1-1-2015

The Development, Use, and Optimization of Sortase-Tag Expressed Protein Ligation

Robert Leslie Warden-Rothman
University of Pennsylvania, rwarden@alum.mit.edu

The Development, Use, and Optimization of Sortase-Tag Expressed Protein Ligation

Abstract
Molecular imaging is an emerging field that seeks to combine the mechanistic detail of biochemical assays with the the broad phenotypic data obtained from non-invasive medical imaging. A cornerstone of molecular imaging is the ability to chemically attach a contrast-inducing payload to a targeting ligand, a protein that binds tightly and specifically to the targeted cell population. There are currently many approaches to bioconjugation. Chemical approaches have broad applicability, but lack the site-specificity and efficiency desired for contrast agent generation. Enzymatic techniques are quantitative, efficient, and site-specific, but often have limited scope and can add significant bulk to the targeting ligand or require additional downstream purification steps. In this work, we create a novel bioconjugation and protein purification technique, Sortase-Tag Expressed Protein Ligation (STEPL). By fusing the active domain of S. aureus Sortase A to the targeting ligand as part of a greater chimeric protein, we are able to combine protein purification and the C-terminal bioconjugation of that protein into a single step. Using a mass-action kinetics model, we prove that this results in a modular, efficient system capable of producing high-purity conjugated protein. We then use the system to generate contrast agents for fluorescent and magnetic imaging studies. The modularity of the STEPL system allows us to create an array of monospecific and bispecific targeting ligand dimers, linked by two different spacer regions. The dimers' binding properties are probed and they are used to successfully bind and label cells, with the bispecific proteins enhancing contrast of a cell line that is positive for both targets. Finally the STEPL system is subjected to a number of directed evolution approaches and screened for a clone that could further improve STEPL's efficiency and yield.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Bioengineering

First Advisor
Andrew Tsourkas

Keywords
antibody drug conjugates, bioconjugation, contrast agents, molecular imaging, sortase

Subject Categories
Biochemistry | Biomedical

The Development, Use, and Optimization of Sortase-Tag Expressed Protein Ligation

Robert L. Warden-Rothman

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Dr. Andrew Tsourkas, Professor of Bioengineering
Supervisor of Dissertation

Dr. Jason Burdick, Professor of Bioengineering
Graduate Group Chairperson

Dissertation Committee
Dr. Mark Goulian, Edmund J. and Louise W. Kahn Endowed Term Professor of Biology
Dr. David Issadore, Assistant Professor of Bioengineering
Dr. Michael Lampson, Associate Professor of Biology
Dr. Donald Siegel, Professor of Pathology and Laboratory Medicine

THE DEVELOPMENT, USE, AND OPTIMIZATION OF
SORTASE-TAG EXPRESSED PROTEIN LIGATION

This thesis is dedicated to my husband, Craig.

He has always been my beacon, guiding me through the stormy waters of life and graduate school. This endeavor would not have been possible without his light and love.

Acknowledgments

The past six years have been some of the most challenging, frustrating, fulfilling, inspiring, and joyful of my life. They were made possible by an amazing network of people who all deserve my gratitude and respect.

First and foremost, I need to thank my advisor, Dr. Andrew Tsourkas. He was always available and willing to offer aid and assistance, coming through lab almost daily to share in our trials and triumphs. In addition to being a role model, he also provided a number of opportunities for me to develop my own skills in teaching, outreach, and mentorship.

I also need to recognize the current and past members of my dissertation committee: Dr. Mark Goulian, Dr. David Issadore, Dr. Michael Lampson, Dr. Donald Siegel, and Dr. Casim Sarkar. In particular, I want to thank Don for teaching me phage display, always being available to answer questions, and guiding me through the initial stages of that chapter of my graduate experience.

There are many members of the Tsourkas lab who deserve a great deal of thanks. I especially need to thank Dr. Julie Czupryna, who went from being my protein engineering mentor to the czar of small animal imaging at Penn, all the while being one of the best friends anyone could ask for. I also want to thank Dr. Drew Elias, Dr. Ajlan Al-Zaki, James Hui, Beth Higbee, Dr. Dan Thorek, Dr. Antony Chen, Dr. Sam Crayton, Dr. Zhiliang Cheng, Dr. Ching-Hui Huang, Dr. Kido Nwe, Dr. Yang Song, Dr. Xuemei Zhang, and all of the collaborators who came through the Tsourkas lab while I was here.

Even before I signed my offer letter, I had begun making the friendships that gave levity and joy to my time here at Penn. Dr. Drausin Wulsin and I met at our interview weekend before living together for three years while our future spouses pursued their own educational endeavors. My other roommate, Dr. Iris Kim, was always up for an adventure and I thank her for that. I was incredibly fortunate to have sat on the border of the Hammer Lab, sharing every day with Dr. Randi Saunders,

Dr. Neha Kamat, and Dr. Laurel Hind, who became my closest confidant at Penn and even member of my wedding party. Drs. G. Aaron Dominguez and Alice Walsh took me under their wing when I first arrived in Philadelphia, and I will always remember the fun and friendship they provided.

For literally as long as I can remember, my parents have pushed me to strive for excellence in everything I do. I cannot thank them enough for the incredible sacrifices they have made to ensure that I had the best education possible: paying for a private school when first grade seemed too easy, making sure I always had a working car and full tank of gas so that I could drive the hour commute each day to my summer internships at Collins & Aikman, following me around the state and country to watch and work FIRST robotics tournaments, and financing two years at one of the most expensive colleges in the country, just to name a few. I hope I've made you proud and that it was all worth the effort. I also need to thank my Philadelphia-based in-laws for truly being a second family. Thank you for everything from finding my first apartment in the city to providing much-needed days at the beach down the shore.

Most importantly, I need to thank my incredible husband, Craig. I can honestly say that I would not be finishing a Ph.D. in Bioengineering without the support and encouragement that Craig has given me every day since I was a wide-eyed freshman at MIT. After my horrific first summer in lab, Craig was somehow able to convince me to drop the introductory mechanical engineering courses and continue with biological engineering. Even way back then, he seemed to know me better than I know myself. Craig has kept me safe and sane throughout this whole process. There are no words that could possibly thank him enough.

ABSTRACT

THE DEVELOPMENT, USE, AND OPTIMIZATION OF SORTASE-TAG EXPRESSED PROTEIN LIGATION

Robert L. Warden-Rothman

Dr. Andrew Tsourkas

Molecular imaging is an emerging field that seeks to combine the mechanistic detail of biochemical assays with the the broad phenotypic data obtained from non-invasive medical imaging. A cornerstone of molecular imaging is the ability to chemically attach a contrast-inducing payload to a targeting ligand, a protein that binds tightly and specifically to the targeted cell population. There are currently many approaches to bioconjugation. Chemical approaches have broad applicability, but lack the site-specificity and efficiency desired for contrast agent generation. Enzymatic techniques are quantitative, efficient, and site-specific, but often have limited scope and can add significant bulk to the targeting ligand or require additional downstream purification steps. In this work, we create a novel bioconjugation and protein purification technique, Sortase-Tag Expressed Protein Ligation (STEPL). By fusing the active domain of *S. aureus* Sortase A to the targeting ligand as part of a greater chimeric protein, we are able to combine protein purification and the C-terminal bioconjugation of that protein into a single step. Using a mass-action kinetics model, we prove that this results in a modular, efficient system capable of producing high-purity conjugated protein. We then use the system to generate contrast agents for fluorescent and magnetic imaging studies. The modularity of the STEPL system allows us to create an array of monospecific and bispecific targeting ligand dimers, linked by two different spacer regions. The dimers' binding properties are probed and they are used to successfully bind and label cells, with the bispecific proteins enhancing contrast of a cell line that is positive for both targets. Finally the STEPL system is subjected to a number of directed evolution approaches and screened for a clone that could further improve STEPL's efficiency and yield.

Contents

List of Tables

List of Figures

List of Synthetic Peptides

List of Plasmid Maps

Chapter 1

Introduction and Scope

The field of molecular imaging uses targeted contrast agents to study and diagnose disease [1]. By revealing the molecular basis of a disease, molecular imaging can help the medical community understand what causes the disease, what treatment options are available, and how well a therapy progresses. The workhorses of molecular imaging are targeted contrast agents, which can seek out and interact with diseased cells in a specific fashion as well as influence some physical property on or near the cells that can alter a medical imaging signal enough that a radiologist can pinpoint where the contrast agent has found cells to bind. These two functions are usually mediated by separate chemical entities, so the ability to efficiently link a targeting molecule to a contrast-producing payload is a cornerstone of molecular imaging.

The importance of the linking reaction has led many groups to develop techniques that accomplish this goal. When both the targeting and contrast agents are small molecules, organic chemistry provides a plethora of efficient linking reactions. However, the targeting ligands are often derived from proteins, which makes the process much more complicated. A number of chemical crosslinking approaches have been described that are capable of attaching a broad range of chemical functionalities to a protein, but they generally suffer from inefficiency and random placement. Enzymatic systems have produced better results because they tend to be site-specific, efficient, and stoichiometric, which are all important for generating imaging agents [2]. However, enzymes are often limited in their substrates, add large amounts of bulk to the imaging agent, or require additional downstream purifications.

In this work, we develop and validate a novel, site-specific enzymatic ligation technique and illustrate its applications to molecular imaging. We create a system that is both capable of accepting most, if not all, categories of potential targeting proteins and payloads. In addition, our system minimizes waste and avoids difficult purification workups, such as separating conjugated and unconjugated proteins whose chemical differences may be as small as a single functional group. The use of protein engineering and molecular biology tools enable us to build this ideal system.

This thesis is organized into six chapters. This first chapter provides a brief motivation and summary of the other chapters. In Chapter 2, we review relevant literature pertaining to the fields of molecular imaging, bioconjugation, and protein engineering. This will provide a solid background for my thesis work, which lies at the intersection of all three fields. In Chapter 3, we describe Sortase-Tag Expressed Protein Ligation (STEPL), a novel bioconjugation and protein expression technique, which we believe has advantages over many existing bioconjugation technologies. In Chapter 4, we use STEPL's inherent modularity to create an array of dimeric targeting ligands and show how these ligands enhance specificity in molecular imaging. In Chapter 5, we evaluate various strategies to enhance STEPL through directed evolution. We then design and screen libraries to increase protein yield. Finally, in Chapter 6, we review the thesis as a whole and discuss its contributions to the fields of molecular imaging and bioconjugation. We will also describe opportunities to extend and enhance the STEPL system for even broader applicability.

1.1 Literature Review

In Chapter 2, we briefly discuss topics relevant to this work.

Molecular Imaging

We describe the concepts underlying the field of molecular imaging as well as some of the pressing medical issues that the field seems poised to solve. We detail the physical chemistry that drives the five main imaging modalities and how contrast agents can be designed to influence their signals. Emphasis is put on fluorescence and magnetic imaging, as they are used throughout this thesis. We then discuss the concepts of passive and active targeting and how contrast agents are localized to diseases. Next, we define biomarkers. Finally, we look at the emerging science of bispecific targeting and its advantages over using a single ligand in particular scenarios.

Targeting Proteins

This section details the origins of many of today's targeting proteins, as well as the advantages and downfalls of each technology. Emphasis is put on antibodies, and fragments thereof, as well as scaffold-derived targeting ligands.

Bioconjugation

In an effort to establish the need for the novel bioconjugation system that is the centerpiece of this thesis, we describe the history and current challenges in the field of bioconjugation. We discuss traditional organic chemistry approaches to labeling proteins with bio-orthogonal functional groups. We detail the mechanisms of these reactions and how the chemistries and proteins can be engineered to produce better reactants. Additionally, we introduce the concept of "click chemistry," a class of efficient, biocompatible chemical reactions. Next, we discuss the use of enzymes to facilitate bioconjugations. Careful attention is paid to the intein-based Expressed Protein Ligation system and sortase-mediated transpeptidation, as these form the basis of our novel system.

Protein Engineering

Finally, we discuss the principles and techniques employed by the two schools of protein engineering, rational design and directed evolution. We describe the power and precision of rational design approaches, given there is enough known about the subject to use these techniques. Then we describe the common methods used to evolve proteins by applying an artificial selective pressure to create useful proteins even with very poor initial characterization.

1.2 Development and Validation of Sortase-Tag Expressed Protein Ligation

In Chapter 3, we introduce Sortase-Tag Expressed Protein Ligation (STEPL). We discuss the new technology's design principles, optimal reaction conditions, and initial uses in molecular imaging.

STEPL Design Strategy

Our novel bioconjugation system is driven by the desire to combine the best aspects of chemical and enzymatic bioconjugation techniques while avoiding their downfalls. We accomplish this by using a sortase domain in the same way that an intein is used in Expressed Protein Ligation. This results in protein purification becoming dependent on bioconjugation, potentially yielding a pure, conjugated protein directly off of the IMAC column. For each component of the STEPL system, we discuss the choice of sequence and location within the chimeric protein.

Determining the Optimal STEPL Reaction

To validate and optimize the STEPL system, we generate an EGFP-STEPL chimera, which allows us to quantitatively monitor the reaction by observing EGFP's transfer from the metal resin to bulk solution. This is first used to determine the optimal concentration for the necessary calcium trigger ion in order to reduce unwanted random cleavage while maintaining enzyme activity. To optimize the other elements of the reaction (time, temperature, and peptide concentration), a mass-action kinetics model is described and fit to EGFP cleavage data. This allows us to determine the optimal set of conditions for different desired outcomes.

Initial Applications

In order to prove STEPL's utility in the context of molecular imaging, the system is used to express small targeting ligands and conjugate fluorescent or chemically functionalized peptides to that ligand. The fluorescent ligands are used directly for microscopy. The functionalized ligand is efficiently attached to the surface of a magnetic nanoparticle. Both systems are assayed for their ability to provide contrast by specifically labeling the cell population to which they are targeted.

1.3 Making Bi-specific Ligands with STEPL

In Chapter 4, we discuss how STEPL is an ideal system for quick and facile generation of useful bispecific targeting ligands.

Modular Protein Dimerization

STEPL is used to express two different targeting ligands and conjugate them to four different functionalized peptides. Ligands bearing compatible chemistries are reacted with one another to produce a panel of six dimeric ligands, which are confirmed by gel electrophoresis. We then discuss the advantages STEPL has in dimer generation over other available technologies.

Increasing Contrast Agent Specificity

We measure the differential ability of the monomer and dimer panel to bind and target cells displaying one or both of the ligands' antigens. These are then used as fluorescence contrast agents to specifically label the double-positive cell population.

1.4 Enhancing STEPL Yield with Directed Evolution

In Chapter 5, we evaluate different systems for directed evolution that could be capable of enhancing the STEPL reaction by decreasing cleavage during expression via altered calcium dependence. We then discuss library designs and the results of screening those libraries by phage display.

Selecting an Evolution System

First, we explore the use of an *E. coli* display system. We describe potential display constructs and screening protocols. Then we briefly consider a yeast display system and discuss why this is sub-optimal. Finally, we look at the use of phage display by detailing a potential screening procedure and the assays validating that screen.

Library Design

We discuss the different options available for library design. We explore the features within the sortase enzyme's structure that are known to bind the calcium ion and how the choice of degenerate codons in that area could effect the enzyme's structure and overall library size and health. Finally, we look at different methods of generating the library and describe a way to accomplish this without biasing the sequences toward wild-type.

Screening Results

This section details our attempts at finding a better sortase domain through phage display. For each library, we detail the conditions used for the positive and negative screens and the sequences that the screen returns. We then clone the sequences into the EGFP-STEPL construct and assay them for activity. For each round, we evaluate potential problems and justify changes that will be made to the next round of library generation and screening.

1.5 Overall Discussion and Future Directions

Finally, in Chapter 6, we summarize the work presented in the previous chapters and discuss STEPL's place in the fields of molecular imaging and bioconjugate chemistry. We also describe future extensions and applications of the technology that will increase its utility and broaden its impact.

Chapter 2

Targeting Ligands in Molecular Imaging

2.1 Molecular Imaging

The field of molecular imaging developed out of the desire to combine the advantages of medical imaging and biochemical assays. While both imaging and biochemical assays are staples of modern medicine and immensely powerful in their own right, neither can tell the whole story of a patient's disease state. Biochemical assays are performed *ex vivo*, requiring the removal of tissue from the subject. The biopsy process irreversibly damages the cells and tissue in the sample and the subject. It also only provides data from a very small portion of an organ at a single point in time, which is problematic when assessing diseases that have a high degree of heterogeneity [3]. And while medical imaging can describe the extent and phenotype of a disease, standard radiology can rarely uncover the underlying mechanisms driving that disease [1]. Despite the limited information obtained by standard imaging techniques, they are heavily relied on in the clinic.

One area where this information gap is especially pertinent is in cancer diagnosis. Although cancer is classically thought of as a single disease, the scientific community has come to understand that what we call "cancer" is really a collection of hundreds of unique proliferation diseases, each with their own genetic causes, therapeutic interventions, and degrees of pathology [4]. Therefore,

when imaging or another early detection method reveals a tumor, it is standard medical practice to assume the worst possible cancer and begin treating that, even though the patient may have a much milder disease [5]. This led to an overdiagnosis of tumors in recent years [6, 7, 8, 9, 10, 11]. Time and again, meta-analysis of tumor screening studies have shown that the once-touted "early screening and diagnostics" recommended by oncologists and cancer advocates do more harm than good [7]. Recently, the overdiagnosis controversy has spilled out of medical journals into mainstream media [12, 13, 14]. One study casting doubt on the benefits of annual mammograms even landed on the front page of *The New York Times* [15, 16]. That study followed 89,835 women over 25 years and found no significant reduction of the mortality rate between those who received yearly mammograms and those who received yearly physicals. This problem is unlikely to go away unless early detection screens begin yielding much more information.

This same problem is echoed in the increasingly specific indications for oncology therapeutics. It is not uncommon for an antineoplastic drug to be effective in fewer than half of its recipients [17]. This can make it incredibly difficult for a drug to pass clinical trials, requiring vast numbers of patients to prove statistical efficacy. More importantly, low-response-rate drugs take a toll on the health of nonresponders, both in unnecessary side effects and time not spent in an effective therapeutic regimen. To combat these economic, regulatory, and human costs, many pharmaceutical and biotechnology companies are using mechanistic knowledge about their drugs to target likely responders by requiring the patients' tumors to present specific biomarkers, molecular components of the cells used to identify a specific disease [18, 19]. Use of biomarkers in Phase I and II clinical trials nearly doubled between 2001 and 2009 [19], with more than 20% of those trials using a biomarker as inclusion criteria [20]. The increasing popularity of targeted cancer therapies will require facile and noninvasive ways to determine which patients fall under a drug's dosing indications.

Molecular imaging stands to solve these problems by connecting the information derived from fundamental cancer research with the utility and broad appeal of medical imaging. Rather than imaging the electronic, nuclear, or physical structures inherent to biology, molecular imaging employs man-made contrast agents that alter the image signal in their vicinity. The contrast agent is then targeted to interact with a specific cellular phenotype. In this way, the contrast agent alters the image only where cells exhibit that phenotype, making it possible to visualize microscopic pathology at a macroscopic level.

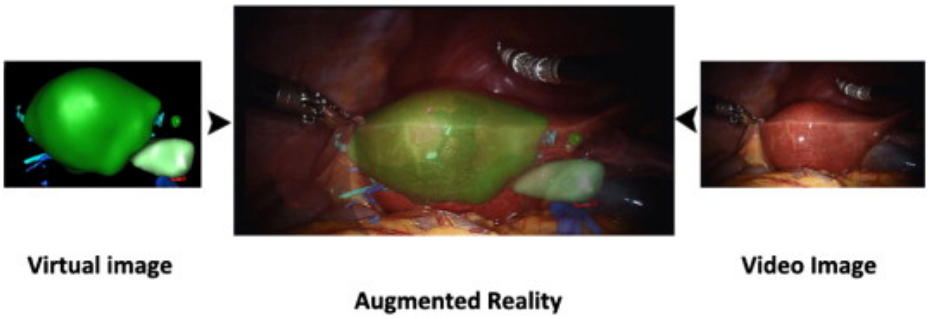


Figure 2.1: Fluorescence Guided Surgery. Injection of a fluorescent contrast agent allows a computer to create a virtual model of the targeted tissue. That model is then superimposed on the real video feed so that the surgeon can be sure they are resecting the entirety of the tissue. Reprinted from [28] with permission from Elsevier.

2.1.1 Modalities

Molecular imaging contrast agents have been produced for all of the major imaging modalities: fluorescence [21, 22], magnetic resonance [23], positron emission tomography/single-photon emission computed tomography (PET/SPECT) [24], X-ray computed tomography (CT) [25], and ultrasound [26]. There is no "ideal" modality. Each must be evaluated for its applicability based on the resolution, sensitivity, and safety requirements of the project at hand. In this thesis, we label cells with fluorescent and magnetic agents, which are described below.

Fluorescence

Fluorescence is a chemical phenomenon that occurs due to the excitation and relaxation of a molecule's electronic structure. When a compound interacts with a photon of compatible energy, the electron in its highest occupied molecular orbital can be promoted to a higher-energy, unoccupied molecular orbital. After promotion, it very quickly relaxes to the lowest vibrational state in the new orbital, releasing the energy as heat. Even then, the heightened electronic state is unstable, and the molecule will relax down to its ground state by converting the excess electronic energy into chemical, thermal, or radiant energy. Radiant energy release is observed as a photon. Thus, fluorescence is when a compound releases a photon in response to photoexcitation. Due to the intermediate vibrational relaxation, emitted photons usually have a lower energy than the exciting photon, a phenomenon known as the Stokes shift [27]. By filtering out the high-energy photons, a clear image can be obtained of the fluorophore's location with an optical microscope or fluorescence scanner.

In media that does not bend or scatter photons, fluorescence has the highest resolution of the medical imaging modalities, as low as the wavelength of the emitted photon (and even more precise with computer-aided ultra-resolution imaging), allowing the visualization of a wide range of biological targets, from sub-cellular organelles to whole organs. Unfortunately, visible excitation photons are readily absorbed and converted to heat in most biological tissues, limiting fluorescence to the top few centimeters of tissue [29]. This makes the modality extremely useful for *in vitro* and small-animal *in vivo* imaging, but greatly limits clinical use for diagnosis. However, one clinical application that has recently emerged is fluorescence-guided surgery [28]. As shown in Figure 2.1, fluorescence and white-light video feeds are overlayed on each other, allowing the surgeon to see the presence of a fluorescent contrast agent in real-time. By showing surgeons the contrast agent, they are able to remove all traces of a targeted cell population, even small tumor foci not visible to the human eye.

Magnetic Resonance

Magnetic Resonance Imaging (MRI) has become a mainstay of medical imaging due to its ability to acquire high-resolution anatomical images without exposing the patient to ionizing radiation [1]. MR images are obtained by recording the bulk magnetic moment of water protons within a given voxel. By carefully timing this measurement in relation to a series of applied magnetic fields, signal differences can be observed between protons belonging to different tissues. MR contrast agents alter how quickly the protons react to the external fields, resulting in a higher [30] or lower [31] signal than the unlabeled tissue.

Higher signals, or positive contrast, are obtained with T_1-weighted contrast agents, generally Gd^{3+} [32]. The paramagnetic ion has seven unpaired electrons in the *d*-orbital, speeding up the magnetic relaxation of coordinated water molecules. This results in more protons aligned with the base magnetic field and a brighter image. Unfortunately, free Gd^{3+} is a potent calcium channel inhibitor and is therefore highly toxic [33]. This can be mediated by chelating the ion with agents such as diethylene triamine pentaacetic acid (DTPA) or 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid (DOTA), which makes the Gd^{3+} biologically inert but also occupies seven of its eight potential coordination sites. This greatly reduces the chelated ion's ability to bind and relax nearby water, requiring very high Gd^{3+} concentrations (10 - 100 μM) in order to produce contrast [34]. However, visible signal can be created by chelating many Gd^{3+} ions into dendrimer nanoparticles.

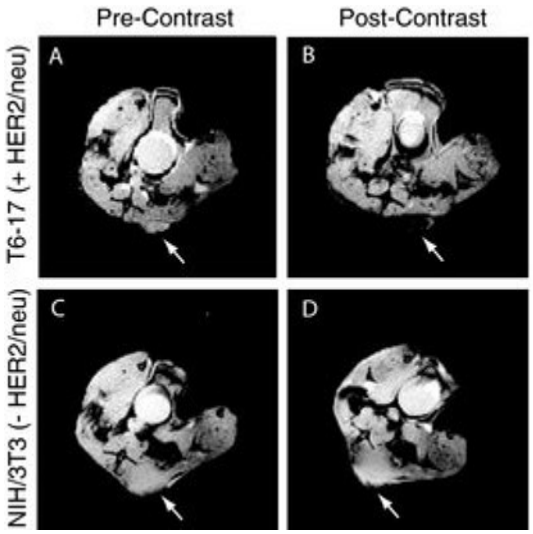


Figure 2.2: SPIO-Mediated Tumor Contrast. MR images of mice before and after administration of a Her2/neu targeted SPIO nanoparticle. Nanoparticle accumulation in the targeted tumor (B) greatly decreases signal in the tumor tissue compared to the pre-administration image (A). Contrast is not see on in the tumor that does not display Her2, either before (C) or after (D) administering the nanoparticle. Arrows point to the location of the xenograft tumor. Adapted from [2] with permission from John Wiley and Sons.

Our lab has loaded up to 300,000 Gd^{3+} ions onto a single dendrimer nanocluster, which successfully labeled folate-receptor positive tumors in a murine model [35].

Lower signals, or negative contrast, are usually obtained with T_2-weighted magnetic nanoparticles, like superparamagnetic iron oxide (SPIO) crystals [32]. SPIO crystals align nearby water protons to their own magnetic field, lowering the number aligned with the base magnetic field and creating a darker image where SPIO is present, as shown in Figure 2.2. This effect can be enhanced by concentrating many SPIO crystals into the core of a nanoparticle [36]. Our lab also uses SPIO particles, and has successfully labeled Her2-positive cells *in vivo* and *in vitro* on a number of occasions [2, 37, 38, 39].

Nuclear Imaging

Radioactive nuclei can be detected using one of two techniques, positron emission tomography (PET) or single-photon emission computed tomography (SPECT), depending on their decay mechanism. For PET, radiohalides that undergo positron emission (β^+) decay, like ^{18}F and ^{64}Cu, spontaneously emit positrons. These quickly collide with their anti-particle, the electron, and annihilate, creating

two 511 keV electrons traveling in exactly opposite directions. The electrons are detected and their detection time and position are used to calculate their origin, thereby locating the agent that produced them. By far, the most common PET agent used in the clinic is ^{18}FDG, a glucose analog that is taken up by cells with high metabolic activity. While a highly active glycolysis pathway is unusual for healthy tissues, it is a common trait shared by many tumors, making ^{18}FDG an effective broad-spectrum tumor imaging agent [1]. The other nuclear imaging technology, SPECT, detects radiohalides that decay by gamma emission, like ^{99m}Tc and ^{123}I. Because these nuclei only emit a single photon, their source's position is harder to determine, making them orders of magnitude less sensitive than PET agents. They also require halides with longer half-lives, further limiting SPECT's clinical relevance due to radio-exposure concerns [1].

