure and CnaB2.

4.3.2 Using Structural Data and Rational Design to Study SpyCatcherΔN1:

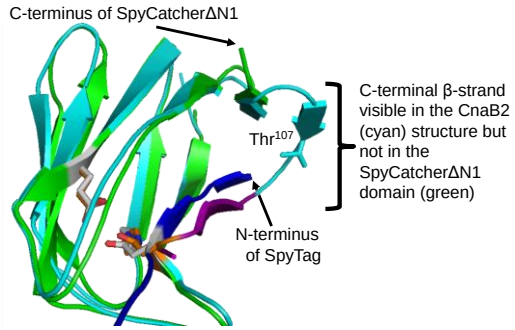
A)

```
SpyCatcher              MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDE 60
SpyCatcherΔN1           MSYYHHHHHHDYDIPTTENLYFQG---------------SG------DSATHIKFSKRDE 39
SpyCatcherΔN1ΔC1        MSYYHHHHHHDYDIPTTENLYFQG---------------SG------DSATHIKFSKRDE 39
SpyCatcherΔN1ΔC2        MSYYHHHHHHDYDIPTTENLYFQG---------------SG------DSATHIKFSKRDE 39
T107L-SpyCatcherΔN1     MSYYHHHHHHDYDIPTTENLYFQG---------------SG------DSATHIKFSKRDE 39
T107V-SpyCatcherΔN1     MSYYHHHHHHDYDIPTTENLYFQG---------------SG------DSATHIKFSKRDE 39

SpyCatcher              DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 120
SpyCatcherΔN1           DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 99
SpyCatcherΔN1ΔC1        DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 99
SpyCatcherΔN1ΔC2        DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 99
T107L-SpyCatcherΔN1     DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 99
T107V-SpyCatcherΔN1     DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 99

SpyCatcher              NEQGQVTVNGKATKGDAHI 139
SpyCatcherΔN1           NEQGQVTVNGKATKGDAHI 118
SpyCatcherΔN1ΔC1        NEQGQVTVNGKATKG---- 114
SpyCatcherΔN1ΔC2        NEQGQVTVNG--------- 109
T107L-SpyCatcherΔN1     NEQGQVTVNGKALKGDAHI 118
T107V-SpyCatcherΔN1     NEQGQVTVNGKAVKGDAHI 118
```

B)



Structures of CnaB2 (PDB 2X5P) aligned with SpyCatcherΔN1-SpyTag (PDB 4MLS) (Dr. Long Li)
CnaB2 (SpyCatcher Highlighted),C-terminal β-strand of CnaB2 (SpyTag),SpyCatcherΔN1,SpyTag peptide

C)

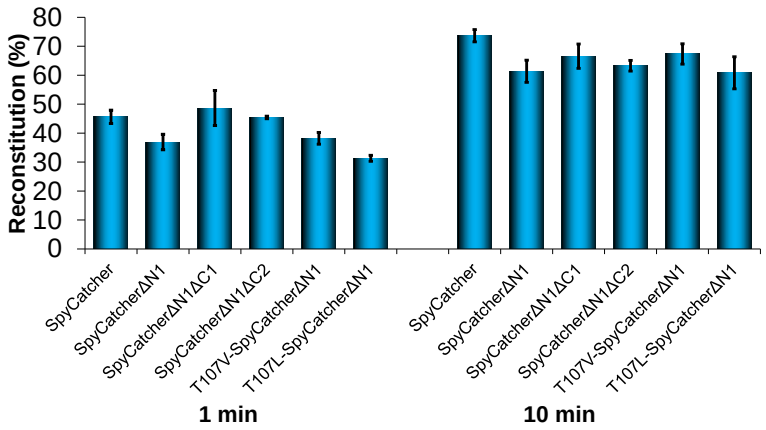


Figure 4.9: Variants based on the structure of SpyCatcherΔN1-SpyTag (PDB 4MLS) and their reconstitution rates. A) Sequence alignments of the SpyCatcherΔN1 mutants. The underlined sections of SpyCatcherΔN1 gave clear electron density in the crystal structure. Residues in green represent the minimized 'core domain' of SpyCatcher. Residues in red are the point mutants of Thr107. B) Overlaid structures of CnaB2 (PDB 2X5P) and SpyTag-SpyCatcherΔN1; SpyTag (blue), SpyTag β-strand of CnaB2 (purple). C) Reconstitution rates of SpyCatcher variants with SpyTag-MBP, analyzed by SDS-PAGE with Coomassie staining. Mean of triplicate ± 1 standard deviation.

The mutants generated based on the structure of SpyCatcherΔN1-SpyTag fall into two categories: truncation and point mutants (Figure 4.9A). The truncation mutants of SpyCatcherΔN1 were focused on the C-terminal region that was not visible in the crystal structure (Figure 4.9A and Figure 4.8 top panel). The amino acids of SpyCatcherΔN1 that are visible in the crystal structure are underlined in the alignment in Figure 4.9A. Two mutants of SpyCatcherΔN1 were generated in which the last 4 residues (SpyCatcherΔN1ΔC1) or the last 9 residues (SpyCatcherΔN1ΔC2) were deleted (Figure 4.9A) to check whether this flexible C-terminal section played an important role in the docking of SpyTag (Figure 4.8, top panel & Figure 4.9B). The next mutants generated were focused on Thr107 (Figure 4.9B), not visible in the crystal structure of SpyCatcherΔN1-SpyTag (Figure 4.9A), but seen in the structure of CnaB2. Thus Thr107 was mutated to Val and Leu to create a more hydrophobic interaction in the hopes to stabilize the secondary structure in this region of SpyCatcher.

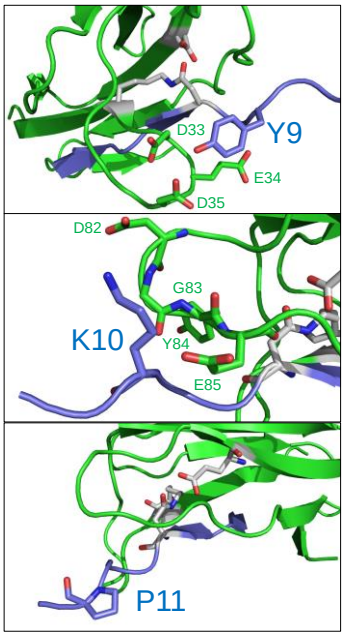
Comparing SpyCatcherΔN1 reaction at 1 min to the Thr107 mutants shows that the T107V mutation was inconsequential (P=0.55, not significant, n=3, unpaired t-test) and the T107L (P=0.028, n=3, unpaired t-test) was marginally detrimental (Figure 4.9C). The C-terminal truncations of SpyCatcherΔN1ΔC1 (P=0.038, n=3, unpaired t-test) and SpyCatcherΔN1ΔC2 (P=0.0058, n=3, unpaired t-test) at 1 min incubation both showed an improvement compared to SpyCatcherΔN1. At 10 min all the SpyCatcherΔN1 variants (SpyCatcherΔN1ΔC1, SpyCatcherΔN1ΔC2, T107V-SpyCatcherΔN1, and T107L-SpyCatcherΔN1) showed similar activity to SpyCatcherΔN1 (Figure 4.9C).

4.3.3 Using Structural Data and Rational Design to Study SpyTag:

A)

```
SpyTag       AHIVMVDAYKPTK 13
Y9R-SpyTag   AHIVMVDARKPTK 13
K10N-SpyTag  AHIVMVDAYNPTK 13
P11G-SpyTag  AHIVMVDAYKGTK 13
```

B)



C)

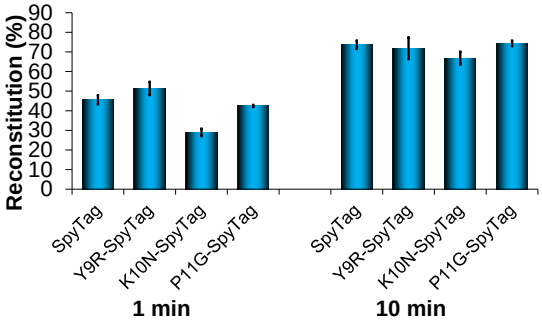


Figure 4.10: Mutants of SpyTag based on the SpyCatcherΔN1-SpyTag structure (PDB 4MLS). A) Alignment of the three SpyTag point mutants compared to SpyTag. B) Reconstituted SpyTag structure of residues 9-11 (labeled in blue) and possible polar contact candidates (residues labeled with green type) with the SpyCatcherΔN1 domain. Images based on the SpyCatcherΔN1-SpyTag structure (PDB 4MLS). C) Reconstitution rate of SpyCatcher for SpyTag and SpyTag mutants fused to MBP, analyzed by SDS-PAGE with Coomassie staining. All reactions were performed in triplicate and error bars represent ± 1 standard deviation.

The SpyTag mutants generated from the structure of SpyTag reconstituted with SpyCatcherΔN1 (PDB 4MLS) are shown aligned in Figure 4.10A. The series of point mutants on SpyTag were generated by looking at the localized interactions which could be possibly improved upon (Figure 4.10B). Tyr9 (AHIVMVDAYKPTK) is the first residue situated outside the docking cleft of SpyCatcher. Tyr9 was mutated to Arg with the goal of generating possible salt bridges with the nearby negatively charged loop (D33, E34, D35) (Figure 4.10B top panel). The next residue mutated was Lys10 (AHIVMVDAYKPTK) to Asn, and was designed to increase the number of hydrogen bonding opportunities with SpyCatcher (Figure 4.10B middle panel). The last SpyTag mutant generated was Pro11 (AHIVMVDAYKPTK) to Glycine, with the intent to increase the flexibility of the SpyTag such that the residues downstream in the SpyTag could make new interactions with SpyCatcher (Figure 4.10B bottom panel).

The effects of the generated SpyTag mutants were generally subtle (Figure 4.10C). However, an important point was that the Y9R mutation was not detrimental, as Arg seems capable of replacing the interactions that Tyr normally would have made with SpyCatcher. K10N-SpyTag was the most detrimental after 1 min incubation with SpyCatcher. The P11G mutant of SpyTag only minimally affected reconstitution. K10N-SpyTag was the unique case of being detrimental to the reconstitution yield (Figure 4.10C).

4.4 Discussion:

The initial series of mutants of SpyCatcher generated based solely on the CnaB2 crystal structure had a range of effects on the reconstitution yield of SpyCatcher. The first set of mutants generated were the N-terminal truncation mutants of SpyCatcher (Figure 4.1A). The reconstitution rate of the N-terminal truncations of SpyCatcher suggests that the N-terminal region is helping SpyTag docking or stabilizing the SpyCatcher structure for reaction, but the effect is still relatively minimal.

The second set of mutants based on the CnaB2 structure was a series of point mutants around the C-terminus and exposed surface hydrophobics of SpyCatcher. The reconstitution rates of the Ile113 mutants suggest that Ile113 is not essential for the docking of SpyTag. The varying loss of activity of the Ile113 mutants can be categorized by the size of the residue, with the greatest loss of activity with the smallest residues Alanine and Glycine, followed by Valine, and finally the sterically largest residue Isoleucine in the original amino acid sequence granted the highest reconstitution rate (Figure 4.3B). Single point mutants of surface hydrophobics on SpyCatcher, regardless of being polar or charged, caused a minimal loss of activity. The single point mutants were then combined in various permutations and their effects observed on the SpyCatcher reconstitution rate with SpyTag-MBP (Figure 4.4C). Mutations of surface hydrophobics were generally well tolerated, but the I58K-V63K-V76D-SpyCatcher was poorly active. Val63 and Ile58 are adjacent to the reactive Lys31, while Val76 is next to the catalytic Glu77, and so perhaps the isopeptide triad conformation is changed in the mutants and isopeptide formation cannot occur efficiently.

The next series of mutants were modifications of SpyTag based on the CnaB2 structure. The point mutants of SpyTag were double mutants which replaced the Lysines at position 10 and 13 of SpyTag (AHIVMVDAYKPTK) (Figure 4.5A). Reconstitution of these 2xR/H/Q-SpyTag-MBP mutants showed that indeed the positive charge was important in that section of SpyTag (Figure 4.6).

The second category of SpyTag mutants were circular permutations of SpyTag that joined a part of the N-terminal region of SpyCatcher/CnaB2 onto the C-terminal of SpyTag (Figure 4.7) and were tested with all the N-terminal truncation variants (Figure 4.1A). Regardless of the N-terminal truncation variant of SpyCatcher, the activity of permuted SpyTag variants was severely hampered.

Following the rational design campaign based on the CnaB2 structure, the structure of SpyCatcherΔN1 reconstituted with SpyTag was solved and used to develop another series of mutants to better understand the SpyCatcher-SpyTag interaction.

The mutants generated of SpyCatcherΔN1 were point mutants at Thr107 abutting the docking area of SpyTag and C-terminal truncations (Figure 4.9A & B). The T107V/L-SpyCatcherΔN1 mutants generally showed negligible effects on the rate of reconstitution with SpyTag (Figure 4.9B). In regard to the C-terminal truncation mutants of SpyCatcherΔN1, both cases showed mild improvements at 1 min incubation (Figure 4.9C). The possible reasoning for the slight improvement is that the C-terminus without the N-terminus to interact with leads to a destabilization and disordering of the C-terminal β-strand, which could explain why that section is not visible in the SpyCatcherΔN1-SpyTag structure but is visible in the CnaB2 structure (Figure 4.8, top panel). Furthermore the C-terminal truncations of SpyCatcherΔN1 demonstrates that the SpyCatcher domain can be minimized by 32 residues to a core domain of 83 amino acids (Figure 4.9A, green colored residues), with minimal effect on the reconstitution capacity (Figure 4.9C). The importance of reducing size is that SpyCatcher can now be feasibly generated via solid-phase synthesis and makes SpyCatcher ~9 kDa, which should be a less intrusive fusion on a target protein, expanding SpyCatchers uses in cellular systems.

A final series of point mutants of SpyTag was generated based on the SpyCatcherΔN1-SpyTag structure (Figure 4.10A) to generate better polar contact of SpyTag with SpyCatcher (Figure 4.10B). K10N-SpyTag was the most injurious in reconstitution rate as the positive charge on the C-terminal tail of SpyTag seems to play an important role, as replacing Tyr on SpyTag with Arg slightly improves activity with SpyCatcher (Figure 4.10C). These results should be followed up in future work with

thorough phage display screens, to improve the interaction between SpyCatcher/SpyTag such that it move towards the diffusion limit of SpyTag association with SpyCatcher.

Chapter 5: SpyLigase: Creating a Tripartite Split Protein for Peptide-Peptide Ligation

Sections of this chapter have been submitted in a manuscript with Mr. Gianluca Veggiani of the Howarth group.

Chapter 5: SpyLigase: Creating a Tripartite Split Protein for Peptide-Peptide Ligation

The value of peptide tags includes a broad range of uses, such as purification, labeling, and directing protein-protein interaction (Brizzard, 2008; Huh et al., 2003; Jarvik and Telmer, 1998; Terpe, 2003). The usage of peptide tags for synthetic or engineering purposes is limited due to fact that peptide tags have a small surface area in which restricted and weak interactions can occur between the peptide tag and a larger interacting protein domain (Guignet et al., 2007; Houk et al., 2003; Huang et al., 2009b). Therefore developing a system in which only peptide tags could be used to create a covalent interaction that could withstand high forces between proteins, such as those driven by molecular motors (Morris et al., 2001) or on the surface of a cell (Jain et al., 2013; McCloskey et al., 2003) could be of great use for biotechnology. Developing a system to covalently lock peptide tags may seem redundant with the development of SpyCatcher-SpyTag. However, fusing a full protein domain is undesirable in the context of many proteins or protein-protein interactions.

Therefore we aimed to develop a system to ligate two proteins using minimally disruptive peptide tags. The process of creating a peptide-peptide ligation system was accomplished with the tripartite split of the CnaB2 domain to create what we called the SpyLigase system. The basis of the SpyLigase system can be viewed as an evolutionary step from SpyCatcher-SpyTag. The SpyCatcher-SpyTag system took the original chemistry of the CnaB2 domain, in which a reactive Lysine and Aspartate would form a spontaneous isopeptide bond that was catalyzed by a strategically positioned Glutamate (Figure 5.1A and B) (Hagan et al., 2010). The CnaB2 domain was then split such that the reactive Lysine and catalytic Glutamate were located on a domain (SpyCatcher), and the reactive Aspartate was located on the excised β-strand which became a peptide tag (SpyTag) (Zakeri et al., 2012). SpyLigase builds upon the lessons of SpyCatcher-

SpyTag, by splitting CnaB2 into three separate parts, two peptide tags and one protein domain. The process involved excising the β-strand with the reactive Lysine to create a new peptide tag, KTag, reusing SpyTag, and leaving the catalytic Glutamate on the CnaB2's scaffold called SpyLigase (Figure 5.1A & C) (Appendix Figure A.5). It was hypothesized that the three parts could reconstitute the 'isopeptide triad', such that SpyLigase would ligate SpyTagged and KTagged proteins by spontaneous amide bond formation. In this chapter the ligation process was shown to be feasible and the reaction was characterized. Furthermore an application for this peptide-peptide ligation system in protein polymerization and rare cell isolation was explored.

5.1 Tripartite Splitting of CnaB2 to Create SpyLigase:

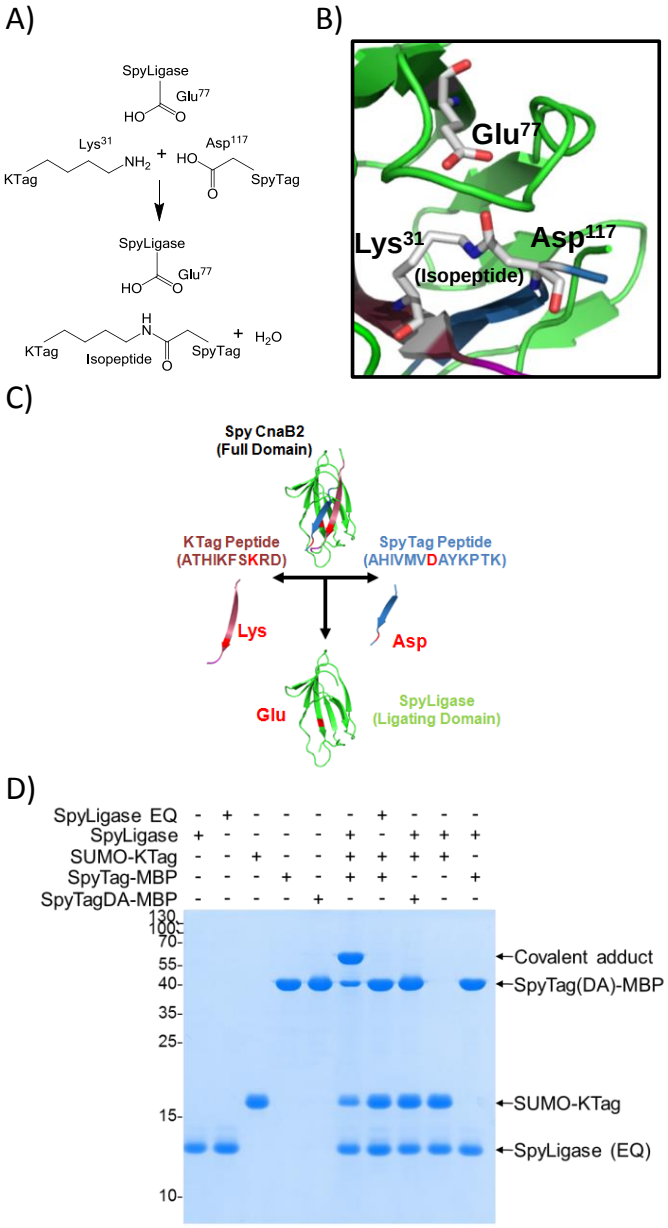


Figure 5.1: The chemistry and splitting of CnaB2 to generate SpyLigase. A) The chemistry of the isopeptide bond formation for SpyLigase. B) Isopeptide structure in CnaB2 (PDB 2X5P). C) Splitting of CnaB2 domain into KTag, SpyTag, and SpyLigase domain. D) Ligation of SUMO-KTag and SpyTag-MBP via SpyLigase. Ligation is shown to be dependent on the presence of the reactive Lysine on KTag, the reactive Aspartate on SpyTag, and the catalytic Glutamate on SpyLigase. Negative controls used were SpyLigase EQ which has the E→Q mutation to inactivate SpyLigase and SpyTag(DA)-MBP which has the reactive Aspartate mutated to Alanine. Reaction conditions of SpyLigase based ligation were 1.5 M trimethylamine N-Oxide (TMAO), 40 mM Na_2HPO_4 and 20 mM Citric acid, 10 µM SpyTag(DA)-MBP, 10 µM SUMO-KTag, and 40 µM SpyLigase (EQ). Reactions were incubated 24 h at 4°C, stopped by the addition of SDS loading dye, boiled for 5 min at 95°C, and run on SDS-PAGE with Coomassie staining.

The process of generating SpyLigase followed in the foot-steps of creating SpyCatcher-SpyTag (Zakeri et al., 2012). The chemistry involved was the same as with SpyCatcher, based on the CnaB2 domain's capacity to form an isopeptide bond using a catalytic Glutamate and reactive Lysine and Aspartate (Figure 5.1A, B). The hypothesis was that if each of the key residues were present on separate components, peptide-peptide ligation would then be feasible via KTag, SpyTag, and SpyLigase reconstituting (Figure 5.1C) and forming the isopeptide bond to covalently link KTag to SpyTag. Figure 5.1D shows that 24 h incubation of SUMO-KTag, SpyTag-MBP, and SpyLigase creates a new band that is ~57 kDa, matching the combined mass of SpyTag-MBP (~40 kDa) and SUMO-KTag (~17 kDa), and has survived boiling with SDS. The ~57 kDa band indicates that a covalent bond has formed between the MBP and SUMO partners. The dependency of the covalent adduct formation on the reactive Aspartate of SpyTag and the catalytic Glutamate of SpyLigase (Figure 5.1A, C) is shown when SpyTagDA-MBP (lacking the reactive Aspartate) or SpyLigase EQ (lacking the catalytic Glutamate) is mixed in, no covalent adduct is visible (Figure 5.1D).

5.2 Mass Spectrometry Analysis of SpyLigase Ligation of SUMO-SpyTag-SUMO-KTag Covalent Adduct:

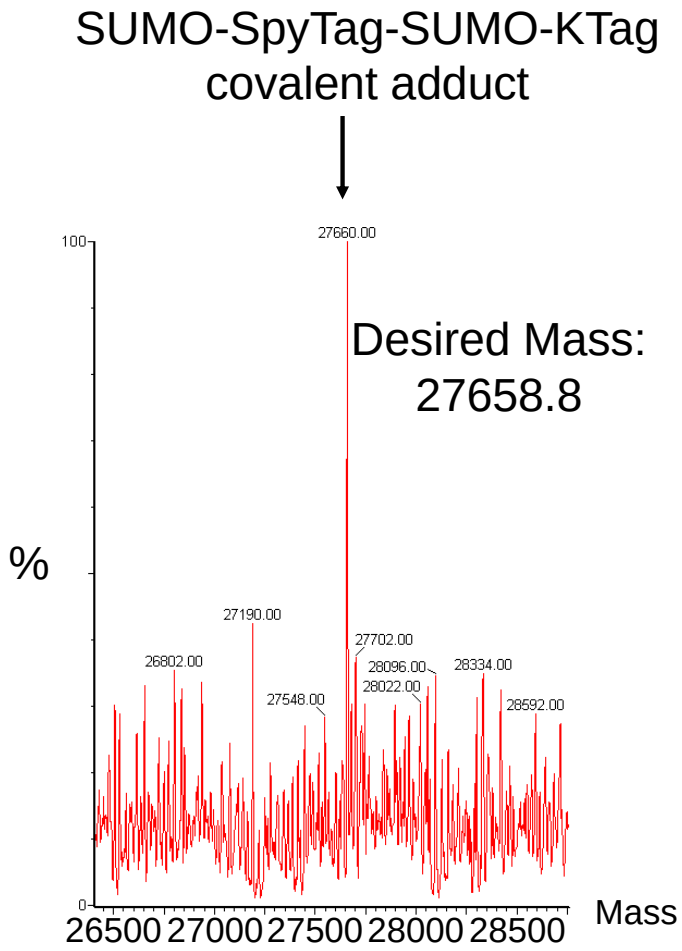


Figure 5.2: ESI-MS of SpyLigase-catalyzed ligation of SUMO-SpyTag with SUMO-KTag. The theoretical mass of the covalent adduct would be the mass of SUMO-KTag (13703.2), SUMO-SpyTag (13973.6), and H_2O subtracted as the leaving group: 27658.8 Da (=27676.8-18). Theoretical masses were calculated without the fMet residue, using ProtParam.

To confirm that SpyLigase is actually forming a covalent interaction between SpyTag and KTag, the covalent adduct of SpyTag-KTag was put through mass spectrometry analysis. In this case (Figure 5.2) the covalent adduct was formed between SUMO-SpyTag and SUMO-KTag, as it was found that a smaller covalent product was better detected. The mass spectrum in Figure 5.2 shows a clear peak for the combined

mass of SUMO-SpyTag (13973.6) and SUMO-KTag (13703.2), with 18 Da less to account for the loss of H_2O.

5.3 Titration of SpyLigase to Determine the Optimal Concentration for Peptide-Peptide Ligation:

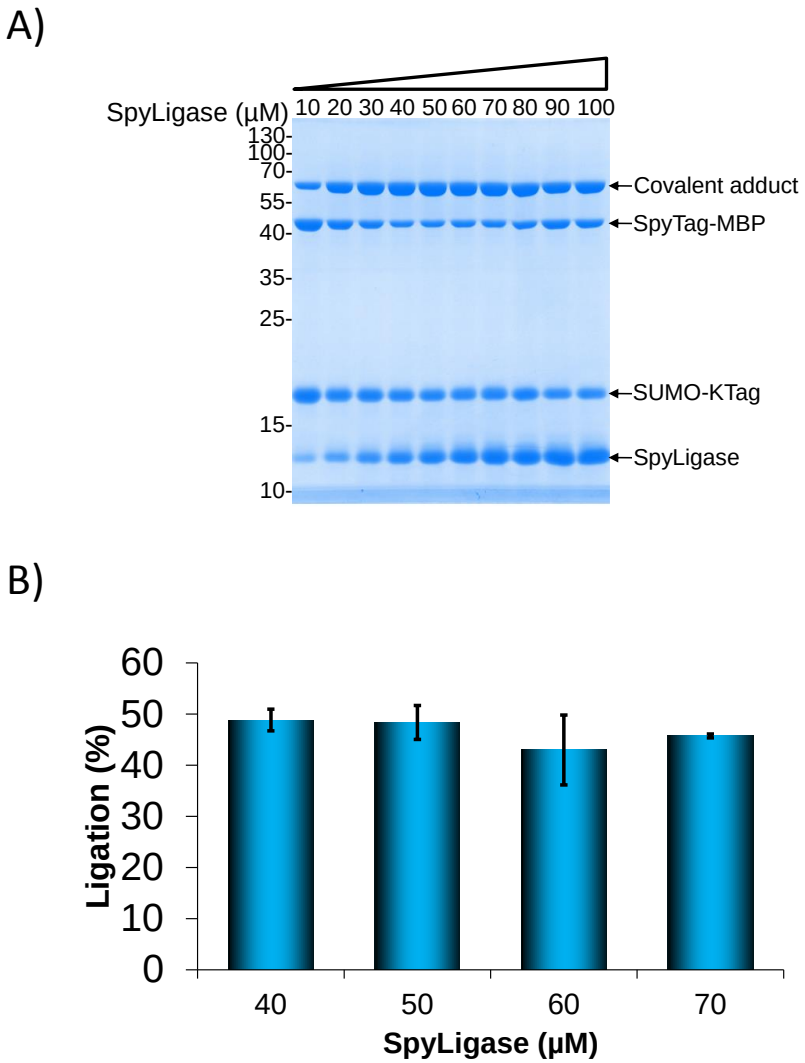


Figure 5.3: Testing the optimal concentration of SpyLigase in ligation. A) Titration of SpyLigase between 10-100 μM in the ligation of SpyTag-MBP and SUMO-KTag; samples were analyzed via SDS-PAGE and Coomassie staining. B) Quantification of ligation yield with SpyLigase at 40-70 μM (mean of triplicate ±1 standard deviation). Reaction conditions were 10 μM SpyTag-MBP, 10 μM SUMO-KTag, 1.5 M TMAO, 40 mM Na_2HPO_4, 20 mM Citric acid, at 4°C for 24 h.

Peptide-peptide ligation was indeed promoted by SpyLigase, but required optimization in numerous respects to achieve the maximum production of the covalent

adduct. Figure 5.3A is the titration of SpyLigase from 10-100 µM to test whether a higher concentration of SpyLigase could assist in covalent adduct formation. The amount of covalent adduct plateaus in the range of 40-70 µM SpyLigase (Figure 5.3A). To determine which SpyLigase concentration between 40-70 µM was optimal, the yield from this concentration range was quantified. In Figure 5.3B there was no dramatic difference between the SpyLigase concentrations in the 40-70 µM range. Thus out of convenience of using less protein, 40 µM of SpyLigase became the standard concentration for covalent adduct ligation when using 10 µM each of a SpyTag- and KTag-fusion. The SpyLigase titration (Figure 5.3) also implies that a certain proportion of SpyLigase and SpyTag and/or KTag are inactive. The reason that a proportion of SpyLigase is shown inactive is that in Figure 5.3A, there is a clear improvement in covalent adduct formation when increasing SpyLigase concentration from 10 µM to 40 µM. Though after this point the amount of covalent adduct formation plateaus at ~50% and not in proportion with the amount of active SpyLigase available. This signals that a certain amount of either or both SpyTag and KTag are inactive, acting as the limiting factor in covalent adduct formation.

5.4 Titration of Chemical Chaperone Trimethylamine N-Oxide (TMAO) to Determine the Optimal Concentration for Covalent Adduct Yield:

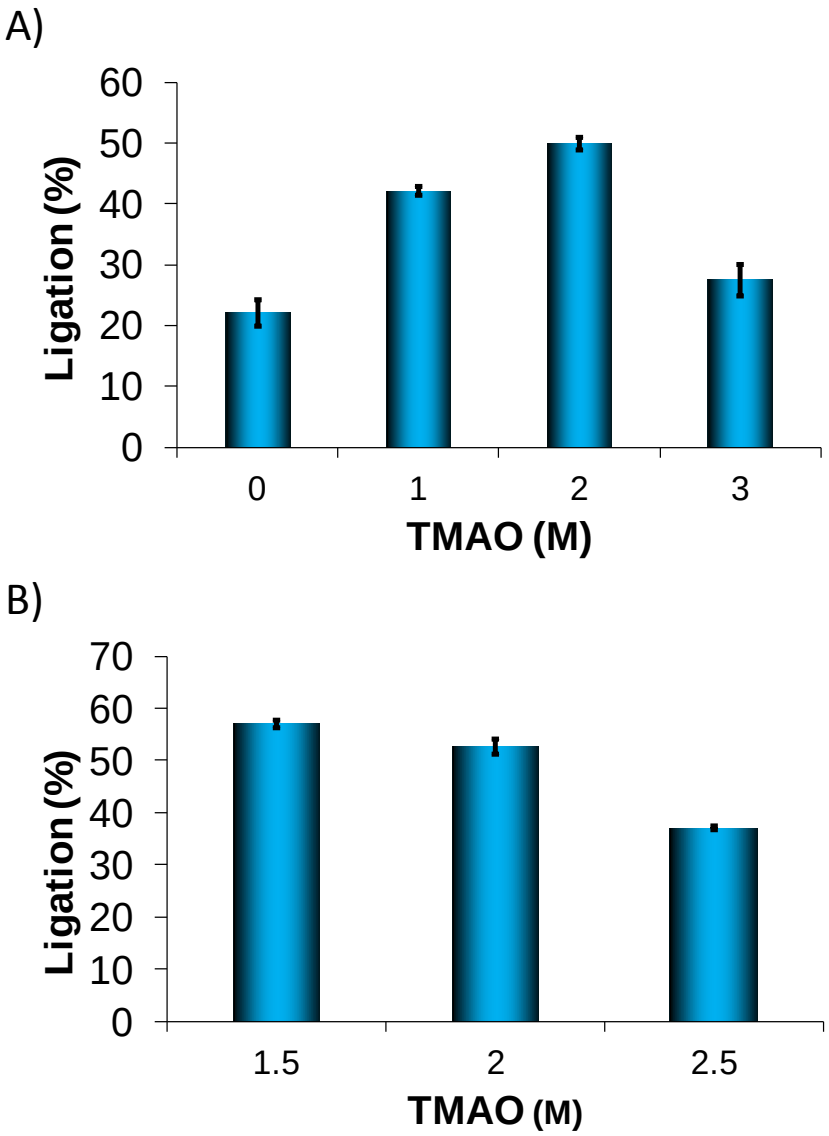


Figure 5.4: Optimizing concentration of TMAO for covalent adduct production between SpyTag-MBP and SUMO-KTag. A) Quantification of covalent adduct formation of SpyLigase with 0, 1, 2, and 3 M of TMAO to determine the general range of effectiveness of TMAO. B) Quantification of intermediate concentrations of TMAO (1.5 and 2.5 M) to determine the optimal concentration for covalent adduct yield by SpyLigase. All reactions show the mean of triplicate ±1 standard deviation. Reactions were performed with variable amounts of TMAO, 40 µM SpyLigase, 10 µM SUMO-KTag, 10 µM SpyTag-MBP, 40 mM Na_2HPO_4, 20 mM Citric acid, and reactions were incubated at 4°C for 24 h. Samples were then analyzed via SDS-PAGE with Coomassie staining.

I attempted to improve the SpyLigase reaction via the use of various additives and chemical chaperones (Appendix Figure. 6A). TMAO was the most effective (Appendix Figure. 6B). Subsequently to gain highest yields I performed an initial broad titration of TMAO between 1-3 M (Figure 5.4A). The results in Figure 5.4A show that 2 M TMAO was initially the optimal concentration. I then focused on a narrower range and showed that 1.5 M was the optimal concentration for covalent adduct formation via SpyLigase (Figure 5.4B).

5.5 Temperature Gradient of SpyLigase-Catalyzed Covalent Adduct Formation:

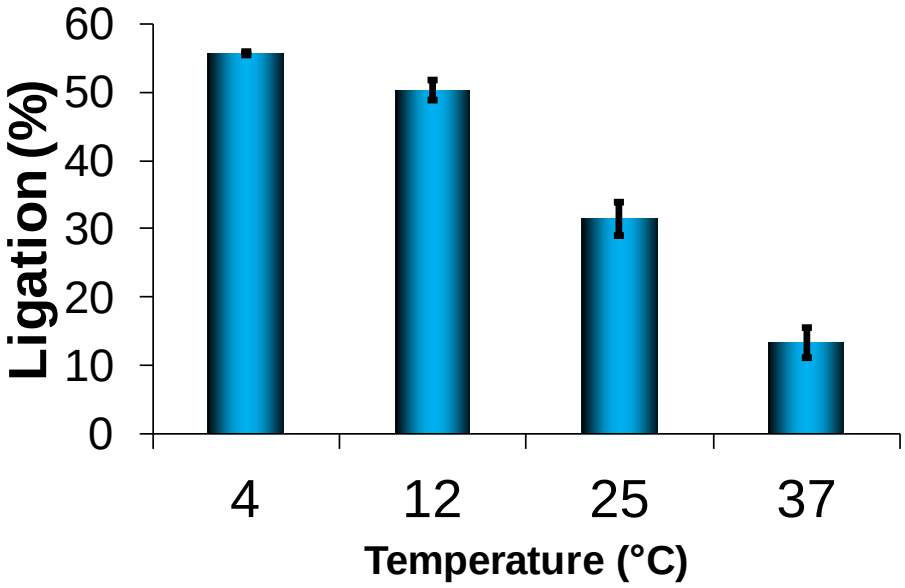


Figure 5.5: Temperature gradient of the ligation of SUMO-KTag-SpyTag-MBP via SpyLigase. All reactions were the mean of triplicate, ±1 s.d. Reactions conditions, were: 40 µM SpyLigase, 10 µM SUMO-KTag, 10 µM SpyTag-MBP, 40 mM Na_2HPO_4 and 20 mM Citric acid, 1.5 M TMAO, and reactions were incubated for 24 h at 4, 12, 25, and 37°C. Samples were analyzed via SDS-PAGE with Coomassie staining.

Figure 5.5 shows a clear preference for the SpyLigase-catalyzed ligation to occur at 4°C, with a maximum yield of ~55%. As the temperature is increased, the worse is the performance of SpyLigase. Therefore the incubation temperature of 4°C was taken as the standard condition.

5.6 Time Course of SpyLigase-Catalyzed Covalent Adduct Formation:

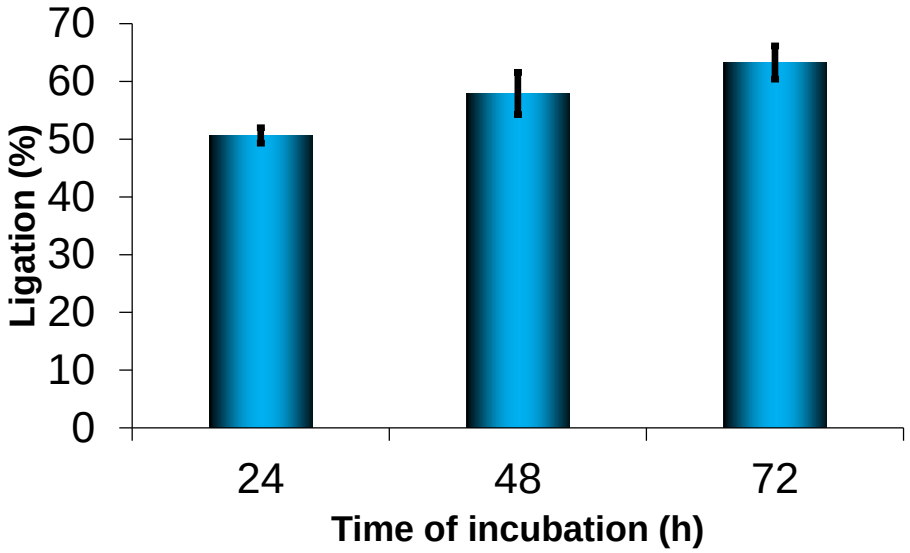


Figure 5.6: Time course of SpyLigase-catalyzed ligation of SUMO-KTag and SpyTag-MBP. All reactions were mean of triplicate ±1 standard deviation. Reactions conditions: 40 μM SpyLigase, 10 μM SUMO-KTag, 10 μM SpyTag-MBP, 40 mM Na_2HPO_4, 20 mM Citric acid, 1.5 M TMAO, and reactions were incubated at 4°C for 24, 48, and 72 h. Samples were then analyzed via SDS-PAGE with Coomassie staining.

Reaction yield increases over 3 days but the majority of covalent adduct formation occurs during the initial 24 h incubation (Figure 5.6).

5.7 Location-Dependence of SpyTag and KTag for SpyLigase-Catalyzed Ligation:

This section examines how ligation by SpyLigase depends on peptide tag location. I tested different combinations of N- and C-terminal SpyTag/KTag fusions and determined whether SpyTag-KTag could form a covalent adduct with SpyTag located on a linker between two protein domains and KTag fused to a separate protein.

5.7.1 SpyLigase-Catalyzed Ligation of N- or C-Terminal SpyTag/KTag:

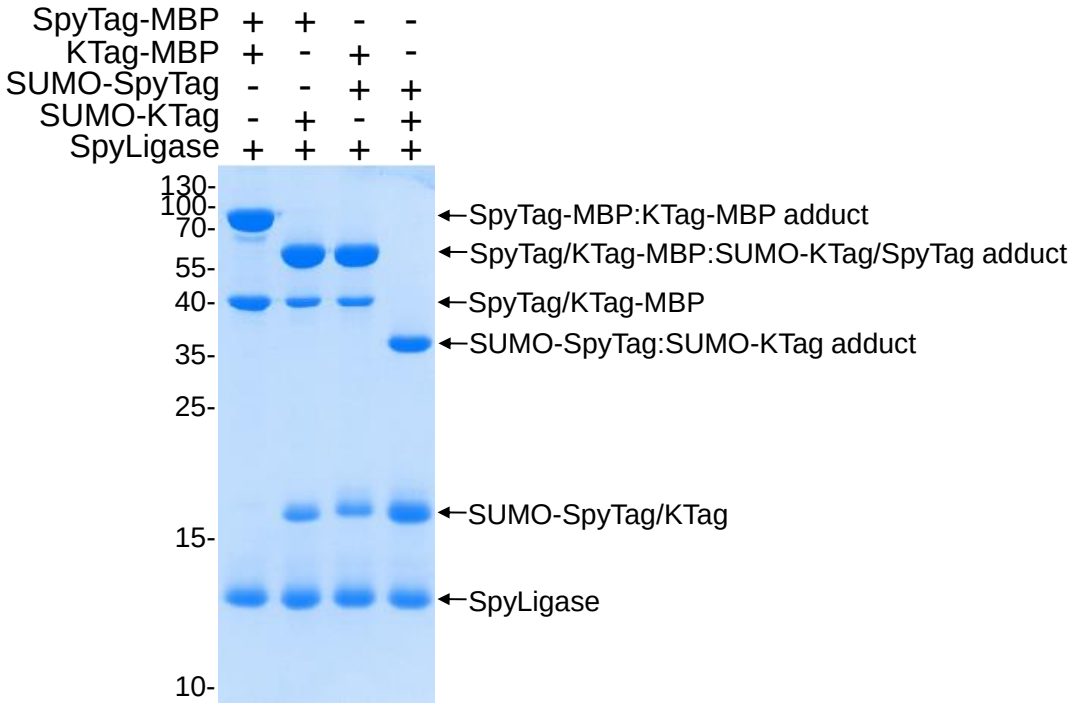


Figure 5.7: Testing SpyTag/KTag location-dependence for SpyLigase-catalyzed ligation. SpyTag/KTag were fused to either the N-terminus of MBP (SpyTag/KTag-MBP) or the C-terminus of SUMO (SUMO-SpyTag/KTag). Proteins were incubated with SpyLigase for 24 h at 4°C and analyzed by 16% SDS-PAGE with Coomassie staining. Reaction conditions were 40 µM SpyLigase, 10 µM SUMO-SpyTag/KTag, 10 µM SpyTag/KTag-MBP, 40 mM Na_2HPO_4, 20 mM Citric acid, and 1.5 M TMAO.

To test the versatility of SpyLigase, we determined whether SpyTag or KTag had location-dependence as many peptide tags must be at specific termini (Brizzard, 2008; Lichty et al., 2005; Terpe, 2003). Thus, four different fusion proteins were used to test whether SpyLigase had any preference in terms of termini. SpyTag/KTag-MBP had the peptide tags fused to the N-terminus of the MBP domain, while SUMO-SpyTag/KTag had the peptide tags located on the C-terminus. In Figure 5.7 SpyLigase catalyzed the ligation of fusion proteins N- to N-terminus, N- to C-terminus, or C- to C-terminus. Therefore if SpyTag or KTag are exposed on either termini, SpyLigase can ligate the two proteins together.

5.7.2 SpyLigase-Catalyzed Ligation of SpyTag Located on a Linker between Two Protein Domains with N- or C-terminal KTag:

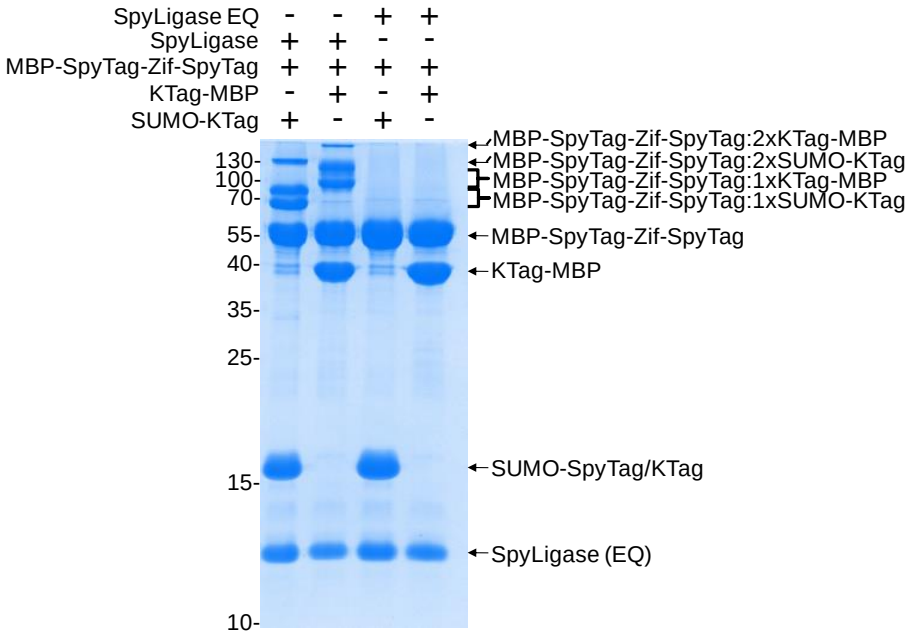


Figure 5.8: Testing SpyLigase location-dependence with peptide tags located between two protein domains. MBP-SpyTag-Zif-SpyTag was incubated with KTag fused on the N-terminus (MBP fusion) or the C-terminus (SUMO fusion). SpyLigase EQ acted as the negative control. 40 μM of SpyLigase (EQ) and 10 μM MBP-SpyTag-Zif-SpyTag were incubated for 24 h at 4°C with either 20 μM KTag-MBP or SUMO-KTag. Reactions were stopped by the addition of SDS loading dye and analyzed by 16% SDS-PAGE with Coomassie staining.

I determined if SpyLigase could ligate a SpyTag located on a linker between two protein domains, with KTag located on either the N- or C-terminus of a fusion protein. In Figure 5.8 MBP-SpyTag-Zif-SpyTag (Zakeri et al., 2012), which has a SpyTag located between MBP and the Zinc finger region from Zif268 (Christy and Nathans, 1989) and a second SpyTag located on the C-terminus, was incubated with KTag-MBP (N-terminally located) or SUMO-KTag (C-terminally located). In each ligation reaction with active SpyLigase, MBP-SpyTag-Zif-SpyTag:SUMO-KTag and MBP-SpyTag-Zif-SpyTag:KTag-MBP, there are three different covalent adducts formed. In the case of the MBP-SpyTag-Zif-SpyTag:SUMO-KTag mixture (left-most lane of Figure 5.8), at ~70 kDa two bands can be seen which represent a single SUMO-KTag ligated to either SpyTag site of MBP-SpyTag-Zif-SpyTag. The location of the isopeptide leads the ligated proteins to run

differently on SDS-PAGE. At ~130 kDa a band of MBP-SpyTag-Zif-SpyTag with two ligated SUMO-KTag molecules can be seen. Similar species are visible in the case where MBP-SpyTag-Zif-SpyTag was incubated with KTag-MBP (second lane from the left of Figure 5.8). Therefore the singly ligated MBP-SpyTag-Zif-SpyTag with KTag-MBP occupies the mass range ~100-120 kDa, and the doubly ligated species is located at >130 kDa. Figure 5.8 shows that SpyLigase can create a covalent adduct with a SpyTag located on a linker between two protein domains, with the KTag located at either the N- or C-terminus of a separate protein. Another conclusion when visually examining the band intensities of the singly ligated MBP-SpyTag-Zif-SpyTag with either KTag fusion protein is that the bands are of approximately equal intensity (Figure 5.8), signifying that there is no clear preference for either SpyTag site.