X-ray computed tomography

Like a standard X-ray radiograph, X-ray computed tomography (CT) measures the attenuation of high-energy X-ray photons while traveling through the patient's body. However, CT is capable of producing a three-dimensional reconstruction of the patient's anatomy, making it much easier for a radiologist to delineate anatomical structures. Because the ionization energy of iodine's K-shell electrons is similar to that of the X-ray photons, iodine has a large X-ray attenuation coefficient and iodinated chemicals can be used to produce CT contrast [40]. These contrast agents generally have low sensitivity, limiting their clinical utility due to the high amounts of iodinated contrast agent necessary to alter the bulk X-ray attenuation of a tissue. CT is also limited by the number of times a patient can be exposed to the ionizing X-ray photons in a given time period to minimize tissue damage [1].

Ultrasound

Clinical ultrasound (US) detects the mechanical responses of biological tissues to high-frequency, 1 to 20 MHz, sound waves transmitted through that tissue. The US transducer emits a sonic pulse which then travels along a straight line in the tissue. When that pulse reaches an interface between tissue types, some of the mechanical energy is reflected back toward the transducer, which then detects the reflection and calculates the distance to the interface based on the time it took to receive the signal. The strength of the reflection is determined by the tissue types at the interface, allowing some level

of delineation. Molecular contrast can be produced by targeted microbubbles, gas-filled liposomes or polymerosomes. These bubbles can resonate with the US field, making them orders of magnitude more reflective than normal tissues, enabling detection of micron-sized cell populations [1].

2.1.2 Targeting

Contrast agents need to be directed to their target cell population in some way. This can be done passively or actively. Passive targeting relies solely on the physical properties of the agent and target. The most common passive targeting approach utilizes the enhanced permeability and retention (EPR) effect [41, 42]. The EPR effect is a phenomenon that occurs when a tumor induces rapid angeogenesis to oxygenate its core. These blood vessels are made quickly and poorly, allowing solutes in the blood to "leak" into the tumor interstitium. In addition, tumors generally have poor lymphatic drainage, increasing the retention time of the solutes that leaked in from the blood. As a whole, the EPR effect allows nanoparticles to "target" a tumor, simply because they preferentially accumulate in that tissue.

In order for a nanoparticle to take advantage of the EPR effect, it must have a long circulation time so that it can randomly leak into the tumor [43]. This can be achieved by coating the particle with hydrophilic, biocompatible, inert materials, such as poly(ethylene glycol) (PEG). The sterics of the PEG chains make it difficult for monocytes to phagocytose the particles, slowing their excretion. Many groups have PEGylated their nanoparticles in order to target them via EPR [44, 45, 46].

A better approach for agents that require high concentrations for detection is active targeting. Actively targeted agents rely on the contrast agent carrying a chemical group that binds tightly and specifically to the target population. This binding event confers a few advantages on actively targeted contrast agents. Active agents do not require poor lymphatic drainage to accumulate, as their affinity to the tumor cells themselves prolongs their retention time [47]. They are able to be internalized by recycled receptors, which greatly increases their local concentration and signal contrast. On a decorated particle, the binding is multivalent, which can trigger internalization even if the target biomarker is not recycled [48].

Targeting ligands can take a number of forms. Popular small-molecule targeting ligands include folate [35, 38, 49, 50], estrone [51, 52], and TRX-20 [53, 54], which target cell surface folate receptors, estrogen receptors, and chondroitin sulfate, respectively. These are easily attached by standard

organic chemistry to nanoparticles or other contrast agents, making them an attractive choice. In addition, due to their small size, a large number of small molecule ligands can be packed onto a nanoparticle surface, increasing particle avidity and internalization.

However, small molecule ligands do not exist for most relevant biomarkers (including those detailed in Section 2.1.3), so more complex targeting proteins are required. Protenaceous targeting ligands must have a high affinity and specificity for their target biomarker. Sometimes these proteins can be found in nature, such as using the epidermal growth factor to target the epidermal growth factor receptor [55], but they are usually antibody derivatives or evolved binding scaffolds. Section 2.2 describes these proteins in detail. Attachment of proteins to the contrast agent is much more difficult than to small molecules, but several strategies have emerged, which are discussed in Section 2.3. Although multivalency is also possible with protein-decorated nanoparticles, avidity is inhibited at high protein densities and an optimal ligand density must be determined for each particle [38].

2.1.3 Biomarkers

A biomarker is any objectively measurable characteristic that gives insight into a biological process or therapeutic response [56]. They can fall into one (or more) of four main categories. First are diagnostic markers, which can be prognostic, indicating the presence or absence of disease, and/or predictive, suggesting potential therapeutic treatments. Second, disease activity biomarkers report the current disease severity, an important quantity when assessing therapeutic efficacy. Third, drug effect biomarkers show how well a drug is interacting with its intended target and are used to determine dosing. Note that this is not the same as a disease activity marker because a drug reaching its target does not mean it will alter the course of the disease. Finally, there are drug kinetics biomarkers, which measure prodrug processing and transport to the desired tissue. Biomarkers can also be classified by their location in the tissue: genomic, cytoplasmic, cell surface, or systemic.

In molecular imaging, the focus is on diagnostic biomarkers, like those listed in Table 2.1 and detailed below, that detect genomic or proteomic differences from healthy cells. In oncology, this is often in the form of genetic amplifications, which lead to heightened levels of a protein in the blood or on the surface of a cell. *In situ* hybridization, quantitative PCR, sequencing, ELISA, and immunohistochemistry are commonly used to quantify the amount of biomarkers in patient samples. These biomarker levels can be highly predictive of a patient's response to a given drug, especially

Table 2.1: Diagnostic Biomarkers

Marker	Disease	Type	References
AFP	Liver Cancer; Germ Cell Tumors	Overexpression	[57]
ALK	Non-small cell lung cancer; anaplastic large cell lymphoma	Splice Variants	[58, 59]
B2M	Multiple myeloma; Chronic Lymphocytic Leukemia; Peripheral Arterial Disease	Overexpression	[60]
BCR-ABL	Chronic Myeloid Leukemia	Translocation	[59]
Beta-hCG	Choriocarcinoma; Testicular Cancer	Overexpression	[61]
BRAF	Cutaneous Melanoma; Colorectal Cancer	Mutation	[58, 59]
CA125	Ovarian Cancer	Overexpression	[62]
CA15-3	Breast Cancer	Overexpression	[63]
CA19-9	Pancreatic Cancer; Gallbladder Cancer; Bile Duct Cancer; Gastric Cancer	Overexpression	[64, 65]
CD20	Non-Hodgkin Lymphoma	Overexpression	[1]
CEA	Breast Cancer; Colorectal Cancer	Expression	[1, 66]
CgA	Neuroendocrine Tumors	Overexpression	[67]
Cytokeratin	Pancreatic Cancer	Splice Variants	[68]
EGFR	Mutliple Solid Tumors	Overexpression; Mutations	[1, 3, 56, 58, 59]
ER/PR	Breast Cancer	Overexpression	[58, 59]
HE4	Ovarian Cancer	Overexpression	[66, 69]
Her2	Breast Cancer; Gastric Cancer	Overexpression	[1, 3, 58, 59]
Integrin $\alpha_v\beta_3$	Solid Tumors	Overexpression	[1]
KIT	Gastrointestinal Stromal Tumor; Mucosal Melanoma	Mutation	[58, 59]
KRAS	Colorectal Cancer; Non-Small Cell Lung Cancer	Mutation	[58, 59]
Mammaprint	Breast Cancer	Multiple Overexpressions	[56]
MMP-2	Melanoma	Activity	[1, 70]
MMP-9	Melanoma	Activity	[1, 70]
Oncotype DX	Breast Cancer	Multiple Mutations	[56, 59]
Ova1	Ovarian Cancer	Multiple Overexpressions	[66, 71]
PDGFR	Malignant Pleural Mesothelioma	Overexpression	[72]
PK11195	Cerebral inflammation	Expression	[73]
PSA	Prostate Cancer	Overexpression	[1]
Thyroglobulin	Thyroid Cancer	Overexpression	[74]
uPA	Breast Cancer	Activity	[75]
VEGFR	Angiogenesis	Overexpression	[3, 70]

for targeted therapeutics [58]. Biomarkers that can be measured by immunohistochemistry, such as cell-surface receptors, are prime candidates for molecular imaging because they are accessible to targeting ligands [1].

EGFR

The epidermal growth factor receptor (EGFR) is the primary member of the ErbB family of receptor tyrosine kinases, and plays an important role in cellular replication and differentiation [76]. Due to its importance in regulating proliferation, overexpression of EGFR is common in many solid tumors, including lung, breast, colorectal, and squamous cell tumors, among others [77]. Many studies have targeted imaging agents and therapeutics to EGFR-positive tumors [78, 79].

Her2/neu

Human epidermal growth factor receptor 2 (HER2) is another receptor tyrosine kinase in the ErbB family. It is overexpressed in approximately one third of all breast cancers due to an amplification of the Her2/neu gene. This results in an overactive proliferation signal and generally aggressive

cancers with poor clinical prognoses [80, 81]. Due to its extreme clinical relevance, a number of therapeutics and diagnostics have appeared for Her2-positive tumors both in the clinic [82] and in academia [83, 84, 85, 86].

2.1.4 Bispecific Ligands

Sometimes, one biomarker is not enough to selectively label a diseased cell. Many biomarkers, like the ones described in the previous section, are displayed on a wide range of cells, both healthy and diseased. And while the diseased cells may express the marker at a much higher level, even low quantities of a surface receptor can mediate the uptake of a targeted particle [87]. One method to further increase selectivity for a target cell population is to employ a bispecific ligand, a targeting ligand that can bind to two different biomarkers. Many cancers express multiple disease biomarkers, often in unusual combinations [88]. By requiring both markers, it is possible to greatly increase the signal-to-noise ratio of a contrast agent or the therapeutic index of a pharmaceutical. For example, Yan *et al.* targeted a ^{18}F labeled bispecific peptide to GRPR (a marker of lung, colon, and prostate cancers) and integrin $\alpha_v\beta_3$ (a general marker of solid tumors) [89]. They were then able to use MicroPET to specifically label double positive PC-3 cells in a murine tumor model. In a different study, Nie *et al.* targeted a PEG-PEI polyplex to the often-overexpressed transferrin receptor (TfR) and integrin $\alpha_v\beta_3$ to deliver a luciferase gene. Not only did they see increased selectivity for tumor cells, they also saw a synergistic effect, where the dual-targeted polyplex transfected cells more efficiently that an equal dose of mixed TfR-targeted polyplex and integrin $\alpha_v\beta_3$-targeted polyplex [90]. Synergy has been seen in a number of other studies, as well [91, 91, 92, 93].

A different application of bispecific ligands is to have one binding domain find a biomarker and the other bind to the imaging agent/therapeutic itself. This is often done with bispecific antibodies because the long circulation time and poor clearance of the antibody become problematic with toxic imaging payloads [88]. For instance, Yazaki *et al.* made a bispecific antibody where one arm bound the carcinoembryonic antigen (CEA) and the other variable domain bound the chelator DOTA [94]. They injected the antibody into nude mice bearing a CEA-positive tumor and allowed the antibody to circulate, bind, and clear from the blood. They then gave an injection of ^{77}Lu $\cdot$ DOTA and allowed that to bind and clear. Biodistribution studies showed that by using the pretargeting approach, the group was able to increase the tumor-to-blood ratio of ^{77}Lu from 3:1 with a directly labeled antibody

to 199:1 with the pretargeted bispecific antibody.

Another use of bispecific ligands is to deliver agents across the blood-brain barrier. Unlike most other tissues in the body, solutes in the blood do not easily diffuse into the brain. For example, IgG content in the brain interstitial fluid is only 0.1%-0.7% of what is found in the blood [95]. This makes it very hard to use molecular imaging to study neurological disorders. However, Yu *et al.* discovered that a low-affinity protein targeted to the transferrin receptor was able to undergo transcytosis and enter the brain compartment, so long as there is a sufficiently high dose of the agent in the blood to drive binding on that side of the barrier [96]. By making bispecific antibodies with one arm having a low affinity to transferrin receptor, they were able to separately image both the neuronal marker NeuN and β-secretase in mouse brains (with the other arm of the antibody localizing to those targets).

Bispecifics can also be used to crosslink two different cell types. This is most commonly done is the context of bispecific T-cell engager (BiTE) ligands, which target CD3 on killer T-cells and a tumor biomarker. Thus, the bispecific ligand recruits $CD3^+$ cells to the tumor environment and activates their cytolytic pathways. The first BiTE therapeutic, Blinatumomab, was recently approved for use in the United States, in December 2014, for patients with Philadelphia-chromosome-negative relapsed/refractory B-cell precursor acute lymphoblastic leukaemia (BCP-ALL) [97]. It targets both CD3 and CD19, a leukemia biomarker. Not only did the drug put most of its patients into complete remission within the first treatment cycle, it also removed PCR-detectible traces of residual disease in 78% of those patients [98]. If these results can be replicated in other tumor types, BiTE may become a powerful new class of anti-cancer agents.

2.2 Classes of Targeting Proteins

Targeting proteins have a wide array of origins, which influences the final protein's size, shape, stability, and manufacturability. Just as there is no ideal imaging modality, there is no best targeting ligand. Each one must be chosen for the desired application. Desirable targeting proteins will have high affinity and selectivity for their targets, be small enough to allow polyvalency on a particle, and simple enough to be produced robustly and cheaply.

Targeting proteins can be sorted into two major categories: those that are derived from antibodies

and those that are evolved from a scaffold. Antibody-derived targeting proteins consist of portions of an antibody that has been raised in an animal model. Scaffold-derived targeting proteins are derived from mutated portions of natural structural or binding protein domains, such that they have a high affinity for their new target and low affinity for their natural binding partners. Pertinent examples of both are described below.

2.2.1 Antibodies

Antibodies, specifically IgGs, are the most common class of targeting protein. This is mainly due to their high affinity and long history of utility in biomedical research [99, 100]. While costly and complex to manufacture, the process of producing monoclonal antibodies from hybridomas has been used since the late 1970s, diminishing these factors' importance [101]. A larger problem to molecular imaging and targeted therapeutics has been immunogenicity [102]. However, the steadily dropping cost of DNA sequencing and rise of genetic engineering has made it possible to "humanize" an antibody by cloning the antigen-binding regions of a non-human antibody into the human IgG sequence [103]. The final factor to consider is size. Antibodies are the largest targeting proteins we will discuss, with a hydrodynamic radius of about 5 nm [104]. This limits the number that can be packed onto a nanoparticle, which may only have a radius of 20 or 30 nm itself [2]. It also gives IgG a half-life in the blood of 23 days [105], much longer than is desirable for imaging applications because of extended exposure to toxic imaging agents and prolonged times between imaging studies.

2.2.2 Antibody Fragments

Many of the problems with monoclonal antibodies can be avoided by using antibody fragments. Protease digestions and molecular cloning have produced a plethora of antibody segments, which can be seen in Figure 2.3. These smaller antibody pieces have many properties that are better suited to imaging agents: faster clearance, improved tissue penetration, and increased tumor-to-blood ratios [99]. However, these improvements are often counterbalanced by lower affinities, manufacturing difficulties, slower uptake, and aggregation [106]. Still, antibody fragments have shown some clinical success. Although scFvs were the first monovalent fragments to be developed, they are cleared from the blood too fast to be good imaging agents on their own. They are ideal for nanoparticle

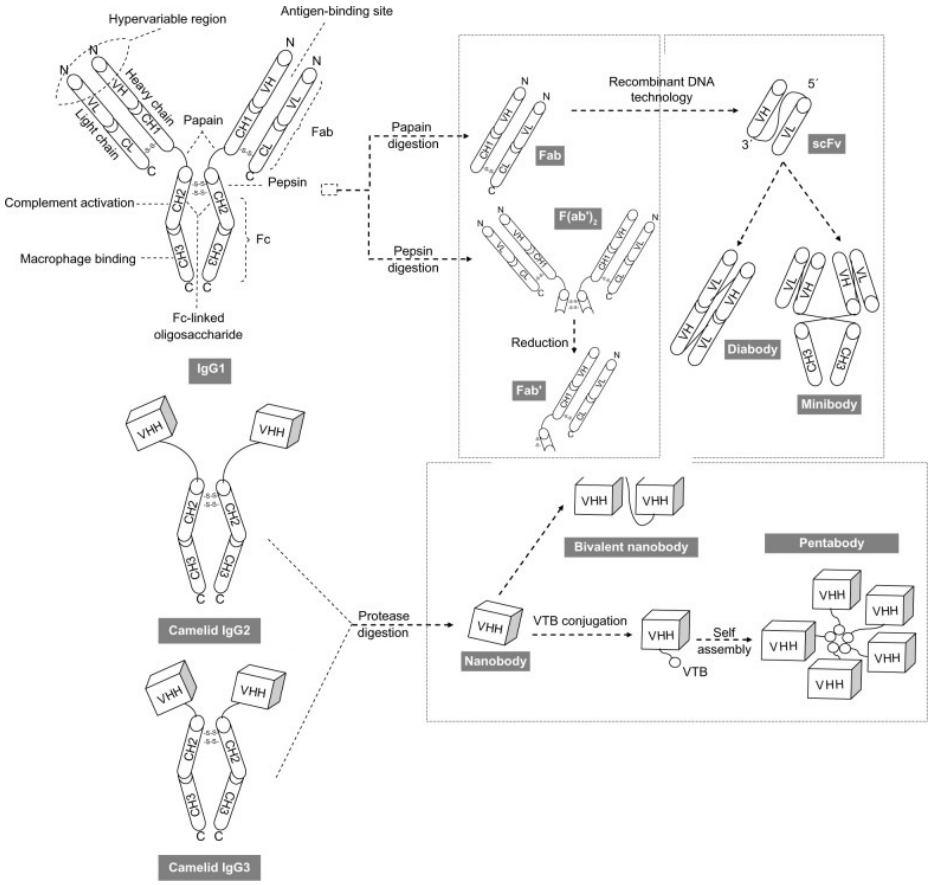


Figure 2.3: Antibody Fragments. Protease digestion and recombinant DNA technologies can produce many useful fragments of both standard and camelid IgGs. Adapted from [106] with permission from DOVE Medical Press.

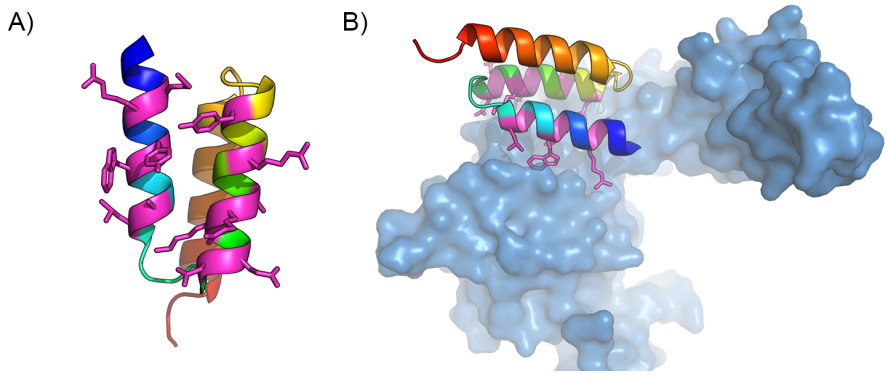


Figure 2.4: Affibody Structure. A) The three-helix bundle structure of an affibody. The thirteen mutable side chains are shown and their positions colored magenta. B) The mutated face of the Z_{Her2} affibody forms the binding interface with the extracellular domain of Her2. The images are produced using crystal structures determined by Eigenbrot *et al.* [107].

conjugations, though, as their small size allows for a much denser packing on the particle surface than an antibody and the particle mass ensures longer circulation times. Diabodies and minibodies, whose structures are detailed in Figure 2.3, seem to be big enough to have an acceptable half-life and have been developed into clinically relevant PET/SPECT agents [1].

A currently emerging class of antibody fragments is derived from camelid, or heavy-chain, IgGs. Usually produced by inoculating llamas, camelid IgGs lack a light chain, disulfides, and glycosylation, making them much easier to manufacture. Due to the single-domain structure of the VHH region, the antigen binding domain is much more soluble and can even be expressed in *E. coli* [106].

2.2.3 Affibodies

Affibodies are small (6.6 kDa) targeting proteins derived from a portion of *S. aureus* protein A (SPA) that naturally binds IgG [108]. The 58-amino-acid B domain of SPA was mutated to increase chemical stability, creating protein Z, shown in Figure 2.4A [109]. Protein Z has a number of useful properties. The small size and lack of structural disulfides make affibodies ideal candidates for expression in bacterial systems. Protein Z has also been shown to have remarkable heat resistance and structural integrity, so much so that it has the fastest refolding time of any measured protein at just 3 μs [110]. The small size also allows for high ligand to nanoparticle ratios, which increases particle avidity and uptake [38].

Most importantly, thirteen of the amino acids on one face of protein Z can be simultaneously

mutated to any other residue without altering the stability and structure of the protein [108]. This allows the protein Z scaffold to be evolved to bind tightly to a target biomarker (see Section 2.4 for details). These matured proteins are known as affibodies. Affibodies have been evolved to target a wide array of biomarkers, including Her2, transferrin, amyloid β peptide, EGFR, Factor VIII, CD25, HIV gp120, and CD28 [111]. They have very high affinities (in the mid- to low-nanomolar range) for their targets. Affibodies have a standardized nomenclature of a capitol Z with the target molecule as a subscript (Z_{TARGET}).

An affibody often seen in the literature is Z_{Her2}, which can be seen binding its target in Figure 2.4B. This affibody has been particularly useful due to its extremely high affinity (K_D = 22pM, [107]) and specificity for Her2. Z_{Her2} has proven highly successful in a large number of therapeutic, imaging, and nanoparticle studies [2, 37, 38, 55, 112, 113, 114, 115].

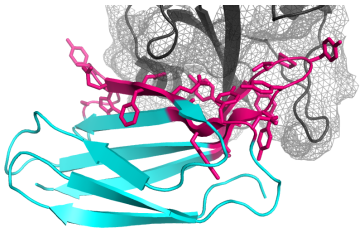
2.2.4 Other Scaffolds

More than 50 different scaffold domains have been employed in creating targeting proteins over the past few decades. Most of these scaffolds were never developed after their initial publication. However, a few, like affibodies, have stood out in the literature and continue to be developed by labs around the world [116].

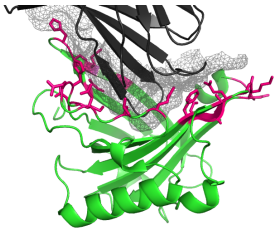
Monobodies, also called Adnectins, are based on the 10^{th} extracellular domain of human fibronectin III (^{10}Fn3) [121]. Monobody domains have structural homology to IgG. The 94 residues form a β-sandwich fold with 2 or 3 exposed loops, which are often the randomized region. Unlike IgG however, monobodies lack a central disulfide bridge, making them much easier to express in prokaryotic systems.

Anticalins are derived from lipocalins [122]. They are the largest of the common affinity scaffolds, forming a conical β-barrel fold comprised of 180 amino acids, which is still less than one third the size of an antibody. As shown in Figure 2.5B, the anticalin binding interface is created by randomizing the loops and internal residues near the wide end of the barrel. The domain shows a high level of plasticity, often flexing from its ground-state conformation when binding to its target [116].

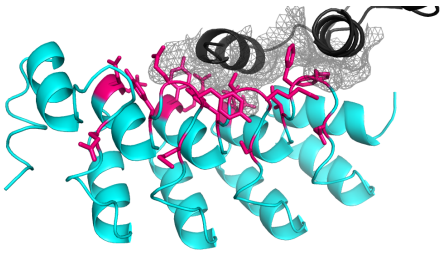
Designed ankyrin repeat domains (DARPins, Figure 2.5C) are affinity scaffolds that are made of five stacked, short coiled-coil domains, usually totaling 166 residues [123]. Libraries are created by randomizing 6 residues along the same side of each repeat structure. The structures are quite rigid,



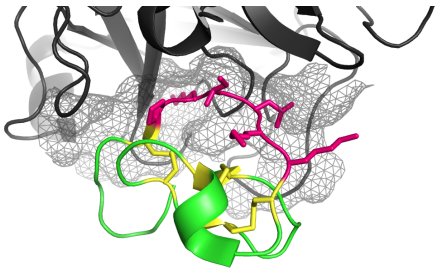
(A) Monobody (PDB ID: 4JE4 [117])



(B) Anticalin (PDB ID: 3BX7 [118])



(C) DARPin (PDB ID: 2BKK [119])



(D) Knottin (PDB ID: 4GUX [120])

Figure 2.5: Other Scaffolds. Scaffold proteins are shown in **blue** or **green**. Residues that had been randomized to create a binding interface are colored **pink**. Disulfide bonds are colored **yellow**. Target molecules are displayed in **black**. An electron density map, represented by a **gray** mesh, shows how well the evolved residues of the affinity ligand fit with the target protein.

and usually bind their target with a lock-and-key mechanism [116].

Another common structural motif is the knottin. These are the smallest of the scaffolds, usually around 30 amino acids [124]. Six cysteine residues, forming three disulfide bridges (shown in yellow in Figure 2.5D), are required so that a protein this small can have a stable tertiary structure. These interlocking disulfides also make knottin ligands exceptionally stable with regards to boiling, low pH, and many proteases. Unfortunately, the large number of disulfides make them notoriously difficult to express and manufacture efficiently.

2.3 Targeting Ligand Conjugation

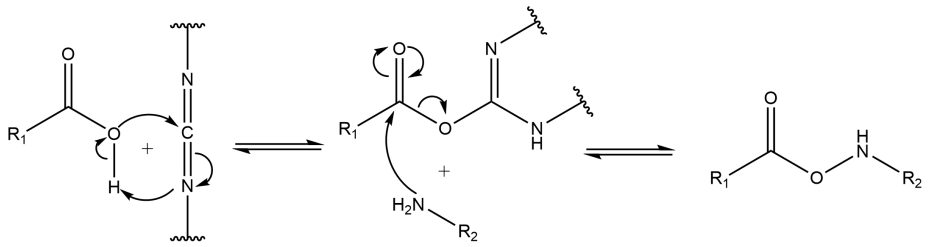
Bioconjugation is the process of forming a chemical bond between a biological agent, such as a protein or nucleic acid, and a synthetic agent. This is critical to molecular imaging because the targeting and imaging aspects of a contrast agent are rarely entirely biological or synthetic. ^{18}FDG and chimeric fluorescent proteins are notable exceptions. As such, bioconjugations are widely used in the field of molecular imaging, and the existence of quantitative, efficient conjugate chemistries are of the utmost importance [1, 125].

2.3.1 Chemical Ligations

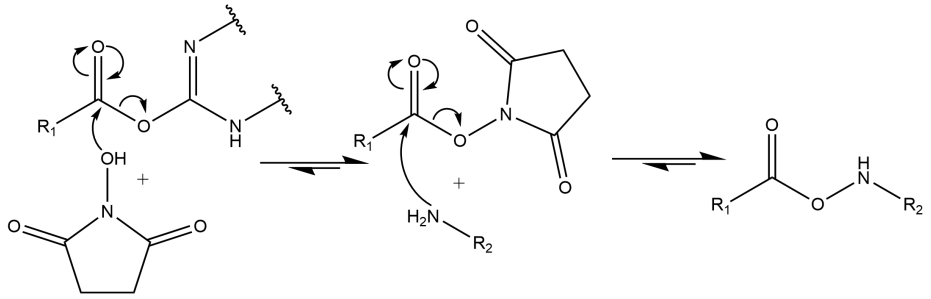
A number of chemical methods have been developed to label ligands with cargo, including maleimide, N-hydroxysuccinimide, carbodiimide, and click chemistries, all of which will be described below [50, 126, 127, 128]. However, many of these approaches suffer from poor reaction efficiencies and indiscriminate labeling of nucleophilic residues (e.g. lysines and cysteines) on the targeting ligand. Random labeling of targeting ligands is problematic because a poorly placed cargo can greatly reduce a ligand's affinity for its target. Additionally, each targeting ligand may be labeled with any number of functional moieties, eliminating their ability to be quantitative or stoichiometric.