5.8 Determining Whether SpyLigase Releases the Covalent Adduct After Isopeptide Bond Formation:

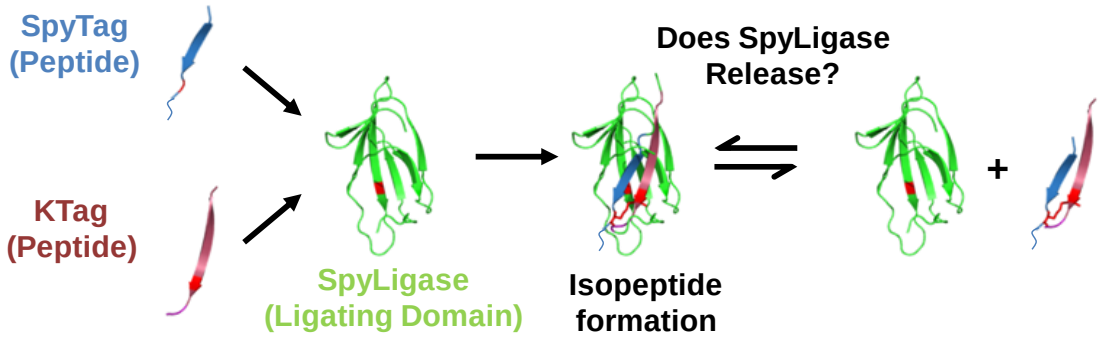


Figure 5.9: Demonstrative image of SpyLigase-catalyzed ligation with SpyTag and KTag, questioning whether the covalent adduct is released by SpyLigase.

Figure 5.9 is used to illustrate the question of whether SpyLigase actually releases the SpyTag-KTag covalent adduct. Non-release of the covalent adduct by SpyLigase may not always be an issue for *in vitro* ligation, but it would be desirable for the covalent linkage to have as small a 'recombinant footprint' as possible. SpyLigase release cannot be tested based on SDS-PAGE since SpyLigase's interaction is non-

covalent and released by boiling with SDS. In this section we determined whether SpyLigase remained on the covalent adduct by native PAGE and gel filtration.

5.8.1 Native PAGE to Determine Whether SpyLigase is still bound to the Covalent Adduct:

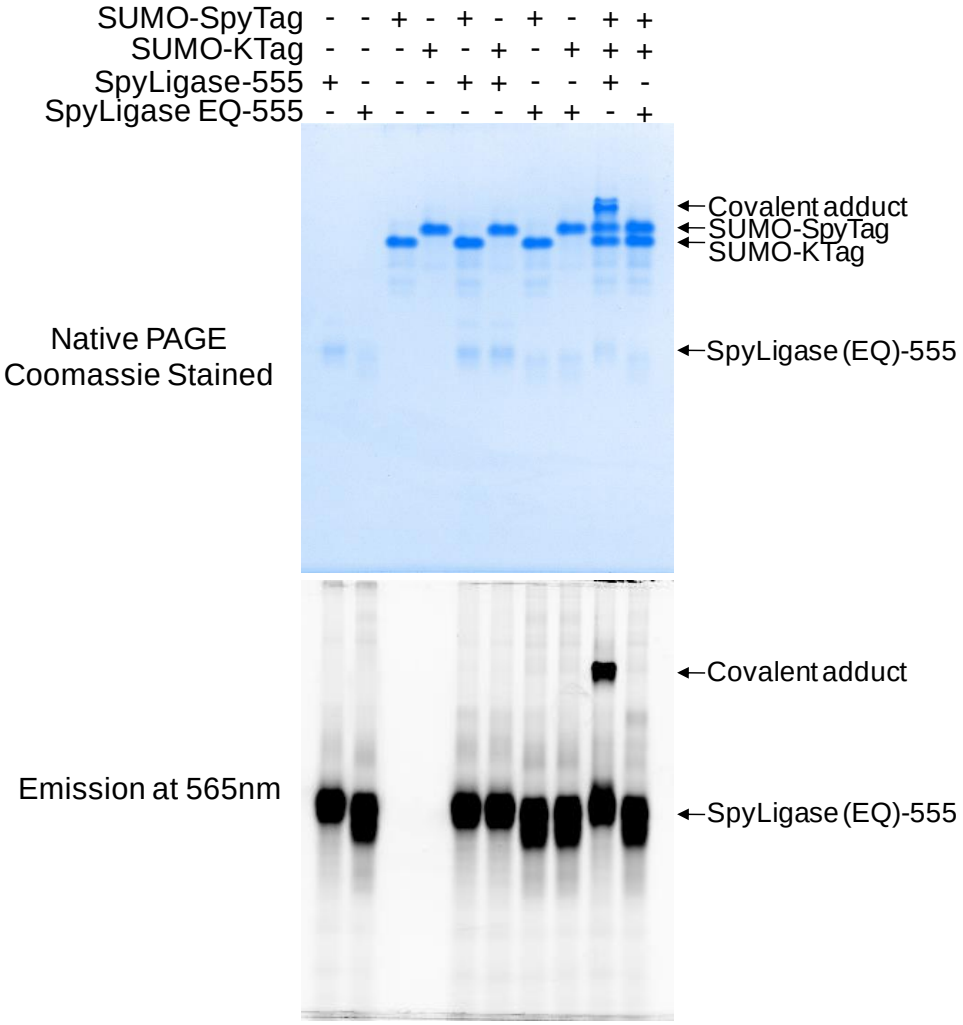


Figure 5.10: Determining whether SpyLigase releases the covalent adduct after isopeptide bond formation. Upper panel: SpyLigase labeled with Alexa Fluor-555 was incubated with SUMO-SpyTag and SUMO-KTag. After 24 h of incubation, samples were run on native PAGE and Coomassie-stained. Lower panel: Same gel except excited at 532 nm and emission was measured at 565 nm to determine the location of SpyLigase-555. Reaction conditions were 40 μM SpyLigase (EQ)-555, 10 μM SUMOKTag, 10 μM SUMOSpyTag, 1.5 M of TMAO, 40 mM Na_2HPO_4, 20 mM Citric Acid, incubated for 24h at 4°C.

Native PAGE was used to determine whether the covalent adduct was released by SpyLigase, after labeling SpyLigase (EQ) with Alexa Fluor-555. The labeled

SpyLigase-555 acted as a way to explicitly identify whether SpyLigase-555 was still non-covalently bound to the SpyTag-KTag adduct (Figure 5.10). The identity of the highest molecular weight band in the Coomassie-stained gel, which had SpyLigase-555, SUMO-KTag, and SUMO-SpyTag incubated together, is not clear. However, in the lower panel of Figure 5.10, when the gel was excited at 532 nm and emission checked at 565 nm, it is clear that SpyLigase-555 is still bound to the KTag-SpyTag covalent adduct. There is only one high molecular weight band seen in Figure 5.10 lower panel, when SpyLigase-555 is mixed with SUMO-KTag and SUMO-SpyTag. Where SpyLigase EQ-555 was mixed with SUMO-SpyTag and SUMO-KTag, no large complex is seen. The formation of the isopeptide bond is important for reconstituting the structure. Though the interaction between SpyLigase and the covalent adduct is non-covalent, the isopeptide bond is an important core for helping to stabilize other interactions within this split CnaB2 domain.

5.8.2 Gel Filtration of SpyLigase-Catalyzed Ligation to Determine Whether SpyLigase is Still Bound to the Covalent Adduct:

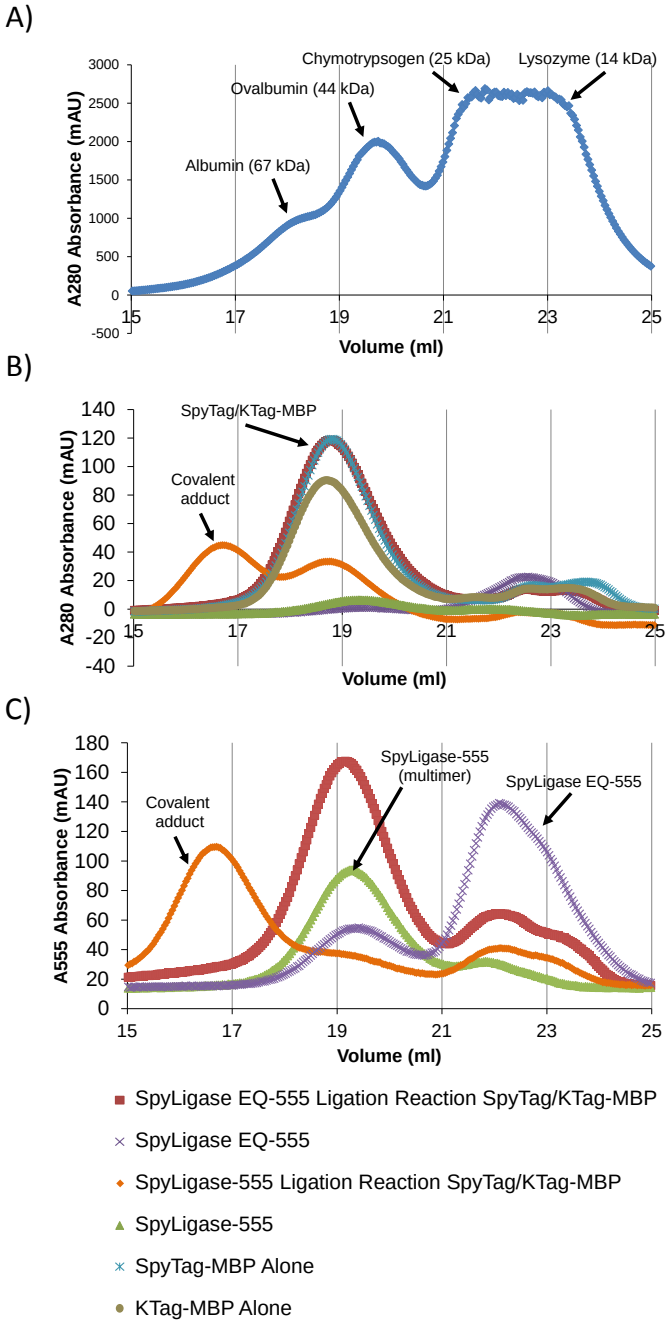


Figure 5.11: Gel Filtration of SpyLigase (EQ)-555 ligation reaction of SpyTag-MBP with KTag-MBP, to determine if SpyLigase (EQ)-555 releases the SpyTag-MBP-KTag-MBP covalent adduct. A) Molecular weight calibration for the gel filtration. Albumin (67 kDa), Ovalbumin (44 kDa), Chymotrypsin (25 kDa), and Lysozme (14 kDa) were all mixed and run through the gel filtration column to give an approximation of the relation between molecular weight and elution volume. B) 280 nm absorbance of SpyLigase (EQ)-555 ligation reaction of SpyTag-MBP-KTag-MBP. C) 555 nm absorbance of SpyLigase (EQ)-555 ligation reaction of SpyTag-MBP-KTag-MBP.

As an alternative assay to analyze whether SpyLigase remained bound to the covalent adduct, gel filtration was used. The procedure, much like in Figure 5.10, used SpyLigase (EQ) labeled with Alexa Fluor-555 to compare absorbances at 280 nm (protein) and 555 nm (dye-labeled SpyLigase (EQ)). In Figure 5.11B when the reaction of SpyTag/KTag-MBP mixed with SpyLigase EQ-555 was run, the major peak eluted at the same volume as the SpyTag/KTag-MBP alone peak at ~19 ml. For SpyLigase-555 mixed with SpyTag/KTag-MBP, there is a new peak not seen in any of the controls at ~16 ml, which would signify the covalent adduct (Figure 5.11B). Is the peak at ~16 ml simply the SpyTag-MBP-KTag-MBP covalent adduct without the SpyLigase? Figure 5.11C measured the absorbance at 555 nm, where the ~16 ml peak is still visible, signifying that SpyLigase is not releasing the covalent adduct. SpyLigase EQ-555 mixed with SpyTag-MBP and KTag-MBP (Figure 5.11C) still shows a 555 nm peak at the volume of ~19 ml, the same as the KTag/SpyTag-MBP only control in Figure 5.11B, suggesting that either SpyTag-MBP or KTag-MBP could be forming a complex stable to gel filtration with SpyLigase EQ-555.

Furthermore in Figure 5.11C SpyLigase-555 alone control shows a peak at ~19 ml, approximately the same size as the KTag/SpyTag-MBP alone controls in Figure 5.11B. Showing that SpyLigase-555 is in a multimeric state, though it is difficult to be certain whether this is two or three units of SpyLigase, as the shape of the multimer could influence the Stokes radius (Horiike et al., 1983). Also the reason for the multimerization is not the dye: SpyLigase EQ-555 elutes at ~22 ml, consistent with a monomeric form (Figure 5.11C).

5.9 Applications of SpyLigase:

We applied SpyLigase to create functional polymers to capture circulating tumor cells (CTCs) (Zhao et al., 2012). CTCs are tumor cells present in the blood that may help

early detection of cancer (Yu et al., 2011). However, CTCs are rare cells, with sometimes only one or two amidst a billion leukocytes (Miller et al., 2010). Thus we developed protein polymers to enhance magnetic bead interaction with CTCs. These polymers were based on affibodies, which are small domains that can be expressed in *E. coli* composed of three α-helices, with antibody-like affinity for a target (Friedman et al., 2007, 2008). The affibody targeted the epidermal growth factor receptor (EGFR), commonly over-expressed in cancer (Friedman et al., 2007, 2008). The hypothesis was that polymerization of the affibodies would improve the isolation of EGFR-expressing cells, allowing for a more effective way to isolate CTCs for detection of tumor growth.

5.9.1 Principle of Polymerizing Anti-EGFR Affibody via SpyLigase:

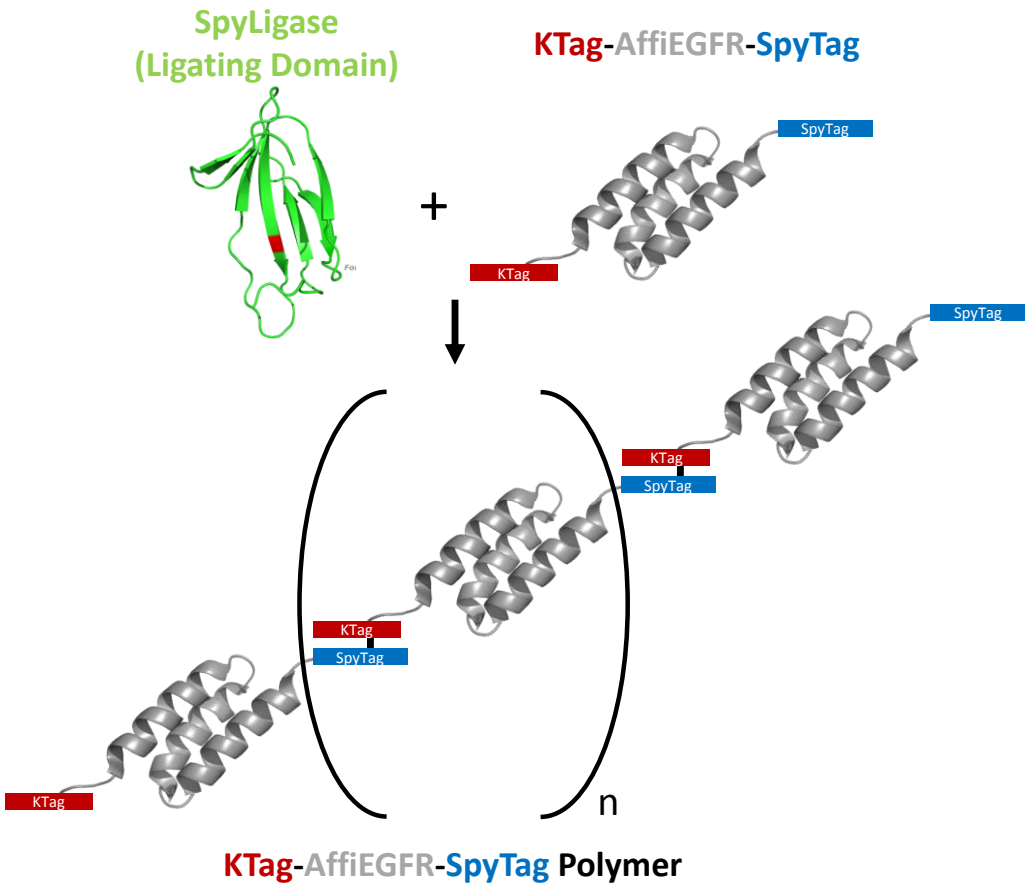


Figure 5.12: Cartoon of the polymerization reaction of KTag-AffiEGFR-SpyTag by SpyLigase.

The basis of forming protein polymers required two modifications of the affibody: KTag was fused to the N-terminus and SpyTag to the C-terminus of the anti-EGFR affibody (Figure 5.12). The fusion of KTag and SpyTag created a new minimally modified fusion protein called KTag-AffiEGFR-SpyTag which could act as a monomer. Then, following incubation with SpyLigase, affibody polymers could form with a specific head-to-tail orientation.

5.9.2 Polymerization of KTag-AffiEGFR-SpyTag:

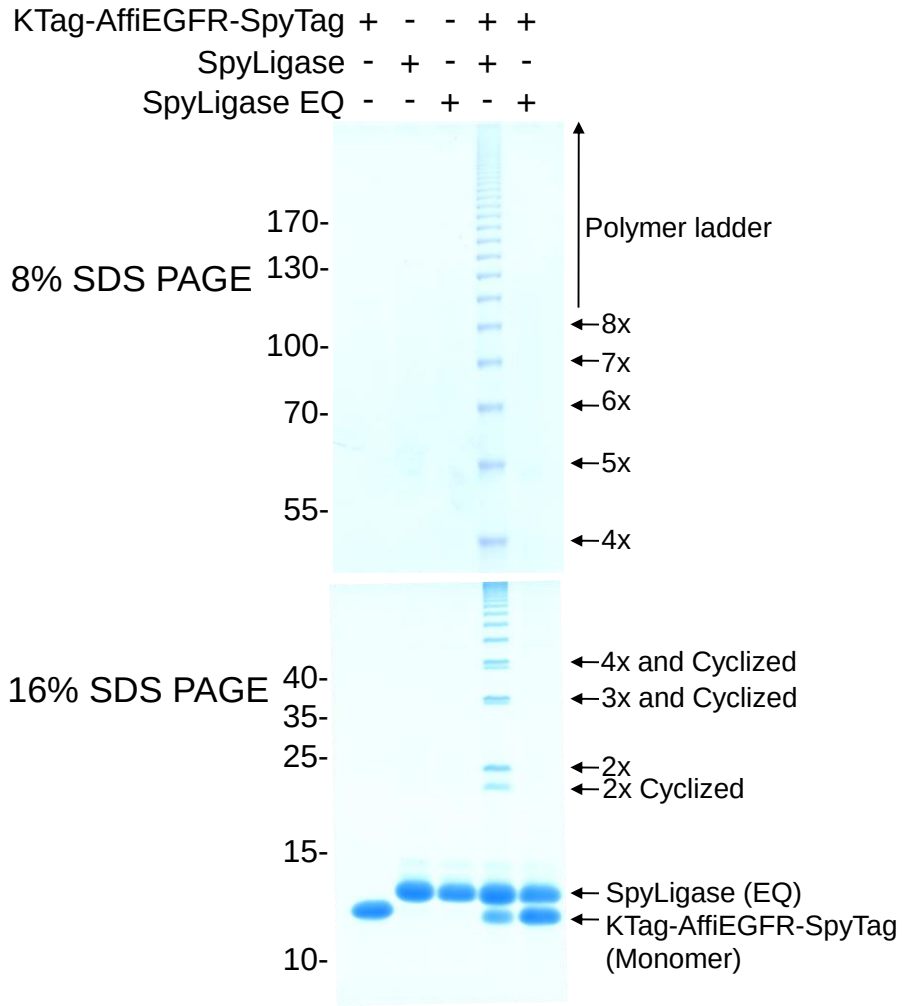


Figure 5.13: Polymerization of KTag-AffiEGFR-SpyTag by incubation with SpyLigase. Lower panel is a 16% SDS-PAGE to show lower molecular weight polymers, while the upper panel is an 8% SDS-PAGE to show the higher molecular weight polymers. Reaction conditions were 1 M TMAO, 40 mM Na_2HPO_4, 20 mM Citric Acid, 100 µM SpyLigase (EQ) and 37.5 µM KTag-AffiEGFR-SpyTag. Samples were incubated at 4°C for 24 h, with the reactions stopped with SDS loading dye, run on SDS-PAGE, and Coomassie stained for analysis.

The polymerization of KTag-AffiEGFR-SpyTag was effective (Figure 5.13). The polymers gave distinct bands on SDS-PAGE and the polymers reached molecular weights >170 kDa (Figure 5.13). There seems to be side-reactions, most likely cyclization as the bands run nearly at the same molecular weight but have different mobility's as seen in the lower panel for the 16% SDS-PAGE (Figure 5.13), with the dimer, trimer, and tetramer forms. For higher molecular weight polymers only one band at each height can be seen. One possibility is that the resolution with the SDS-PAGE at 16% is too low to separate between linear and cyclized polymers of ≥5 units. The second possibility is that once over the pentamer size, the probability of the ends of the same linear polymer meeting becomes small.

5.9.3 Time Course for the Polymerization of KTag-AffiEGFR-SpyTag by SpyLigase:

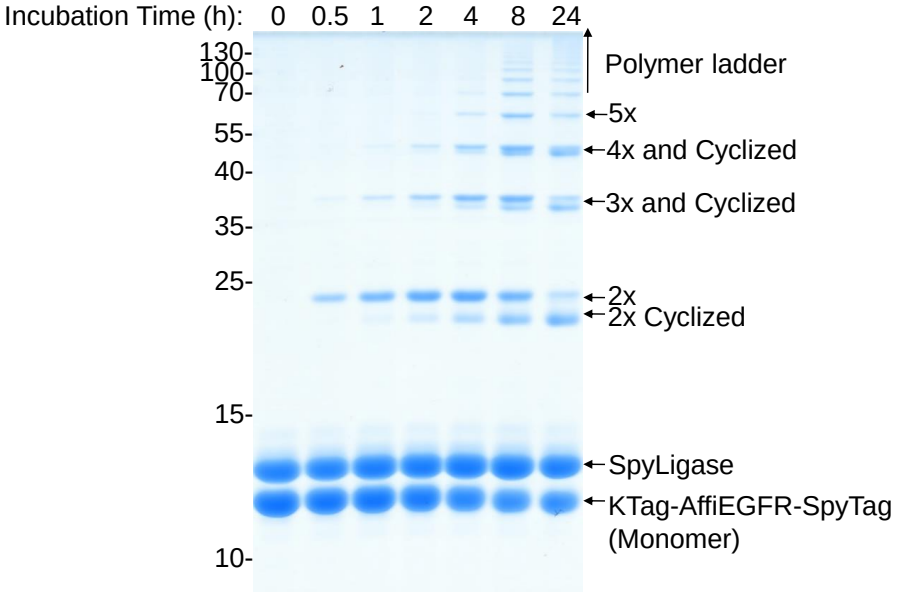
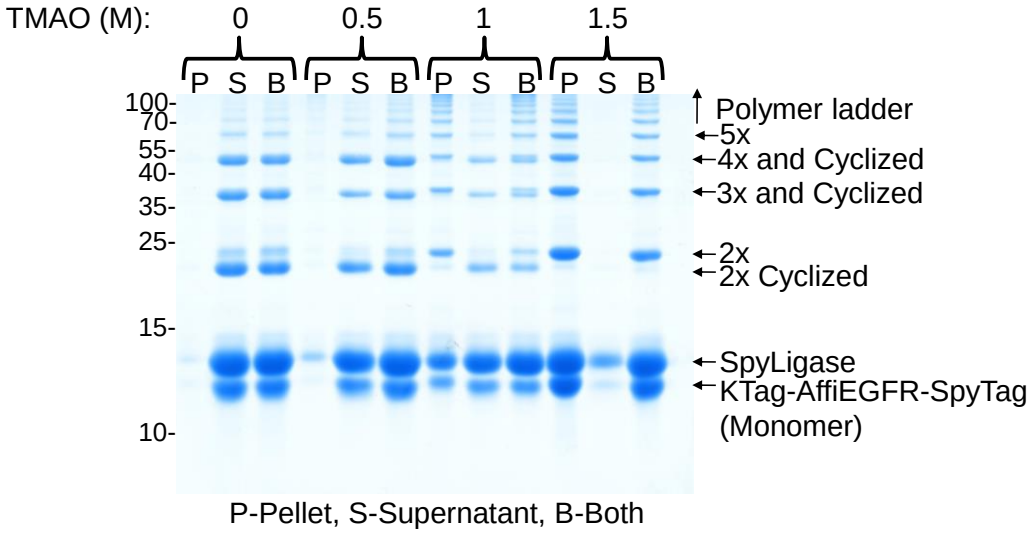


Figure 5.14: Time course of the polymerization of KTag-AffiEGFR-SpyTag catalyzed by SpyLigase. The reaction conditions were 1 M TMAO, 40 mM Na_2HPO_4, 20 mM Citric Acid, 100 µM SpyLigase (EQ) and 37.5 µM KTag-AffiEGFR-SpyTag at 4°C. The samples were analyzed by 16% SDS-PAGE and Coomassie staining.

The polymerization of KTag-AffiEGFR-SpyTag goes through a stepwise process (Figure 5.14). KTag-AffiEGFR-SpyTag forms dimers after 30 min and larger polymeric species then appear. The shorter time frame seems to favors the linear form of the

KTag-AffiEGFR-SpyTag polymers. Comparing 8 and 24 h incubation (Figure 5.14), at 24 h there is more of the putative cyclized versions of dimer, trimer, and tetramer. The longer incubation may give more time for the ends of a linear polymer to meet.

5.9.4 Influence of TMAO Concentration on Polymerization of KTag-AffiEGFR-SpyTag and Solubility of KTag-AffiEGFR-SpyTag Polymers:



P-Pellet, S-Supernatant, B-Both

Figure 5.15: Titration of TMAO to determine influence on affibody polymerization. Reactions were performed with 40 mM Na_2HPO_4, 20 mM Citric Acid, 100 µM SpyLigase and 38 µM KTag-AffiEGFR-SpyTag for 24 h at 4°C. For the pellet and supernatant lanes the reaction tube was spun down at max speed for 10 min to pellet a visible white solid found during the polymerization. 'B' for both was the same composition as the Pellet/Supernatant reactions except that the SDS loading dye was directly added to the KTag-AffiEGFR-SpyTag polymerization after the 24 h incubation period.

Initial work with the polymerization of KTag-AffiEGFR-SpyTag revealed a precipitate at the end of the incubation period. Therefore, a titration of TMAO was used to determine whether the chemical chaperone was involved. Figure 5.15 shows that as the TMAO concentration is increased, more of the KTag-AffiEGFR-SpyTag, both monomer and polymeric forms, are found in the pellet. On the other hand the increasing amount of TMAO also helps to create larger polymers of KTag-AffiEGFR-SpyTag, as with 0 and 0.5 M TMAO in the polymerization reaction there is barely any polymer chain over four units long. Once the threshold of 1 M TMAO is crossed, higher molecular

weight species can be seen, but are predominantly in the pellet at 1 M TMAO and only detectable in the pellet at 1.5 M TMAO. This leads to a quandary, as to achieve large polymeric species it would require the polymer chains to have low solubility. Also how is it possible to continue to grow the affibody polymers if the KTag-AffiEGFR-SpyTag polymers have aggregated? The N- and C-termini may still be accessible to SpyLigase while other parts of the polymer have bundled together. Therefore polymer solubility required further investigation.

5.9.5 Incubation of KTag-AffiEGFR-SpyTag Polymers at 37°C to Observe the Effect on the Insoluble Fraction of Polymers:

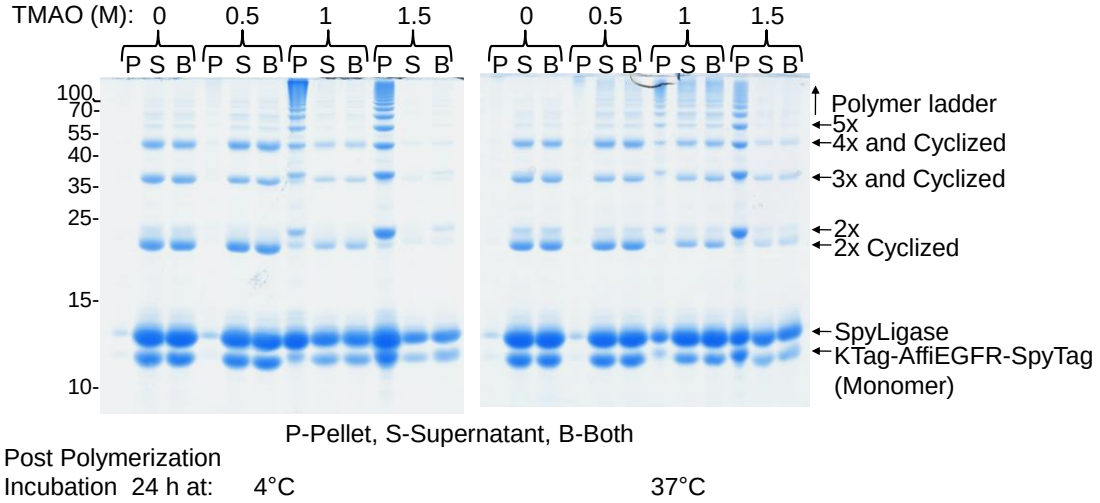


Figure 5.16: Determining whether the low incubation temperature of the polymerization of KTag-AffiEGFR-SpyTag is leading to accumulation of high molecular weight polymers in the insoluble fraction. Titration of TMAO for the polymerization reaction as in Figure 5.15. Following the initial 24 h incubation at 4°C, samples were then either kept for another 24 h at 4°C (left panel) or incubated at 37°C (right panel).

The issue of the KTag-AffiEGFR-SpyTag high molecular weight polymers becoming insoluble led me to characterize the polymers further. I repeated Figure 5.15 but after the initial 24 h polymerization at 4°C, samples were either incubated for an additional 24 h at 4°C (Figure 5.16 left panel) or 37°C (Figure 5.16 right panel). For 0 and 0.5 M TMAO the 24 h incubation at 37°C did not have an effect. For 1.5 and more clearly 1 M TMAO, the additional incubation at 37°C helped make some of the higher

molecular weight polymers soluble (Figure 5.16 right panel). Enhancing solubility by the minor change of incubation at 37°C signals that the insoluble fraction previously observed is not an aggregate akin to an inclusion body, which would require more elaborate methods to solubilize (Dashivets et al., 2009; Georgiou and Valax, 1999). The more likely scenario is that KTag-AffiEGFR-SpyTag polymers form a complex without any protein unfolding but with numerous weak interactions leading to the formation of an insoluble solid.

For Figure 5.16 lanes labeled ‘Both’, it was challenging to obtain samples with reproducible total protein amounts. The problem is most likely that simple pipetting up-and-down is not enough to evenly break up the aggregates. Another aspect is that the aggregates may stick to the surface of the tube.

5.9.6 Application of KTag-AffiEGFR-SpyTag Polymers in Improving the Capture of a Cancer Cell Line with Low Expression of EGFR:

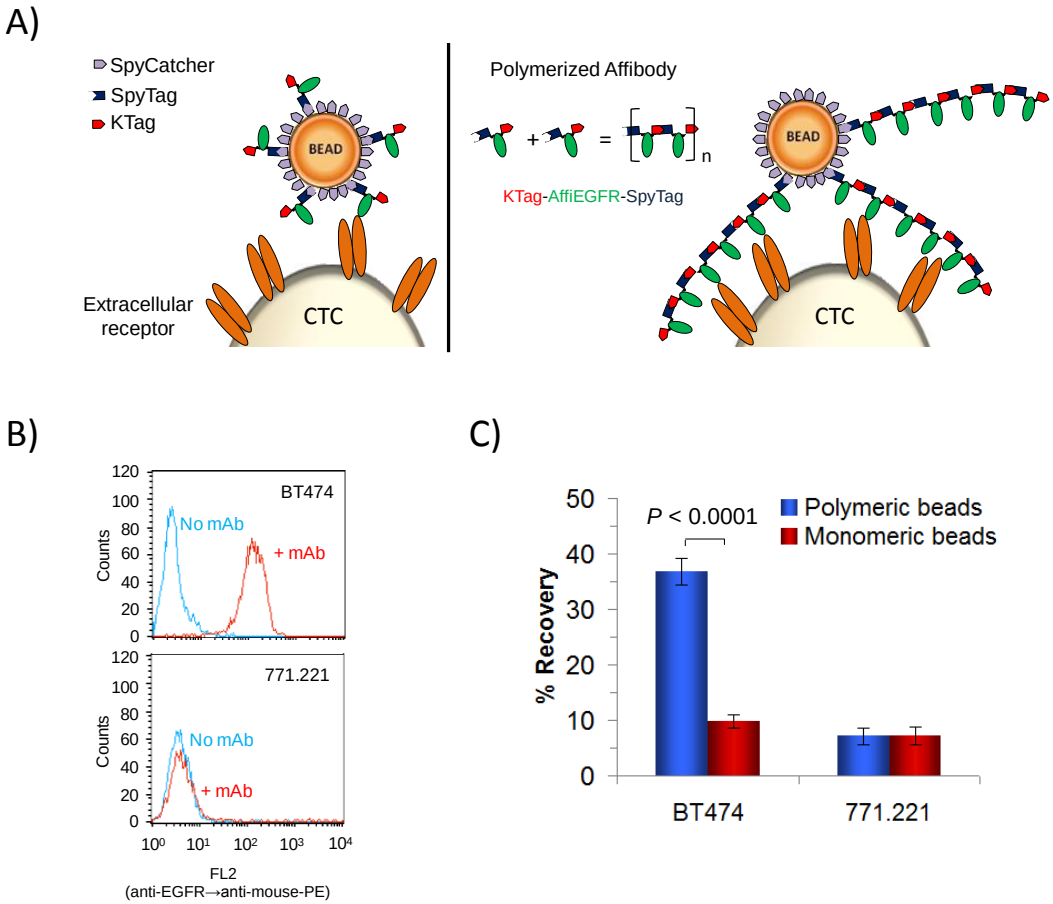


Figure 5.17: Improving isolation of cancer cell-line BT474 with low EGFR expression levels. A) Cartoon of the hoped effect of using polymerized anti-EGFR affibodies compared to the usual monomer forms, attached to magnetic beads for cell capture. B) Flow cytometry to determine expression levels of EGFR on the different cancer cell-lines. The blue line represents where no primary anti-EGFR antibody was used (only secondary), and the red line represents when the primary anti-EGFR antibody was included. C) Comparing the recovery of cancer lines using magnetic beads labeled with KTag-AffiEGFR-SpyTag polymers (blue bars) or with only monomer KTag-AffiEGFR-SpyTag (red bars). Mean of triplicate ±1 standard deviation. SpyLigase and KTag-AffiEGFR-SpyTag proteins were supplied by me. However, all experimental work of polymerizing the affibody on the magnetic beads, flow cytometry, capture experiments, and cartoon drawing were done by Mr. Gianluca Veggiani in the Howarth laboratory.

The principle of the application of polymerized KTag-AffiEGFR-SpyTag done by Mr. Gianluca Veggiani can be seen in Figure 5.17A. The idea was to change from the standard means of CTC isolation via magnetic beads, in which only a monolayer of affibodies (or more commonly antibodies) (Jain et al., 2013) coat the magnetic bead

(Figure 5.17A left panel), to polymerized affibodies to increase the avidity between the target cell and magnetic beads (Figure 5.17A right panel). KTag-AffiEGFR-SpyTag was shown to effectively polymerize on magnetic beads (Appendix Figure A.9). Polymerization was an ordered linear process, ligating KTag-to-SpyTag and ensuring that the polymers were all the same orientation on the magnetic beads; the beads themselves were coated with SpyCatcher, to anchor the SpyTag end of the KTag-AffiEGFR-SpyTag polymers (Figure 5.17A right panel).

Then the cancer cell lines were tested for EGFR expression levels. Figure 5.17B shows that breast cancer cell-line BT474 have low levels of EGFR (in comparison to another breast cancer cell-line such as MDA-MB-231), while 771.221 cells (B cell-line) have undetectable levels of EGFR, making them an effective negative control (Jain et al., 2013). The cells were then tested via a magnetic capture assay: the KTag-AffiEGFR-SpyTag-labeled magnetic beads were either incubated with SpyLigase to create affibody polymer coated beads (Figure 5.17A right panel) or SpyLigase EQ to act as a negative ligation control where the beads would only have monomer KTag-AffiEGFR-SpyTag coated on the magnetic beads (Figure 5.17A left panel). Then both polymeric and monomeric magnetic beads were incubated with the BT474 and 771.221 cells. Following magnetic isolation, the isolated cells were counted (Figure 5.17C). In the case of the low expressing BT474 cell line, there was a large improvement due to the use of the polymerized KTag-AffiEGFR-SpyTag relative to the monomeric form ($P<0.0001$, $n=3$, unpaired t-test) (Figure 5.17C). The monomer KTag-AffiEGFR-SpyTag-covered beads can barely capture the BT474 cell line above background levels, when comparing to the negative control 771.221 (Figure 5.17C). Therefore the polymerization of the anti-EGFR affibody helped to improve the capture of the low expressing cell lines.

5.10 Discussion:

The development of a three-part split protein (Figure 5.1C) for ligation of peptides is a new development in terms of chemical biology tools. The system functions to create an isopeptide bond using the same 'isopeptide triad' chemistry as SpyCatcher-SpyTag (Zakeri et al., 2012). There are alternative systems for protein ligation. The primary examples are split inteins, which allow for covalent ligation between the N- and C-termini of two fragments which self-cleave without a trace (Lockless and Muir, 2009). Sortagging is an additional method, using coincidentally another Gram-positive bacterial protein, sortase. Sortase can ligate the C-terminal Sortag, LPXTG, onto an N-terminally located oligoglycine-tagged target protein (Popp et al., 2007; Tsukiji and Nagamune, 2009). Other options exist based on the chemistry of the sulfhydryl group of Cysteine. Native chemical ligation fuses two target proteins via a thioester transfer to create a regular peptide bond (Kent, 2009). Lastly a covalent option based on disulfide bridges exists: the 'dock-and-lock' system uses the kinase A domain and its corresponding anchor domain, which has been mutated to include Cysteines that covalently lock the two proteins together after binding of the kinase A domain with its anchor domain (Chang et al., 2007).

However, there are many factors about SpyLigase that distinguishes it from these aforementioned protein ligation systems. The methods described above require that each reactive section is specifically at the N- or C-terminus of the target protein (Chang et al., 2007; Kent, 2009; Lockless and Muir, 2009; Popp et al., 2007; Tsukiji and Nagamune, 2009). The SpyLigase system is more flexible as SpyLigase can ligate the SpyTag and KTag pair regardless of whether one or both are located at the N- or C-termini (Figure 5.7). SpyLigase can also ligate proteins with SpyTag located on linkers between protein domains (Figure 5.8). The value of SpyLigase's flexibility is it can be used as a protein-based crosslinker, such as if studying how a supramolecular structure

is formed by allowing for the placement of SpyTag/KTag at varying locations in a monomer unit to observe interactions and possible contacts (Preston et al., 2012). The other advantage of SpyLigase is that Sortagging, Native chemical ligation, 'dock-and-lock', and many split inteins depend on an appropriate oxidation state of Cysteine (Chang et al., 2007; Fu et al., 2007; Kent, 2009; Popp et al., 2007; Shah and Muir, 2011; Tsukiji and Nagamune, 2009). As the CnaB2 domain is devoid of cysteine residues and the reaction has been done in reducing conditions with SpyCatcher-SpyTag, it is likely that reducing agents would not influence the SpyLigase reaction (Oke et al., 2010; Zakeri et al., 2012).

On the other hand it should be clear that SpyLigase is not without fault as a protein-protein ligation system. SpyLigase must be in excess compared to the KTag and SpyTag fusions (Figure 5.3), though a saving grace is that SpyLigase expresses well with ≥9 g/L of protein per culture purification. The requirement of TMAO for high yield is not ideal. Even though for many *in vitro* reactions TMAO would likely not interfere, there is always a preference to keep reactions as simple as possible and TMAO is not practical for cellular applications. An additional consideration is the fact that SpyLigase acts as a single-turnover catalyst, in which SpyLigase does not release the covalent adduct (Figure 5.10). A ~13 kDa domain wrapped around a SpyTag/KTag ligation point could interfere with protein-protein interactions. Bound SpyLigase could also possibly act as a point where new protein interactions could occur inside cells. Though activity at 4ºC has the advantage of being gentle for purified proteins, the system would be more versatile if SpyLigase was efficient up to 37ºC. The final issue is the rate of reaction, which is ~50% yield in 24 h (Figure 5.6). The rate is slow especially when comparing to SpyCatcher-SpyTag, which can accomplish 50% reconstitution within ~1 min (Zakeri et al., 2012). Also, certain split inteins have been shown to go to near-completion within 2 min (Shah et al., 2012).

However, even with these limitations, SpyLigase was still capable of being applied in a useful fashion for polymerizing proteins. There are obvious uses in terms of biomaterials with this simple process of polymer creation. These potential applications range from tissue engineering to create environments for cells to grow and develop within, to the more industrial applications of acting as a scaffold for various enzymes to create catalytic nano-hubs (Amiram et al., 2011; Heyman et al., 2007; Vasita and Katti, 2006). However, we focused on creating multivalent protein polymers, using tagged affibodies, analogous to polymerized DNA aptamers which were used for cell capture (Zhao et al., 2012). The polymerization of the recombinant KTag-AffiEGFR-SpyTag proved simple and robust (Figure 5.13). Though the polymers themselves pellet upon centrifugation (Figure 5.15), it is not clear that the polymers are forming an insoluble aggregate (Figure 5.16). SpyLigase was still able to extend these oligomers to create longer polymers. From 8% SDS-PAGE we were not able to determine a size limit, as the polymers reached well over 170 kDa (or over fourteen KTag-AffiEGFR-SpyTag units long). The interaction between KTag-AffiEGFR-SpyTag polymer chains is not strong since the interaction seems to be partially broken by incubation at 37°C (Figure 5.16). In addition, when KTag-AffiEGFR-SpyTag was polymerized onto magnetic beads (Appendix Figure A.9), the KTag-AffiEGFR-SpyTag polymers were still functional (Figure 5.17).

The value of polymerized KTag-AffiEGFR-SpyTag can be seen when recovering cancer cell-lines with low expression of EGFR (Figure 5.17C). What is the mechanism helping to improve the binding? One possibility is by casting a ‘wider net’, whereby each bead has more opportunities for interactions with EGFR (relating to a change in on-rate). The second possible effect of polymerization is increasing the number of interaction events per cell, creating a stronger binding of the magnetic bead with the cancerous cell (relating to a change in off-rate) (Figure 5.17A right panel). Even without fully

understanding the mechanism, SpyLigase-mediated polymerization represents a simple way to improve the cellular binding of affibodies (and possibly antibodies in the future).

Despite the fact that this current iteration of SpyLigase has shown some value in cell capture, the question remains as to how to improve the SpyLigase system. Though there has already been much protein engineering on SpyLigase (Appendix Figure A.3 and A.4), these variants of SpyLigase only reached just over 50% yield. The answer I believe may be like the leap from the Pilin-C/Isopeptag to SpyCatcher-SpyTag system (Zakeri and Howarth, 2010; Zakeri et al., 2012). Like Pilin-C, SpyLigase is a good proof of principle and simply the first successful iteration. However, to truly develop a protein ligation system equivalent in speed to other protein-protein ligation systems such as split inteins (Shah et al., 2012) one may need to explore different isopeptide containing protein domains. One major issue with SpyLigase is that the two peptide tags are not equally distributed across the CnaB2 domain: the β-strand that comprises SpyTag is located directly at the C-terminus of CnaB2. The β-strand comprising KTag is located more internally in the sequence and structure of CnaB2 (Appendix Figure A.5). Previous work attempted to split CnaB2 to create a 'reverse-SpyCatcher' or orthogonal system, KCatcher, where the protein domain containing the Aspartate and Glutamate could react with the Lysine on KTag. KTag and KCatcher had slow and low yielding reconstitution (Appendix Figure A.10). Only through engineering the SpyLigase domain (Appendix Figure A.3 and A.4), addition of a chemical chaperone (Figure 5.4), and various other improvements was the SpyLigase domain able to achieve >50% yield. Thus the lesson that was learned in the development of the SpyLigase system is that maintaining the core structure of the domain is crucial to achieve activity based on split protein reconstitution. Therefore, the data signals that the most effective way to improve the activity of SpyLigase is to find an alternative protein domain. The main feature that would be emphasized for this new SpyLigase protein domain is that the two reactive residues

of the ‘isopeptide triad’ (Lysine and Aspartate/Aspargine (Hagan et al., 2010; Kang et al., 2007)) would not be on regions that would be disruptive to the overall folding of the domain if removed. This may be a difficult requirement when examining all the spontaneous isopeptide-containing protein domains (Vengadesan and Narayana, 2011). Therefore the development of an improved SpyLigase system may require a combined approach of trying multiple domains and then also performing circular permutations to observe whether shifting the N- and C-terminal ends of the domains can stabilize the double splitting (Yu and Lutz, 2010).

However, a parallel approach to improve upon the current SpyLigase would be a two-part directed evolution campaign (Jäckel et al., 2008). Here both SpyCatcher/SpyTag and KCatcher/KTag would be evolved separately via phage display (Burton, 1995) and the information gleamed from the improved variants would be applied to SpyLigase. It may be asked why not evolve SpyLigase directly? That would be problematic as SpyLigase-induced ligation is unlike an enzymatic reaction, where an alternative substrate can be used to create an easy reporter signal (Wong et al., 2004).

Though there is much to improve in SpyLigase, SpyLigase was a useful exploration and proof of concept of the feasibility of creating multi-split proteins that can reconstitute and then ligate proteins.

Chapter 6: General Conclusions

Chapter 6: General Conclusions

This thesis had two focuses. The first was the characterization of the SpyCatcher-SpyTag covalent reconstitution. The second was the generation of a new molecular tool, the SpyLigase domain that can perform peptide-peptide ligation of SpyTag to a new peptide tag, KTag. Both approaches were based on the natural protein chemistry available from the collagen adhesion binding domain (CnaB2) from *S. pyogenes* (Hagan et al., 2010; Oke et al., 2010; Zakeri et al., 2012). The value of these protein-domain (SpyCatcher-SpyTag) and peptide-peptide ligation (SpyLigase) systems is that they fill in gaps where currently available peptide tags are limited and also create flexible alternatives to other parallel protein ligation systems.

6.1 Characterization of SpyCatcher-SpyTag:

The initial work of splitting and engineering the CnaB2 domain along with some fundamental characterizations were performed by Dr. Bijan Zakeri. I followed up Dr. Zakeri's work with further extensive characterization.

The *in vitro* aspects tested of the SpyCatcher-SpyTag system fell under two categories, SpyTag location-dependence and reconstitution capacity in reducing conditions. SpyTag was shown able to reconstitute covalently with SpyCatcher regardless of whether SpyTag was located on the N-terminus, C-terminus, or on an exposed linker between two protein domains (Section 3.3). The *in vivo* aspects tested were that the two components could be expressed *in vivo* via well-established induction means without any requirement for cell membrane permeation, *in vivo* reconstitution can also be expanded to a more controlled interaction via the inclusion of a secondary induction system such as with arabinose (Khlebnikov et al., 2000). Still SpyCatcher-SpyTag reconstitution in the *E. coli* cytosol will still need to be better understood before becoming a fully-fledged *in vivo* labeling system in bacteria, as the silver staining and

immunoblotting of cell lysate containing co-expressed SpyCatcher-SpyTag showed that there was one side-reaction and some degradation of the reconstituted adduct (Section 3.4). In future work it would be important to test other *E. coli* strains to determine how to limit or circumvent the degradation problem. SpyCatcher-SpyTag also shows promise in the area of mammalian cell surface labeling, demonstrated by the HA-SpyTag-ICAM-EGFP and TfR-sfGFP-SpyCatcher fusion proteins. In both cases the fusions of SpyTag and SpyCatcher indicated covalent interaction on the surface of HeLa cells. The capacity to remain functional after traveling through the secretory pathway also indicates that SpyCatcher-SpyTag could function as *in vivo* labeling system in mammalian cells, and would be interesting to test as a secretory pathway labeling system.

ITC analysis showed that even though the SpyCatcher-SpyTag pair did not evolve for reconstitution, their dissociation constant with an inactive form of SpyTag(DA) was ~200 nM, which is equivalent to other peptide-protein dissociation constants (Houk et al., 2003). And that improving the rate of the SpyCatcher-SpyTag interaction requires that both the docking cleft and the microenvironment in which the 'isopeptide triad' function in must be considered.