Carboxyl-Amine Crosslinkers

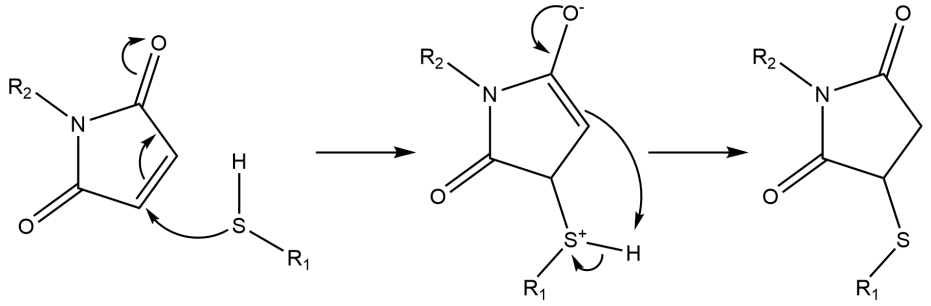
Carbodiimides are a class of chemicals that can activate carboxylic acids by readily undergoing condensation reactions and then serving as excellent leaving groups for amide formation with free amines. The mechanisms are shown in Figures 2.6A and 2.6B. The carbodiimide group is relatively



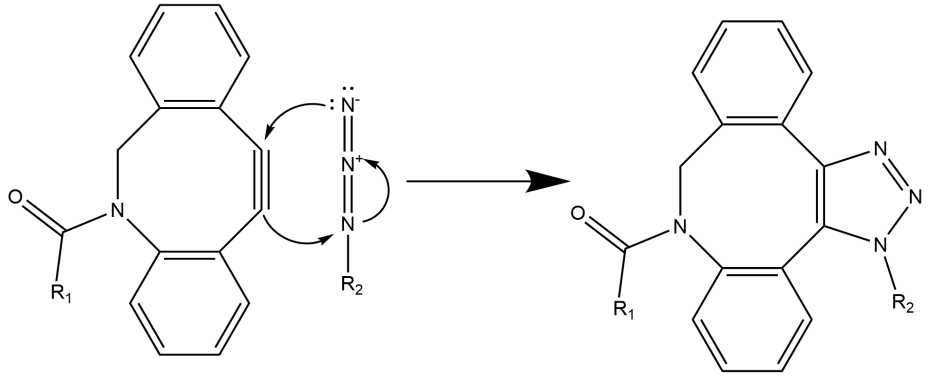
(A) Carbodiimide Heterobifunctional Ligation



(B) Carbodiimide-NHS Heterobifunctional Ligation



(C) Maleimide Ligation



(D) Strain-Assisted Azide-Alkyne Cycloaddition

Figure 2.6: Chemical Ligations

unstable [129], thus, it is often used in combination with *N*-Hydroxysuccinimides (NHS) to improve reaction efficiencies.

Typically, the reactive carboxylate resides on the imaging agent. Carboxylic acids are much more common on proteins than primary amines [130], so activation of the imaging agent will decrease the heterogeneity of the conjugation. Pre-activated agent will then be mixed with the targeting ligand, which will form amide bonds with the imaging agent, producing a targeted contrast agent. Although this conjugation strategy is popular, it is not efficient. Reaction yields are commonly between 1 and 20% when attaching a targeting ligand to a nanoparticle [129]. It also produces links at random sites on the protein, which is deleterious to imaging agent efficacy [38].

Thiol-Amine Crosslinkers

Another popular chemistry for heterobifunctional crosslinking is the maleimide-thiol reaction. The groups undergo an addition reaction in the presence of even a modest base, as shown in Figure 2.6C [131]. The maleimide is generally only part of a crosslinker, not a part of the imaging agent itself. The commonly used crosslinker is NHS-maleimide, which acts as a heterobifunctional thiol-amine crosslinker. Part of the popularity of this linking reaction is due to the fact that while thiol groups (cysteines) are present in most proteins, they are even less abundant than primary amines [132]. Some proteins, like Annexin A5 (an apoptosis biomarker) have no natural cysteine residues at all, allowing Fonge *et al.* to introduce a single cysteine by site-directed mutagenesis [133]. They were then able to conjugate their cys-Annexin A5 site-specifically and quantitatively to ^{99}Tm−HYNIC. Thiols can also be added to a protein via treatment with N-succinimidyl S-acetylthioacetate (SATA), which converts free amines to sulfhydryl groups. Caution must be used while adding free thiol groups, however, because they can spontaneously form disulfide bonds that lead to unwanted dimerization and protein aggregation.

The converse, cysteine removal, is also possible. This type of cysteine engineering is often used in the context of antibody drug conjugates. Antibodies naturally have eight cysteine residues that form four disulfide bridges. Oddly, antibodies assemble correctly and retain their antigen affinity with these bonds removed, although they lose some effector functions [134]. Thus, they can be used as bioconjugation handles, allowing quantitative addition of eight drug molecules [135]. However, a 2-3 fold increase in therapeutic window is achieved by reducing the number of drugs per antibody to

2-4 [136]. Reducing the number of conjugates has the unfortunate effect of introducing heterogeneity to the antibody-drug conjugate, which is undesirable from efficacy, manufacturing, and regulatory standpoints. Because the eight cysteine residues are not required, McDonagh *et al.* were able to mutate some of the residues to serine without changing the structure or binding function of an α-CD30 monoclonal antibody [137]. They were able to create homogeneous, effective antibody-drugs conjugates by cysteine engineering and maleimide chemistry.

Azide-Alkyne Cycloaddition

A chemistry growing in popularity is the azide-alkyne cycloadditon, which creates a covalent linkage in the form of a 1,2,3-triazole (shown in Figure 2.6D). The reaction is reliable, high yielding, biocompatible, and bio-orthogonal. In fact, this reaction is efficient enough to be deemed a "click" chemistry, which was defined by Sharpless *et al.* in 2001 [138] to be reactions that are: "modular, wide in scope, give very high yields, generate only inoffensive byproducts that can be removed by nonchromatographic methods, stereospecific (but not necessarily enantioselective)." They must also have the following process characteristics: "simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation." In the presence of a copper catalyst, several groups have shown cycloaddition yields of greater than 90% in less than 24 hours at standard temperature and pressure [139, 140]. Unfortunately. the copper catalyst has also shown cytotoxicity, limiting its usefulness [141]. To increase the biocompatibility while retaining reaction efficiency, the linear alkyne functional group was replaced by a cyclooctyne [142]. The cyclic structure creates enormous ring strain and primes the triple bond for cycloaddition. Not only did this allow the click reaction to proceed without a catalyst, it also accelerated the reaction compared to the copper-catalyzed incarnation [142]. The acceleration effect was further enhanced by electrophilic substitutions at C3 and C8 around the cyclooctyne ring. Two of these substituted rings, dibenzocyclooctyne (DBCO) and difluorocyclooctyne (DIFO) show triazole formation rates 60-fold faster than their unsubstituted counterparts [143, 144]. The major drawback of azide-alkyne cycloadditions in bioconjugation is that neither functional group is found in nature, so they must be conjugated to a protein by another method, such as NHS-azide or DBCO-maleimide, prior to the click reaction.

2.3.2 Enzymatic Ligations

Enzymatic ligation techniques overcome many of the shortcomings attributed to chemical ligations because they are generally site-specific and stoichiometric [128]. However, many naturally-derived ligases are limited to natural, or minimally modified, substrates like biotin or lipoic acid [128]. These additives can then be reacted with the desired cargo. Additionally, the enzymes need to be purified from the conjugate product, a potential difficulty. Other enzymatic systems, such as SNAP tags, fuse the enzyme itself to the protein of interest, adding significant and potentially disruptive bulk to the targeting ligand [126].

Expressed Protein Ligation

Inteins are a class of enzymes that catalyze their own extraction from a mature protein by using a cysteine or serine residue (the first amino acid within the intein domain) to displace the peptide bond immediately preceding it, as shown in Figure 2.7. A transesterification then occurs between the newly formed (thio)ester and a nucleophilic residue immediately following the intein domain. Finally, a glutamine in the intein cleaves the intein domain from the branched polypeptide and the ester rearranges to form a peptide bond, leaving an intact protein.

Many intein domains can be split to act in *cis*, allowing them to ligate together two separate proteins. An intein particularly adept at this is the intein domain from *Nostoc punctiforme* DnaE [145]. It can be split into a 123 a.a. N-terminal domain (IN) and a 36 a.a. C-terminal domain (IC) that spontaneously associate and restore intein activity. The C-terminal domain is small enough to be produced by solid-state peptide synthesis, making it amenable to the inclusion of bio-orthogonal chemistries or non-canonical amino acids. Borra *et al.* took advantage of this possibility to site-specifically label a transcription factor *in situ* [146]. They synthesized an IC peptide containing a FITC fluorophore C-terminal to the intein region and a dabcyl within the intein, which showed 99% quenching and improved their signal 30-fold. They transfected the quenched peptide into cell lines expressing the YY1 transcription factor fused to the IN domain. The split intein reacted in the cytosol to create a fluorescent YY1 that they tracked to show the effect of active vs. passive nuclear transport on YY1.

By removing catalytic residues either N-terminal or C-terminal to the intein domain, a crippled

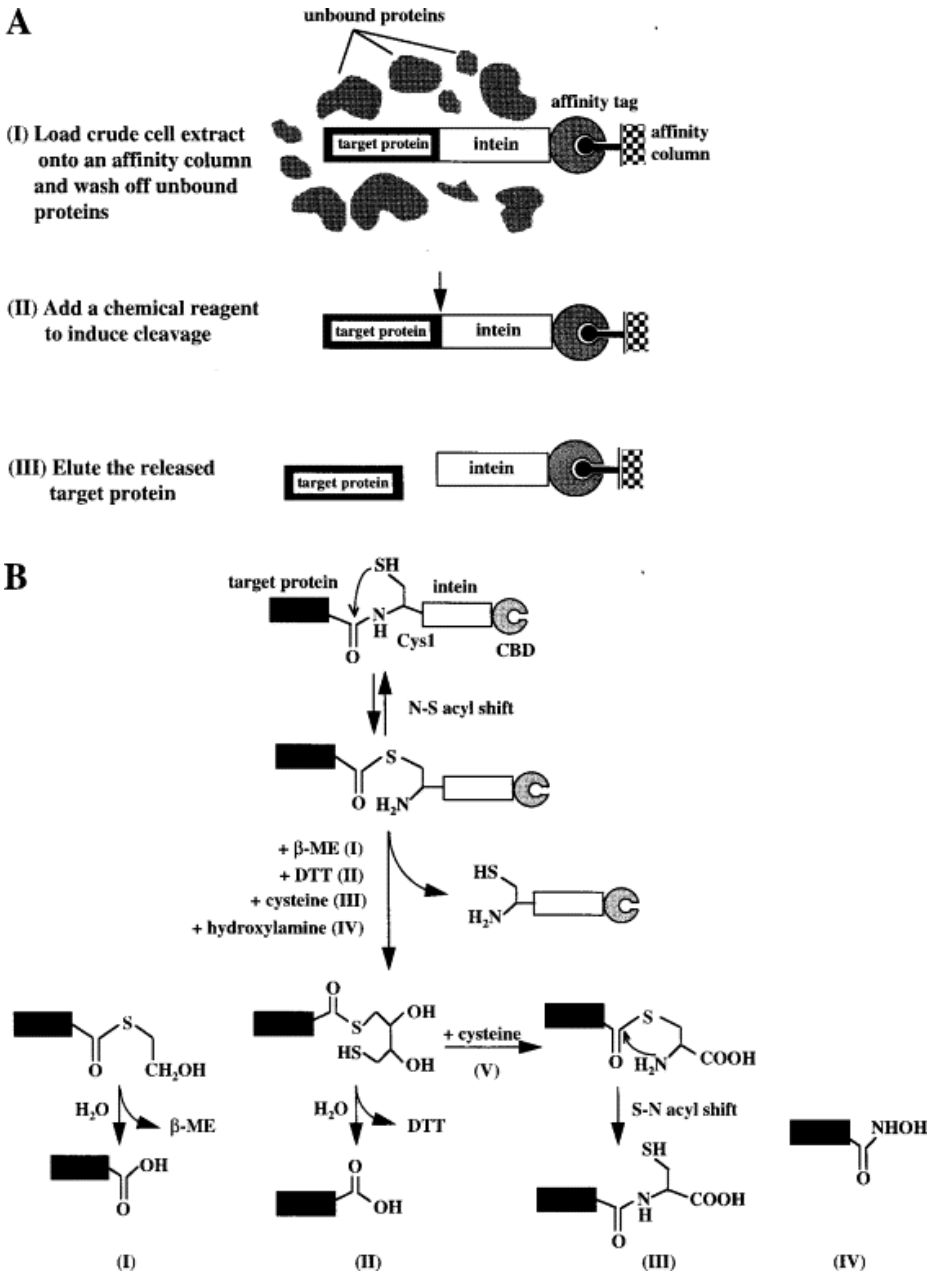


Figure 2.7: Expressed Protein Ligation. A) Schematic of the intein purification system. The fusion protein is isolated by affinity chromatography and cleaved off of the column with addition of a sulfhydryl group. B) The molecular mechanism of intein purification. Four different purification options are presented. EPL is pathway III, with the cysteine being the N-terminal residue of a synthetic peptide. Reprinted from [147], Copyright 1997, with permission from Elsevier.

intein is formed. These can only partially complete their mechanism, leaving a primed thioester that can be displaced by a different sulfur nucleophile, generally mercaptoethanesulfonic acid (MESNA) [145]. As detailed in Figure 2.7, Expressed protein ligation (EPL) takes advantage of this by fusing one end of a crippled intein to a chitin binding domain and the other end to the target protein. Because the target can be fused to either side of the intein, it can be labeled on either the N or C terminus. The chimera is expressed in *E. coli*, bound to a chitin column, and washed to remove impurities. MESNA is then added overnight to cleave the target off of the intein and column. A cysteine-containing peptide can then displace the MESNA, labeling the target with whatever else the peptide contains.

Our group has used EPL extensively to attach affibodies and antibody-linked Protein Z to a variety of nanoparticles [2, 37, 38, 39]. EPL allows us to attach an alkyne moiety to the C-terminus of our targeting ligand. This is then covalently bound to particles displaying azide groups via highly efficient click chemistry. We have also shown that using EPL instead of traditional chemistries has significant advantages in particle targeting, presumably because the C-terminal label gives the targeting ligand orientation relative to the particle [2]. For the same reason, Venter *et al.* used EPL to produce antigens for a nanoparticle-based vaccine. Maltose-binding protein was expressed and linked to a peptide containing an aminooxy group, which was subsequently linked to benzaldehyde-displaying virus-like particle. The aniline-catalyzed ligation produced a particle displaying antigen in a defined orientation with icosahedral symmetry, which greatly increased the B-cell response over antigen alone.

Sortase-Tag

Sortase is a transpeptidase found in gram-negative bacteria. It catalyzes the cleavage of the Thr-Gly peptide bond within its LPXTG recognition motif and the subsequent formation of a new peptide bond between the now-C-terminal threonine of the motif-containing protein and the primary amine of a glycine residue (Figure 2.8A). This is done by a reverse-protonated, mono-iso bi bi ping pong reaction mechanism with a hydrolytic shunt [148], as depicted in Figure 2.8B. At the enzyme's optimal pH of 7.8, only a small fraction of the catalytic residues C189 and H120 are correctly protonated. Therefore, very little of the enzyme is in an active state, about 0.06% [148]. Despite its low activity, sortase has proven to be one of the most useful enzymes for bioconjugations [149, 150].

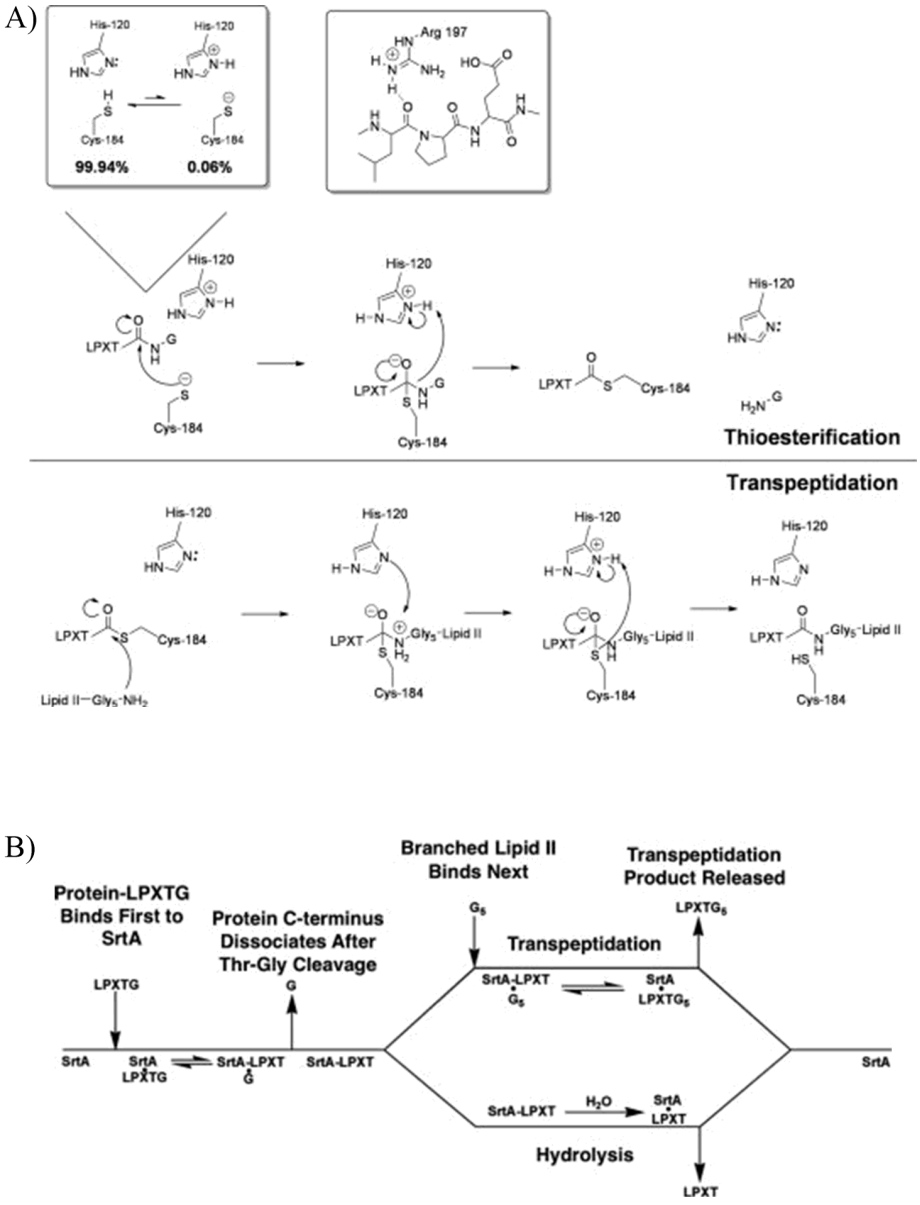


Figure 2.8: Sortase A Mechanism. A) Reaction diagram of the sortase catalytic mechanism. First, Cys184 cleaves the LPXTG motif between the T and G via tholysis. Then, a peptide beginning with glycine cleaves the LPXT from the acyl-enzyme intermediate, regenerating the sortase enzyme. B) Sortase reaction schematic featuring both the transpeptidation and hydrolysis pathways. Adapted from [148] with permission from John Wiley and Sons.

The only requirements for sortase tagging, or "sortagging", seem to be an LPXTG motif and an N-terminal glycine, although using three to five glycine residues has been shown to improve reaction efficiency [151]. Calcium is also needed to activate the enzyme. Either can be preceded or followed by a fully-folded protein or synthetic peptide. As such, sortagging has proven to be a flexible and broad approach to enzymatically tagging the N- or C- termini of proteins.

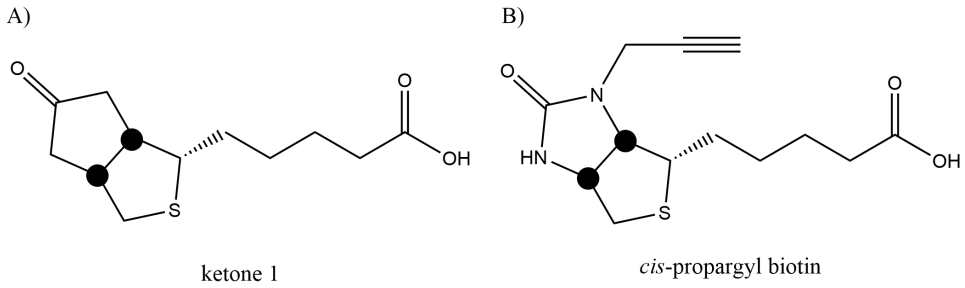
Many groups have successfully used the soluble domain of Sortase A (SrtA) from *S. aureus* in concert with a fused LPETG motif. That system has been used to label full-length antibodies with radiohalides for imaging studies [152], an isotopically-labeled, insoluble SH3 domain with a non-labeled GH1 solubility tag for protein NMR [153], bacterial adhesins to fluorescent microspheres to study bacterial adhesion to host cells [154], as well as many other applications that have been previously reviewed [149, 150].

One challenge in sortagging has been that the ligation regenerates an LPXTG motif and produces a new glycyl-amine, which can lead to the production of the original motif-containing protein and reduce reaction efficiency. To mitigate this reverse reaction, some have run the ligation in dialysis tubing to selectively remove the unwanted product [153]. More recently, the Liu group has replaced the glycyl-amine of the peptide tag with a hydrazine [155]. The modification is still recognized by the enzyme for ligation but the product is no longer able to serve as a cleavage substrate. Thus, they produced an irreversible sortagging reaction that led to higher yields and increased purity.

Because the sortase mechanism includes both a cleavage and ligation, the enzyme can be used to concurrently purify and conjugate a target protein. Chilkoti *et al.* fused an LPETG motif between their target protein and an elastin-like-peptide (ELP) [156]. By raising the temperature, they aggregate the ELP fusions and purify their protein from bulk lystate. They then add an ELP-tagged sortase, calcium, and a synthetic peptide, which cleaves the ELP tag off of the target protein while ligating the synthetic peptide to the target. By again inducing ELP aggregation, the ligated target protein is separated from the unligated target, cleaved ELP tags, and enzyme.

Biotin Ligase

Biotinylation is an attractive method for labeling proteins due to the small size and high specificity of the tag as well as the femtomolar interaction between biotin and avidin [125]. The *E. coli* biotin ligase, BirA, catalyzes the ligation of adenylated biotin to the lysine residue within the 15-amino-



ketone 1 *cis*-propargyl biotin

Figure 2.9: Biotin Analogs. Both ketone 1 (A, [157]) and *cis*–propargyl biotin (B, [162]) can be ligated to the biotin acceptor peptide and directly enable bio-orthogonal chemistries.

acid biotin acceptor peptide (BAP) sequence, GLNDIFEAQ**K**IEWHE. BirA is highly specific for the BAP sequence; overexpressing the enzyme in mammalian cells does not result in biotinylation of any endogenous proteins [157]. Only one protein is naturally modified by BirA in *E. coli* cells, the biotin carboxyl carrier protein (BCCP).

The interaction between biotin and avidin-family proteins is well-characterized and the most common method of functionalizing a BirA-modified protein. Fluorescently-labeled streptavidin has been used to image EGFR and AMPA trafficking [157, 158], nanoparticle biodistributions [159, 160], and even track single neurotrophic factors moving along an axon in real time [161].

While BirA is exquisitely selective for the BAP sequence, it tolerates some promiscuity in the biotin moiety itself. For example, Chen *et al.* determined that wild-type BirA will ligate ketone 1 (Figure 2.9A), allowing direct modification of labeled protein with a hydrazine [157]. Slavoff *et al.* tested biotin ligases from eight species for activity with a number of biotin analogs [162]. They found that the enzyme from *Pyrococcus horikoshii* accepts desthiobiotin azide and *cis*-propargyl biotin (which has an alkyne group, Figure 2.9B), allowing them to directly modify proteins for use in click-chemistry reactions.

Transglutaminase

Microbial transglutaminase (mTG), most often from *Streptomyces mobaraensis*, catalyzes the transamination reaction of the γ-carbonyl amide of a glutamine with the ε-amine of a lysine or another primary amine supplied in excess. mTG exhibits broad specificity toward its substrate glutamines, which is convenient if genetic manipulation of the target protein is impractical. Although there are no specific sequence requirements for mTG, it generally only recognizes one or two of a protein's

surface-exposed glutamine residues. This common belief of limited active glutamines may be due to publishing bias, however [163]. mTG's specificity is instead due to local secondary structure (the Gln must be on a loop) and the properties of nearby residues. Thus, the active glutamines must be empirically determined for each protein of interest.

A number of approaches have been developed to limit mTG activity to a single site on a target protein. For instance, producing human IgG1 without glycosylation allows only a single γ-carbonyl amide to be recognized in the constant region of the heavy chain. Thus, Dennler *et al.* were able to produce homogenous antibody-drug conjugates with enzymatically deglycosylated commercial Trastuzumab (Herceptin) [164]. Mero *et al.* took advantage of the structural specificity of mTG by running the enzymatic labeling reaction in PBS with an organic co-solvent, which generally alters protein structure by disrupting hydrophobic collapse. Adding 30% (v/v) DMSO and 50% (v/v) ethanol led to site-specific PEGylation of salmon calcitonin and human growth hormone, respectively, which each have two natural mTG sites [165].

If a target protein has no natural mTG sites, a Q-Tag (PKPQQFM) or K-Tag (MKHKGS) can be engineered into the protein sequence [166, 167]. For example, Kitaoka *et al.* attached a K-tagged alkaline phosphatase to DNA containing glutamine labeled dUTP to produce a sensitive DNA hybridization probe that could detect as little as 10 fg of target DNA [168].

Aldehyde Tag

Formylglycine-generating enzyme (FGE) can be used to create a unique aldehyde functional group on a protein. Posttranslationally, FGE oxidizes the cysteine residue within a CxPxR motif to create a formylglycine (fGly) residue, which can be further functionalized by hydrazone or oxime formation [169]. Because aldehyde groups are not present in any of the canonical amino acids, the conversion of a cysteine to formylglycine creates a site-specific, genetically-encoded conjugation handle. Additionally, FGE is expressed naturally in the cytoplasm of many prokaryotes and eukaryotes [170] and can be artificially expressed in secretion pathways [171], making the aldehyde tag accessible to nearly all expression systems.

Another major advantage of the aldehyde tag is that it is not limited to the N- or C-terminus of the target protein, but can be present anywhere that is surface accessible. This has proven particularly useful when engineering monoclonal antibodies to display a specific glycosylation pattern because

the sugar moiety can be placed in its usual location along the Fc molecule [172]. In a different study, Drake *et al.* created a small library of monoclonal antibodies displaying an aldehyde tag in different locations [173]. After conjugating a drug to the aldehyde, they found that the location not only affected conjugation efficiency, but also the pharmacodynamics and pharmacokinetics of the antibody-drug conjugate.