6.2 Lessons about the SpyCatcher-SpyTag Interaction Based on Crystal Structures:

The crystal structures of CnaB2 (Oke et al., 2010) and SpyCatcherΔN1 reconstituted with SpyTag became the basis of a mutational campaign (Chapter 4). The purpose was to better understand the SpyCatcher-SpyTag interaction and trying to improve the reconstitution rate. Though the mutants of both SpyCatcher and SpyTag failed to achieve dramatic improvements in rate, SpyCatcher was greatly decreased in size and there is a better understanding of the interaction between SpyCatcher-SpyTag.

Even though Dr. Zakeri had great success with the I34E-M69Y mutations in the initial engineering of the SpyCatcher domain (Zakeri et al., 2012), all the subsequent

mutants I tried were either negligible in effect or injurious to activity. Thus for any future mutational campaigns it would be advisable to use a library-based evolution approach.

Though the mutational campaign brought limited results, the truncation of SpyCatcher was the real achievement in that the core domain of SpyCatcher was determined. Considering the reconstitution rates of the double N- and C-terminus truncated SpyCatchers (SpyCatcherΔN1ΔC1 and SpyCatcherΔN1ΔC2) along with the ordered regions in the crystal structure of SpyCatcherΔN1 reconstituted with SpyTag, the core domain of SpyCatcher is 83 residues (amino acids 22-104 of the original SpyCatcher). The shrinking of SpyCatcher to a minimal sequence that still retains nearly the same activity as the original SpyCatcher makes the domain more applicable as a fusion protein for cell biology applications and feasible to synthesize via solid-phase, allowing for the inclusion of unique chemistries, increasing SpyCatcher's versatility.

6.3 Developing a Tripartite Split Protein that can Ligate Peptides:

The process of splitting a protein into two parts such that they can reconstitute and restore protein function is not new, as it has been done with TEV protease (Wehr et al., 2006), GFP (Cabantous et al., 2005), β-Lactamase (Galarneau et al., 2002), and other proteins (Ozawa, 2006; Shekhawat and Ghosh, 2011). The splitting of proteins into three components has been done as well with GFP (Kellermann et al., 2013), Ribonuclease (Lin et al., 1970), inteins (Sun et al., 2004), and Cytochrome C (Juillerat et al., 1980). What is unique in the case of SpyLigase is that three separate components have been generated that allow for the ligation of two different peptide tags.

Even with SpyLigase's stringent reaction condition requirements, SpyLigase showed itself to be a useful and distinctive peptide-peptide ligation system, with the capacity to polymerize affibodies. The affibody polymers proved to be a useful method to increase magnetic capture of cell-lines expressing low levels of EGFR. These results

indicate the possible capacity of affibody polymers to isolate CTCs with lower tumor antigen expression (Miller et al., 2010; Yu et al., 2011).

6.4 Structural Lessons when Comparing and Contrasting Bipartite and Tripartite Splits of CnaB2:

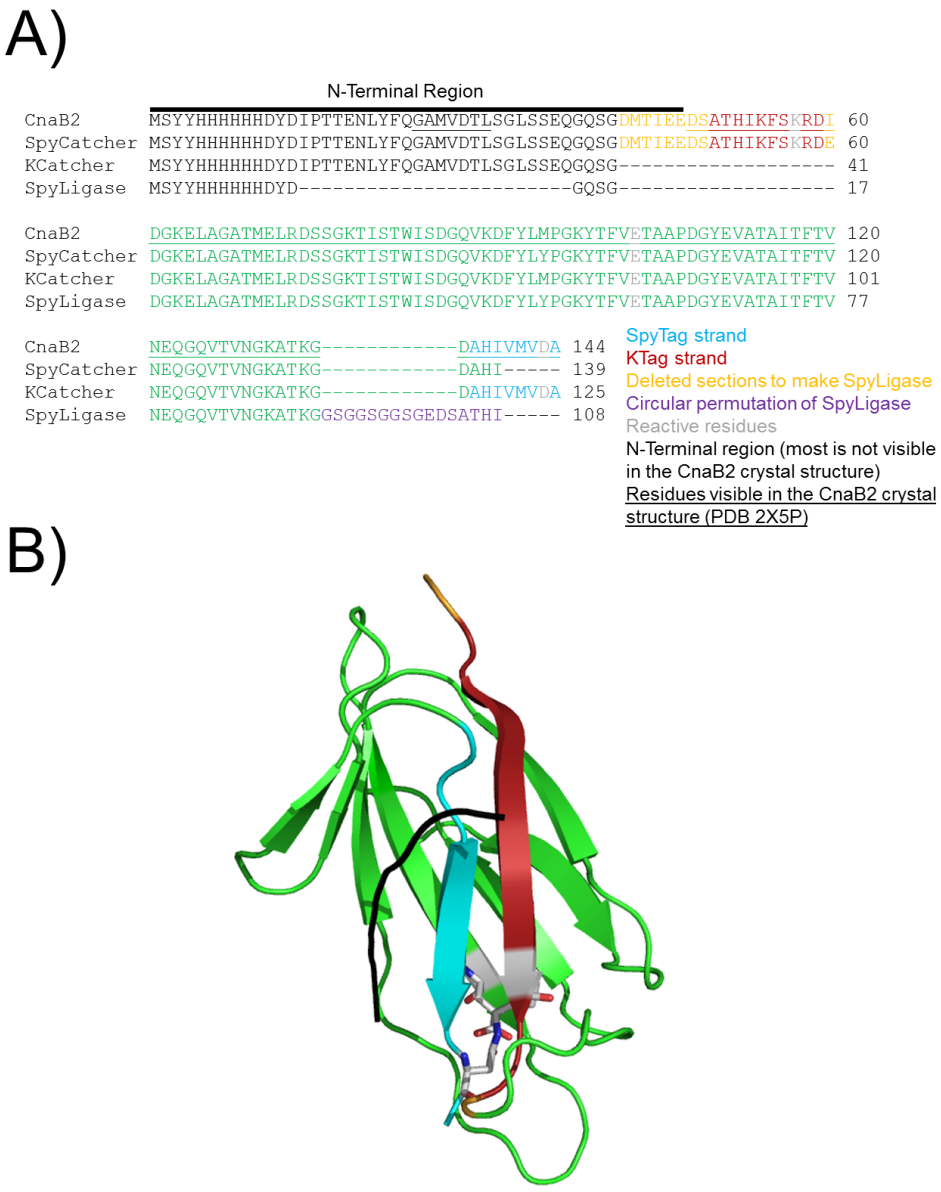


Figure 6.1: Highlighting the differences between the structures based on the CnaB2 domain (PDB 2X5P). A) Sequence alignment of CnaB2, SpyCatcher, KCatcher, and SpyLigase colored to distinguish the different features of each domain. B) CnaB2 crystal structure colored to emphasize the differences between CnaB2, SpyCatcher, KCatcher, and SpyLigase.

SpyCatcher-SpyTag and SpyLigase are based upon the same domain, CnaB2. Therefore the reactions use the same isopeptide triad. Thus why is SpyCatcher-SpyTag so much more robust and faster compared to SpyLigase? The obvious point is that SpyCatcher-SpyTag is bimolecular and SpyLigase is trimolecular. Therefore reconstitution of SpyCatcher-SpyTag requires simply that these two components find one another, dock, and then have the 'isopeptide triad' residues form an amide bond. The process for SpyLigase is more complicated and entropically less favorable compared to SpyCatcher.

While 2 versus 3-component aspect is part of the explanation for the differences in the reconstitution rates, it still does not explain why 50% reconstitution/ligation takes ~1 min for SpyCatcher-SpyTag (Zakeri et al., 2012) and 24 h for SpyLigase. The 2 versus 3-component aspect also does not explain why SpyCatcher-SpyTag can function in a range of buffers, temperatures, detergents, and only requires a 1:1 ratio of each component (Zakeri et al., 2012) while SpyLigase has stringent reaction conditions. An important consideration is the structural differences between the two domains. The KTag β-strand is a destabilizing truncation on the CnaB2 domain structure, as demonstrated with KCatcher-KTag which is quintessentially the 'reverse SpyCatcher-SpyTag' (Appendix Figure A.10A, B). KCatcher-KTag with the reactive Aspartate and catalytic Glutamate on the KCatcher domain and the reactive Lysine on the KTag peptide was inefficient with a yield of ~10% reconstitution after 10 h (Appendix Figure A.10C). The deleterious effect on the fold of the KTag β-strand truncation was shown further by CD (Appendix Figure A.11).

The likely reason why the removal of the KTag β-sheet is so detrimental to the domain structure is due to the geometry of CnaB2. During the ribosomal synthesis, there is co-translational folding (Fedorov, 1997). Therefore the truncation of the KTag β-strand could be causing problems with the folding pathways of KCatcher and SpyLigase

domains. Also the SpyTag β-strand is located adjacent to the KTag β-strand (Figure 6.1B). Thus during the translation of the CnaB2 domain the KTag β-strand could be acting as a 'guide' for the SpyTag β-strand positioning and structure. With the KTag β-strand removed it leads to improper folding and disordering of the SpyTag β-strand or its remnants in the case of SpyLigase, and subsequently destabilization of the KCatcher and SpyLigase domains. The requirement of SpyLigase to be in excess could hint that there are different folded populations of SpyLigase.

The major lesson that can be taken away from comparing bipartite with tripartite split CnaB2 domains is that the geometry of the domain itself is essential when considering the splitting of a protein domain.

6.5 Future Applications of Peptide-Peptide Ligation Technology:

The use of SpyCatcher-SpyTag as an alternative detection system to antibody-epitopes *in vitro*, possible use within the cytosol of *E. coli*, and a cell surface labeling system are already clearly shown (Zakeri et al., 2012).

Therefore I will focus on the SpyLigase system (Chapter 5), which although it had an application with the polymerizing of affibodies and improvement in CTC isolation, it is still important to expand on the possible uses of this system. In Figure 6.2 I have highlighted ideas for SpyLigase applications I consider feasible within the near future, especially with an enhanced SpyLigase system.

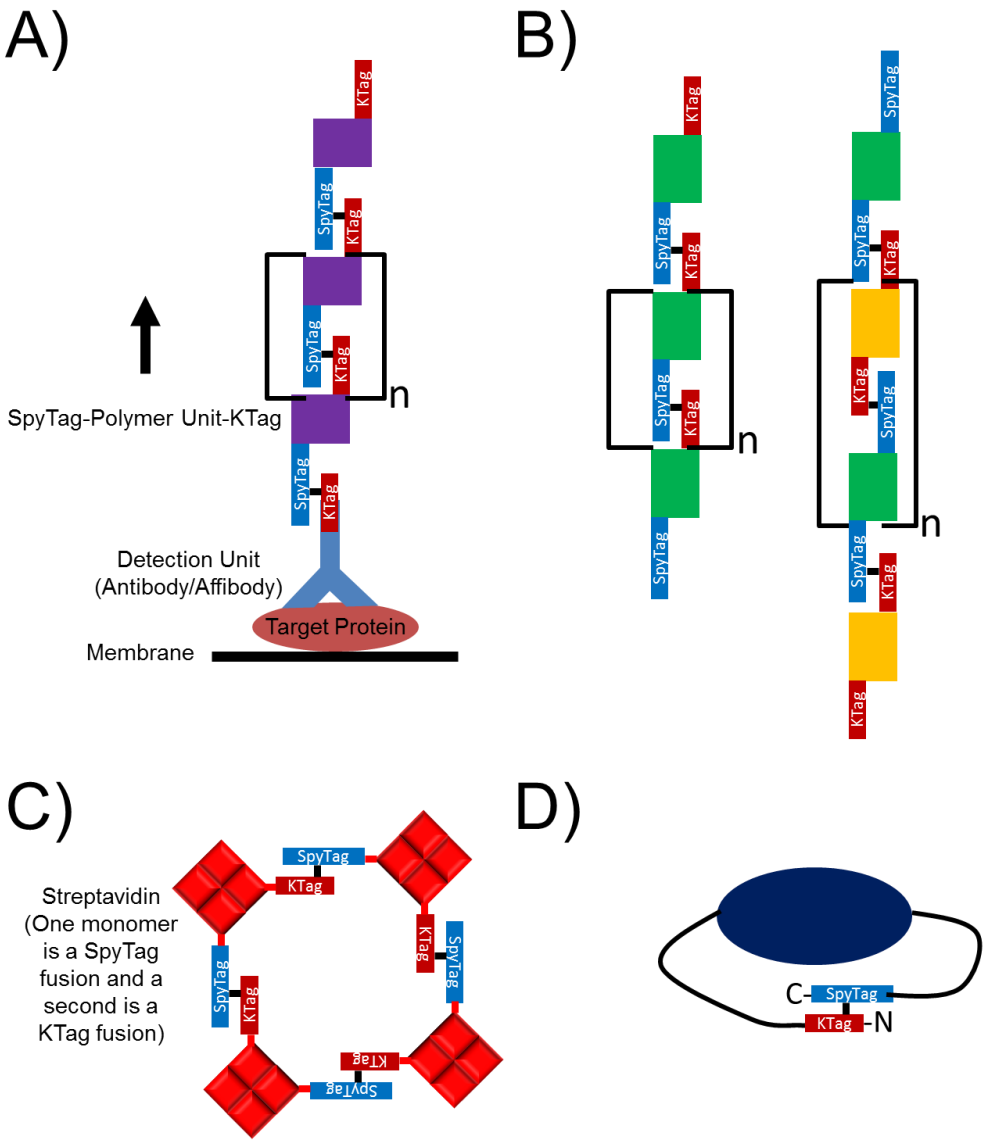


Figure 6.2: Cartoons of possible future applications of SpyLigase. A) Using SpyLigase to polymerize proteins for detection in immunoblots. B) Polymerization of one or two subunits to make different types of polymers and co-polymers. C) SpyTag-KTag fusions to different units of Streptavidin to develop unique architectures. D) Fusing KTag and SpyTag to the termini of a protein/peptide for cyclization.

Many of the uses of SpyLigase would be based upon the polymerization capacity already shown with affibodies. The first case would be to act as a new system in visualizing proteins for Western blotting (Figure 6.2A). Here an antibody/affibody labeled with SpyTag or KTag would be used for the initial detection of the desired protein.

Following the initial binding a protein acting as a monomer unit could be polymerized onto the detection unit. The advantage here is that technically any protein could be a monomer unit, ranging from an enzyme such as β-galactosidase for colorimetric detection (Porstmann et al., 1985) to GFP for amplification of a fluorescence signal (Tsien, 1998).

The next application would be again to use the polymerization capacity of SpyLigase for generating biomaterials (Figure 6.2B). The SpyLigase system could be used to create ordered single polymers or co-polymers based on protein domains (Figure 6.2B) (Bhushan and Schricker, 2013). Although many covalently-linked protein-based biomaterials have been generated previously, they tend to be crosslinked using chemical crosslinkers such as glutaraldehyde which would make it difficult to create an ordered polymer due to the wide availability of reactive groups on a protein surface (Sengupta and Heilshorn, 2010). Though it should be noted that ordered covalent polymers have also been produced with the consensus tetratricopeptide repeat (CTPR) monomer using a combination of native chemical ligation and expressed protein ligation (Phillips et al., 2012). However, the biomaterials must be based upon the CTPR monomer unit and reducing agent was needed to generate large amounts of polymers >60 kDa (Phillips et al., 2012). Thus SpyLigase could be filling in a gap of creating new types of protein-based biomaterials which are highly ordered and independent of the protein units composing the polymers. However, this is not the only application for protein-based polymers, as there is value for crosslinked enzymes, in particular CLEAs or crosslinked enzyme aggregates. CLEAs are enzymes that have been stabilized by crosslinking one enzyme to another in a large cluster to increase their robustness under a wide array of conditions (Häring et al., 1999). Polymerizing of enzymes could help to stabilize the structures and as shown by Mr. Gianluca Veggiani it is possible to anchor the polymers to magnetic beads, therefore setting an interesting precedent for anchoring

protein polymers to other solid matrices, which would be useful in separating an enzyme-catalyst from a substrate (Figure 5.17 and Appendix Figure A.9).

Though polymerization of enzymes is a very tempting target, the inclusion of the peptide tags or the polymerization itself could inhibit the enzymatic activity. Therefore it would be of interest to create a protein scaffold that would be robust yet easily modifiable and can possibly form unique 3D architectures. In the case of Figure 6.2C Streptavidin would be the ideal scaffold. The primary advantage of Streptavidin, besides being a common tool (Diamandis and Christopoulos, 1991; Green, 1990), is that it is modifiable. While Streptavidin is a tetramer, the composition of the tetramer can be modified such that only one of the subunits is active, creating a monovalent tetramer (Howarth et al., 2006). Dr. Michael Fairhead from the Howarth group has made it possible to efficiently isolate different combinations of inactive and active units and it would not be a large technical gap to create tetramer units composed of KTag and SpyTag fusions mixed with wild type units (Fairhead et al., 2013). The new Streptavidin tetramers with SpyTag and KTag fusion subunits could then be used as molecular building blocks to create unique scaffolds, which could then be functionalized with nearly any biotinylated protein. As an alternative to Streptavidin, unique architectures could also be generated through the peptide polymerization system of the self-assembling fiber (SAF) system, which can form long self-assembling coiled-coil rod polymers (Papapostolou et al., 2007; Smith et al., 2004). SAF could be used in conjunction with the SpyLigase system, as the isopeptide bond would allow for perpendicular extensions from the main SAF chain (Section 5.7.2). There has already been inclusion of peptide tags (FLAG) in the SAF system with Fiber Recruiting Peptides (FiRe) (Ryadnov and Woolfson, 2004). The system could also be further expanded due to the work of mixing and matching coiled-coils allowing for precision assembly of three dimensional spheres (Fletcher et al., 2013). Thus it could be feasible to create lateral and covalently linked

alignments of fibrils as a kind of ‘molecular chain linked fence’ or functionalized macromolecular spheres with specific modifications attached onto the surface irreversibly.

The final application I will consider is the possibility of SpyLigase-mediated protein cyclization (Figure 6.2D). Cyclizing a protein can increase the stability of a protein and resistance to proteolysis (Camarero and Muir, 1999; Clark and Craik, 2010; Conlan et al., 2010). Although work has already been done on cyclizing proteins using split inteins and native chemical ligation (Camarero and Muir, 1999; Clark and Craik, 2010; Xu and Evans, 2001), it would also be of interest to perform protein locking that is not only head-to-tail, as not all proteins have their N- and C-termini conveniently located adjacent to one another (Kent, 2009; Lockless and Muir, 2009). Thus SpyLigase could be used to ‘staple’ disordered termini or even loops onto the surface of a domain to increase the structural stiffness and stability of different enzymes and therefore expand the number of possible proteins for cyclization.

6.6 General Strategies of Improving the Reconstitution and Ligation Capacity of Bipartite and Tripartite Split Proteins:

The process of improving reconstitution and ligation of bipartite and tripartite-split isopeptide-containing domains is not straightforward. One problem is that there is limited background information in terms of splitting isopeptide-containing domains.

The first strategy is directed evolution using phage display (Joyce, 2007; Winter et al., 1994). The advantage of an *in vitro* evolution system is that it can be unbiased. The disadvantage of phage display is that, at least initially, the system would have to be focused on only bipartite splits. Directed evolution of a tripartite system using an alternative method such as double emulsion technology (Aharoni et al., 2005) would be just as challenging due to the requirement of developing a high-throughput assay of peptide-peptide ligation. To improve SpyLigase it would be more feasible to initially

perform directed evolution on SpyCatcher-SpyTag and KCatcher-KTag. Then using the mutational information and improved variants, these could act as a basis to improve SpyLigase.

The second strategy would be a general splitting screen of isopeptide-containing domains. Isopeptide bonds are known in several Gram-positive bacterial pilin and adhesin proteins (Vengadesan and Narayana, 2011). Therefore a screen of splitting many domains containing intramolecular isopeptides and testing the reconstitution (bipartite splits) or ligation (tripartite splits) would be a valuable means of increasing our insight.

The third strategy would be to generate *de novo* isopeptide bond formation in a domain. One would use a computational biology approach of screening other proteins for appropriate environments that could house the 'isopeptide triad'. Thus the screen would consider not only primary sequence information but also the nearby residues that create the hospitable environment for the amide bond formation. The next step would then be mutating into the target proteins the 'isopeptide triad' to see if it is possible to get a *de novo* reaction. Unlike an enzyme, where it would be possible to detect even low activity, in this case unless a certain number of the expressed domains were able to form the isopeptide, it would be difficult to see a band shift even with silver staining (Merril et al., 1988).

6.7 Final Remarks:

The purpose of this thesis was the characterization and development of technologies based on the splitting of the CnaB2 domain, which were SpyCatcher-SpyTag and SpyLigase. Therefore the importance of this work was characterizing and creating chemical biology tools that will allow for new kinds of biological experiments and the creation of protein architectures that were not previously feasible.

References

Abe, H., Wakabayashi, R., Yonemura, H., Yamada, S., Goto, M., and Kamiya, N. (2013). Split Spy0128 as a potent scaffold for protein cross-linking and immobilization. Bioconjug. Chem. *24*, 242–250.

Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-throughput screening of enzyme libraries: thiolactonases evolved by fluorescence-activated sorting of single cells in emulsion compartments. Chem. Biol. *12*, 1281–1289.

Alegre-Cebollada, J., Badilla, C.L., and Fernández, J.M. (2010). Isopeptide bonds block the mechanical extension of pili in pathogenic Streptococcus pyogenes. J. Biol. Chem. *285*, 11235–11242.

Amiram, M., Quiroz, F.G., Callahan, D.J., and Chilkoti, A. (2011). protein polymers. Nat. Mater. *10*.

Anraku, Y., Mizutani, R., and Satow, Y. (2005). Protein splicing: its discovery and structural insight into novel chemical mechanisms. IUBMB Life *57*, 563–574.

Bailey, A. (1991). The chemistry of natural enzyme-induced cross-links of proteins. Amino Acids *1*, 293–306.

Bhagawati, M., Lata, S., Tampé, R., and Piehler, J. (2010). Native laser lithography of His-tagged proteins by uncaging of multivalent chelators. J. Am. Chem. Soc. *132*, 5932–5933.

Bhushan, B., and Schricker, S.R. (2013). A review of block copolymer based biomaterials that control protein and cell interactions. J. Biomed. Mater. Res. A 1–30.

Brizzard, B. (2008). Epitope tagging. Biotechniques *44*, 693–695.

Burton, D. (1995). Phage display. Immunotechnology *1*, 87–94.

Busch, G.G.K., Tate, E.W.E., Gaffney, P.P.R.J., Rosivatz, E., Woscholski, R., and Leatherbarrow, R.J. (2008). Specific N-terminal protein labelling: use of FMDV 3C pro protease and native chemical ligation. Chem. Commun. (Camb). 3369–3371.

Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. Nat. Biotechnol. *23*, 102–107.

Camarero, J.A., and Muir, T.W. (1999). Biosynthesis of a Head-to-Tail Cyclized Protein with Improved Biological Activity. 5597–5598.

Chang, C.-H., Rossi, E. a, and Goldenberg, D.M. (2007). The dock and lock method: a novel platform technology for building multivalent, multifunctional structures of defined composition with retained bioactivity. Clin. Cancer Res. *13*, 5586s–5591s.

Chin, J.W., Santoro, S.W., Martin, A.B., King, D.S., Wang, L., and Schultz, P.G. (2002). Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J. Am. Chem. Soc. *124*, 9026–9027.

Chiu, J., March, P.E., Lee, R., and Tillett, D. (2004). Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. Nucleic Acids Res. *32*, e174.

Chivers, C.E., Crozat, E., Chu, C., Moy, V.T., Sherratt, D.J., and Howarth, M. (2010). A streptavidin variant with slower biotin dissociation and increased mechanostability. Nat. Methods *7*, 391–393.

Christy, B., and Nathans, D. (1989). DNA binding site of the growth factor-inducible protein Zif268. Proc. Natl. Acad. Sci. U. S. A. *86*, 8737–8741.

Clark, R.J., and Craik, D.J. (2010). Native chemical ligation applied to the synthesis and bioengineering of circular peptides and proteins. Biopolymers *94*, 414–422.

Conlan, B.F., Gillon, A.D., Craik, D.J., and Anderson, M. a (2010). Circular proteins and mechanisms of cyclization. Biopolymers *94*, 573–583.

Crampton, N., and Brockwell, D.J. (2010). Unravelling the design principles for single protein mechanical strength. Curr. Opin. Struct. Biol. *20*, 508–517.