One drawback of the hydrazone and oxime chemistry generally used with the aldehyde tag is that it is reversible, complicating robust and quantitative protein labeling. To address this issue, Agarwal *et al.* created tryptamine derivatives that first react with an aldehyde to create a traditional oxime and then continued with a Pictet-Spengler reaction that creates a stable carbon-carbon bond [174]. After being incubated for a week at room temperature, proteins ligated by the Pictet-Spengler method remained largely conjugated, while those conjugated using oxime chemistry showed nearly complete hydrolysis.

Sugar Remodeling

IgG carries two well-characterized N-linked glycans. These sugar trees can be trimmed and regrown post-purification. Because the glycosyltransferases necessary to rebuild the glycan are known to tolerate chemical modification of their sugar-nucleotide substrates, it is possible to include small bio-orthogonal functional groups in the remade glycosylation pattern [125]. For example, Li *et al.* used a galactosyltransferase (GalT) to remove the outer galactose and sialic acid from a monoclonal antibody and then religate the galactose residue [175]. They then added recombinant sialyltransferase (ST6GalI) and azido-modified sialic acid to finish the glycosylation pattern and incorporate four azide groups per antibody. They then used strain-assisted azide-alkyne click chemistry to add a variety of chemical functions to the antibody in a site-specific and quantitative manner. While this technique is powerful and has been used by a number of groups [176, 177, 178], it is limited to proteins that have well-defined surface glycans.

Fusion Protein Tags

The SNAP tag consists of the active domain of the human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT) [179]. In the cell, hAGT transfers a methyl group from an O^6-alkylguanine, irreversibly generating a guanine residue and converting the hAGT active site cysteine

into a methionine. Keppler *et al.* correctly hypothesized that they could use this mechanism to permanently ligate other alkyl chains to the hAGT active site. They found that the protein was incredibly tolerant of O^6-benzylguanine (BG) derivatives, allowing them to conjugate biotin and a number of fluorophores to hAGT. And by genetically fusing hAGT to the N- or C-terminus of a target protein, they were able to directly and specifically ligate their BG derivatives to the target protein in any cellular compartment. The technique's ease of use and broad applicability made it popular in the scientific literature. Largely used for fluorescence imaging, including advanced techniques like super-resolution imaging [180] and total internal reflection fluorescence microscopy [181].

The popularity of the SNAP tag led Gautier *et al.* to evolve the CLIP tag [182]. Eight mutations were able to alter the specificity of hAGT from O^6-benzylguanine to O^6-benzylcytosine (BC), while maintaining promiscuity in the O^6-alkyl chain. In fact, the change in substrate specificity was so complete that SNAP and CLIP tags can be used in tandem for multichannel labeling of cellular proteins. An example SNAP/CLIP system is provided by Maurel *et al.*, who used SNAP- and CLIP-tagged proteins to show membrane protein interactions by independently labeling them with fluorophores capable of performing time-resolved fluorescence resonance energy transfer [183].

The major drawback of all fusion tags is their ability to alter the functions of the target protein. This was specifically put to the test by Landgraf *et al.* by fusing their protein of interest, a Clp protease to a variety of fusion tags [184]. Prior to the Landgraf study, the Clp family of proteases had widely been reported to form foci in the *E. coli* cytoplasm. They hypothesized that after cell division, an observable (fluorescent) protein should be degraded by Clp faster in the daughter cell that retained the Clp foci than the daughter cell that needed to synthesize a new one. However, they saw no difference in degradation rates between daughter cells with untagged ClpX. This led them to fuse more than 20 visualization tags to ClpX, including the SNAP tag. The result, shown in Figure 2.10, shows the incredible differences in ClpX localization due entirely to their fusion partner. So while fusion tags can be very powerful tools, the consequences of adding a large fusion domain must also be kept in mind.

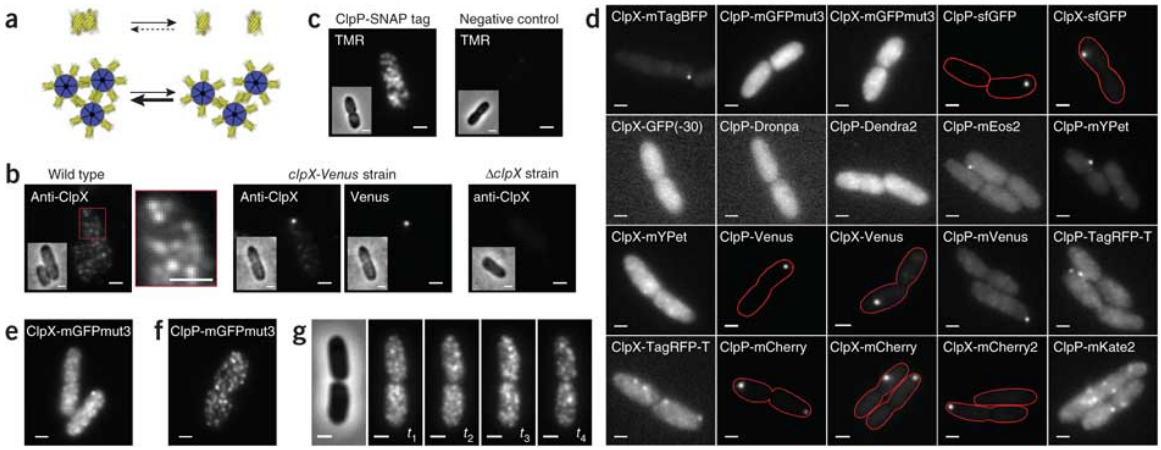


Figure 2.10: Effects of Fusion Domains. Landgraf *et al.* fused the protein ClpX to a number of different fusion tags, mostly fluorescent proteins, which caused the proteins to congregate in foci within the *E. coli* cell. From the paper: "(a) Schematic of antiparallel fluorescent protein dimer (yellow) dissociating into monomers (top) and of avidity effects potentially clustering tagged ClpX hexamers (blue; bottom). For simple monomers, re-association is slow because the molecules diffuse apart; for tagged oligomers, the complex acts as a scaffold to facilitate re-association. (b) Immunofluorescence microscopy with an antibody to ClpX (anti-ClpX) in the wild type (left; magnification of the boxed region is shown on the right), clpX-Venus strain (middle; Venus fluorescence is shown on the right) and ΔclpX strain (right). Insets, phase-contrast images. (c) Fluorescence images of bacteria expressing ClpPSNAP tag labeled with tetramethylrhodamine (TMR), compared to wild type (negative control). Insets, phase-contrast images. (d) Fluorescence images of bacteria expressing the indicated constructs. Cell outlines (red) are shown for cells with weak cytoplasmic signal. (e,f) HILO microscopy of gently fixed cells expressing ClpX-mGFPmut3 (e) and ClpP-mGFPmut3 (f). (g) Phase-contrast (left) and live-cell HILO microscopy fluorescence images (right) of cells expressing ClpP-mGFPmut3 taken one second apart at time points t_1-t_4. Scale bars, 1 μm."
Reprinted by permission from Macmillan Publishers Ltd: Nature Methods [184], copyright 2012.

2.4 Engineering Ligands and Enzymes

Nature has provided a host of interesting proteins that can be used for both targeting and ligations. However, native proteins are often unsuitable for man-made systems, as was previously discussed in Section 2.2. When this happens, it is the province of protein engineers to deliver a custom protein tailored to the system at hand. The techniques used to develop novel proteins fall into two broad categories: rational design and directed evolution. As the name implies, proteins created by the rational design approach are the result of deliberate mutations, fusions, or deletions that are hypothesized to add or remove specific functionality. In contrast, directed evolution involves introducing a large number of mutations, in excess of 10^7, to a coding sequence. The resulting proteins are expressed in parallel with some link to their coding sequence and then subjected to rounds of rigorous screening. Only the proteins that meet the conditions of the screen are recovered. Those sequences then receive additional mutations and are put through another screen; a cycle that is repeated until a protein that completes its objective with the desired quality is recovered. Both schools of thought have significant benefits and downfalls, which will be discussed.

2.4.1 Rational Design

The rational design of proteins leverages basic science discoveries in structural biology, systems biology, and enzymology to create proteins with novel functionalities. Rational design approaches have the advantage of being hypothesis-driven and testable, making it easy to determine efficacy. Because so much is known about the systems chosen for rational design, they have a higher likelihood of success. A heavy dependence on biophysical data is a major downfall of the approach, due to relative scarcity of well-characterized protein systems and the extensive time and cost required to obtain that data. Even with those limitations, rational design has been producing useful proteins for as long as we have been able to read and modify the genetic code.

Chimeric Proteins

Like the mythical beast of ancient Anatolia that had the heads of a lion, goat, and snake, protein chimeras consist of globular domains from different proteins that have been fused together into a single polypetide chain. Utilizing molecular cloning techniques or direct DNA synthesis, the engineer

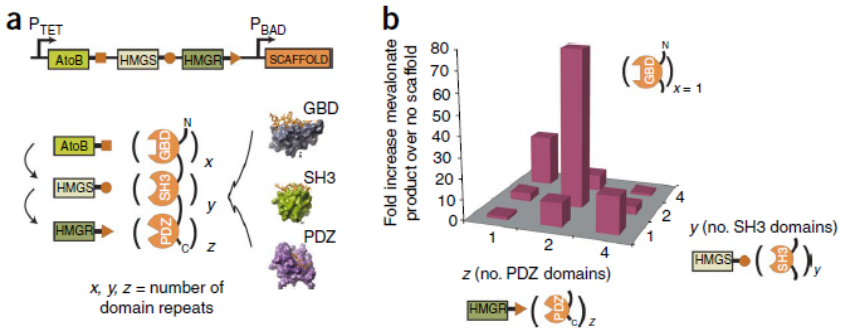


Figure 2.11: Chimeric Scaffolds Enhance Biosynthetic Pathways. A) Dueber *et al.* fuzed enzymes in the mevalonate biosynthetic pathway to protein binding domains and created a scaffold protein to co-localize the enzymes in specific ratios. B) Mevalonate synthetic flux is altered by changed the scaffold:enzyme ratios while holding enzyme expression constant. Reprinted by permission from Macmillan Publishers Ltd: Nature Biotechnology [186], copyright 2009.

is able to create a coding sequence containing the genetic instructions for each desired domain in series and in frame. By doing this, we can combine the functions of disparate proteins.

Often, these combinations are used to alter the localization, regulation, or binding partners of a useful enzyme. For example chimeric antigen receptors (CARs) are the fusion of one of the targeting ligands mentioned in Section 2.2 with the transmembrane and signaling domain of CD3ζ [185]. When transfected into killer T-cells, CARs redirect the cells' cytotoxic effect toward a cell population determined by the targeting ligand. This effect was later enhanced by fusing additional signaling domains (CD28 and CD134) C-terminal to CD3ζ. This allowed the CAR to simultaneously activate co-stimulatory pathways that maintain the activated, cytolytic phenotype for longer periods of time, making the cells much more effective.

Another common use of chimeras is to fuse a binding domain to a protein of interest to direct its localization inside or outside of the cell. The simplest example of this is the hexahistidine tag; six histidine residues fused to the N- or C-terminus of a protein imparts high affinity for certain transition metals. This is used to localize proteins in a cell lysate slurry to added metal-coated beads, purifying the target protein from the bulk cell solution. This same concept can enable much more complex interactions. For instance, Dueber *et al.* fused the enzymes AtoB, HMGS, and HMGR to orthogonal, transgenic protein-binding domains [186]. They then expressed chimeras of the partner domains in varying ratios to create scaffold proteins (Figure 2.11). The scaffold significantly increased the effective concentration of each enzyme in the biosynthetic pathway, allowing them to produce 77-fold

more product without the additional metabolic burden of overexpressing the enzymes individually.

Due to the innate modular nature of proteins, each globular domain is often fully functional in the new context of a chimera because globular domains tend to fold independently of one another. To ensure the independent folding and functionality of each of a chimera's domains, linker regions can be included between domains. While linkers are not generally a primary design focus, their selection can play an important role in protein stability and function [187]. Variations of the glycine-serine linker, $(GGS)_n$, are the most frequently used linkers due to their flexibility, solubility, and resistance to proteolysis. On the other end of the spectrum, α-helical linkers, $(EAAAK)_n$, are rigid and can be used to deter interactions between domains.

Crystal Structures and Computational Analysis

Crystal structures are one of the most important tools in rational design. Using a suite of *in silico* tools, such as PyMol [188] and Rosetta [189], the effects of mutations, deletions, and fusions can be modeled and analyzed before ever reaching the wet lab. Homology modeling allows us to approximate the tertiary structure of chimeric proteins and predict interactions between domains [190]. Modeled binding interfaces enabled Han *et al.* to create a decoy TLR4 receptor specific to the MD2 ligand [191]. By using a crystal structure of the bound proteins and predicting the change in free energy resulting from mutations to TLR4, they increased the affinity twenty-fold.

As our understanding of protein structure and function matures, *de novo* protein design software is becoming more common and useful. For example, Richter *et al.* used the Rosetta3 enzyme modeling suite [192] to design a custom esterase [193]. They used the software to first scan and score 214 potential scaffold enzymes. Next, Rosetta produced a series of mutations to try and optimize enzyme binding to the transition state of the thiolysis reaction, creating a library of theozymes (theoretical enzymes). Quantum mechanical calculations were then run on the theozyme mechanisms to predict reaction rates and cull the library down to 55 candidates, which were cloned and screened in the wet lab for valid enzymes.

Site-Directed Mutagenesis

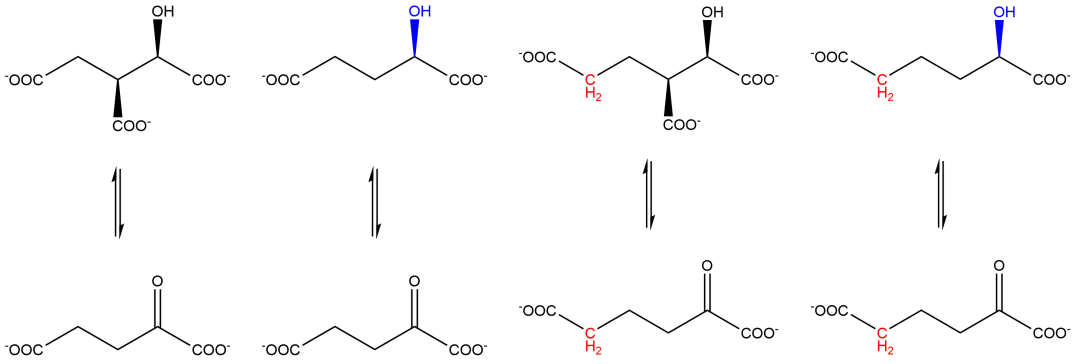
Site-directed mutagenesis (SDM) is a simple method that can make profound changes to a protein's stability or function. In SDM, a protein engineer makes specific, targeted changes to the DNA

encoding a protein, resulting in an altered primary structure. The mutations are generally driven by knowledge of the protein's crystal structure or the enzyme's mechanism.

One of the first applications of SDM in protein engineering was to improve protein stability. A primary source of instability is oxidative stress, especially if the protein has been removed from its natural environment, which often occurs with industrial enzymes. Subtilisin, a promiscuous protease found in the *Bacillus* family, is one such enzyme. It was known that exposure to oxidative stress led to the conversion of Met222 to methionine sulfoxide and complete inactivation of the enzyme within minutes [194]. Using SDM, Estell *et al.* were able to mutate M222 to all 19 other amino acids and measure their activity and peroxide tolerance [195]. They found that while the M222A mutation reduced activity by 47%, it also made the enzyme impervious to even 1M H_2O_2. The mutated enzyme is also resistant to inactivation by bleach and is now a common ingredient in laundry detergent [196].

Perhaps the most common modern application of SDM is a loss-of-function mutation. Most enzyme families have one or two residues that are necessary for their function. These can be easily identified as they are highly conserved when comparing primary structures. For example, the C-terminal asparagine is strictly conserved in all intein domains [197]. Mutation of the critical asparagine residue to alanine results in a branched protein because the intein is unable to complete its splicing function, resulting in a crippled intein. These partially inactivated inteins form the basis of the expressed protein ligation technique described in Section 2.3.2 by leaving the N-extein attached to the protein by a thioester that can be cleaved by an external sulfide group [198].

Less common is a gain-of-function mutation. It is much harder to intuit how altering one or two amino acids will add a function than how removal of a reserved residue will remove functionality from a protein. It can be done, however by studying naturally occurring mutations that result in a change of substrate, product, or binding partner. For example, Reitman *et al.* mined cancer mutation data to identify candidates for an enzyme to synthesize (R)-2-hydroxyadipate [199]. They found that a mutation in the enzyme isocitrate dehydrogenase (IDH) catalyzes a reaction whose product is only one carbon shorter than (R)-2-hydroxyadipate (Figure 2.12B). The wildtype IDH enzyme is closely related to homoisocitrate dehydrogenase (HIDH), which serves the same catalytic function as IDH, but with a metabolite one carbon longer (Figure 2.12C). They correctly hypothesized that making the analogous mutation to HIDH would lead to the formation of (R)-2-hydroxyadipate, giving the



(A) Wild-type IDH (B) Mutant IDH (C) Wild-type HIDH (D) Mutant HIDH

Figure 2.12: Gain of Function Mutation. Reitman *et al.* determined the mutation required to alter the IDH substrate from (2R,3S)-isocitrate (A) to 2-oxoglutarate (B). They then mapped that mutation onto the homologous enzyme HIDH, enabling the desired conversion of 2-oxoadipate to (R)-2-hydroxyadipate (D). Differences in IDH and HIDH substrates are drawn in **red**. Difference between the wild-type and mutant substrates are drawn in **blue**. Adapted with permission from Macmillan Publishers Ltd: Nature Chemical Biology [199], copyright 2012.

enzyme a new function with only a single mutation.

Recently, SDM has also been gaining traction as a way to introduce unnatural amino acids. Life has done amazing things with the current twenty canonical amino acids since the standard genetic code was set at least 2.5 billion years ago [200]. However, the natural amino acids exclude large portions of the chemistry that can be performed in a lab environment. To bridge this gap, many research groups have devised ways of incorporating synthetic amino acids with novel functionality into a cell's genetic code. The most common way to do this is by suppressing the amber stop codon (UAG), which accounts for only 7% of *E. coli* stop codons and is rarely the terminator of critical genes [201]. By co-transforming cells with orthogonal tRNA and tRNA synthetases, it is possible to introduce a 21st amino acid into the genetic code. SDM is then used to mutate a natural codon to TAG, which will be translated by the ribosome into the unnatural amino acid. An example comes from Hui and Tsourkas, who used amber suppression to replace residues in Protein Z with p-benzoyl-L-phenylalanine (BPA), a phenylalanie analog containing a benzophenone photocrosslinker [202]. They used SDM to incorporate BPA into 11 unique sites and then measured their ability to form covalent bonds with different IgGs. These created stable, site-specific IgG-protein Z conjugates that could never exist using the repertoire of natural amino acids. Researchers have also incorporated BODIPYFL-lysine, photocaged lysine, and F-para-phenylalanine into proteins to create a FRET

sensor, activate a signaling pathway with a UV pulse, and study the effects of π interactions on ligand binding, respectively [203, 204, 205, 206].

2.4.2 Directed Evolution

Directed evolution approaches are useful when there are no known or characterized proteins that serve a needed function. They require little, if any, biochemical knowledge of the desired outcome. As a result, many surprising and unpredictable sequences have arisen from evolutionary techniques. For example, when screening for affibodies with high affinity for Amyloid β, the characteristic biomarker of Alzheimer's disease, Grönwall *et al.* found that their best binder was actually an affibody dimer linked by a disulfide bond [207]. No one has ever purposely built a bridged affibody homodimer to bind a single target molecule, so directed evolution provided a solution that had never been considered.

There are three fundamental aspects of every directed evolution experiment. First, a library encoding a large number of protein coding sequences quickly creates the genetic diversity that would occur over eons in a natural setting. Second, a screen that recovers the desired phenotypes provides an artificial selective pressure that culls the non-functional phenotypes. Finally, a way to link each sequence in the genetic library to the specific version of the protein it encodes. The most important of these aspects is the phenotype-genotype link, as that will determine the maximum library size and available screening methods. The major classes of links are described in detail below.

The major downside of this approach is a large front-end time and capital investment with no guarantee of success. The largest library may still not contain a functional sequence. The best sequence may simply not express well in its host organism. The screen may be passable by an unexpected, undesired phenotype that out-competes the desired phenotype. Still, directed evolution has been immensely popular and useful.

Cell Surface Display

Cell surface display co-opts the secretory pathways of bacteria and yeast to produce library-encoded proteins that are anchored to the extracellular surface of the cell membrane. When selecting for targeting ligands, the cells are incubated with fluorescently labeled target protein and sorted via fluorescence activated cell sorting (FACS), which reads the fluorescence signal of each cell individually

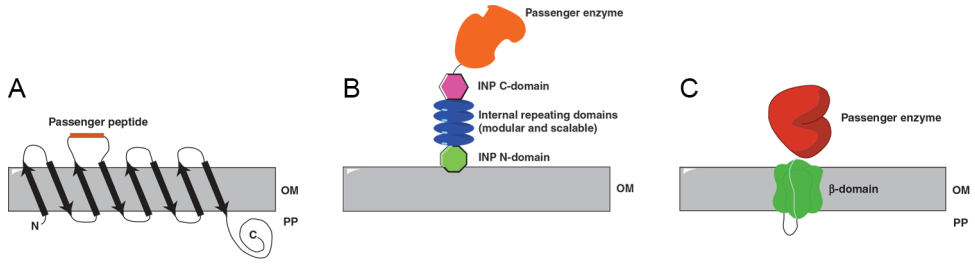


Figure 2.13: Bacterial Surface Display Strategies. A) Proteins and peptides can be cloned into the loops of OmpA, a well-studied *E. coli* surface protein. B) Fusions to ice nucleation proteins allow flexible spacing from the cell surface by including internal repeating domains. C) Autotransporters express the protein in the periplasm and then actively pull the protein through a channel for it to refold outside the cell. Adapted from [208], with permission from Elsevier.

and sorts high-signal and low-signal cells into separate collection tubes. The highly fluorescent cells are subsequently expanded and screened again under more stringent conditions until a suitable binding partner is found. Although the selection methods for enzyme maturation are less generalizable, surface display is necessary for precise control of substrate concentrations.

Proteins can be trafficked to the surface of *E. coli* by a number of methods, shown in Figure 2.13 [208]. In one example, Binder *et al.* found the autotransporter EspP was the most robust method of displaying their Anticalin library [209]. They then produced a library of 2×10^9 Anticalin variants with error-prone PCR and sorted for cells that retained fluorescence after one hour of competition with unlabeled target. The library was able to increase the dissociation half-life of the natural binder by ten-fold.

Yeast display is generally done with *S. cerevisiae* by fusing the library protein to Aga2p [210]. During secretion, Aga2p is linked by disulfide bonds to the integral membrane protein Aga1p, which carries it to the surface. Thus, the displayed protein is transiently in the endoplasmic reticulum and is subject to ER post-translational modifications like glycosylation and proteolysis. This can be a disadvantage if displaying an enzyme or cytosolic protein, as glycosylation can disrupt activity and participate in binding. On the other hand, if the end product will be produced solubly in a eukaryotic system, yeast display is preferred because it will mimic the end product [211]. Yeast have been glyco-engineered to reproduce human N-linked glycosylation patterns, making them the system of choice for selecting and displaying humanized antibodies [212].

Unfortunately, cell surface display systems usually have the lowest library sizes (generally 1×10^7

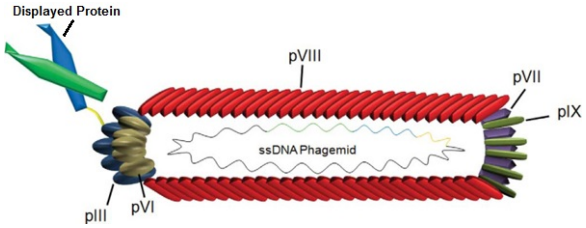


Figure 2.14: Bacteriophage M13. Adaptations of figures from [214] are allowed under the Creative Commons License v4.0.

to 1×10^9) [213]. The libraries are limited by the inherent inefficiency of stable genetic transformation. However, this is balanced by the relevance of the display system to the downstream application [212]. It is much more likely for a targeting ligand evolved in yeast to express robustly in yeast.

Phage Display

Perhaps the most common display technique is phage display, specifically filamentous phage display. The filamentous bacteriophage M13 (see Figure 2.14) has a simple, circular single-stranded DNA (ssDNA) genome that encodes only eleven proteins. The life cycle of the phage is well characterized. It first binds to the F pilus on *E. coli*. Then it injects its genome and replication enzymes into the bacterium, which synthesizes the second strand of the genomic plasmid. The eleven genes are then expressed and assembled in the bacterial periplasm. Newly syntheisized ssDNA genomes are transported into the assembling phage and extruded from the cell in the pVIII coat protein. Finally, five copies of pIII are used to cap the phage and it is released into the media. The *E. coli* cell survives the infection, albeit with a slower growth rate. For some strains of M13, such as KO7, the genome's packaging signal contains mutations that lower its affinity for the coat protein. The phage is thus capable of preferentially incorporating the ssDNA of a separate plasmid, known as a phagemid, rather than its genome. When the resulting phage infects a bacterium, the *E. coli* receives a copy of the phagemid and will not produce any phage. To create more phage with the phagemid, the transformed bacteria must be superinfected with M13KO7, which provides the necessary structural and assembly proteins. If the phagemid encodes one of the phage's coat proteins, the encoded version will be randomly incorporated into the daughter phage along with the phagemid that encodes it. This is the phenotype-genotype link in phage display [215]. Due to the extremely small size of bacteriophage, libraries of up to 1×10^{11} unique sequences can be screened [213].

Commonly, the protein to be evolved is fused to the N-terminus of the cap protein pIII. pIII chimeras have been used to evolve many different proteins, but affinity targeting proteins are the most common. Antibodies [216, 217], scFvs [218], affibodies [108], monobodies [219], and short peptides [220] have all been evolved by phage display, among many others. Phage have even been used to select for high-affinity D-peptides by screening against synthesized D-antigens and relying on the mirror-image hypothesis, which makes an L-peptide binding a D-target equivalent to a D-peptide binding an L-target [221]. The screen involved in all of these is a binding assay. Briefly, phage libraries are expressed and purified from *E. coli* and bound to antigen-coated plates. After a sufficient amount of time has passed for binding, unbound phage are washed away. Bound phage are then eluted with acid, protease, or excess antigen. While acid or protease elution can ensure recovery of all bound phage, eluting with excess antigen can be used to find phage with a particular binding affinity [222]. The eluted phage infect fresh *E. coli* cells, which can be grown up for sequencing, phage production, further cloning, or storage. Phage may also be subjected to a negative screen, where they are panned against similar antigens, such as different isoforms of the antigen, to remove non-specific phage.

One of the major advantages of phage display is the robustness of the phage themselves. Bacteriophage M13 is heat and acid stable and can only infect *E. coli* bearing the F plasmid, making them essentially orthogonal to mammalian systems. Of the major display techniques, phage display is the only one capable of being screened directly on cultured cells or even *in vivo*. When screening against cultured cells, it is imperative to have a good negative screen consisting of cells that differ from the positive line by only the target antigen. *In vivo*, the negative screen is the rest of the body and the target antigen is an unknown molecule on the target tissue [223]. In one example, Arap *et al.* injected a phage library into immunodeficient mice bearing a MDA-MB-435 xenograft breast tumors [224, 225]. They then recovered the phage for amplification and sequencing. The peptides were tested for affinity and selectivity to determine the best sequence. Finally, they conjugated the optimal peptide to the chemotherapeutic doxorubicin and administered the targeted drug to tumor-bearing mice, which showed improved survival over the untargeted drug.

To speed up phage display, Esvelt *et al.* developed a technique called phage assisted continuous evolution (PACE) [226]. As shown in Figure 2.15, PACE encodes its library onto an M13 genome that is missing pIII. Phage infect cells that have been transformed with a plasmid with an error-

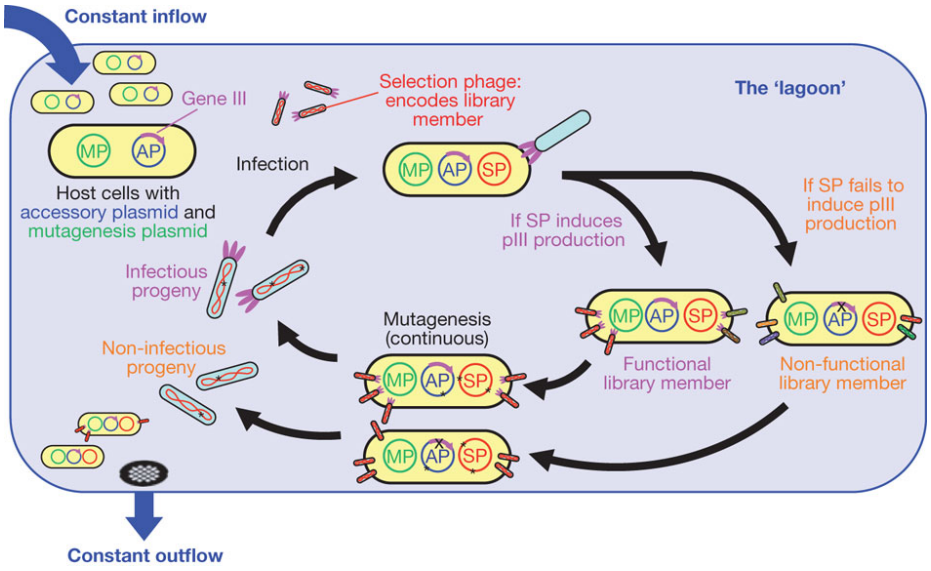


Figure 2.15: Phage Assisted Continuous Evolution. Host cells continuously flow through a lagoon, where they are infected with selection phage (SP) encoding library members. Functional library members induce production of pIII from the accessory plasmid (AP) and release progeny capable of infecting new host cells, whereas non-functional library members do not. Increased mutagenesis is triggered through induction of the mutagenesis plasmid (MP). Host cells flow out of the lagoon on average faster than they can replicate, confining the accumulation of mutations to replicating phage. Reprinted by permission from Macmillan Publishers Ltd: Nature [226], copyright 2011.

prone polymerase for mutagenesis and a plasmid that codes for an intercellular screen, using the phage protein pIII as a reporter gene. Because pIII is necessary for phage infection, only phagemids that code for a passing protein can produce infectious phage. Constructs that cannot pass the screen are therefore not infective because they lack the necessary cap protein. And because the mutagenic plasmid is constantly introducing mutations in the phage genome, the library will continue evolving as the screen is running. Continuous systems such as PACE represent a major advance in directed evolution, as a few days of PACE can be the equivalent of hundreds of rounds of evolution and screening. However, PACE is limited to being intercellular screens, which are only applicable to the evolution of targeting proteins if both the ligand and biomarker can be expressed in *E. coli*.

Ribosome Display

Ribosome display is a fully *in vitro* technique. Because it does not require an extremely inefficient transformation step, it can produce the largest libraries, containing up to 1×10^{13} unique sequences [213]. The key to ribosome display is that the screened sequences must not contain a stop codon. During protein synthesis, the stop codon recruits release factor proteins that dissociate the ribosomal subunits from the messenger RNA and nascent peptide chain. By expressing these libraries via *in vitro* protein synthesis, the ribosome will stall when it reaches the end of the mRNA because it is unable to disassemble without a release factor. Thus, the ribosome itself will act as a link between the coding mRNA and translated protein. The genotype is recovered by simply amplifying the mRNA with a reverse transcriptase and running a standard PCR.

While ribosome display can screen libraries 100-fold larger than the most diverse phage display libraries, it is not nearly as robust [227]. Nature treats external RNA as a threat, so enzymes that degrade RNA, ribonucleases, are everywhere in our environment [228]. And due to the high specific activity of ribonucleases, a small contamination can destroy an entire sample. In addition, the genotype-phenotype linkage of a ribosome is not covalent, meaning that inhospitable buffer conditions such as high/low pH, high temperature, detergents, or organic solvent can disrupt the ribosome complex and destroy the sample. Thus, ribosome display can only be performed *in vitro* in nuclease-free wells with purified antigen.

Library Generation

None of the above screening techniques can work without a good library of genes for them to screen. A common way of producing libraries is through random mutagenesis [229]. By amplifying a gene of interest with *Taq* polymerase in error prone conditions, which include the presence of $MnSO_4$ and uneven dNTP ratios, it is possible to randomly introduce mutations during the PCR reaction. The introduction of random mutations mimics natural evolution and generally produces a stable library. However, stability is traded for lack of diversity, as error-prone PCR only introduces a few mutations per kilobase. In order to introduce a large number of mutations, like the 13 randomized amino acids of an affibody, the DNA must be synthesized. Degenerate oligonucleotides allow for any mixture of nucleic acids at a specific location, making it easy to quickly produce large, targeted libraries. These libraries of oligonucleotides can be incorporated into the gene in a number of ways, including PCR, overlap extension, site-directed mutagenesis, and Kunkel mutagenesis [219, 230]. And with the rapid advance of DNA synthesis technologies, entire genes can simply be purchased with degenerate positions, although this is currently too expensive for routine use.

When designing degenerate oligonucleotides, an important consideration is codon choice [231]. There are 64 codons, but only 20 amino acids, so the NNN codon (N = A, T, C, or G) skews the library toward the more commonly represented amino acids, like leucine. By instead using NNK (K = G or T), the codon space is reduced by half, while maintaining the amino acid space. By further paring down the possible codons, it is possible to design libraries with positions skewed toward codons more likely to produce a stable result, such as heavily biasing an amino acid toward being charged or hydrophobic [231, 232, 233, 234, 235].

Chapter 3

Sortase-Tag Expressed Protein Ligation

Abstract

Efficient labeling of protein-based targeting ligands with various cargos (drugs, imaging agents, nanoparticles, etc.) is essential to the fields of molecular imaging and targeted therapeutics. Many common bioconjugation techniques, however, are inefficient, nonstoichiometric, not site-specific, and/or incompatible with certain classes of protein scaffolds. Additionally, these techniques can result in a mixture of conjugated and unconjugated products, which are often difficult to separate. In this study, a bacterial sortase enzyme was utilized to condense targeting ligand purification and site-specific conjugation at the C-terminus into a single step. A model was produced to determine optimal reaction conditions for high conjugate purity and efficient utilization of cargo. As proof-of-principle, the sortase-tag expressed protein ligation (STEPL) technique was used to generate tumor-specific affinity ligands with fluorescent labels and/or azide modifications at high purity (>95%) such that it was not necessary to remove unconjugated impurities. Click chemistry was then used for the highly efficient and site-specific attachment of the azide-modified targeting ligands onto nanoparticles.

3.1 Introduction

As described in Section 2.3.2, EPL and sortagging have a number of advantages over chemical (and many enzymatic) bioconjugation strategies. Both are site-specific, stoichiometric, and can attach a wide array of payloads to the target protein. Despite their advantages, however, the techniques still fall short of an ideal system for conjugating imaging cargo to a targeting ligand. Ligation of a small cargo, such as a fluorophore, often has minimal effect on the protein's physical properties (mass, charge, hydrophobicity, etc.), making it difficult to separate conjugated products from unligated reactants in either system. EPL relies heavily on thiol chemistry, so it is incompatible with classes of targeting ligands that possess disulfide bonds, such as knottins (Section 2.2.4) and antibodies (Section 2.2.1). In Sortase-mediated transpeptidation, the conjugated product must be purified from the enzyme as well as the unconjugated proteins.

In this chapter, we set out to create a system that retains all of the advantages of EPL while making the technique more efficient, expanding the pool of potential targeting ligands, and simplifying downstream purification.

3.2 Design Strategy

To mitigate the shortfalls of sortase labeling and expressed protein ligation, we combine the strategies into a single conjugation and purification technique called Sortase-Tag Expressed Protein Ligation (known hereafter as STEPL). In a single chimeric protein, we express a targeting ligand in frame with the sortase recognition motif LPXTG, a $(GGS)_5$ linker, SrtAΔ59, and a His_6 tag (Figure 3.1). After binding the chimera to a metal resin, addition of calcium and any protein/peptide with an N-terminal glycine (and attached cargo, if desirable) activates the sortase domain, ligating the protein of interest to the peptide while simultaneously cleaving it from the rest of the sortase chimera. Thus the conjugate is released while the sortase enzyme is retained on the column via the His-tag. By making purification and conjugation codependent, STEPL remains site-specific and stoichiometric in nature but does not require any additional steps to remove SrtA from the purified protein sample. Further, large excesses of peptide are not essential since only correctly ligated product is released from the affinity column and conditions can be optimized to nearly exhaust any added peptide. The rationale behind this design is described below:

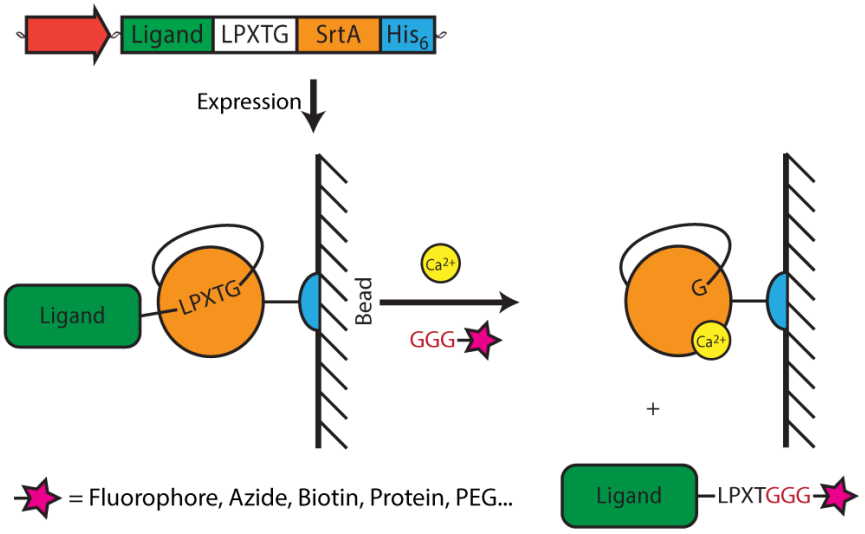


Figure 3.1: Sortase-tag expressed protein ligation scheme: Ligands are cloned in series with the amino acid sequence LPXTG, a $(GGS)_5$ linker, SrtA, and a hexahistidine tag, respectively. The chimeric protein is expressed and isolated on a nickel column. The addition of calcium and a peptide with an N-terminal glycine (and optimally three glycines) allows the SrtA enzyme to simultaneously catalyze ligand release and peptide ligation. This allows any cargo (e.g. a fluorophore, azide, biotin, PEG, proteins, etc., represented by the star) that is attached to the triglycine peptide to be site-specifically conjugated to the ligand.

S. aureus SrtAΔ59

Sortase A is uniquely suited to a system that links conjugation to purification because its catalytic mechanism (described in detail in Section 2.3.2) involves the thiolysis of one peptide bond followed by the formation of another. Both steps are necessary for transpeptidation, creating a firm link between substrate cleavage and ligation. By expressing a ligand in *cis* with a tethered sortase domain, we can instead link ligand release (cleavage) to ligand bioconjugation (ligation). Additionally, sortase activity can be modulated by calcium, allowing the chimera to stay intact through expression until release is desired.

In nature, sortase A is an extracellular protein consisting of an N-terminal transmembrane domain and a C-terminal soluble domain [237]. By deleting the first 59 amino acids, we are left with only the soluble catalytic domain that can be used *in vitro* [149, 150].

LPETG Motif

Although sortase recognizes the ambiguous LPXTG motif (where X is any amino acid), the bioconjugation literature quickly converged on X=E as the motif of choice [150]. Thus, LPETG is used as the sortase recognition motif in this and all subsequent studies.

N-Terminal Ligand

Placing the ligand N-terminal to the sortase motif and catalytic domain is the key differentiator between STEPL and the Mao sortase expression system [238]. In the previous system, the protein of interest was expressed at the C-terminus of the chimera so that only the thiolyis step is required for release and a single N-terminal glycine is left as an expression scar. By reversing the order of domains in STEPL, the initial thiolysis of the T$-$G bond keeps the protein of interest attached to the catalytic domain (and therefore the solid substrate) via an acyl-enzyme intermediate. Only when a glycine-initiated peptide is used to reform a T$-$G peptide bond is the protein of interest actually released from the enzyme. This mechanism requires the protein of interest to have a C-terminal conjugation to be recovered from the column. STEPL has a larger expression scar (LPET), but this is irrelevant since a larger C-terminal conjugate is the desired outcome.

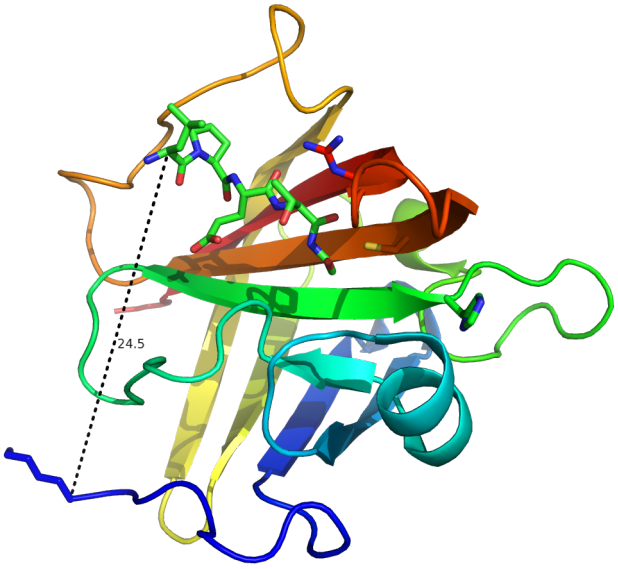


Figure 3.2: STEPL Flexible Linker Design. A) Crystal structure of Sa-SrtA active domain showing 24.5Å between the N-Terminus of the domain and C_α of the LPETG ligand.

$(GGS)_5$ Linker

Five GGS repeats were chosen for this fusion construct because the crystal structure (shown in Figure 3.2) reports a length of 24.5Å between the N-terminus of the sortase domain and its active site [239], corresponding to the length of approximately three GGS repeats (8.8Å each). Thus a $(GGS)_5$ linker was expected to provide sufficient spatial flexibility for the sortase domain to recognize and bind the LPETG motif.

His_6 Affinity Tag

The hexahistidine tag for metal ion affinity chromatography enjoys widespread use in biology because it is a small, soluble tag which binds tightly to inexpensive nickel or cobalt resins. Thus, His_6 was chosen for STEPL due to its ubiquity and simplicity. Notably, STEPL avoids one of the major drawbacks of the His_6 tag, generally "sticky" proteins; because the column is never eluted with imidazole, the impurities remain bound to the column.

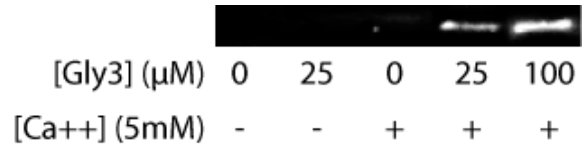


Figure 3.3: Western Blot of EGFP released from an affinity column under various conditions. An EGFP · STEPL fusion protein was expressed and washed on an affinity column. The column was then treated with various concentrations of triglycine and Ca^{2+}. Released GFP was detected by western blot.

Triglycine Peptide

Only one N-terminal glycine is necessary for sortase-mediated ligation. However, beginning the peptide with three glycine residues allows the sortase to accommodate the peptide more easily and results in greater efficiency [150]. Thus, all STEPL peptides described will begin with GGG.

3.3 Determination of Calcium Dependence

A valuable feature of the STEPL system is that it allows for the site-specific labeling of recombinantly expressed proteins without requiring any steps in addition to what is normally required for protein purification. Under optimal conditions, all of the recombinant protein that is released from the affinity column would be labeled with the desired cargo as a result of the SrtA-mediated ligation reaction. To evaluate the efficiency of this ligation reaction and to assess the extent of any non-specific cleavage of the LPXTG motif, in the absence of ligation, a model system was designed with EGFP as the "ligand" (EGFP · STEPL). This allowed for quantitative monitoring of protein release from the affinity column in the presence and absence of triglycine and calcium. Notably, peptides with two or more glycines are typically preferred for SrtA-mediated ligations since they exhibit significantly improved binding and catalysis [238].

Initial studies with the EGFP-STEPL system, in the presence and absence of Ca^{2+} (5mM) and triglycine ([GGG], 25μM or 100 μM), revealed that release of the GFP from the affinity column increased with triglycine concentration (Figure 3.3). However, it was also observed that Ca^{2+} alone could lead to some non-specific release of GFP, albeit at lower levels than when triglycine was also present. In the absence of Ca^{2+}, no GFP was released from the affinity column, with or without triglycine. These results suggested that it was important to identify an optimal Ca^{2+} concentration

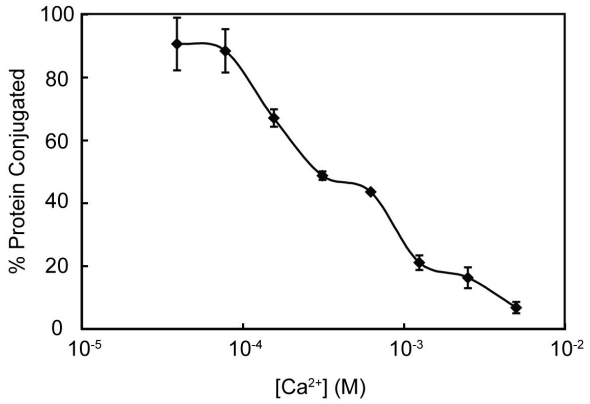


Figure 3.4: Effect of calcium on the efficiency of protein ligation. An EGFP · STEPL fusion protein was expressed and washed on an affinity column. The column was then treated with various concentrations of Ca^{2+}, in the presence of triglycine, at 37°C for 4hrs. These experiments provided a measure of the total amount of ligated and unligated product ($[EGFP]_{Total}$) released from the affinity column for each Ca^{2+} concentration. The amount of unligated product ($[EGFP]_{unligated}$) was determined by performing analogous experiments in the absence of triglycine. The percent protein conjugated ($[EGFP]_{ligated}$) was then calculated as $([EGFP]_{Total} - [EGFP]_{unligated})/([EGFP]_{Total}) * 100$.

that would maximize the ratio of ligated (i.e. GFP-triglycine conjugates) to unligated recombinant protein. Therefore, GFP-STEPL was performed in the presence of a fixed concentration of triglycine (25μM) and increasing concentrations of Ca^{2+} (0 to 5 mM). These experiments provided a measure of the total amount of ligated and unligated product released from the affinity column for each Ca^{2+} concentration. The amount of unligated product was determined by performing analogous experiments in the absence of triglycine. The maximum percent of ligated product occurred at Ca^{2+} concentrations below 100μM (Figure 3.4). Therefore, a Ca^{2+} concentration of 100μM was used for all subsequent experiments.

3.4 Modeling and Optimization of the STEPL Reaction

In order to further optimize the GFP-STEPL procedure, a systematic study on the effect of triglycine concentration, reaction temperature, and reaction time on the amount of recombinant protein released from the affinity column was conducted (Figure 3.5). As expected, the rate of protein release increased with triglycine concentration and reaction temperature. All experimental data was fit with

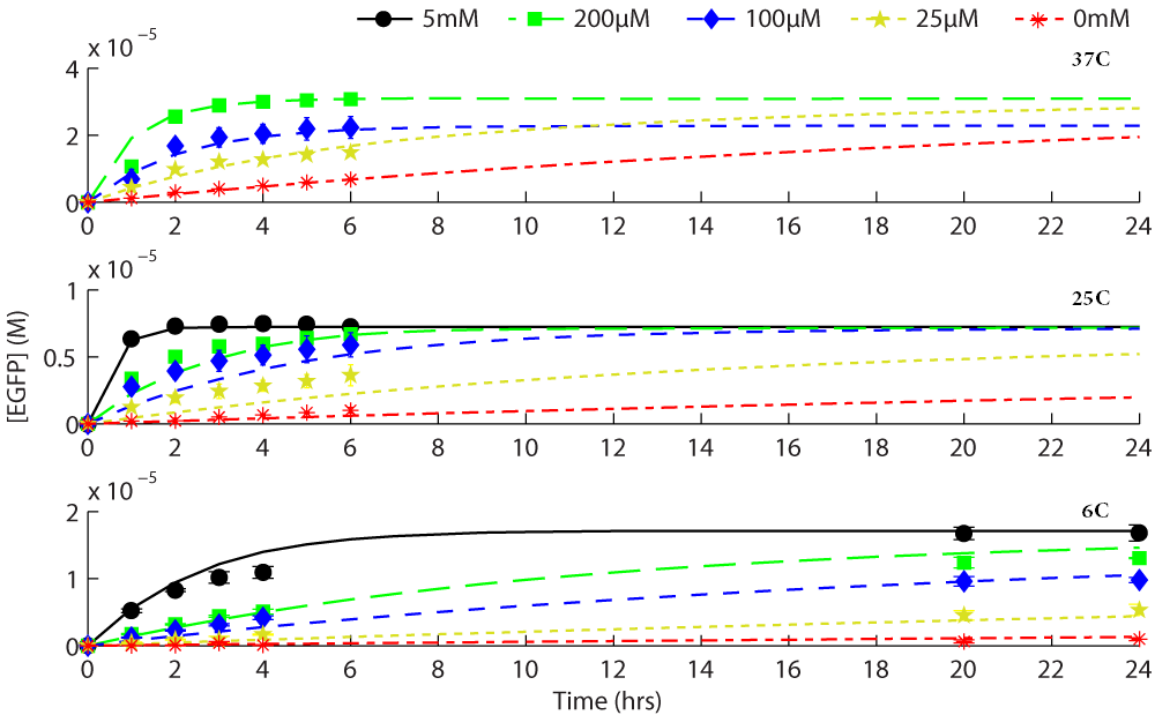


Figure 3.5: Modeled and actual EGFP release from an affinity column as a function of temperature, triglycine concentration, and time. An EGFP · STEPL fusion protein was expressed and washed on an affinity column. The column was then treated with 0mM (asterisk), 25μM (star), 100μM (diamond), 200μM (square), or 5mM (circle) triglycine and 100μM Ca^{2+}. EGFP release was monitored as a function of time. Protein release experiments were conducted at 37°C(top), 25°C(middle), and 6°C(bottom). All data was fit using a kinetic model of EGFP cleavage that takes into account both triglycine-dependent and triglycine-independent pathways. Modeled GFP release (lines) has been superimposed onto the recorded data (symbols).

a kinetic model of the reaction, detailed below in Reaction C 3.1.

$$\mathrm{EGFP} + \mathrm{SrtA} \underset{k_1}{\longleftarrow} \mathrm{EGFP} \cdot \mathrm{SrtA} \xrightarrow[k_2]{+\mathrm{GGG}} \mathrm{EGFP} \cdot \mathrm{GGG} + \mathrm{SrtA} \tag{C 3.1}$$

This model assumes that transpeptidation is the rate-limiting step of the glycine-dependent pathway and therefore collapses reversible peptide binding and transpeptidation into a single second-order rate constant. To include temperature dependence, the model also assumes that the rate constants can be modeled by the Arrhenius equation. Thus, the model's parameters are the preexponential constants and activation energies of the two pathways (Table 3.1). As shown in Figure 3.5, the model provides an acceptable fit to the observed data.

The kinetic model was used to predict the effect of various reaction conditions on three outcomes: (i) the percentage of released protein that is ligated to the peptide, (ii) the percentage of peptide consumed in the reaction, and (iii) the yield of ligated protein (i.e. the amount of ligated protein normalized by the total amount of protein initially bound to the affinity column) (Figure 3.6). The value for each of the desired outcomes was determined for reaction times of 2, 4, 6, and 24 h, with 100μM $CaCl_2$, initial triglycine peptide concentrations of 1 to 1000 μM, reaction temperatures of 4, 25, or 37°C, and assuming 100μM initial protein concentration on the affinity column. It was determined that the purity of the ligated protein was independent of reaction time, as it is simply a ratio of the two rate constants. As a result, adding excess peptide could be used to drive the ligation reaction and overwhelm the basal cleavage rate. Overall, if a highly pure ligated product is desirable, >95% purity can be achieved by simply using 2-fold or greater molar excess of triglycine-containing peptide compared with the concentration of total column-bound protein (at 37°C, Figure 3.6C). This is significantly lower than the 10-fold excess of peptide typically required for efficient intein-based EPL [2].

If it is more desirable to exhaust all of the triglycine peptide than achieve high purity of the ligated product, perhaps because the peptide is functionalized with a cargo that is cost-prohibitive,

Table 3.1: Model Parameters. The kinetics model determined Arrhenius pre-exponential constants and activation energies for the glycine-free (k_1) and glycine-dependent pathways (k_2).