Crozat, E., Meglio, A., Allemand, J.-F., Chivers, C.E., Howarth, M., Vénien-Bryan, C., Grainge, I., and Sherratt, D.J. (2010). Separating speed and ability to displace roadblocks during DNA translocation by FtsK. EMBO J. *29*, 1423–1433.

Cumming, R.C., Andon, N.L., Haynes, P. a, Park, M., Fischer, W.H., and Schubert, D. (2004). Protein disulfide bond formation in the cytoplasm during oxidative stress. J. Biol. Chem. *279*, 21749–21758.

Darwin, K.H., and Hofmann, K. (2010). SAMPyling proteins in archaea. Trends Biochem. Sci. *35*, 348–351.

Dashivets, T., Wood, N., Hergersberg, C., Buchner, J., and Haslbeck, M. (2009). Rapid matrix-assisted refolding of histidine-tagged proteins. Chembiochem *10*, 869–876.

Deivanayagam, C.C., Rich, R.L., Carson, M., Owens, R.T., Danthuluri, S., Bice, T., Höök, M., and Narayana, S. V (2000). Novel fold and assembly of the repetitive B region of the Staphylococcus aureus collagen-binding surface protein. Structure *8*, 67–78.

Diamandis, E.P., and Christopoulos, T.K. (1991). The biotin-(strept)avidin system: principles and applications in biotechnology. Clin. Chem. *37*, 625–636.

Ellgaard, L. (1999). Setting the Standards: Quality Control in the Secretory Pathway. Science (80-.). *286*, 1882–1888.

Fairhead, M., Krndija, D., Lowe, E.D., and Howarth, M. (2013). Plug-And-Play Pairing Via Defined Divalent Streptavidins. J. Mol. Biol.

Fedorov, a. N. (1997). Cotranslational Protein Folding. J. Biol. Chem. *272*, 32715–32718.

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. a, Lerner, R. a, and Wigler, M. (1988). Purification of a RAS-responsive adenylyl cyclase complex from Saccharomyces cerevisiae by use of an epitope addition method. Mol. Cell. Biol. *8*, 2159–2165.

Fletcher, J.M., Harniman, R.L., Barnes, F.R.H., Boyle, A.L., Collins, A., Mantell, J., Sharp, T.H., Antognozzi, M., Booth, P.J., Linden, N., et al. (2013). Self-assembling cages from coiled-coil peptide modules. Science *340*, 595–599.

Friedman, M., Nordberg, E., Höidén-Guthenberg, I., Brismar, H., Adams, G.P., Nilsson, F.Y., Carlsson, J., and Ståhl, S. (2007). Phage display selection of Affibody molecules with specific binding to the extracellular domain of the epidermal growth factor receptor. Protein Eng. Des. Sel. *20*, 189–199.

Friedman, M., Orlova, A., Johansson, E., Eriksson, T.L.J., Höidén-Guthenberg, I., Tolmachev, V., Nilsson, F.Y., and Ståhl, S. (2008). Directed evolution to low nanomolar affinity of a tumor-targeting epidermal growth factor receptor-binding affibody molecule. J. Mol. Biol. *376*, 1388–1402.

Fritze, C.E., and Anderson, T.R. (2000). Epitope tagging: general method for tracking recombinant proteins. Methods Enzymol. *327*, 3–16.

Fu, J., Schoch, R., and Stevens, A. (2007). A patterned anisotropic nanofluidic sieving structure for continuous-flow separation of DNA and proteins. Nature *2*, 121–128.

Galarneau, A., Primeau, M., Trudeau, L.-E., and Michnick, S.W. (2002). Beta-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein protein interactions. Nat. Biotechnol. *20*, 619–622.

Gautier, A., Juillerat, A., Heinis, C., Corrêa, I.R., Kindermann, M., Beaufils, F., and Johnsson, K. (2008). An engineered protein tag for multiprotein labeling in living cells. Chem. Biol. *15*, 128–136.

Georgiou, G., and Valax, P. (1999). Isolating inclusion bodies from bacteria. Methods Enzymol. *309*, 48–58.

Giese, C., Zosel, F., Puorger, C., and Glockshuber, R. (2012). The most stable protein-ligand complex: applications for one-step affinity purification and identification of protein assemblies. Angew. Chem. Int. Ed. Engl. *51*, 4474–4478.

Green, N.M. (1990). Avidin and streptavidin. Methods Enzymol. *184*, 51–67.

Greenberg, C.S., Birckbichler, P.J., and Rice, R.H. (1991). Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. FASEB J. *5*, 3071–3077.

Griffin, B. a., Adams, S.R., and Tsien, R.Y. (1998). Specific covalent labeling of recombinant protein molecules inside live cells. Science *281*, 269–272.

Griffin, M., Casadio, R., and Bergamini, C.M. (2002). Transglutaminases: nature's biological glues. Biochem. J. *368*, 377–396.

Grignani, F., Kinsella, T., Mencarelli, A., Lanfrancone, L., Peschle, C., Nolan, G.P., Pelicci, P.G., Valtieri, M., and Riganelli, D. (1998). High-efficiency gene transfer and selection of human hematopoietic progenitor cells with a hybrid EBV/retroviral vector expressing the green fluorescence protein. Cancer Res. *58*, 14–19.

Guignet, E.G., Segura, J.-M., Hovius, R., and Vogel, H. (2007). Repetitive reversible labeling of proteins at polyhistidine sequences for single-molecule imaging in live cells. Chemphyschem *8*, 1221–1227.

Hackbarth, J.S., Lee, S.-H., Meng, X.W., Vroman, B.T., Kaufmann, S.H., and Karnitz, L.M. (2004). S-peptide epitope tagging for protein purification, expression monitoring, and localization in mammalian cells. Biotechniques *37*, 835–839.

Hagan, R.M., Björnsson, R., McMahon, S. a., Schomburg, B., Braithwaite, V., Bühl, M., Naismith, J.H., and Schwarz-Linek, U. (2010). NMR Spectroscopic and Theoretical Analysis of a Spontaneously Formed Lys-Asp Isopeptide Bond. Angew. Chemie *122*, 8599–8603.

Hancock, S.M., Uprety, R., Deiters, A., and Chin, J.W. (2010). Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. J. Am. Chem. Soc. *132*, 14819–14824.

Häring, D., Schreier, P., and Hubland, A. (1999). Cross-linked enzyme crystals. Curr. Opin. Chem. Biol. *3*, 35–38.

Hendrickx, A.P.A., Budzik, J.M., Oh, S.-Y., and Schneewind, O. (2011). Architects at the bacterial surface - sortases and the assembly of pili with isopeptide bonds. Nat. Rev. Microbiol. *9*, 166–176.

Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annu. Rev. Biochem. *67*, 425–479.

Heyman, A., Levy, I., Altman, A., and Shoseyov, O. (2007). SP1 as a novel scaffold building block for self-assembly nanofabrication of submicron enzymatic structures. Nano Lett. *7*, 1575–1579.

Hirel, P.H., Schmitter, M.J., Dessen, P., Fayat, G., and Blanquet, S. (1989). Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. Proc. Natl. Acad. Sci. U. S. A. *86*, 8247–8251.

Hodneland, C.D., Lee, Y.-S., Min, D.-H., and Mrksich, M. (2002). Selective immobilization of proteins to self-assembled monolayers presenting active site-directed capture ligands. Proc. Natl. Acad. Sci. U. S. A. *99*, 5048–5052.

Homola, J., Yee, S.S., and Gauglitz, G. (1999). Surface plasmon resonance sensors: review. Sensors Actuators B Chem. *54*, 3–15.

Horiike, K., Tojo, H., Yamano, T., and Nozaki, M. (1983). Interpretation of the stokes radius of macromolecules determined by gel filtration chromatography. J. Biochem. *93*, 99–106.

Houk, K.N., Leach, A.G., Kim, S.P., and Zhang, X. (2003). Binding affinities of host-guest, protein-ligand, and protein-transition-state complexes. Angew. Chem. Int. Ed. Engl. *42*, 4872–4897.

Howarth, M., and Ting, A.Y. (2008). Imaging proteins in live mammalian cells with biotin ligase and monovalent streptavidin. Nat. Protoc. *3*, 534–545.

Howarth, M., Chinnapen, D.J., Gerrow, K., Dorrestein, P.C., Grandy, M.R., Kelleher, N.L., El-Husseini, A., and Ting, A.Y. (2006). A monovalent streptavidin with a single femtomolar biotin binding site. Nat. Methods *3*, 267–273.

Huang, J., Nagy, S.S., Koide, A., Rock, R.S., and Koide, S. (2009a). A peptide tag system for facile purification and single-molecule immobilization. Biochemistry *48*, 11834–11836.

Huang, J., Makabe, K., Biancalana, M., Koide, A., and Koide, S. (2009b). Structural basis for exquisite specificity of affinity clamps, synthetic binding proteins generated through directed domain-interface evolution. J. Mol. Biol. *392*, 1221–1231.

Huh, W.-K., Falvo, J. V, Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Jäckel, C., Kast, P., and Hilvert, D. (2008). Protein design by directed evolution. Annu. Rev. Biophys. *37*, 153–173.

Jain, J., Veggiani, G., and Howarth, M. (2013). Cholesterol loading and ultrastable protein interactions determine the level of tumor marker required for optimal isolation of cancer cells. Cancer Res. *73*, 2310–2321.

Jarvik, J.W., and Telmer, C. a (1998). Epitope tagging. Annu. Rev. Genet. *32*, 601–618.

Johnson, E.C.B., and Kent, S.B.H. (2006). Insights into the mechanism and catalysis of the native chemical ligation reaction. J. Am. Chem. Soc. *128*, 6640–6646.

Joyce, G.F. (2007). Forty years of in vitro evolution. Angew. Chem. Int. Ed. Engl. *46*, 6420–6436.

Juillerat, M., Parr, G.R., and Taniuchi, H. (1980). A biologically active, three-fragment complex of horse heart cytochrome c. J. Biol. Chem. *255*, 845–853.

Kaluz, S., Kaluzová, M., and Pastorek, J. (1999). Inverse PCR-generated internally deleted constructs for direct characterization of promoter regulatory regions. Biotechniques *26*, 446–450.

Kang, H.J., and Baker, E.N. (2009). Intramolecular isopeptide bonds give thermodynamic and proteolytic stability to the major pilin protein of Streptococcus pyogenes. J. Biol. Chem. *284*, 20729–20737.

Kang, H.J., and Baker, E.N. (2011). Intramolecular isopeptide bonds: protein crosslinks built for stress? Trends Biochem. Sci. *36*, 229–237.

Kang, H.J., Coulibaly, F., Clow, F., Proft, T., and Baker, E.N. (2007). Stabilizing isopeptide bonds revealed in gram-positive bacterial pilus structure. Science *318*, 1625–1628.

Kang, H.J., Paterson, N.G., Gaspar, A.H., Ton-That, H., and Baker, E.N. (2009). The Corynebacterium diphtheriae shaft pilin SpaA is built of tandem Ig-like modules with stabilizing isopeptide and disulfide bonds. Proc. Natl. Acad. Sci. U. S. A. *106*, 16967–16971.

Kellermann, S.J., Rath, A.K., and Rentmeister, A. (2013). Tetramolecular fluorescence complementation for detection of specific RNAs in vitro. Chembiochem *14*, 200–204.

Kent, S.B.H. (2009). Total chemical synthesis of proteins. Chem. Soc. Rev. *38*, 338–351.

Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003). A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat. Biotechnol. *21*, 86–89.

Khlebnikov, A., Risa, Ø., Skaug, T., Trent, A., Keasling, J.D., and Carrier, T.A. (2000). Regulatable Arabinose-Inducible Gene Expression System with Consistent Control in All Cells of a Culture Regulatable Arabinose-Inducible Gene Expression System with Consistent Control in All Cells of a Culture.

Kline, K.A., Dodson, K.W., Caparon, M.G., and Hultgren, S.J. (2010). A tale of two pili: assembly and function of pili in bacteria. Trends Microbiol. *18*, 224–232.

Korndörfer, I.P., and Skerra, A. (2002). Improved affinity of engineered streptavidin for the Strep-tag II peptide is due to a fixed open conformation of the lid-like loop at the binding site. Protein Sci. *11*, 883–893.

Kumar, K.S.A., Spasser, L., Ohayon, S., Erlich, L.A., and Brik, A. (2011). Expeditious chemical synthesis of ubiquitinated peptides employing orthogonal protection and native chemical ligation. Bioconjug. Chem. *22*, 137–143.

Kurien, B.T., and Scofield, R.H. (2003). Protein blotting: a review. J. Immunol. Methods *274*, 1–15.

Kurien, B.T., and Scofield, R.H. (2006). Western blotting. Methods *38*, 283–293.

Lai, T.-S., Weisel, J.W., Greenberg, C.S., Grant, P.J., and Ariëns, R.A.S. (2002). Role of factor XIII in fibrin clot formation and effects of genetic polymorphisms. Blood *100*, 743–754.

Lang, K., Davis, L., Torres-kolbus, J., Chou, C., Deiters, A., and Chin, J.W. (2012a). Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nat. Chem. *4*, 298–304.

Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D.J., Blackman, M.L., Fox, J.M., and Chin, J.W. (2012b). Genetic Encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. J. Am. Chem. Soc. *134*, 10317–10320.

Lee, H.S., Dimla, R.D., and Schultz, P.G. (2009). Protein-DNA photo-crosslinking with a genetically encoded benzophenone-containing amino acid. Bioorg. Med. Chem. Lett. *19*, 5222–5224.

Lee, J.-S., Shukla, A., Schneider, J., Swanson, S.K., Washburn, M.P., Florens, L., Bhaumik, S.R., and Shilatifard, A. (2007). Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell *131*, 1084–1096.

Leenaars, M., and Hendriksen, C.F.M. (2005). General Introduction : Protocols for. *46*.

Li, L. Personal Communication.

Li, L., Fierer, J.O., Rapoport, T. a., and Howarth, M. (2013). Structural Analysis and Optimization of the Covalent Association between SpyCatcher and a Peptide Tag. J. Mol. Biol.

Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., and Tan, S. (2005). Comparison of affinity tags for protein purification. Protein Expr. Purif. *41*, 98–105.

Lin, M.C., Gutte, B., Moore, S., and Merrifield, R.B. (1970). Regeneration of activity by mixture of ribonuclease enzymically degraded from the COOH terminus and a synthetic COOH-terminal tetradecapeptide. J. Biol. Chem. *245*, 5169–5170.

Ling, J.J., Policarpo, R.L., Rabideau, A.E., Liao, X., and Pentelute, B.L. (2012). Protein thioester synthesis enabled by sortase. J. Am. Chem. Soc. *134*, 10749–10752.

Listgarten, J.N., Maes, D., Wyns, L., Aguilar, C.F., and Palmer, R.A. (1995). Structure of the crystalline complex of deoxycytidylyl-3',5'-guanosine (3',5'-dCpdG) cocrystallized with ribonuclease at 1.9 A resolution. Acta Crystallogr. D. Biol. Crystallogr. *51*, 767–771.

Lockless, S.W., and Muir, T.W. (2009). Traceless protein splicing utilizing evolved split inteins. Proc. Natl. Acad. Sci. U. S. A. *106*, 10999–11004.

Los, G. V, Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., et al. (2008). HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem. Biol. *3*, 373–382.

Luna, E.J. (2001). Biotinylation of proteins in solution and on cell surfaces. Curr. Protoc. Protein Sci. *Chapter 3*, Unit 3.6.

MacPhee, D.J. (2010). Methodological considerations for improving Western blot analysis. J. Pharmacol. Toxicol. Methods *61*, 171–177.

Malhotra, A. (2009). Tagging for protein expression. (Elsevier Inc.).

Marraffini, L.A., Dedent, A.C., and Schneewind, O. (2006). Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. Microbiol. Mol. Biol. Rev. *70*, 192–221.

Mazmanian, S.K., Liu, G., Ton-That, H., and Schneewind, O. (1999). Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. Science *285*, 760–763.

McCloskey, K.E., Chalmers, J.J., and Zborowski, M. (2003). Magnetic cell separation: characterization of magnetophoretic mobility. Anal. Chem. *75*, 6868–6874.

McDonald, C. a, Yang, J.Y., Marathe, V., Yen, T.-Y., and Macher, B. a (2009). Combining results from lectin affinity chromatography and glycocapture approaches substantially improves the coverage of the glycoproteome. Mol. Cell. Proteomics *8*, 287–301.

Melchior, F. (2000). SUMO--nonclassical ubiquitin. Annu. Rev. Cell Dev. Biol. *16*, 591–626.

Merril, C. (1990). Gel staining techniques. eLS *182*, 477–488.

Merril, C.R., Bisher, M.E., Harrington, M., and Steven, a C. (1988). Coloration of silver-stained protein bands in polyacrylamide gels is caused by light scattering from silver grains of characteristic sizes. Proc. Natl. Acad. Sci. U. S. A. *85*, 453–457.

Van der Merwe, P.A. (2000). Protein–Ligand Interactions: Hydrodynamics and Calorimetry. In Practical Approach Series, S. Harding, and P. Chowdhry, eds. (Oxford: Oxford University Press), pp. 137–170.

Van der Merwe, P.A., and Dushek, O. (2011). Mechanisms for T cell receptor triggering. Nat. Rev. Immunol. *11*, 47–55.

Miller, M.C., Doyle, G. V, and Terstappen, L.W.M.M. (2010). Significance of Circulating Tumor Cells Detected by the CellSearch System in Patients with Metastatic Breast Colorectal and Prostate Cancer. J. Oncol. *2010*, 617421.

Morris, P.D., Tackett, a J., and Raney, K.D. (2001). Biotin-streptavidin-labeled oligonucleotides as probes of helicase mechanisms. Methods *23*, 149–159.

El Mortaji, L., Terrasse, R., Dessen, A., Vernet, T., and Di Guilmi, A.M. (2010). Stability and assembly of pilus subunits of Streptococcus pneumoniae. J. Biol. Chem. *285*, 12405–12415.

Motoki, M., and Seguro, K. (1998). Transglutaminase and its use for food processing. Trends Food Sci. Technol. *9*, 204–210.

Muralidharan, V., and Muir, T.W. (2006). Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat. Methods *3*, 429–438.

Nathani, R.I., Moody, P., Chudasama, V., Smith, M.E.B., Fitzmaurice, R.J., and Caddick, S. (2013). A novel approach to the site-selective dual labelling of a protein via chemoselective cysteine modification. Chem. Sci. *4*, 3455.

Nordlund, H.R., Hytönen, V.P., Hörhä, J., Määttä, J.A.E., White, D.J., Halling, K., Porkka, E.J., Slotte, J.P., Laitinen, O.H., and Kulomaa, M.S. (2005). Tetravalent single-chain avidin: from subunits to protein domains via circularly permuted avidins. Biochem. J. *392*, 485–491.

Noren, C.C.J.C., Wang, J., and Perler, F.B.F. (2000). Dissecting the Chemistry of Protein Splicing and Its Applications. Angew. Chem. Int. Ed. Engl. *39*, 450–466.

Oke, M., Carter, L.G., Johnson, K. a, Liu, H., McMahon, S. a, Yan, X., Kerou, M., Weikart, N.D., Kadi, N., Sheikh, M.A., et al. (2010). The Scottish Structural Proteomics Facility: targets, methods and outputs. J. Struct. Funct. Genomics *11*, 167–180.

Ozawa, T. (2006). Designing split reporter proteins for analytical tools. Anal. Chim. Acta *556*, 58–68.

Papapostolou, D., Smith, A.M., Atkins, E.D.T., Oliver, S.J., Ryadnov, M.G., Serpell, L.C., and Woolfson, D.N. (2007). Engineering nanoscale order into a designed protein fiber. Proc. Natl. Acad. Sci. U. S. A. *104*, 10853–10858.

Paterson, G.K., and Mitchell, T.J. (2004). The biology of Gram-positive sortase enzymes. Trends Microbiol. *12*, 89–95.

Phillips, J.J., Millership, C., and Main, E.R.G. (2012). Fibrous nanostructures from the self-assembly of designed repeat protein modules. Angew. Chem. Int. Ed. Engl. *51*, 13132–13135.

Pointon, J.A.J. a J.A., Smith, W.D.W.D., Saalbach, G., Crow, A., Kehoe, M. a M.A., and Banfield, M.J.M.J. (2010). A highly unusual thioester bond in a pilus adhesin is required for efficient host cell interaction. J. Biol. Chem. *285*, 33858–33866.

Ponnuraj, K., and Narayana, S.V.L. (2007). Crystal structure of ACE19, the collagen binding subdomain of Enterococus faecalis surface protein ACE. Proteins *69*, 199–203.

Ponnuraj, K., Xu, Y., Moore, D., Deivanayagam, C.C.S., Boque, L., Hook, M., and Narayana, S.V.L. (2002). Crystallization and preliminary X-ray crystallographic analysis of Ace: a collagen-binding MSCRAMM from Enterococcus faecalis. Biochim. Biophys. Acta *1596*, 173–176.

Popp, M.W., Antos, J.M., Grotenbreg, G.M., Spooner, E., and Ploegh, H.L. (2007). Sortagging: a versatile method for protein labeling. Nat. Chem. Biol. *3*, 707–708.

Porstmann, B., Porstmann, T., Nugel, E., and Evers, U. (1985). Which of the commonly used marker enzymes gives the best results in colorimetric and fluorimetric enzyme immunoassays: Horseradish peroxidase, alkaline phosphatase or β-galactosidase? J. Immunol. Methods *79*, 27–37.

Preston, G.W., Radford, S.E., Ashcroft, A.E., and Wilson, A.J. (2012). Covalent cross-linking within supramolecular peptide structures. Anal. Chem. *84*, 6790–6797.

Rutledge, E. a, Green, F. a, and Enns, C. a (1994). Generation of the soluble transferrin receptor requires cycling through an endosomal compartment. J. Biol. Chem. *269*, 31864–31868.

Ryadnov, M.G., and Woolfson, D.N. (2004). Fiber recruiting peptides: noncovalent decoration of an engineered protein scaffold. J. Am. Chem. Soc. *126*, 7454–7455.

Salichs, E., Ledda, A., Mularoni, L., Albà, M.M., and de la Luna, S. (2009). Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment. PLoS Genet. *5*, e1000397.

Sengupta, D., and Heilshorn, S.C. (2010). Protein-engineered biomaterials: highly tunable tissue engineering scaffolds. Tissue Eng. Part B. Rev. *16*, 285–293.

Sevier, C.S., and Kaiser, C. a (2002). Formation and transfer of disulphide bonds in living cells. Nat. Rev. Mol. Cell Biol. *3*, 836–847.

Shah, N.H.N., and Muir, T.W.T. (2011). Split Inteins: Nature's Protein Ligases. Isr. J. Chem. *51*, 854–861.

Shah, N.H., Dann, G.P., Vila-Perelló, M., Liu, Z., and Muir, T.W. (2012). Ultrafast Protein Splicing is Common among Cyanobacterial Split Inteins: Implications for Protein Engineering. J. Am. Chem. Soc. *134*, 11338–11341.

Shekhawat, S.S., and Ghosh, I. (2011). Split-protein systems: beyond binary protein-protein interactions. Curr. Opin. Chem. Biol. *15*, 789–797.

Siebold, C., and Erni, B. (2002). Intein-mediated cyclization of a soluble and a membrane protein in vivo: function and stability. Biophys. Chem. *96*, 163–171.

Skretas, G., and Wood, D. (2005). Regulation of protein activity with small-molecule-controlled inteins. Protein Sci. 523–532.

Smith, A.M., Acquah, S.F. a, Bone, N., Kroto, H.W., Ryadnov, M.G., Stevens, M.S.P., Walton, D.R.M., and Woolfson, D.N. (2004). Polar assembly in a designed protein fiber. Angew. Chem. Int. Ed. Engl. *44*, 325–328.

Stains, C.I., Furman, J.L., Porter, J.R., Rajagopal, S., Li, Y., Wyatt, R.T., and Ghosh, I. (2010). A general approach for receptor and antibody-targeted detection of native proteins utilizing split-luciferase reassembly. ACS Chem. Biol. *5*, 943–952.

Stanfield, R.L., and Wilson, I. a (1995). Protein-peptide interactions. Curr. Opin. Struct. Biol. *5*, 103–113.

Stephanopoulos, N., and Francis, M.B. (2011). Choosing an effective protein bioconjugation strategy. Nat. Chem. Biol. *7*, 876–884.

Suchanek, M., Radzikowska, A., and Thiele, C. (2005). Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat. Methods *2*, 261–267.

Sugawara, S., Abo, T., and Kumagai, K. (1987). A simple method to eliminate the antigenicity of surface class I MHC molecules from the membrane of viable cells by acid treatment at pH 3. J. Immunol. Methods *100*, 83–90.

Sun, W., Yang, J., and Liu, X.-Q. (2004). Synthetic two-piece and three-piece split inteins for protein trans-splicing. J. Biol. Chem. *279*, 35281–35286.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. *60*, 523–533.

Thorn, K.S., Naber, N., Matuska, M., Vale, R.D., and Cooke, R. (2000). A novel method of affinity-purifying proteins using a bis-arsenical fluorescein. Protein Sci. *9*, 213–217.

Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. Trends Microbiol. *12*, 228–234.

Tsien, R.Y. (1998). The green fluorescent protein. Annu. Rev. Biochem. *67*, 509–544.

Tsukiji, S., and Nagamune, T. (2009). Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. Chembiochem *10*, 787–798.

Vasita, R., and Katti, D.S. (2006). Nanofibers and their applications in tissue engineering. Int. J. Nanomedicine *1*, 15–30.

Vengadesan, K., and Narayana, S.V.L. (2011). Structural biology of Gram-positive bacterial adhesins. Protein Sci. *20*, 759–772.

Voigt, P., and Reinberg, D. (2011). Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. Chembiochem *12*, 236–252.

Wang, H., Xiao, Y., Fu, L., Zhao, H., Zhang, Y., Wan, X., Qin, Y., Huang, Y., Gao, H., and Li, X. (2010). High-level expression and purification of soluble recombinant FGF21 protein by SUMO fusion in Escherichia coli. BMC Biotechnol. *10*, 14.

Wehr, M.C., Laage, R., Bolz, U., Fischer, T.M., Grünewald, S., Scheek, S., Bach, A., Nave, K., and Rossner, M.J. (2006). Monitoring regulated protein-protein interactions using split TEV. Nat. Methods *3*, 985–993.

White, C.J., and Yudin, A.K. (2011). Contemporary strategies for peptide macrocyclization. Nat. Chem. *3*, 509–524.

Wikoff, W.R., Liljas, L., Duda, R.L., Tsuruta, H., Hendrix, R.W., and Johnson, J.E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. Science *289*, 2129–2133.

Williamson, D.J., Fascione, M. a, Webb, M.E., and Turnbull, W.B. (2012). Efficient N-terminal labeling of proteins by use of sortase. Angew. Chem. Int. Ed. Engl. *51*, 9377–9380.

Winter, G., Griffiths, A.D., Hawkins, R.E., and Hoogenboom, H.R. (1994). Making antibodies by phage display technology. Annu. Rev. Immunol. *12*, 433–455.

Wong, T.S., Arnold, F.H., and Schwaneberg, U. (2004). Laboratory evolution of cytochrome p450 BM-3 monooxygenase for organic cosolvents. Biotechnol. Bioeng. *85*, 351–358.

Woof, J.M., and Burton, D.R. (2004). Human antibody-Fc receptor interactions illuminated by crystal structures. Nat. Rev. Immunol. *4*, 89–99.

Xu, M.Q., and Evans, T.C. (2001). Intein-mediated ligation and cyclization of expressed proteins. Methods *24*, 257–277.

Yancey, P.H. (2005). Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. J. Exp. Biol. *208*, 2819–2830.

Yang, Y., Jun, C.-D., Liu, J.-H., Zhang, R., Joachimiak, A., Springer, T. a, and Wang, J.-H. (2004). Structural basis for dimerization of ICAM-1 on the cell surface. Mol. Cell *14*, 269–276.

Yu, Y., and Lutz, S. (2010). Circular permutation: a different way to engineer enzyme structure and function. Trends Biotechnol. *29*, 18–25.

Yu, M., Stott, S., Toner, M., Maheswaran, S., and Haber, D. a (2011). Circulating tumor cells: approaches to isolation and characterization. J. Cell Biol. *192*, 373–382.

Zakeri, B., and Howarth, M. (2010). Spontaneous intermolecular amide bond formation between side chains for irreversible peptide targeting. J. Am. Chem. Soc. *132*, 4526–4527.

Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc. Natl. Acad. Sci. U. S. A. *109*, E690–7.

Zettler, J., Schütz, V., and Mootz, H.D. (2009). The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett. *583*, 909–914.

Zhao, W., Cui, C.H., Bose, S., Guo, D., Shen, C., Wong, W.P., Halvorsen, K., Farokhzad, O.C., Teo, G.S.L., Phillips, J. a, et al. (2012). Bioinspired multivalent DNA network for capture and release of cells. Proc. Natl. Acad. Sci. U. S. A. *109*, 19626–19631.

Appendix

Appendix

A.1 Sequence Alignments of All Variants of SpyLigase Produced During this Thesis:

A)

```
CnaB2       MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDI 60
SpyCatcher  MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDE 60
SpyLigase1  MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSG---------------------- 41
SpyLigase2  MSYYHHHHHHDYD-------------------------GQSG---------------------- 17
SpyLigase2B MSYYHHHHHHDYD-------------------------GQSG---------------------- 17
SpyLigase3  MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSG----------------RDI 44
SpyLigase4  MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSG---------------------- 41
SpyLigase5  MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSG---------------------- 41
SpyLigase   MSYYHHHHHHDYD-------------------------GQSG---------------------- 17

CnaB2       DGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTV 120
SpyCatcher  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 120
SpyLigase1  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 101
SpyLigase2  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 77
SpyLigase2B DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 77
SpyLigase3  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 104
SpyLigase4  DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 101
SpyLigase5  ---ELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 98
SpyLigase   DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 77

CnaB2       NEQGQVTVNGKATKG-----------DAHIVMVDA 144
SpyCatcher  NEQGQVTVNGKATKG-----------DAHI----- 139
SpyLigase1  NEQGQVTVNGKATKG-----------DAHI----- 120
SpyLigase2  NEQGQVTVNGKATKG-----------DAHI----- 96
SpyLigase2B NEQGQVTVNGKATKGSGGSGGSGEDSATHIKF--- 110
SpyLigase3  NEQGQVTVNGKATKG-----------DAHI----- 123
SpyLigase4  NEQGQVTVNGKATKG---------------------- 116
SpyLigase5  NEQGQVTVNGKATKG-----------DAHI----- 117
SpyLigase   NEQGQVTVNGKATKGSGGSGGSGEDSATHI----- 108
```

B)

```
CnaB2                MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDI 60
SpyCatcher           MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDE 60
L4C-H112C-SpyLigase1 MSYYHHHHHHDYDIPTTENLYFQGAMVDTCSGLSSEQGQSG---------------------- 41
T43C-A79C-SpyLigase2 MSYYHHHHHHDYD-------------------------GQSG---------------------- 17
A42C-S59C-SpyLigase2 MSYYHHHHHHDYD-------------------------GQSG---------------------- 17
SpyLigase            MSYYHHHHHHDYD-------------------------GQSG---------------------- 17

CnaB2                DGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTV 120
SpyCatcher           DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 120
L4C-H112C-SpyLigase1 DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 101
T43C-A79C-SpyLigase2 DGKELAGACMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETACPDGYEVATAITFTV 77
A42C-S59C-SpyLigase2 DGKELAGCTMELRDSSGKTISTWICDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 77
SpyLigase            DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 77

CnaB2                NEQGQVTVNGKATKGDAHIVMVDA------- 144
SpyCatcher           NEQGQVTVNGKATKGDAHI---------- 139
L4C-H112C-SpyLigase1 NEQGQVTVNGKATKGDACI---------- 120
T43C-A79C-SpyLigase2 NEQGQVTVNGKATKGDAHI---------- 96
A42C-S59C-SpyLigase2 NEQGQVTVNGKATKGDAHI---------- 96
SpyLigase            NEQGQVTVNGKATKGGSGGSGGSGEDSATHI 108
```

C)

```
CnaB2              MSYYHHHHHHDYDIPTTEN---------------------------------------- 19
SpyCatcher         MSYYHHHHHHDYDIPTTEN---------------------------------------- 19
SUMO-SpyLigase1    MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
SUMO-GS-SpyLigase1 MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
SUMO-SpyLigase2    MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
SUMO-GS-SpyLigase2 MGSSHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME 60
SpyLigase          MSYYHHHHHHDYD---------------------------------------------- 13

CnaB2              ------------LYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFS----------- 56
SpyCatcher         ------------LYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFS----------- 56
SUMO-SpyLigase1    AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGDTLSG----LS 116
SUMO-GS-SpyLigase1 AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGSGSGDTLSGLS 120
SUMO-SpyLigase2    AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGG----------- 109
SUMO-GS-SpyLigase2 AFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGGSGSG------- 113
SpyLigase          ------------------------------------------------------------

CnaB2              ---KRDIDGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVA 113
SpyCatcher         ---KRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA 113
SUMO-SpyLigase1    SEQGQSGDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA 176
SUMO-GS-SpyLigase1 SEQGQSGDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA 180
SUMO-SpyLigase2    ---GQSGDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA 166
SUMO-GS-SpyLigase2 ---GQSGDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA 170
SpyLigase          ---GQSGDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVA 70

CnaB2              TAITFTVNEQGQVTVNGKATKG------------DAHIVMVDA 144
SpyCatcher         TAITFTVNEQGQVTVNGKATKG------------DAHI----- 139
SUMO-SpyLigase1    TAITFTVNEQGQVTVNGKATKG------------DAHI----- 202
SUMO-GS-SpyLigase1 TAITFTVNEQGQVTVNGKATKG------------DAHI----- 206
SUMO-SpyLigase2    TAITFTVNEQGQVTVNGKATKG------------DAHI----- 192
SUMO-GS-SpyLigase2 TAITFTVNEQGQVTVNGKATKG------------DAHI----- 196
SpyLigase          TAITFTVNEQGQVTVNGKATKGGSGGSGGSGEDSATHI----- 108
```

SpyTag AHIVMVDAYKPTK SpyTag strand
KTag ATHIKFSKRD KTag strand
Deleted sections to make SpyLigase
SUMO Domain
Circular Permutation of SpyLigase
Cysteine Point Mutations

Figure A.1: Sequence alignments of all SpyLigase variants. The sequence alignments are separated into three categories: A) SpyLigase truncation and circular permutation variants, B) Disulfide engineering, and C) SUMO fusions.

A.2 Original Ligation reaction of SUMO-KTag with SpyTag-MBP Catalyzed by SpyLigase1:

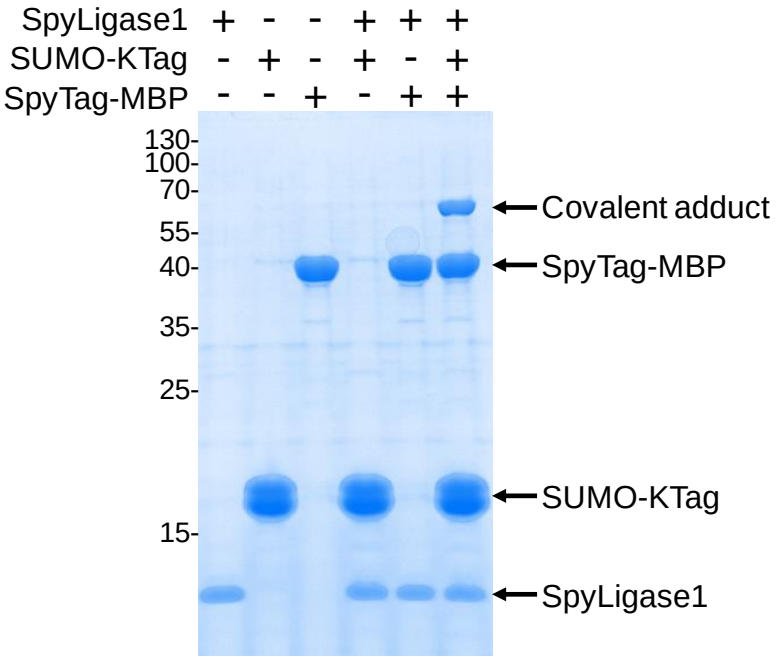


Figure A.2: Original ligation of SUMO-KTag with SpyTag-MBP with the first iteration of SpyLigase, SpyLigase1. The reaction conditions were 40 mM Na_2HPO_4, and 20 mM Citric Acid at pH 7.0, 10 μM of SpyLigase1, 20 μM SUMO-KTag, 20 μM SpyTag-MBP, incubated at 4°C for 24 h.

The first iteration of SpyLigase, SpyLigase1 (Figure A.1A), was simply SpyCatcher with the KTag β-sheet truncated (residues 41-60 in terms of the CnaB2 structure (Oke et al., 2010)). The initial test of SpyLigase1 can be seen in Figure A.2, where after 24 h the ligation was shown first and foremost to actually work. Though the system was inefficient, slow, and required the reactants of SUMO-KTag and SpyTag-MBP to be in excess, the primary milestone of this experiment was that a tripartite split protein could reconstitute and then ligate two components together.

A.3 Comparison of Truncation and Insertion Mutants of SpyLigase1:

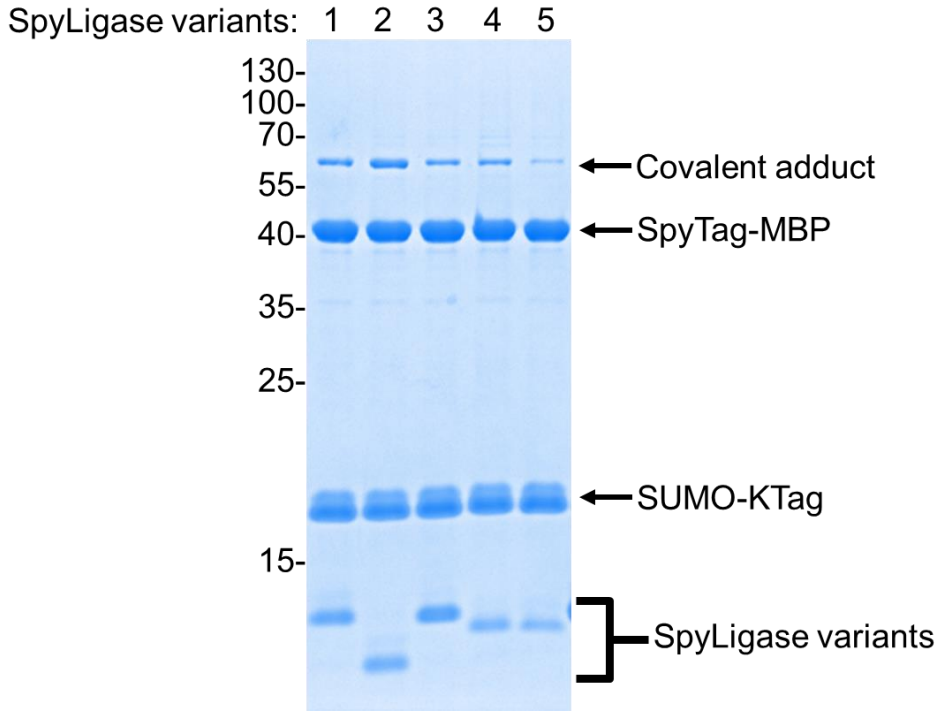


Figure A.3: Comparison of SpyLigase variants for ligation of SpyTag-MBP and SUMO-KTag. The reaction conditions were 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 5.0, 10 µM of SpyLigase variants, 10 µM SUMO-KTag, 10 µM SpyTag-MBP, incubated at 4°C for 24 h.

The development of SpyLigase continued with attempts at improving SpyLigase either by truncation (SpyLigase2, 4, and 5) or insertion (SpyLigase 3) (Figure A.1A). From Figure A.3 it can be observed that the truncation mutant SpyLigase2 produced the largest improvement and became the basis for further work to reach the current iteration of SpyLigase.

A.4 Circular Permutation of SpyLigase2 to Generate the Current Iteration of SpyLigase:

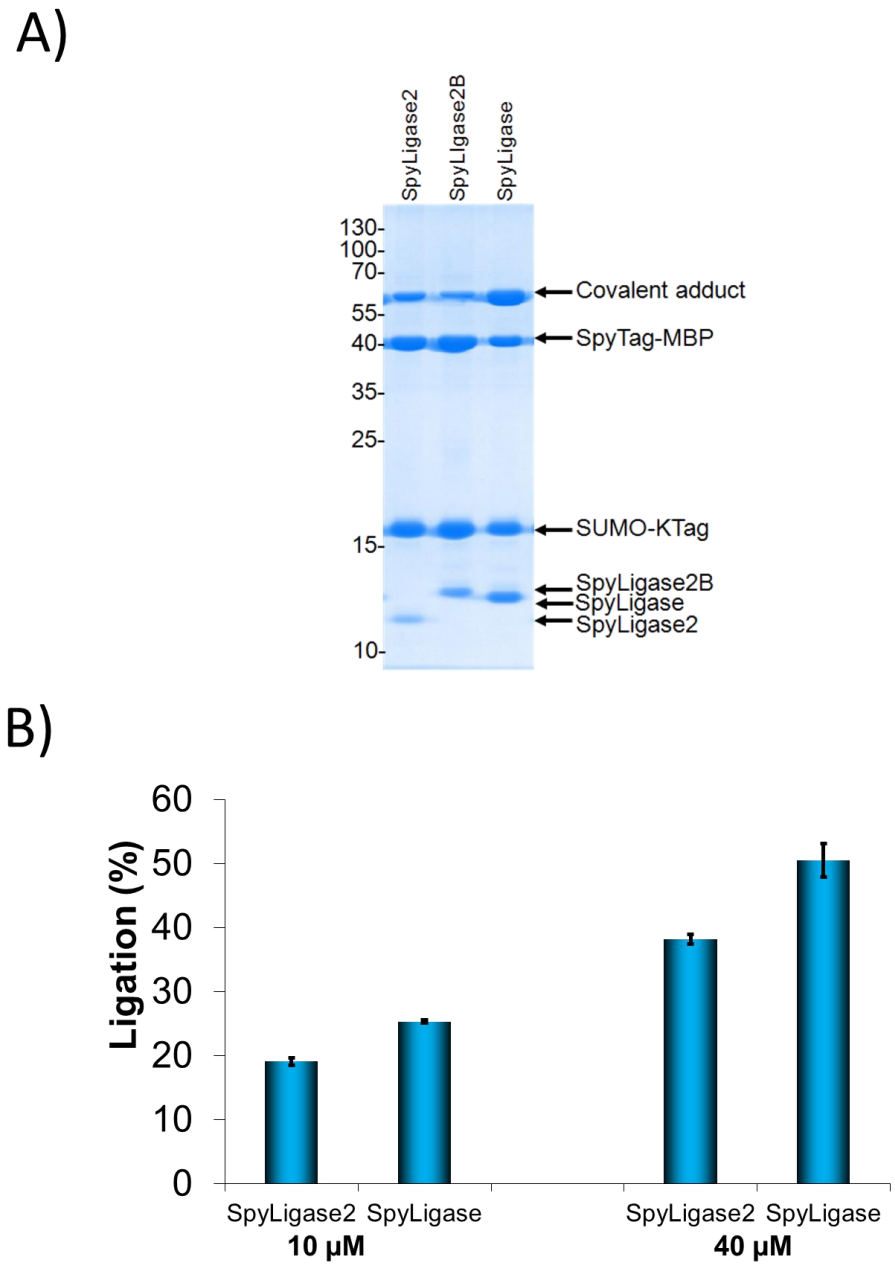


Figure A.4: Circular permutation of SpyLigase2 leading to the current iteration of SpyLigase. A) Comparing amount of ligation with SpyLigase2, SpyLigase2B, and SpyLigase. Reaction conditions were 40 mM Na_2HPO_4, and 20 mM Citric Acid at pH 5.0, 10 µM of SpyLigase variants, 10 µM SUMO-KTag, 10 µM SpyTag-MBP, 2 M TMAO incubated at 4°C for 24 h. B) Quantification of SpyLigase2 and SpyLigase to determine the effectiveness of the circular permutation on SpyLigase2 activity. The reaction conditions were the same as in (B) except that the SpyLigase variants were used at 10 or 40 µM where designated. The quantification was the mean of reactions performed in triplicate and error bars are ±1 standard deviation.

The current amino acid sequence of SpyLigase was reached via the use of circular permutation (Yu and Lutz, 2010). The circular permutations of SpyLigase2 generating SpyLigase and SpyLigase2B, included a GS linker and varying amounts of the KTag β-strand on the C-terminal of SpyLigase2 to create the new variants (Figure A.1A). SpyLigase2, SpyLigas2B, and SpyLigase were compared with an initial ligation of SpyTag-MBP and SUMO-KTag, which showed a clear improvement in the case of SpyLigase (Figure A.4A). Thus SpyLigase2 and SpyLigase were quantified as shown in Figure A.4B, and in both cases of 10 and 40 μM the circular permutation of SpyLigase2 to generate SpyLigase helped to improve the activity. The value of TMAO as a chemical chaperone (Yancey, 2005) had been shown with testing with SpyLigase2 (Figure A.6).

A.5 Sequence Alignment of CnaB2, SpyCatcher, and SpyLigase:

```
CnaB2        MSYYHHHHHH DYDIPTTENL YFQGAMVDTL SGLSSEQGQS GDMTIEEDSA THIKFSKRDI
SpyCatcher   MSYYHHHHHH DYDIPTTENL YFQGAMVDTL SGLSSEQGQS GDMTIEEDSA THIKFSKRDE
SpyLigase    MSYYHHHHHH DYD------- ---------- -------GQS G--------- ----------

CnaB2        DGKELAGATM ELRDSSGKTI STWISDGQVK DFYLMPGKYT FVETAAPDGY EVATAITFTV
SpyCatcher   DGKELAGATM ELRDSSGKTI STWISDGQVK DFYLYPGKYT FVETAAPDGY EVATAITFTV
SpyLigase    DGKELAGATM ELRDSSGKTI STWISDGQVK DFYLYPGKYT FVETAAPDGY EVATAITFTV

CnaB2        NEQGQVTVNG KATKGDAHIV MVDA------ -
SpyCatcher   NEQGQVTVNG KATKGDAHI- ---------- -
SpyLigase    NEQGQVTVNG KATKGGSGGS GGSGEDSATH I

SpyTag AHIVMVDAYKPTK
KTag   ATHIKFSKRD
```

SpyTag strand
KTag strand
Deleted sections to make SpyLigase
Circular Permutation of SpyLigase
Residues mutated during the generation of SpyCatcher

Figure A.5: Sequence alignment of CnaB2, SpyCatcher, and SpyLigase. SpyTag β-sheet is colored in cyan and KTag in red. Sections deleted in the process of engineering SpyLigase are shown in yellow for both CnaB2 and SpyCatcher. The circular permutation on the C-terminus is shown in purple, composed of a GS linker, part of the β-sheet prior to the KTag (EDS), and the first four residues of KTag (ATHI). Underlined residues represent the point mutations made to improve SpyCatcher, I34E and M69Y (Zakeri et al., 2012). Sequence alignments were performed using the ClustalW2 online tool from EMBL.

The sequence alignment of CnaB2, SpyCatcher, and SpyLigase can be seen in Figure A.5. SpyLigase is based on SpyCatcher, though heavily modified by our testing for improving the ligation capacity. The most obvious difference is the removal of the KTag β-strand colored in red, followed by the deletion of a large section of the N-terminal region of the SpyCatcher domain, colored in yellow. Finally the circular permutation was performed on SpyLigase with the addition of a GS linker, parts of the deleted N-terminal region (EDS), and parts of the KTag β-strand (ATHI) fused on the C-terminus of SpyLigase. The circular permutation was designed to improve SpyLigase structural stability by acting as a ‘place holder’: the C-terminus should sit in the gap where the KTag β-sheet is in SpyCatcher, which could help to stabilize the SpyLigase structure.