Pathway	Pre-exponential Constant (A)	Activation Energy ($\Delta G^{\ddagger}$)	Rate Constant at 37°C
Glycine-free	$4.0568 \times 10^2 \mathrm{s}^{-1}$	$3.8463 \times 10^4 \mathrm{J/mol}$	$0.0419 \mathrm{s}^{-1}$
Glycine-dependent	$2.9246 \times 10^{10} \mathrm{s}^{-1}\mathrm{M}^{-1}$	$5.4958 \times 10^4 \mathrm{J/mol}$	$5.0352 \times 10^3 \mathrm{s}^{-1}\mathrm{M}^{-1}$

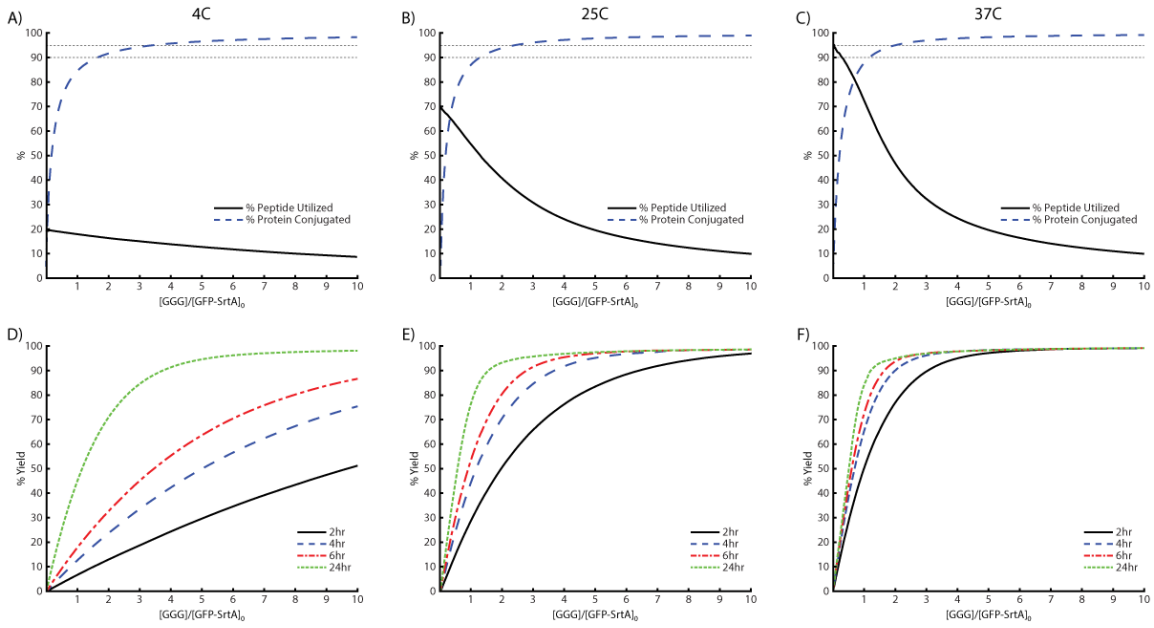


Figure 3.6: Model predictions of STEPL ligation efficiency, triglycine peptide utilization, and the percent yield of expressed GFP that is recovered from the affinity column. The kinetic model of GFP cleavage from the STEPL system was evaluated with initial conditions of 100μM Ca^{2+}, 100μM EGFP $\cdot$ SrtA, and 1μM to 1mM triglycine at 4°C, 25°C, and 37°C for 0 to 24hrs. The percentage of GFP that has been ligated to triglycine (time independent) and the percentage of triglycine peptide consumed during a 6 hr incubation was determined for reaction temperatures of (A) 4°C, (B) 25°C, and (C) 37°C. Dotted lines at 90% and 95% are included for reference. The percentage of ligated GFP recovered after a 2, 4, 6, and 24 hour incubation was determined as a function of excess triglycine (relative to the total amount of EGFP $\cdot$ SrtA) for reaction temperatures of (D) 4°C, (E) 25°C, and (F) 37°C.

Table 3.2: Synthetic Peptides

	Peptide	Molecular Weight (Da)	$\lambda_{ex}/\lambda_{em}$
P 3.1	NH_2-Gly-Gly-Gly-Lys(HiLyteFluor 750)-NH_2	1,327	754/778
P 3.2	NH_2-Gly-Gly-Gly-Lys(5-FAM)-Gly-Gly-Ser-Lys(N_3)-NH_2	1,030	492/518

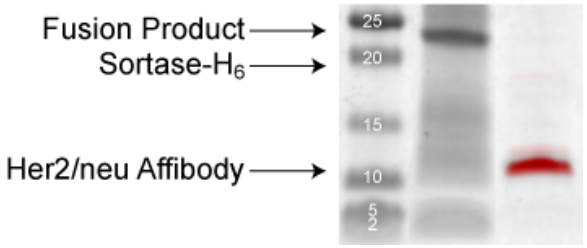


Figure 3.7: Her2/neu affibody expression and ligation. An SDS-PAGE gel was run with (1) marker, (2) raw lysate of bacterially expressed STEPL-Her2 affibody, and (3) Her2 affibody purified using a 2-fold excess of HiLyte 750-labeled triglycine peptide, 100μM Ca^{2+} at 37°C for 6hr. Black signal is from SimplyBlue SafeStain protein stain. Red signal is HiLyte 750 peptide fluorescence.

then >90% peptide consumption can be achieved by adopting reaction conditions whereby the recombinantly expressed protein is in 4-fold or greater molar excess over the peptide (at 37°C, Figure 3.6C). However, this comes at the cost of reduced purity of the ligated product and will likely require additional purification to remove unligated targeting ligands.

Higher reaction temperatures can be used to speed up the reaction and improve peptide utilization, particularly at lower triglycine peptide concentrations, but purity of the ligated product is only marginally improved. When yield is considered, the STEPL system clearly favors higher temperatures regardless of whether high purity of the ligated protein or high peptide utilization is desirable. (Figure 3.6D-F). Of course, some proteins may be unstable at high temperatures, requiring longer reactions to be performed at room temperature or in a cold room.

3.5 Fluorophore Ligation

To demonstrate the utility of STEPL as a general methodology for the site-specific labeling of tumor targeting ligands with imaging agents, the coding sequence for a Her2/neu-targeting affibody (Z_{Her2}) was cloned into the STEPL vector. The affibody was expressed and conjugated to a triglycine peptide containing the near-infrared dye HiLyte Fluor 750 (Table 3.2, P 3.1) using conditions that were expected to result in >95% purity of the fluorescent labeled affibody, based on the kinetic

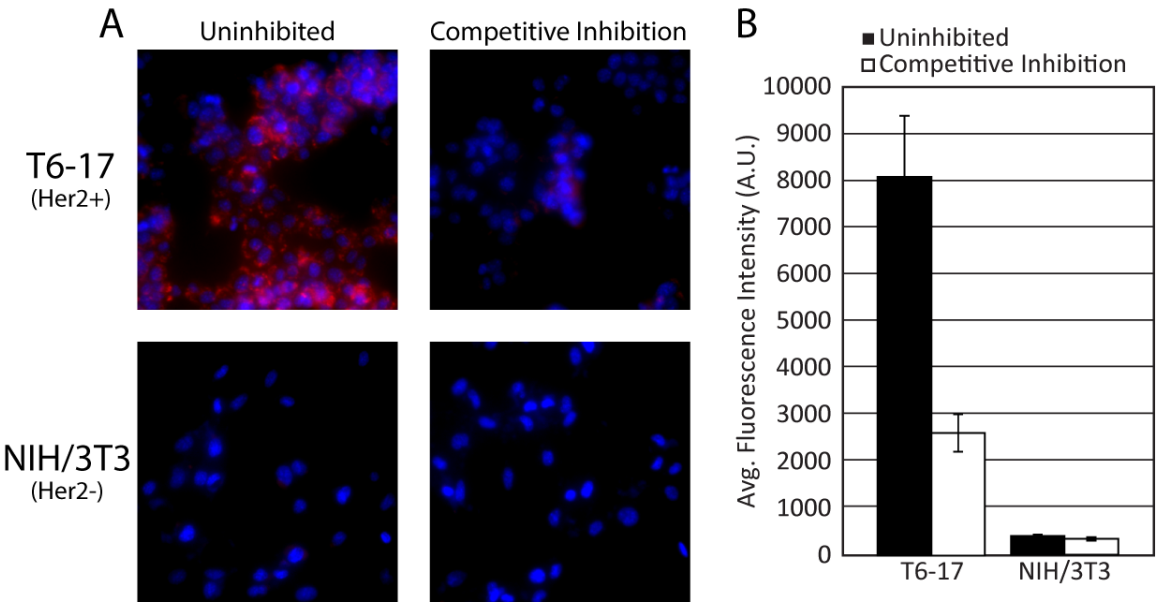


Figure 3.8: Functional evaluation of the Her2 affibody-HiLyteFluor 750 conjugate. (A) Her2/neu positive and negative cells were incubated with Her2/neu-targeted affibodies that were conjugated to HiLyteFluor 750 (red) using the STEPL system. Cells were also stained with Hoechst 33342 (nuclear stain, blue). (B) In-cell western quantification of HiLyteFluor 750 fluorescence.

model established above (2-fold excess peptide, 100μM Ca^{2+}, 37°C, 6hr). Efficient ligation between the affibody and the fluorescently-labeled peptide was confirmed by SDS-PAGE (Figure 3.7). The major band in the 700nm channel (protein stain) co-localized with the single fluorescent band in the 800nm channel (HiLyte Fluor 750), following removal of excess free peptide by dialysis. Only a very faint signal stemming from the unligated protein was observed in the 700 nm channel, at a slightly lower molecular weight than the ligated product. To confirm that the affibody remained functional following the ligation reaction, it was incubated with T6-17 and NIH/3T3 cells in vitro, which are positive and negative for the Her2/neu receptor, respectively (Figure 3.8). As expected, the affibody labeled the T6-17 cells exclusively, with no observable labeling of the NIH/3T3 cells. Further, the addition of excess unlabeled affibody (i.e. cleaved with triglycine) competitively inhibited the binding of the fluorescently labeled affibody to T6-17 cells, suggesting that binding was specific for the Her2/neu receptor. Quantification using an in-cell western assay (Figure 3.8B) corroborated the fluorescence microscopy findings. Similar results were obtained by applying the epidermal growth factor receptor (EGFR)-targeted affibody, Z_{EGFR}, to EGFR-positive and negative cells (Figure 3.9).

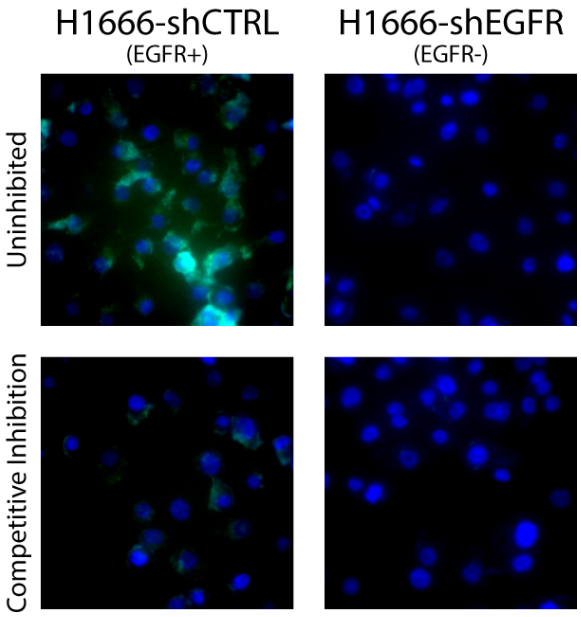


Figure 3.9: Functional evaluation of the EGFR affibody-FAM conjugate. EGFR positive and negative cells were incubated with Her2/neu-targeted affibodies that were conjugated to FAM (green) using the STEPL system. Cells were also stained with Hoechst 33342 (nuclear stain, blue).

3.6 Azide Ligation and Nanoparticle Synthesis

In addition to imaging agents, STEPL can also be used to conjugate various other functional moieties including bio-orthogonal reactive groups (e.g. azide) onto the C-terminus of targeting ligands. The site-specific introduction of azides onto recombinant proteins provides a very favorable approach for the efficient coupling of targeting ligands to nanoparticles using click chemistry. In particular, this approach allows tight control over both ligand orientation and density on the nanoparticle surface. We have previously shown that both of these factors can have a dramatic impact on nanoparticle avidity [38]. As proof-of-principle, the Her2/neu affibody was conjugated to a synthetic peptide containing a green fluorophore (5-FAM) as well as an azide group (Table 3.2, P 3.2). This conjugate was then reacted with superparamagnetic iron oxide (SPIO) nanoparticles functionalized with azadibenzocyclooctyne (ADIBO). ADIBO is a dibenzocyclooctyne derivative capable of copper-free click reactions with azides. The resulting Z_{Her2}-SPIO conjugates were incubated with T6-17 and NIH/3T3 cells. Cell labeling was then assessed by acquiring T_2 relaxation times and T_2^*-weighted images of cell lysates (Figure 3.10). The Her2-positive cells exhibited a marked decrease

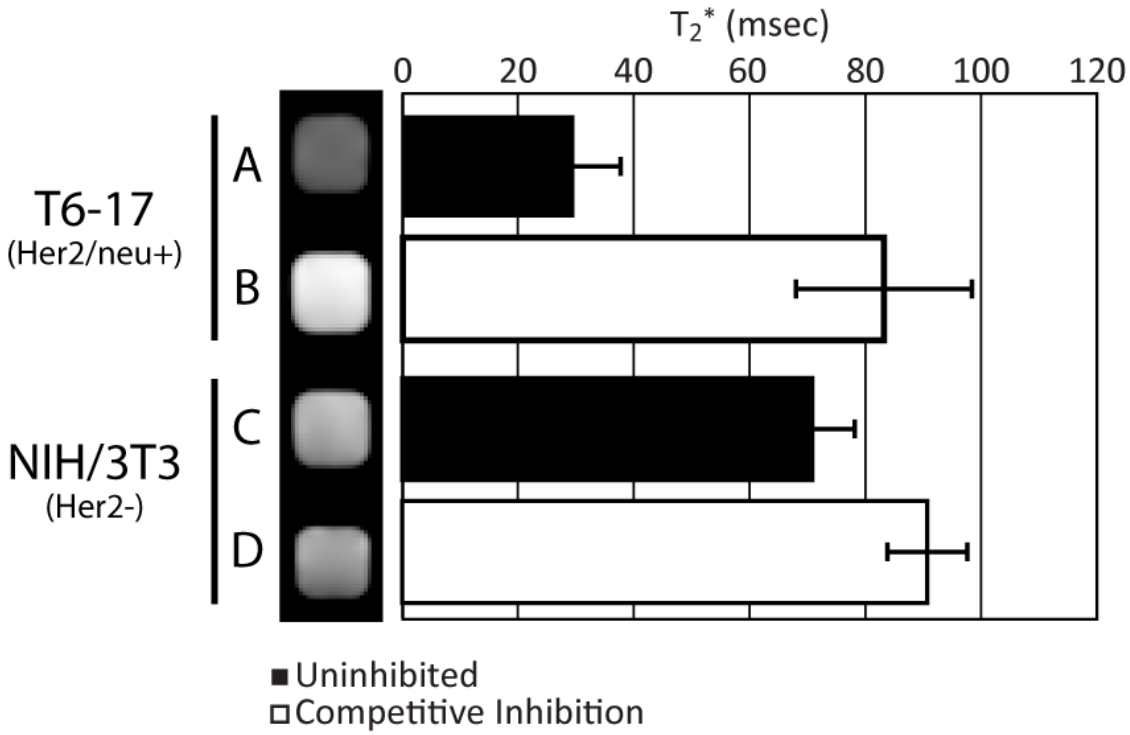


Figure 3.10: Functional evaluation of the Her2 affibody-SPIO conjugates. Her2/neu-positive and Her2/neu-negative cells were incubated with Her2-SPIO conjugates in the presence and absence of excess free affibody. Free affibody served as a competitive inhibitor to confirm specific binding of the Her2/neu receptor. Relaxivity measurements and T_2^*-weighted MR images of each cell suspension were acquired.

in T_2^*-relaxation times, consistent with the presence of SPIO, in comparison to Her2-negative cells. An observable negative contrast was also observed upon MR imaging of the Her2-positive cells. Competitive inhibition, using an excess of unlabeled Z_{Her2}, led to a loss in MR contrast, indicative of receptor-specific binding. Therefore, these results provide clear evidence that STEPL can be combined with click chemistry for the site-specific attachment of targeting ligands onto nanoparticles.

3.7 Discussion

STEPL offers a number of features that make it a very favorable approach for bioconjugation reactions. First and foremost, STEPL combines release of recombinant proteins from the affinity column and bioconjugation into a single step. This greatly simplifies the entire bioconjugation procedure since no subsequent labeling and purification steps are required, saving time, money, and complexity. Second, STEPL allows for the site-specific conjugation of cargo. Site-specific functionalization has been shown to be beneficial in a number of applications including the preparation of protein-drug conjugates, which often exhibit higher efficacy than randomly labeled targeting ligands [240]. It has also been shown that the site-specific attachment of targeting ligands to nanoparticles can improve nanoparticle avidity [162]. Third, STEPL conjugates the peptide to ligand in a stoichiometric manner. This is particularly important when labeling targeting ligands with imaging agents, since it allows for precise quantitative imaging. Fourth, the conditions used to release protein from the affinity column can be manipulated to ensure that essentially all of the recovered protein is conjugated with the desired cargo. This eliminates the often difficult process of purifying conjugated products from unconjugated proteins. Since in many applications a large protein is labeled with low molecular weight drugs or imaging agents, the conjugated and unconjugated forms of the protein can differ by as little as a few hundred to a few thousand Da, potentially without any significant change to hydrophobicity or charge. Fifth, construction of the STEPL system as a single expressible protein removes the additional step of removing sortase from the conjugated product, a common feature of current sortase conjugation systems [150, 156, 241]. Additionally, in systems where unconjugated ligand is easily separable, the reaction conditions could be altered to ensure that expensive synthetic peptides can be exhausted in the ligation reaction. Thus, STEPL is a single-chain, self-cleavable system where high-cost components can be fully utilized; traits highly desirable in industrial protein

production as they reduce overall cost and time [242].

The one identified shortcoming of Sortase A is that it exhibits some cleavage even in the absence of glycine. Previous studies have addressed this problem by making a destabilizing mutation to Trp-194. However, we hypothesized that reducing the calcium concentration would have a similar effect with finer control. This was found to be true, as sub-millimolar calcium concentrations provided a sharp, dose dependent drop in background cleavage. There is also potential for background cleavage during expression due to cytosolic calcium in the bacterium. However, the calcium level inside *E. coli* is estimated to be between 0.1 - 1μM [243], so while there is likely some loss, it should be minimal and could always be discouraged by the aforementioned mutation.

To further optimize and understand the cleavage reaction, a kinetic model was established and its parameters (time, temperature, and initial triglycine concentration) were systematically varied. To simplify the model, the reversible binding of peptide to the enzyme and product conversion were condensed into one second-order rate constant. This is justifiable because the applicable peptide concentrations do not appear to saturate the binding curve and the determined rate constants are well below the diffusion limit [244, 245], implying that product conversion is rate limiting.

The model reveals a fundamental conflict between conjugate purity and peptide utilization. The glycine-independent pathway can be easily overwhelmed by adding a large excess of peptide to drive the glycine-dependent pathway. On the other hand, the peptide can be fully utilized by making it the limiting reactant. Therefore, if product purity is required and downstream purification of unreacted ligand would be difficult, it is optimal to use excess peptide. If peptide utilization is the primary concern, the optimal conditions are 37°C with a 1:4 ratio of synthetic peptide to STEPL protein, although it is important to note that in this latter case additional purification is needed to purify conjugated product from unconjugated proteins. Systems where the peptide enables the conjugate to be immobilized onto a surface or particle are ideal candidates for peptide exhaustion.

In this study, STEPL was used to functionalize affibodies with chemical groups useful for molecular imaging. A near-IR fluorophore was utilized to optically differentiate between cells expressing and lacking the proto-oncogene Her2/neu. The NIR-dyed affibody was used to quantify Her2/neu expression differences between the T6-17 and NIH/3T3 cells, which demonstrates STEPL's utility for in-cell western techniques. Additionally, STEPL was used to conjugate a bio-orthogonal reactive group (an azide) to the Her2/neu affibody. The azide readily reacted to a strained alkyne on the

surface of superparamagnetic iron oxide nanoparticles. Due to the site-specific nature of STEPL, the affibody was linked in a specific orientation, which greatly increases the particle's efficacy in distinguishing between cells expressing and lacking Her2/neu. STEPL has the potential to conjugate many other moieties to its target protein, such as biotin, poly(ethylene-glycol), antibiotics, metal chelates, and photocrosslinkers, all of which have been proven compatible with the sortase enzyme [150].

3.8 Conclusions

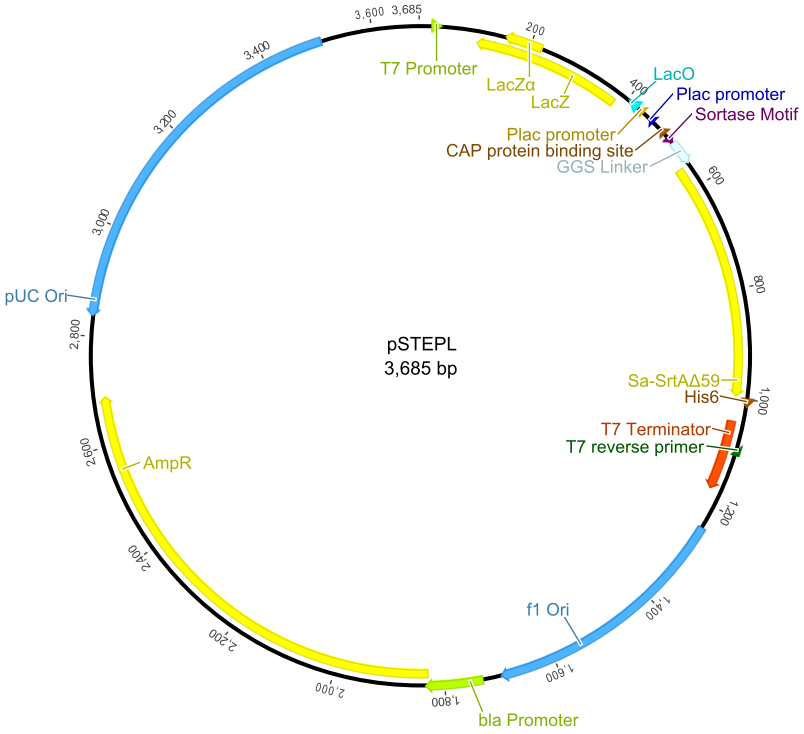
STEPL has proven to be a flexible and efficient system for molecular imaging and targeted therapeutic applications. This study validated and optimized the system for ligand purity and peptide-cargo utilization. STEPL was then used to visualize and quantify Her2/neu and EGFR expression *in vitro*. Moreover, because it has the ability to link virtually any protein expressible in bacteria with any cargo that can be attached to a triglycine peptide, STEPL has potential applications in many fields.

3.9 Materials & Methods

Cloning

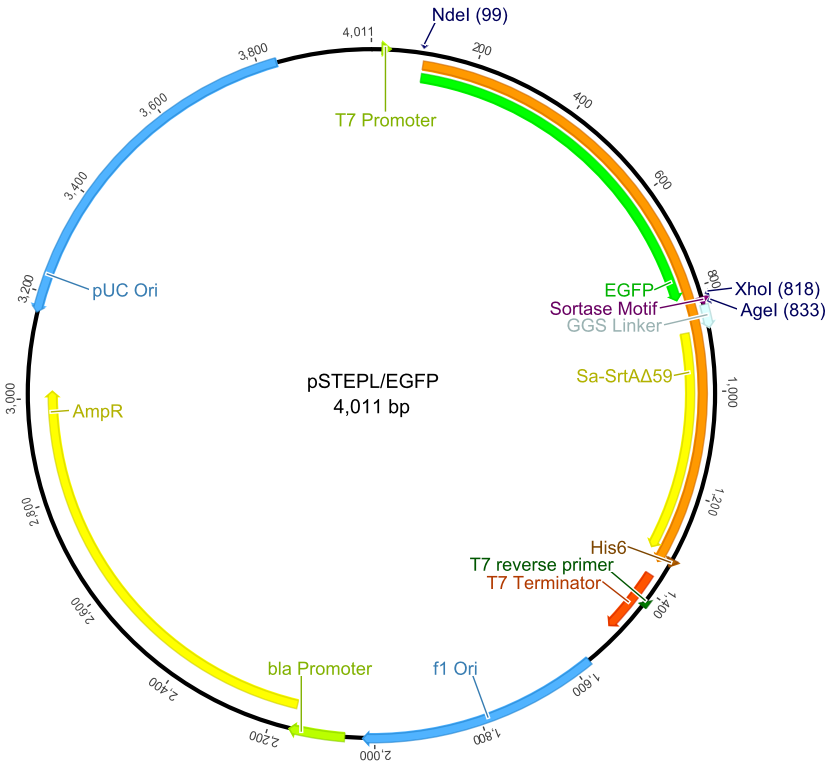
Sa-SrtAΔ59 [238] was amplified from pGMBCS-SrtA (Addgene plasmid 21931 [153]) with an N-terminal $(GGS)_5$ sequence and C-terminal H_6 sequence. To facilitate blue/white screening, the Lac operon was amplified from pUC19 (Invitrogen) in an antisense orientation with a C-terminal sequence coding for the restriction site XhoI, the sortase recognition sequence LPETG, and the $(GGS)_5$ linker. Overlap-extension PCR was used to join the Lac operon product to the Sa-SrtA59 product. The full sequence was then cloned into pRSET-A (Invitrogen) via the NdeI and MluI restriction sites, creating the STEPL vector, pSTEPL (see Plasmid Map 3.1). Sequences were verified by restriction analysis and sequencing. The EGFP sequence, Z_{Her2} sequence [113], and Z_{EGFR} sequence [114] were amplified with 5' NdeI and 3' SalI sites and cloned into pSTEPL via its NdeI and XhoI sites (Plasmid Maps 3.2, 3.3, and 3.4, respectively). White colonies were picked and verified by restriction analysis and sequencing.