A.6 Testing Varying Additives to SpyLigase2 Ligation Reaction:

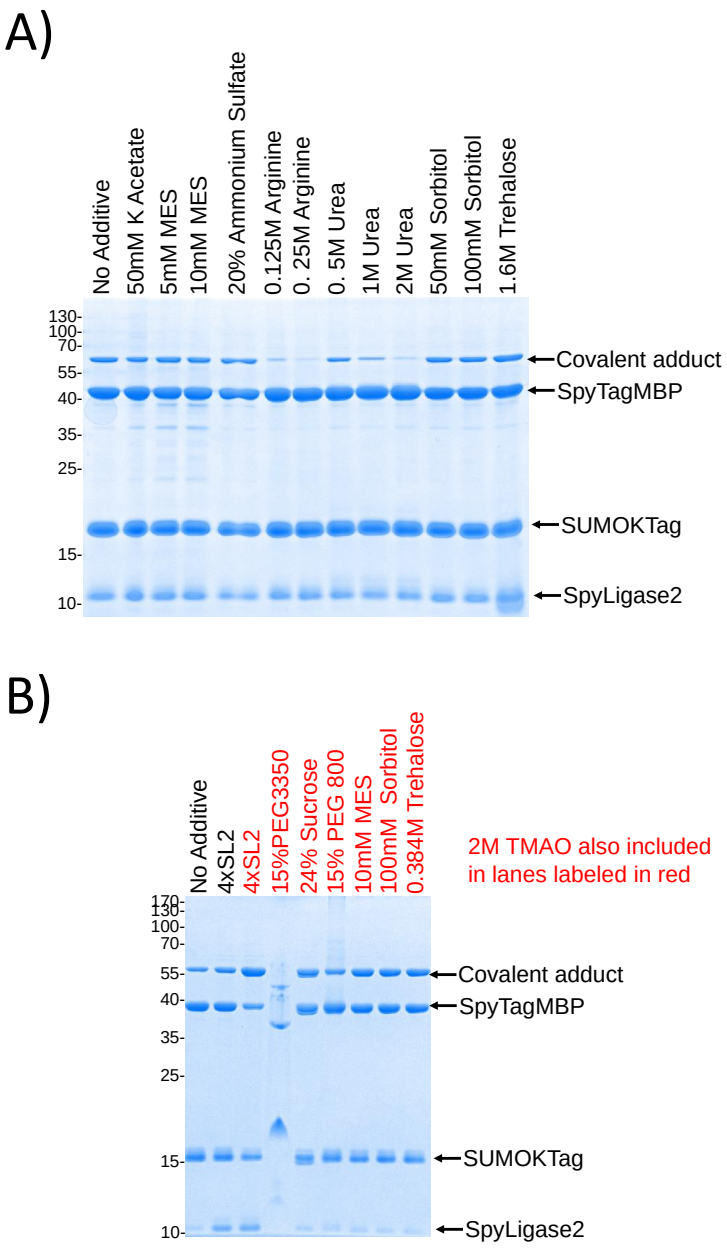


Figure A.6: Various additives to try to improve ligation yield with SpyLigase2. A) Examining the effect of different amounts of potassium acetate, 2-(N-morpholino)ethanesulfonic acid (MES), ammonium sulfate, arginine, urea, sorbitol, and trehalose on reaction between SUMO-KTag and SpyTag-MBP for 24h at 4°C, analyzed by SDS-PAGE with Coomassie staining. B) Effects on SpyLigase2 ligation of another series of additives. However all those highlighted in red also include 2 M trimethylamine N-Oxide (TMAO) to observe whether the covalent adduct yield with TMAO could be improved upon by mixing in an additional compound. Reaction conditions were 40 mM Na_2HPO_4, and 20 mM Citric Acid at pH 5.0, 10 μM of SpyLigase2, 10 μM SUMO-KTag, 10 μM SpyTag-MBP, and the included additive(s), incubated at 4°C for 24 h

Figure A.6 explores a number of the compounds tested to improve the ligation yield of the previous iteration of SpyLigase, SpyLigase2. Numerous additives were tested (Figure A.6A), until it was found that trimethylamine N-oxide was effective. TMAO is a molecular chaperone used by certain fish such as sharks and stingrays to counteract accumulation of Urea due to their nitrogen metabolism pathway (Yancey, 2005). I also examined whether any additive tested previously could help build on this improvement, by mixing those additives with TMAO. No improvements could be observed that were better than TMAO on its own (Figure A.6B).

A.7 Attempts at Improving SpyLigase2 Activity via Disulfide Engineering:

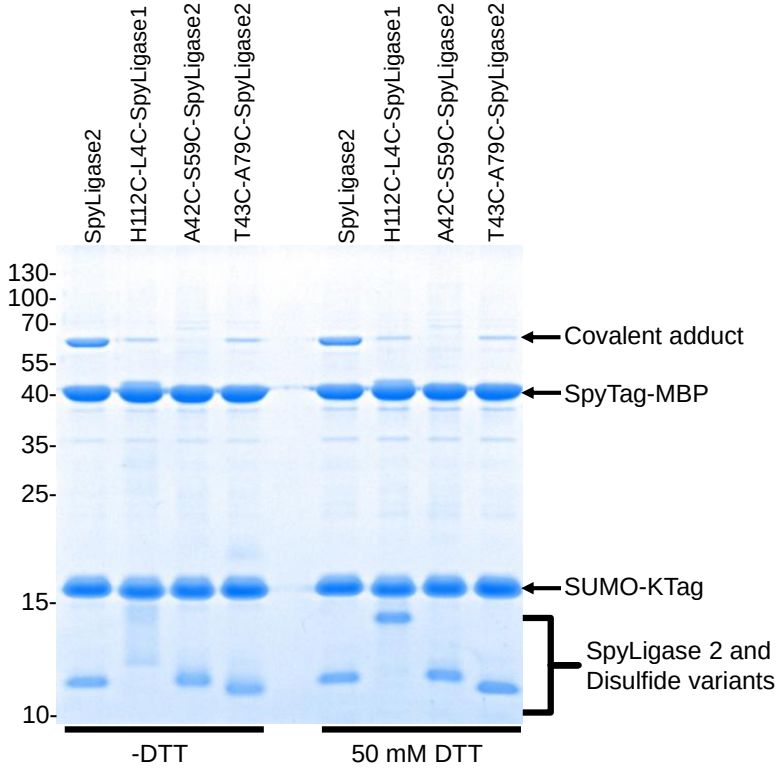


Figure A.7: Disulfide engineering in SpyLigase 1 and 2. Determining the effect of disulfide mutations on SpyLigase1 and SpyLigase2 domains. The four lanes on the left were boiled in SDS loading dye, the four lanes on the right including 50 mM DTT. The reaction conditions were 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 5.0, 10 µM of SpyLigase variants, 10 µM SUMO-KTag, 10 µM SpyTag-MBP, incubated at 4°C for 24 h.

The purpose of introducing disulfides into SpyLigase was to try to create a stable structure, to reduce some of the flexibility introduced by removing two β-strands, using the program Disulfide by Design Version 1.2 (Wayne State University, Dr. Alan Dombkowski) where structure of CnaB2 (Oke et al., 2010) was used as a basis to determine possible disulfide bridges that could be introduced into SpyLigase(1 or 2). The double Cysteine mutants generated by the program can be seen in the alignment in Figure A.1B, which have been underlined. The disulfides cause a massive retardation in the reactivity of SpyLigase1 and SpyLigase2. Therefore engineering of disulfides was not pursued further.

A.8 Attempts at Improving SpyLigase1,2 Activity via Fusion with SUMO Domain:

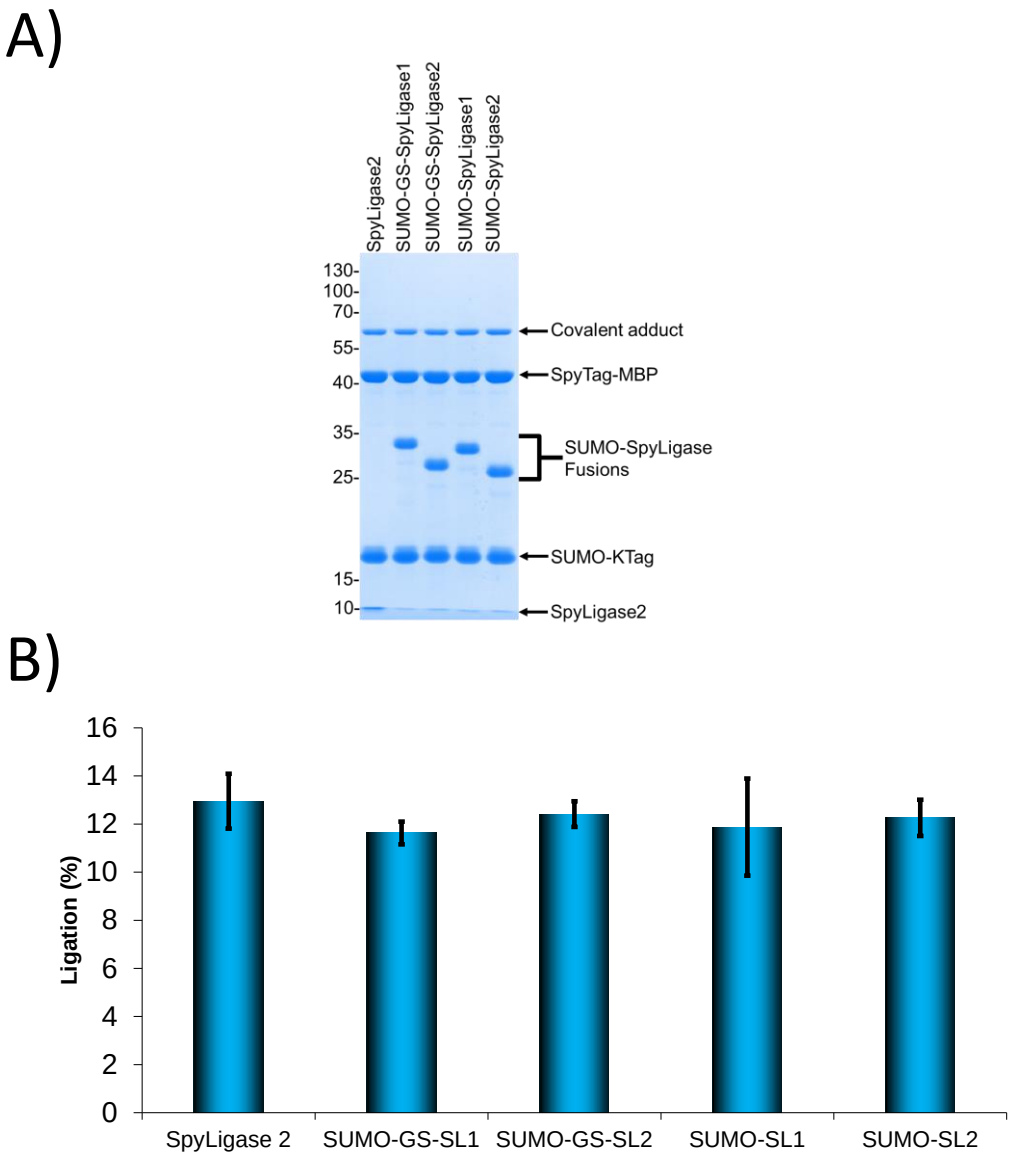


Figure A.8: Attempt at improving SpyLigase1 and SpyLigase2 activity by fusing with a SUMO domain to act as a molecular chaperone. A) SDS-PAGE of SUMO-SpyLigase ligation of SUMO-KTag with SpyTag-MBP compared to SpyLigase2. Reactions were performed at 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 5.0, 10 μM of SpyLigase variants, 10 μM SUMO-KTag, 10 μM SpyTag-MBP, incubated at 4°C for 24 h. B) Quantification of SUMO-SpyLigase fusions compared to SpyLigase2 activity (mean of triplicate ±1 standard deviation).

An alternative method to try to improve SpyLigase was fusing SpyLigase1 and SpyLigase2 to the SUMO domain. SUMO has been shown to improve expression, folding, and protein stability of attached proteins, acting as a molecular chaperone

(Wang et al., 2010). SUMO-GS-SpyLigase versions have a GSGSG linker while the SUMO-SpyLigase versions are only separated from the SUMO domain by the three Glycines on the C-terminus (Figure A.1C). However, from the SDS-PAGE analysis no substantial difference between SpyLigase2 and any of the SUMO fusions can be seen.

A.9 Summary of All SpyLigase Variants' Reactivity:

Table A.1: Summary of the ligation capacity of all the SpyLigase variants and some of the conditions attempted in this thesis. The general conditions in all these listed cases is 10 μM SpyTag-MBP, 10 μM SUMO-KTag, 4°C, and 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 5.0.

	Ligation (%)	Concentration SpyLigase Variant (μM)	TMAO (M)
Truncation mutants			
SpyLigase1	8	10	0
SpyLigase1	17	100	0
SpyLigase2	12	10	0
SpyLigase2	27	100	0
SpyLigase2	40	40	2
SpyLigase3	7	10	0
SpyLigase4	8	10	0
SpyLigase5	3	10	0
2nd round of truncation			
SpyLigase2B	10	10	0
SpyLigase	≥50	40	1.5
Disulfide Engineered Mutants			
H112C-L4C-SpyLigase1	2	10	0
A42C-S59C-SpyLigase2	0	10	0
T43C-A79C-SpyLIgase2	2	10	0
SUMO Fusions			
SUMO-GS-SpyLigase1	12	10	0
SUMO-GS-SpyLigase2	12	10	0
SUMO-SpyLigase1	12	10	0
SUMO-SpyLigase2	12	10	0

A.10 Polymerization of KTag-AffiEGFR-SpyTag on Magnetic Beads:

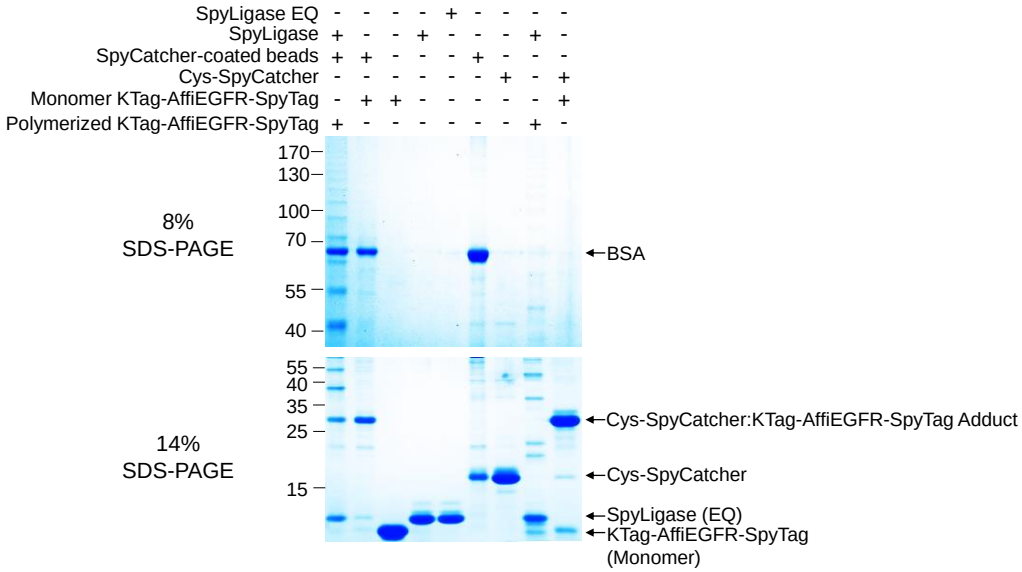
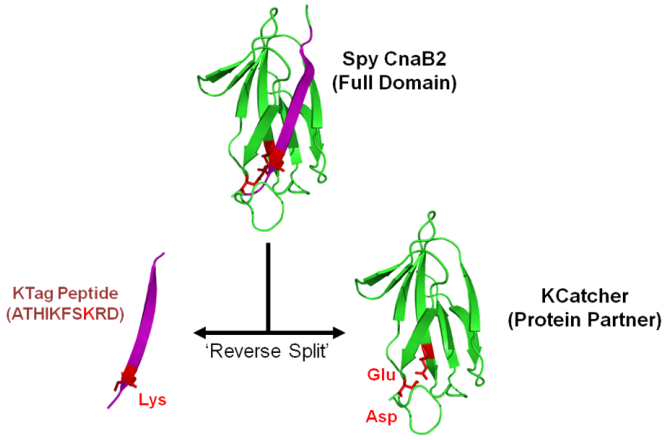


Figure A.9: Coating magnetic beads with monomer or polymerized KTag-AffiEGFR-SpyTag. Cys-SpyCatcher (SpyCatcher with a single exposed Cysteine) was incubated to form a disulfide linkage with the magnetic bead (fourth lane from the right). Then KTag-AffiEGFR-SpyTag was mixed with Cys-SpyCatcher coated beads, with active SpyLigase (left-most lane) showing polymerization. For the negative control, the reaction was mixed with SpyLigase EQ (second lane from the left), showing no polymerization, and acting as the monomer bead control for Figure 5.17C. (KTag-AffiEGFR-SpyTag, SpyLigase, and SpyLigase EQ were supplied by me. All other work related to this figure is the work of Mr. Gianluca Veggiani in the Howarth laboratory.)

The process of polymerizing KTag-AffiEGFR-SpyTag onto magnetic beads is shown in Figure A.9 from the work of Mr. Gianluca Veggiani. The procedure involved activating the magnetic beads for reaction with Cys-SpyCatcher. KTag-AffiEGFR-SpyTag was then successfully polymerized on the magnetic beads, as can be seen by boiling the beads with DTT (Figure A.9 left-most lane). An important note is that the beads are functionalized with a variety of affibody polymer chain lengths, due to the difficulty in controlling the size of the polymers during the polymerization reaction.

A.11 Creating an Orthogonal System to SpyCatcher-SpyTag: KCatcher-KTag:

A)



B)

```
CnaB2          MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDI 60
KCatcher       MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSG------------------- 41
SpyCatcher     MSYYHHHHHHDYDIPTTENLYFQGAMVDTLSGLSSEQGQSGDMTIEEDSATHIKFSKRDE 60

CnaB2          DGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTV 120
KCatcher       DGKELAGATMELRDSSGKTISTWISDGQVKDFYLMPGKYTFVETAAPDGYEVATAITFTV 101
SpyCatcher     DGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTV 120

CnaB2          NEQGQVTVNGKATKGDAHIVMVDA 144   SpyTag strand  SpyTag AHIVMVDAYKPTK
KCatcher       NEQGQVTVNGKATKGDAHIVMVDA 125   KTag strand    KTag  ATHIKFSKRD
SpyCatcher     NEQGQVTVNGKATKGDAHI----- 139
```

C)

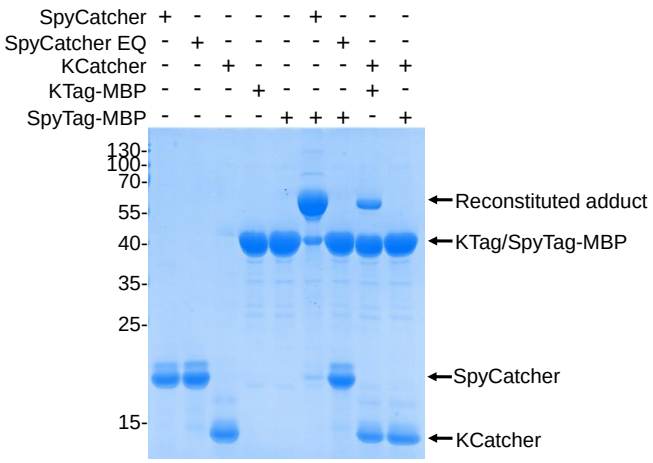


Figure A.10: Creating an orthogonal system to SpyCatcher-SpyTag in which the reactive Lysine β-strand was removed (KTag) and the remnant domain was the 'Catcher' section (KCatcher). A) Cartoon of splitting the CnaB2 domain to form KTag and KCatcher. B) Sequence alignment of CnaB2, KCatcher, and SpyCatcher. C) Comparison of reconstitution capacities between SpyCatcher and KCatcher after 10 h incubation at 4°C. Reactions conditions 40 mM Na_2HPO_4 and 20 mM Citric Acid at pH 7.0, 10 μM K/SpyCatcher, 10 μM K/SpyTag-MBP.

It was decided to create an orthogonal covalent peptide tag system to SpyCatcher-SpyTag. Instead of the reactive Aspartate β-strand being split from the domain, the Aspartate would remain with the Glutamate on the domain, and the reactive Lysine would go on to become its own peptide tag (Figure A.10A). As seen in Figure A.10C after 10 h of incubation, the KCatcher reaction still showed a low yield of reconstituted adduct. One possible explanation of the low yield is that the docking cleft for KTag in KCatcher is not as exposed as the docking cleft for SpyTag in SpyCatcher. Opening of KCatcher's cleft to allow reaction with KTag could require a rarely formed conformation of KCatcher, limiting the reaction rate.

A.12 Circular Dichroism of CnaB2, SpyCatcher, KCatcher, SpyLigase1, and SpyLigase2:

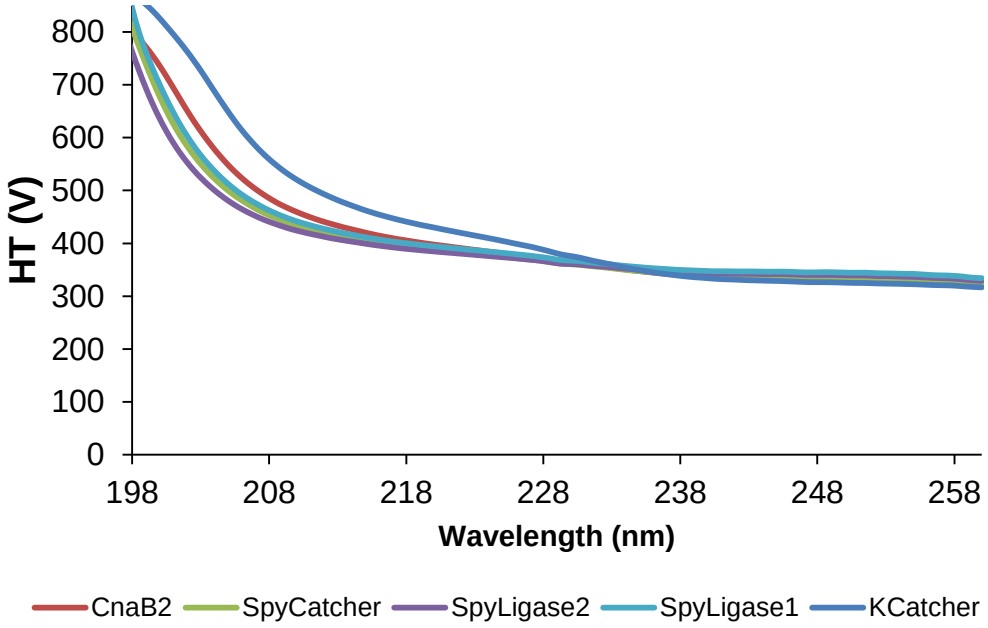


Figure A.11: Circular dichroism spectra of CnaB2, SpyCatcher, KCatcher, SpyLigase1, and SpyLigase 2.

One of the issues of engineering a split protein from CnaB2, either bipartite or tripartite, is the influence on the rest of the structure of removing a β-strand (or two) from the CnaB2 domain. Therefore I examined the secondary structure via circular dichroism

(CD). Figure A.11 shows the CD spectra of CnaB2, SpyCatcher, KCatcher, SpyLIgase1, and SpyLigase2. SpyCatcher has the spectrum closest to CnaB2. KCatcher's spectrum is farthest from CnaB2, with the rest following more closely to CnaB2's structure.

A.13 Information Regarding Crystal Structure of SpyCatcherΔN1 Reconstituted with SpyTag:

Table A.2: Collection of data and statistical refinement of the SpyCatcherΔN1-SpyTag crystal structure by Dr. Long Li.

	SpyTag/SpyCatcherΔN1
Wavelength (Å)	0.97920
Resolution range (Å)	50 - 2.0 (2.03 - 2.0)
Space group	C 2
Unit cell a, b, c (Å) α, β, γ (°)	 71.13 30.47 39.88 90.00 98.93 90.00
Total reflections	16321
Unique reflections	5530
Multiplicity	3.0 (2.9)
Completeness (%)	94.4 (94.6)
Mean I/sigma(I)	12.65 (5.23)
R-merge	0.117 (0.362)
R-work	0.210
R-free	0.237
RMSD (bond lengths, Å)	0.008
RMSD (bond angles, °)	1.11
Ramachandran favored (%)	98.9
Ramachandran outliers (%)	0