Plasmid Map 3.1: pSTEPL



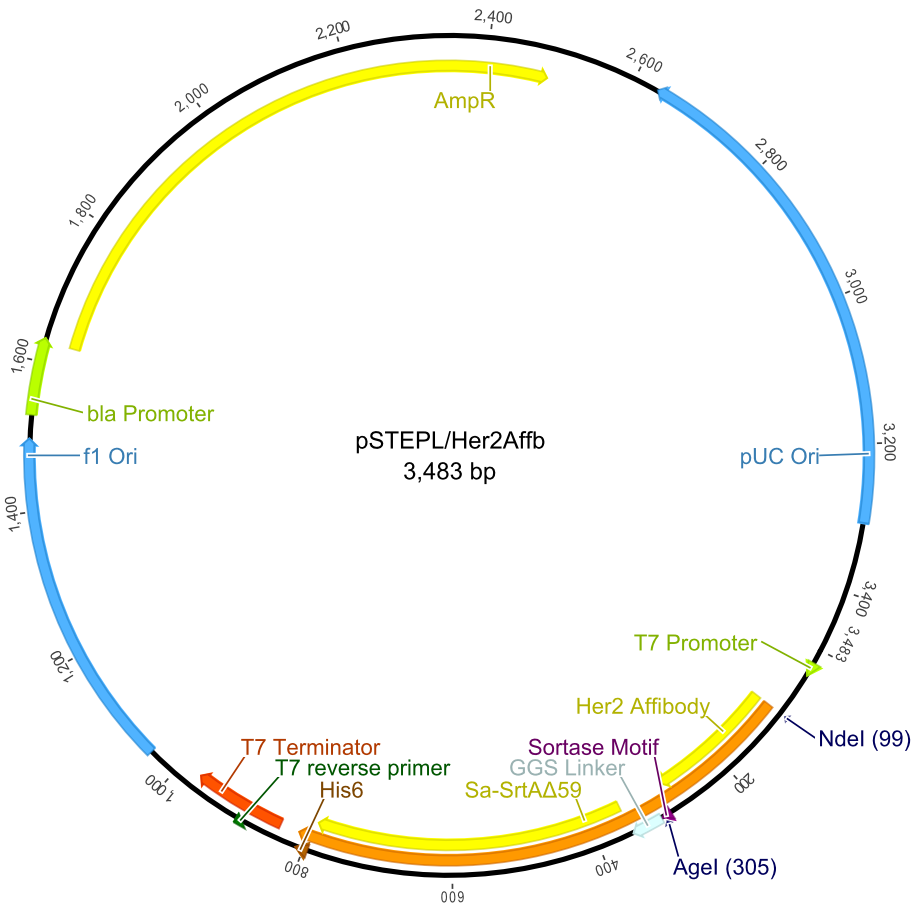
Feature	*Location*
T7 Promoter	20-39
LacZ	384-106
LacZα	222-154
LacO	424-402
P_{lac} -10	434-429
P_{lac} -35	458-453
CAP Binding Site	490-478
Sortase Motif	497-511
$(GGS)_5$ Linker	512-556
Sa-SrtAΔ59	557-997
His_6	998-1015
T7 Terminator	1044-1173
f1 Origin	1244-1699
bla Promoter	1731-1835
Amp^R	1830-2690
pUC Origin	3508-2835

Plasmid Map 3.2: pSTEPL/EGFP



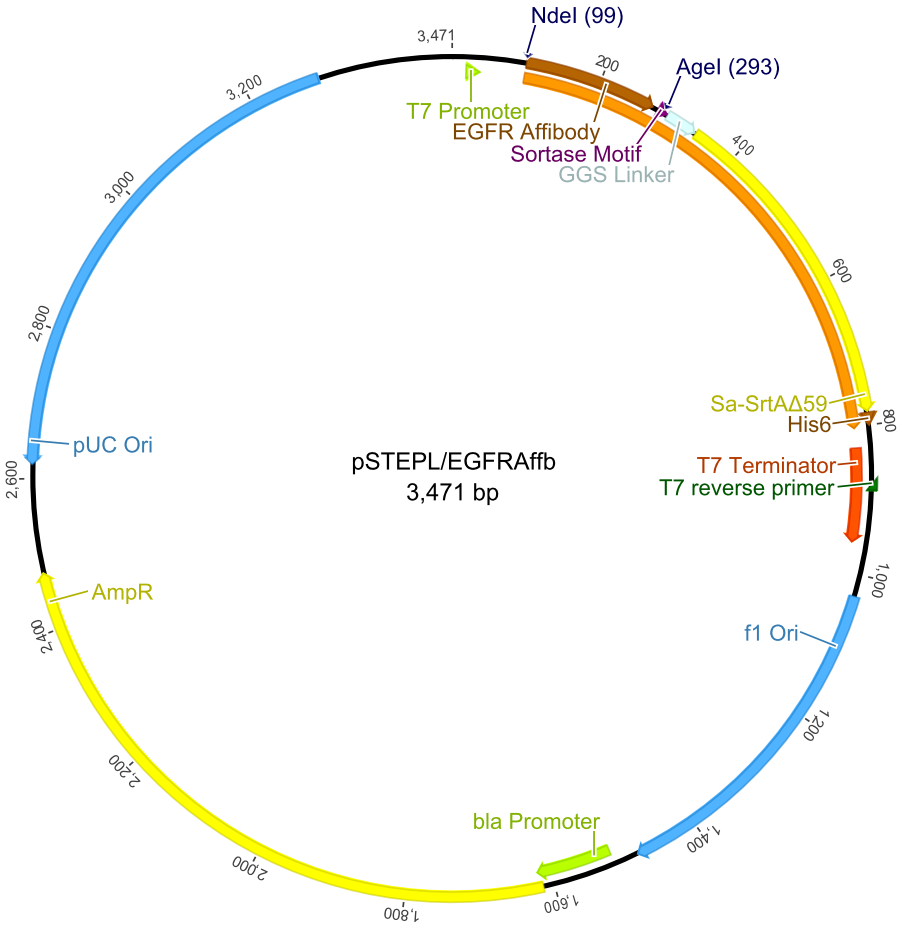
Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-1344
EGFP	100-816
Sortase Motif	823-837
$(GGS)_5$ Linker	838-882
Sa-SrtAΔ59	883-1323
His_6	1324-1341
T7 Terminator	1370-1499
f1 Origin	1570-2025
bla Promoter	2057-2161
Amp^R	2156-3016
pUC Origin	3843-3161

Plasmid Map 3.3: pSTEPL/Her2Affb



Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-816
Her2 Affibody	100-288
Sortase Motif	295-309
$(GGS)_5$ Linker	310-354
Sa-SrtAΔ59	355-795
His_6	796-813
T7 Terminator	842-971
f1 Origin	1042-1497
bla Promoter	1529-1633
Amp^R	1628-2488
pUC Origin	3306-2633

Plasmid Map 3.4: pSTEPL/EGFRAffb



Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-804
EGFR Affibody	100-276
Sortase Motif	283-297
$(GGS)_5$ Linker	298-342
Sa-SrtAΔ59	343-783
His_6	784-801
T7 Terminator	830-959
f1 Origin	1030-1485
bla Promoter	1517-1621
Amp^R	1616-2476
pUC Origin	3294-2621

Protein Expression, Cleavage, & Bioconjugation

Constructs were transformed into the BL12-derived Rosetta2 BL21(DE3) line (EMD Millipore). 50mL starter cultures were grown overnight in LB-Ampicillin. These were added to 450mL of LB and grown to an OD_{600} of 0.8-1 before induction with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were allowed to express for 24hrs at 25°C. Cells were then harvested by centrifugation (6000g, 15min) and resuspended in 10mL of lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 1mg/mL Lysozyme, 1 EDTA-free cOmplete Mini protease inhibitor tablet (Roche), pH 7.5). Lysates were incubated at room temperature for 30min under gentle agitation before freezing overnight at -80°C. Samples were then thawed and incubated for 30min with DNAse I (Roche) under gentle agitation. Lysates were then sonicated and separated by centrifugation (10,000 rpm, 30min).

For optimization experiments, 8mL of clarified lysate for each condition was incubated for 1 hr with 0.6mL Talon resin (Clontech, equilibrated in lysis buffer). The lysate and beads were then added to a column and beads were washed with 6mL STEPL buffer (20mM Tris-base, 50mM NaCl, pH 7.5). Washed beads were resuspended in a total of 1.2mL of STEPL buffer containing the indicated amounts of $CaCl_2$ and triglycine (Sigma-Aldrich) and aliquoted into three 1.5mL microcentrifuge tubes. Samples were shaken at 1,000 rpm, the indicated temperature, and protected from light. At each timepoint, samples were spun down at 3,000 rpm for 5min. Absorbance spectra were taken of the supernatants from 400-600nm using a Cary 100 Bio UV-Visible Spectrophotometer (Varian) and the sample was returned to shaking. At the end of the timecourse, beads were washed three times with 1mL STEPL buffer and incubated for 30 min in 100mM EDTA. Stripped beads were spun down as before and the absorbance spectra were taken of the supernatants from 400 600nm.

For bioconjugation experiments, 8mL of clarified lysate was incubated for 1hr with 0.5mL Talon resin (equilibrated in lysis buffer). The lysate and beads were added to a column and beads were washed with 5mL STEPL buffer. 400μL of STEPL buffer containing 150μM synthetic peptide (Table 3.2) or 5mM triglycine (for labeled and unlabeled preparations, respectively) and 100μM $CaCl_2$ was flowed over the beads until it replaced the wash buffer. Columns were protected from light and reacted for 6hrs at 37°C. 1mL of STEPL buffer was added to the column and the flowthrough collected. To remove unreacted peptide, flowthrough was dialyzed three times against 4L of STEPL buffer at 4°C while protected from light (Slide-A-Lyzer2 cassettes, 3.5K cutoff, Thermo Scientific).

EGFP Release Analysis & Model Design

Absorbance spectra obtained from optimization samples were baseline corrected and the EGFP concentration was determined using the Beer-Lambert law (ε(EGFP, 488nm) = $55,000\text{M}^{-1}\text{cm}^{-1}$, [246]). EGFP concentrations were fit to the sum of equations (3.4) and (3.5) in the following system of ODEs using non-linear least squares:

$$\frac{d\,[\text{EGFP}\cdot\text{SrtA}]}{dt} = -k_1\,[\text{EGFP}\cdot\text{SrtA}] - k_2\,[\text{EGFP}\cdot\text{SrtA}]\,[\text{GGG}] \tag{3.1}$$

$$\frac{d\,[\text{SrtA}]}{dt} = k_1\,[\text{EGFP}\cdot\text{SrtA}] + k_2\,[\text{EGFP}\cdot\text{SrtA}]\,[\text{GGG}] \tag{3.2}$$

$$\frac{d\,[\text{GGG}]}{dt} = -k_2\,[\text{EGFP}\cdot\text{SrtA}] \tag{3.3}$$

$$\frac{d\,[\text{EGFP}]}{dt} = k_1\,[\text{EGFP}\cdot\text{SrtA}] \tag{3.4}$$

$$\frac{d\,[\text{EGFP}\cdot\text{GGG}]}{dt} = k_2\,[\text{EGFP}\cdot\text{SrtA}]\,[\text{GGG}] \tag{3.5}$$

$$k_1 = A_1 T e^{-\Delta G_1^{\ddagger}/RT} \tag{3.6}$$

$$k_2 = A_2 T e^{-\Delta G_2^{\ddagger}/RT} \tag{3.7}$$

where the model fits for $A_1, A_2, \Delta G_1^{\ddagger}$, and $\Delta G_2^{\ddagger}$. Temperatures are on the Kelvin scale. Initial EGFP · SrtA concentration was determined by adding the concentration of EGFP in the final time-point to the concentration of EGFP from the stripped beads. Initial triglycine concentration and temperature were varied experimentally. Initial conditions for [SrtA], [EGFP], and [EGFP · GGG] were zero. Model predictions were produced with the following initial conditions: 100μM Ca^{2+}, 100μM EGFP · SrtA, 1μM - 1μM triglycine, 4° - 37°C.

Cell Culture

NIH/3T3 and T6-17 cells (i.e. NIH/3T3 cells engineered to stably express Her2/neu; kindly provided by Dr. Mark Greene, University of Pennsylvania) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin at 37°C and 5% CO_2. H1666 cells expressing pLKO.shCTRL and pLKO.shEGFR [247] (kindly provided by Dr. Matthew Lazzara, University of Pennsylvania) were maintained in RPMI supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin at 37°C and 5% CO_2.

Fluorescence Analysis of Cell Targeting

NIH/3T3 and T6-17 cells were incubated with 1μM Her2/neu affibody conjugated to HiLyte Fluor 750 for 4 hours in full media with and without a 10-fold excess of unlabeled Z_{Her2}. Cells were washed 3 times with affibody-free media before being imaged in serum-free DMEM. Imaging was performed with an Olympus IX81 inverted fluorescent microscope with a back-illuminated EMCCD camera (Andor) and a SOLA excitation source (Lumencor). Images of HiLyte Fluor 750 were acquired using the filter set (HQ710/75, HQ810/90, Q750LP) (Chroma). A LUC PLAN FLN 40X objective (NA 0.6) was used for all imaging studies. ImageJ was used to merge the fluorescent images and equalize levels. After optical imaging, the plate was scanned by an Odyssey Imaging System (Li-Cor). User-defined regions of interest were drawn within each well and fluorescence within the 800-channel was quantified to determine relative Her2/neu expression.

Relaxation Measurements of Cell Targeting and MR Imaging

Azodibenzocyclooctyne (ADIBO)-functionalized superparamagnetic iron oxide (SPIO) nanoparticles were synthesized as previously described [38]. Azide-modified Z_{Her2} was conjugated to ADIBO-SPIO nanoparticles by combining 5mg Fe/mL SPIO nanoparticles with 30μM affibody. Reactions were mixed overnight at room temperature and affibody-SPIO conjugates were purified on PD-10 columns (GE Healthcare).

T6-17 and NIH/3T3 cells were incubated with 125μg Fe/mL of Her2/neu-targeted SPIO for 45 minutes in full media with and without a 100-fold excess of unlabeled Z_{Her2} in triplicate. Cells were transferred to 1.5mL microcentrifuge tubes and washed with 500μL PBS three times before being resuspended in 300μL RIPA Lysis Buffer (Millipore). T2 measurements were taken using a benchtop relaxometer (Bruker mq60). Following relaxation measurements, cell lysates were combined and 100μL was transferred to wells of a 364-well plate. Images of the cells were taken on a 9.4-T magnet interfaced to a Varian INOVA console using a 70 mm inner diameter volume coil for radiofrequency transmission and reception. T_2^*-weighted gradient echo (GRE) MR images were collected using parameters as follows: repetition time (TR) = 200 ms, echo time (TE) = 5 ms, flip angle = 20°, slice thickness = 0.5 mm, number of acquisitions = 8.

Chapter 4

Making Bispecific Ligands with STEPL

Abstract

Bispecific targeting ligands have proven to be useful in a number of applications, such as targeted immunotherapies, molecular imaging, and delivery across the blood-brain barrier. However, the generation of such agents can be difficult. In this study, we use the Sortase-Tag Expressed Protein Ligation system to produce eight monomeric and dimeric targeting ligands from only two optimized expression cassettes for the affibodies Z_{Her2} and Z_{EGFR}. The ligands are then assayed to determine their apparent affinity for Her2/neu-positive, EGFR-positive, and Her2/neu/EGFR double-positive cell lines. This data is used to predict the ligand concentration that will maximize contrast between the single- and double-positive cell lines, which is supported by fluorescence microscopy.

4.1 Introduction

As described in Section 2.1.4, bispecific targeting ligands have proven useful in increasing the specificity of an imaging agent, recruiting killer T-cells to a tumor, and transporting agents across the blood-brain barrier, among other applications. The Sortase-Tag Expressed Ligation (STEPL) system is particularly adept at creating bispecific ligands due to the inherent modularity and site specificity of its active enzyme, *S. aureus* Sortase A, a calcium-dependent bacterial transpeptidase that cleaves the threonine-glycine peptide bond in its recognition motif, LPXTG, and subsequently ligates any other protein that begins with a glycine residue to restore a threonine-glycine bond [237]. The ability of sortase to link any two proteins, so long as one contains LPXTG and the other an N-terminal glycine, makes sortagging an incredibly powerful and versatile bioconjugation strategy. By making one of the reactant proteins synthetically, non-canonical amino acids can be included that can give the protein bio-orthogonal chemical reactivities, fluorescence, lipid anchors, PEG chains, and antibiotics, among many other functionalizations [149]. Notably, synthetic peptides are not limited to just a single modification, so they can be added in controlled combinations and stoichiometries.

STEPL further improves on traditional sortase-mediated conjugations by expressing the target protein, sortase recognition motif, a flexible linker, the sortase active domain, and a hexahistidine tag as a single fusion protein. This chimeric design allows the active enzyme to be immobilized onto a metal resin along with any unconjugated ligand, eliminating the need for inefficient downstream purifications. The addition of calcium and a synthetic peptide activates the sortase domain and results in the simultaneous cleavage of the target protein off of the column and C-terminal conjugation of the synthetic peptide, resulting in >95% of the collected protein being ligated to the synthetic peptide [236]. In addition to being efficient, STEPL is also modular. Once a ligand has been cloned into the pSTEPL vector and its expression has been optimized, it can be ligated to any number of synthetic peptides. This makes it easy to incorporate different functionalities onto a targeting ligand, such as the bio-orthogonal crosslinking groups necessary for the production of bispecifics.

When crosslinking two biomolecules, the strain-assisted azide-alkyne cycloaddition reaction has emerged as a robust strategy [248]. In addition to being a “click” reaction, which designates it as efficient, biocompatible, and modular [138], the reaction is also bio-orthogonal, which limits potential unwanted products. The azide and dibenzocyclooctyne (DBCO) groups can be incorporated into

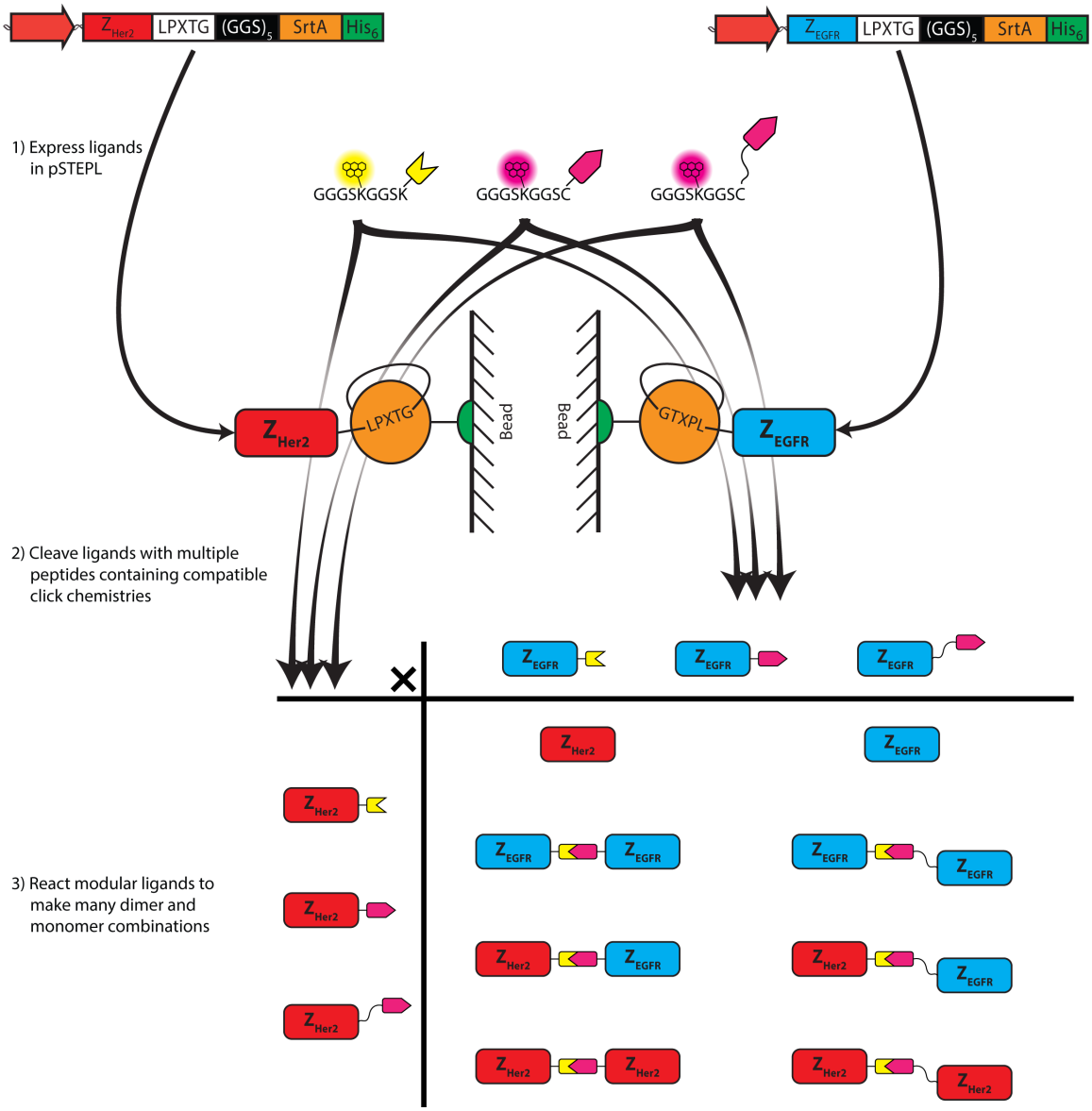


Figure 4.1: Bispecific Ligand Generation using STEPL. Arrays of hetero- and homodimers are easily generated in the STEPL system. First, the targeting ligands must be cloned into pSTEPL and expressed. Second, targeting ligands are cleaved off of separate columns using peptides that contain bio-orthogonal chemistries, such as the azide-alkyne click groups. These peptides may contain multiple lengths and categories of spacer regions. Finally, dimers are created by simply mixing monomers with compatible reactive groups. Although this figure depicts the ligands and peptides used in this study, they may be substituted for any of the wide range of ligands and peptides compatible with the STEPL system with the same effect.

Table 4.1: Synthetic Peptides

	Peptide	Molecular Weight (Da)	$\lambda_{ex}/\lambda_{em}$
P 3.2	NH_2-Gly-Gly-Gly-Lys(5-FAM)-Gly-Gly-Ser-Lys(N_3)-NH_2	1,030	492/518
P 4.1	NH_2-Gly-Gly-Gly-Lys(TAMRA)-Gly-Gly-Ser-Cys-NH_2	1,034	541/568
P 4.2	NH_2-Gly-Gly-Gly-Lys(TAMRA)-Gly-Gly-Ser-Cys(DBCO-maleimide)-NH_2	1,462	541/568
P 4.3	NH_2-Gly-Gly-Gly-Lys(TAMRA)-Gly-Gly-Ser-Cys(DBCO-$(PEG)_4$-maleimide)-NH_2	1,709	541/568

synthetic peptides, making the reaction compatible with the STEPL system as well. By purifying one ligand with a peptide containing an azide and one with a peptide containing DBCO, the ligands can be crosslinked by simply incubating them together in Tris buffer, generating dimeric ligands of controlled specificity and stoichiometry. The modularity inherent to the STEPL system allows for easy alteration of the chemical crosslinking strategy as well as the length of the spacer region between the constituent ligands, making STEPL exquisitely well-suited to the generation of bispecific targeting ligands.

In this study, we produced an array of homodimeric and heterodimeric affibodies with two different spacer regions using the STEPL system (Figure 4.1). The affibodies are then used to compare the effects of dimerization and spacing on targeting specificity and relative affinity.

4.2 Affibody Dimerization

One of the major advantages of the STEPL system is that it is easy to produce the same protein with a number of different modifications. Taking advantage of this capability, we expressed the Z_{Her2}- and Z_{EGFR}-STEPL constructs cloned in Chapter 3, cleaving each with Peptides P 3.2, P 4.2, and P 4.3 (Table 4.1), which contain functional groups compatible with the strain-assisted azide-alkyne cycloaddition click chemistry.

The modified affibodies were then incubated with each other in various combinations to create a variety of hetero- and homodimers (Table 4.2). Affibody dimerization was verified by SDS-PAGE. As shown in Figure 4.2, purified dimers have a higher apparent molecular weight (approximately 20 kDa) than the monomers (6.6 kDa). It was expected that the dimers would exhibit an apparent molecular weight of roughly twice the monomers, around 13 kDa. However, the bulky dibenzocyclooctyne and poly(ethylene-glycol) groups added by the peptide likely inhibit motility through the poly(acrylamide) gel, giving it a higher apparent molecular weight.

In all, we were able to create eight unique affibody constructs while only optimizing two protein

Table 4.2: Affibody Click Reactions. Homo- and heterodimeric affibodies were created by incubating one azide-conjugated affibody with one DIBO-conjugated affibody. Created dimers are indicated with a ✓. Construct names are in parentheses.

	$Z_{Her2} \cdot DIBO$	$Z_{Her2} \cdot (PEG)_4 \cdot DIBO$	$Z_{EGFR} \cdot DIBO$	$Z_{EGFR} \cdot (PEG)_4 \cdot DIBO$	None
$Z_{Her2} \cdot N_3$	✓(HH)	✓(H4H)	✓(HE)	✓(H4E)	✓(H)
$Z_{EGFR} \cdot N_3$			✓(EE)	✓(E4E)	✓(E)

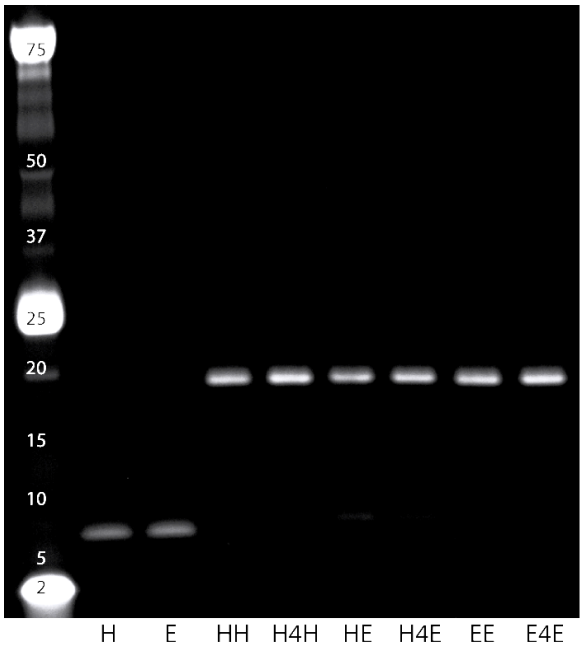


Figure 4.2: Affibody Dimerization. Affibody monomers and dimers were analyzed by SDS-PAGE and imaged by UV fluorescence. Dimers have an apparent molecular weight of nearly 20kD, indicating successful click reactions.

expression protocols. We easily incorporated two different linker lengths, which could potentially alter binding properties. It was also possible to include fluorophores on each of the peptides, which enabled us to quantify the samples to ensure equimolar click reactions and to use both monomers and dimers in downstream imaging and flow cytometry applications. The site-specific and stoichiometric bioconjugation of multiple chemical functionalities would be much more complicated in other chemical and enzymatic bioconjugation systems.

This study gave only a small example of the combinatorial possibilities of creating bispecific ligands with STEPL. There are many biomarkers that already have affinity ligands compatible with the STEPL system (See Section 2.1.3). Expressing just five of these ligands via STEPL enables quick access to ten unique bispecific molecules, five homodimers, and five monomers per linkage. A panel of ten would give 45 bispecific ligands, in addition to the ten monomers and ten homodimers. As the number of potential biomarkers increases, the modularity of STEPL becomes more and more powerful, quickly entering the realm of medium to high-throughput screens.

STEPL's modular approach to creating bispecific proteins would also benefit systems where one of the affinity ligands is always the same. For example, the generation of BiTE agents could be made much more efficient by scaling production of the α-CD_3 ligand. Partner ligands could then simply be expressed in STEPL, cleaved with a compatible peptide, and reacted with stock α-CD_3 ligand.

Many other groups have recognized the utility in a modular approach to creating bispecific targeting ligands and published their own techniques to accomplish this [249, 250, 251, 252]. However, using STEPL to generate modularity has distinct advantages over these previous technologies. Unlike modular systems relying on IgG heavy and light chain interactions, STEPL is not limited to any one class of proteins. Ligand classes could even be mixed and matched using STEPL. It would work just as well linking a nanobody to an scFv as it would dimerizing two affibodies. Also, since the linker is being introduced by conjugation, the system is symmetric, as opposed cloned systems where asymmetry is coded into the protein, such as knob-in-hole bispecific antibodies. There is no need to clone different constructs for the ligand to receive an azide or alkyne functionalization. That is decided post-expression, during purification. Finally, spacer regions have been shown to have a large effect on bispecific functionality [253]. The inclusion of a spacer region in the peptide rather than the cloning sequence allows that to be changed and optimized as easily as the linking chemistry.

One previously reported method for bispecific generation also uses sortase A and azide-alkyne click chemistry [252]. Wagner *et al.* also took advantage of the incredible flexibility of the sortase

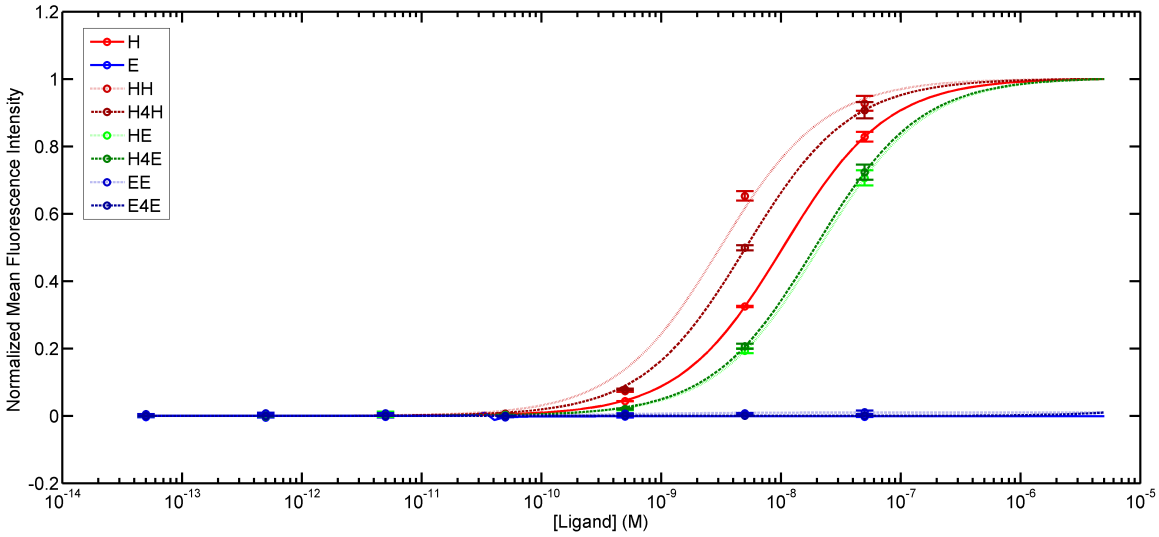


Figure 4.3: T6-17/Affibody Binding Curves. Mean fluorescence intensity as determined by flow cytometry was fit to a ligand binding model. Modeled fractional saturation (lines) were superimposed onto the measured mean intensities (circles) for affibodies that showed affinity for the T6-17 cell line in the measured range. Monospecific Z_{EGFR} ligands showed no binding to T6-17 cells.

enzyme to create a modular system. Although this is very similar to what is described in this study, STEPL still retains all of the advantages over traditional sortagging that we have previously described [236]. Specifically, linking protein purification to sortase activity leads to purer samples, less peptide waste, and simpler downstream processing.

4.3 Determination of Apparent Affibody K_Ds

To quantify the differences between the six dimeric and two monomeric affibodies that were generated, the ligands were applied to the Her2/neu+ cell line T6-17, the EGFR+ cell line KB, and the Her2/neu+/EGFR+ cell line SK-BR-3. 5×10^4 cells were incubated with affibody concentrations ranging from 100 nM to 100 fM before being assayed for cell binding by flow cytometry. The mean fluorescence intensity fluorescence intensity for each sample was then fit to a ligand-binding curve to determine the apparent dissociation constant (K_D) for each affibody-cell line pair. These curves are detailed in Figures 4.3, 4.4, and 4.5. As expected, the Z_{EGFR} homodimers and monomer showed no affinity for the EGFR-negative T6-17 cells, while Z_{Her2} homodimers and monomer showed no affinity for the Her2/neu-negative KB cell line. All of the affibodies bound to the double-positive

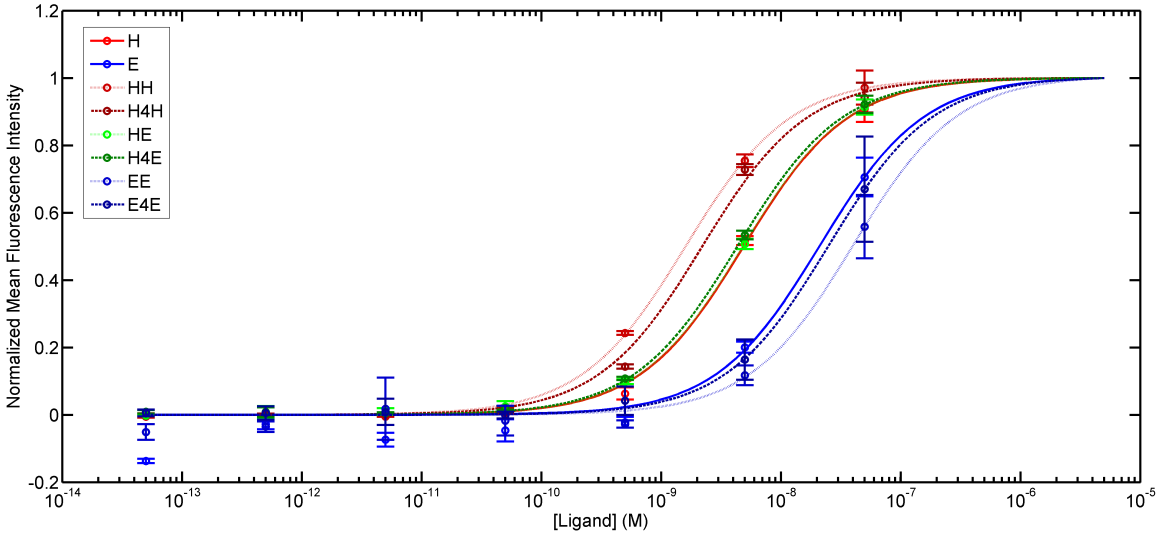


Figure 4.4: SK-BR-3/Affibody Binding Curves. Mean fluorescence intensity as determined by flow cytometry was fit to a ligand binding model. Modeled fractional saturation (lines) were superimposed onto the measured mean intensities (circles) for affibodies that showed affinity for the SK-BR-3 cell line in the measured range, which was all eight.

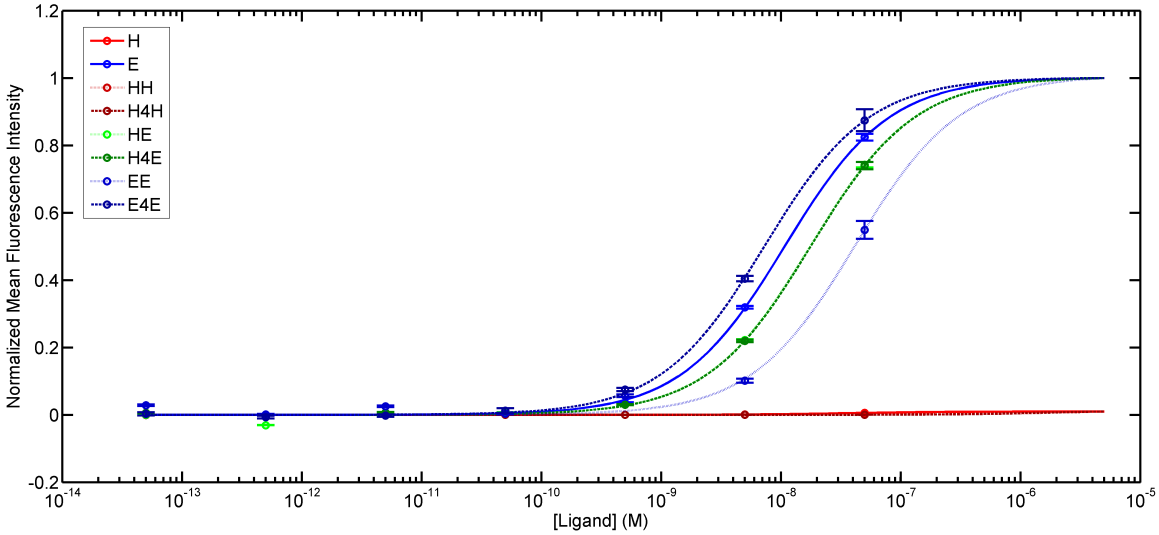


Figure 4.5: KB/Affibody Binding Curves. Mean fluorescence intensity as determined by flow cytometry was fit to a ligand binding model. Modeled fractional saturation (lines) were superimposed onto the measured mean intensities (circles) for affibodies that showed affinity for the KB cell line in the measured range. Monospecific Z_{Her2} ligands showed no binding to KB cells.

Table 4.3: Ligand Binding Parameters. Values and 95% confidence intervals are listed for the ligand K_Ds and receptor copy numbers per cell as determined by the mathematical model.

Ligand K_Ds (nM)	T6-17	SK-BR-3	KB
H	10.45 ± 0.31	4.89 ± 0.94	
E		21.08 ± 43.21	10.18 ± 0.72
HH	3.12 ± 1.41	1.60 ± 0.14	
H4H	5.11 ± 0.37	2.21 ± 0.76	
HE	21.07 ± 1.84	4.82 ± 0.57	17.59 ± 5.63
H4E	19.41 ± 1.19	4.34 ± 0.43	17.83 ± 1.76
EE		40.15 ± 73.91	41.92 ± 61.21
E4E		25.03 ± 8.92	7.29 ± 0.70
Receptor Copy Number			
Her2/neu	7,223,620 ± 69,665	3,081,328 ± 2,304,491	4,994 ± 81,526
EGFR	0 ± 2,958	316,433 ± 359,378	1,806,512 ± 290,473

SK-BR-3 cells to some extent. The Z_{Her2}-derived affibody constructs showed tighter binding than did the Z_{EGFR}-derived constructs, consistent with previous measurements of the monomeric affibodies' affinities [112]. Values of the fitted parameters are listed in Table 4.3. The model did a good job fitting the Z_{Her2}-containing ligands, it did less well at determining the parameters involving EGFR. This is most likely due to a combination of EGFR being expressed in lower amounts than Her2/neu and Z_{EGFR} having a lower molecular affinity for its target than Z_{Her2}.

The ligand binding experiments clearly illustrate the utility in being able to quickly screen dimers with different linker properties. For the Z_{Her2} homodimers, the $(PEG)_4$-containing linker reduces affinity of the ligand by around 50% for both the T6-17 and SK-BR-3 cell lines. The opposite is true for the Z_{EGFR} homodimers, where the addition of the $(PEG)_4$ group doubles the affinity as compared to the E4E protein. And the $(PEG)_4$ group makes no difference for the heterodimers. These conflicting results are not easily predicted, so it is important to be able to easily and modularly incorporate different linker regions so that the best design can be determined empirically.

4.4 Increased Fluorescence Specificity

An enhancement index was determined for each of the affibodies to quantify its selectivity for the double-positive cell line. The enhancement index for a given ligand was defined as the ratio of apparent dissociation constants for that ligand with the double-positive line to the ligand with a single-positive cell line multiplied by the ratio of the available receptors on the double-positive line to

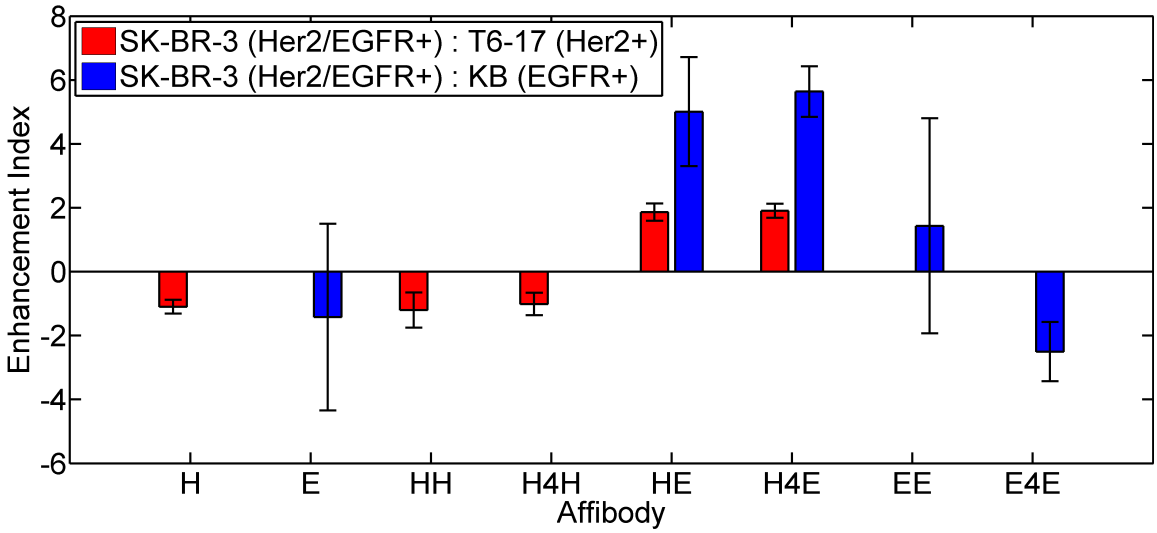


Figure 4.6: Affibody Enhancement Indices. Ratios of the apparent dissociation constants and 95% confidence intervals for each affibody between the double-positive SK-BR-3 cell line and the single-positive T6-17 and KB lines. Positive values indicate stronger binding to the double-positive line.

the available receptors on the single-positive line. Positive enhancement indices indicate a selective advantage for the double-positive cells and negative values mean tighter binding to the single-positive line. The calculated values, presented in Figure 4.6, showed that monomers and homodimers are more selective for the single-positive lines or are statistically indifferent. This is likely because of the higher receptor expression on the T6-17 and KB cell lines. The heterodimers, however, showed selectivity for the SK-BR-3 double-positive cell line. Selectivity for SK-BR-3 cells over KB cells is higher than selectivity for T6-17 cells, probably driven by the higher affinity of the Z_{Her2} component.

Before obtaining images of the cells, the optimal ligand concentrations to enhance selectivity were determined by subtracting the fractional saturation curve of a ligand against the signal-positive cell lines from the fractional saturation curve of the ligand against double-positive cell lines. Shown in Figure 4.7, the concentrations that show the maximum difference vary between affibodies, but are generally near 1×10^{-8} molar. Thus, all of the cell lines were incubated with 10 nM ligand and the images in Figures 4.8, 4.9, and 4.10 were obtained. In agreement with the flow cytometry data, both monomers labeled the single-positive lines more than the double-positive line. Homodimers did not obviously distinguish between single- or double-positive lines, with the exception of the E4E construct that clearly selected for the KB cells, as the enhancement index predicted. However, the bispecific ligands were able to specifically label the SK-BR-3 double-positive cells to a greater extent

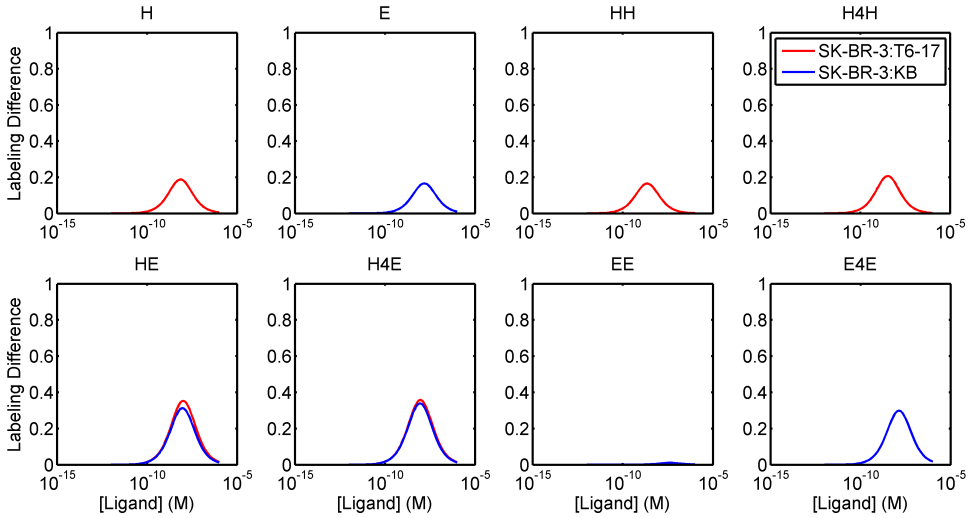


Figure 4.7: Differential Ligand Binding. The difference was calculated between fractional saturation curves between the single- and double-positive cell lines. Maximum values denote the optimal ligand concentration to select for one cell line over the other.

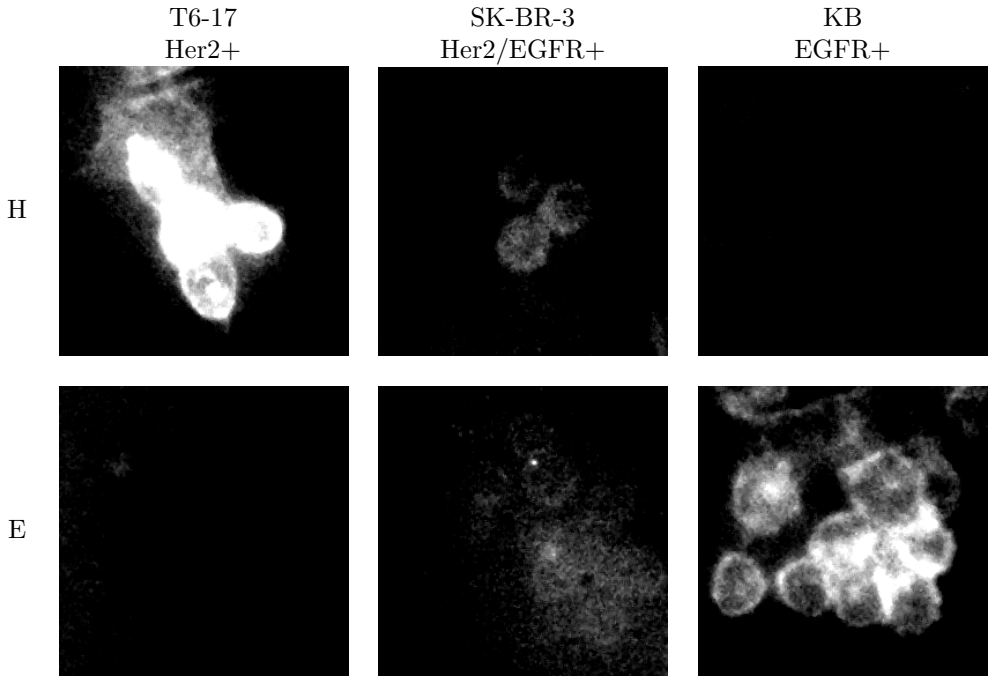


Figure 4.8: Bispecific Fluorescence Imaging: Monomers. Fluorescence images of cell lines treated with monomeric affibodies. Z_{Her2} labels both Her2/neu positive cell lines, while Z_{EGFR} labels both EGFR positive cell lines.

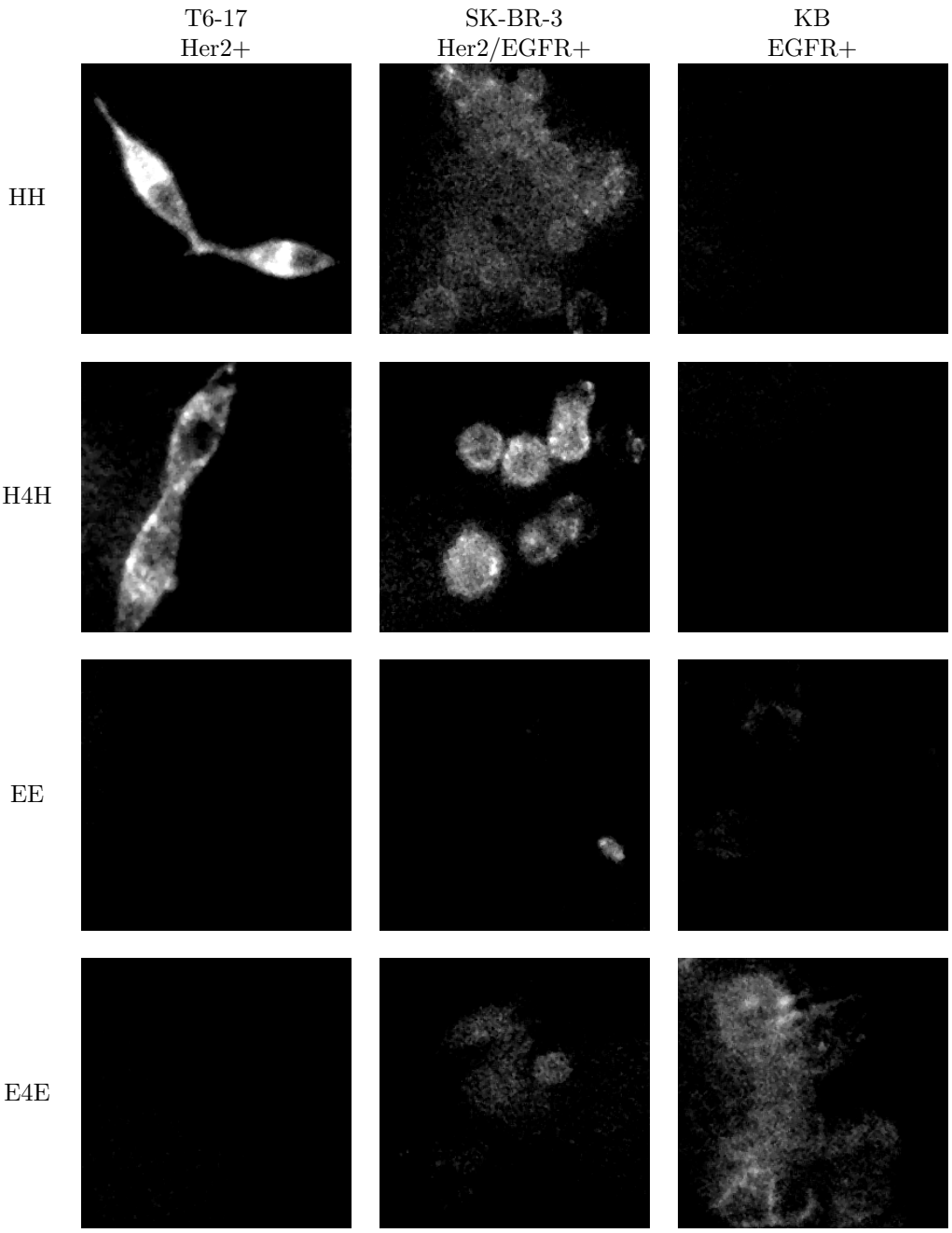


Figure 4.9: Bispecific Fluorescence Imaging: Homodimers. Fluorescence images of cell lines treated with homodimeric affibodies. Z_{Her2} dimers label both Her2/neu positive cell lines, while Z_{EGFR} dimers label both EGFR positive cell lines.

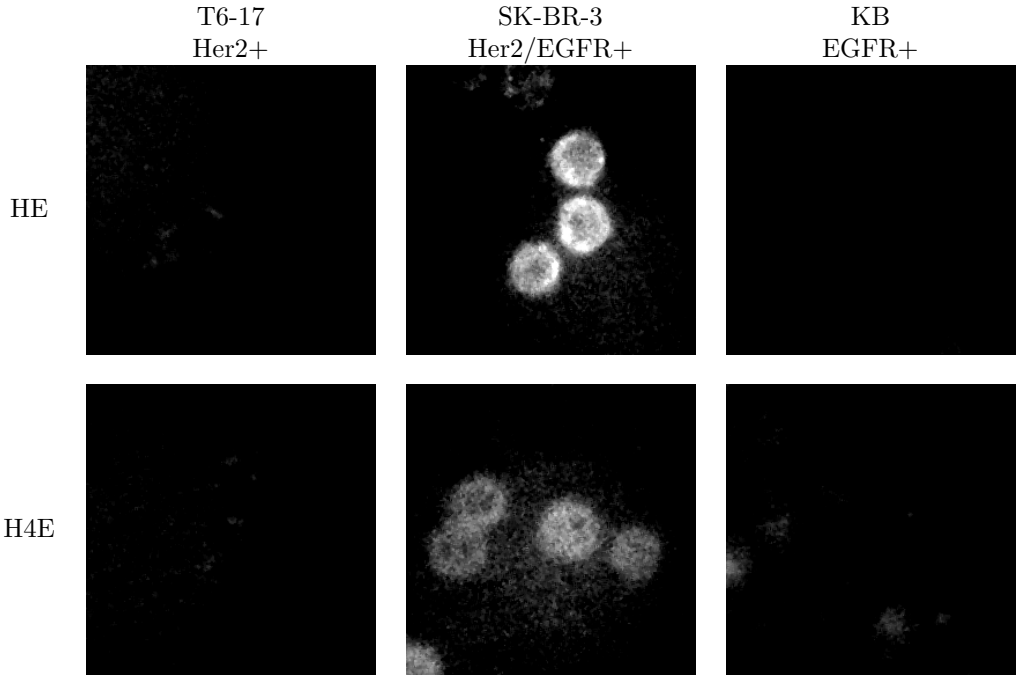


Figure 4.10: Bispecific Fluorescence Imaging: Heterodimers. Fluorescence images of cell lines treated with heterodimeric affibodies. In both cases, the bispecific affibody labels the double-positive line to a greater extent than the single-positive lines.

than the T6-17 Her2/neu-positive cell and the KB EGFR-positive cells. These images validate the hypothesis that the bispecific ligands will provide enhanced selectivity for a cell line presenting both ligands. There may also be a synergistic effect to drive this selectivity despite a lower receptor count on the SK-BR-3 cells.

4.5 Conclusions

STEPL is a facile and modular system for generating bispecific targeting ligands. Using four peptides and two affibody-STEPL chimeras, a panel of eight homo- and heterodimeric affibodies were created in a matter of days. The affinities of these ligands were determined by flow cytometry and used to calculate the optimal ligand concentration to specifically label a double-positive cell line. This simple example demonstrates how the inherent flexibility of the STEPL system to conjugate a wide array of targeting ligands to any synthetic peptide with an N-terminal glycine allows this one system to create ligand dimers with differing specificities, spacers, and multiple chemical functionalities.

4.6 Materials & Methods

Peptide Labeling

Peptide P 4.1 (synthesized by Anaspec, Inc., 6 mg/mL in dimethylformamide (DMF)) was incubated with a 2-fold excess of tris(2-carboxyethyl)phosphine (TCEP) (Sigma, 0.5M) and a 4-fold excess of either dibenzocyclooctyne−maleimide (DIBO-maleimide, Sigma, 5 mg/mL in DMF) or dibenzocyclooctyne−$(PEG)_4$−maleimide (Sigma, 5 mg/mL in DMF) for 4 hours at room temperature. 10 volumes of methyl tert-butyl ether (Sigma) was then added to precipitate the peptide, which was pelleted and resuspended in 0.1M Tris Buffer, pH 8 (Mediatech). DIBO-linked peptide was purified by reverse-phase high performance liquid chromatography (RP-HPLC) (Agilent) and resuspended in STEPL buffer (20mM Tris-base, 50mM NaCl, pH 7.5).

Protein Expression, Cleavage, & Bioconjugation

Z_{Her2} and Z_{EGFR} sequences were cloned into the pSTEPL vector (see Section 3.9). 25mL starter cultures were grown overnight in LB-Ampicillin. These were added to 225mL of LB and grown

to an OD_{600} of 0.8-1 before induction with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were allowed to express for 24hrs at 25°C. Cells were then harvested by centrifugation (3000 RPM, 15min) and resuspended in 10mL of lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 1mg/mL Lysozyme, 1 EDTA-free cOmplete Mini protease inhibitor tablet (Roche), pH 7.5). Lysates were incubated at room temperature for 30min under gentle agitation before freezing overnight at -80°C. Samples were then thawed and incubated for 30min with DNAse I (Roche) under gentle agitation. Lysates were then sonicated and separated by centrifugation (16,000g, 30min).

8mL of clarified lysates was incubated for 1hr with 0.5mL Talon resin (equilibrated in lysis buffer, Clontech). The lysate and beads were added to a column (BioRad) and resin was washed with 5mL STEPL buffer. 600μL of STEPL buffer containing 200μM synthetic peptide and 100μM $CaCl_2$ was flowed over the beads until it replaced the wash buffer. Columns were protected from light and reacted for 6hrs at 37°C. 1mL of STEPL buffer was added to the column and the flowthrough collected. Conjugated affibody was purified from excess peptide by RP-HPLC and resuspended in 0.1M Tris-HCl, pH 8.

Dimerization

For the dimerization reactions, equimolar azide-conjugated and DIBO-conjugated affibody were incubated together for 24 hours. Reactions were then purified by RP-HPLC and resuspended in 0.1M Tris-HCl, pH 8. Dimerization was confirmed by SDS-PAGE (4-12% Gel, Life Technologies). After PAGE, fluorescent bands were imaged by UV transillumination. Then proteins were visualized by incubating the gel with SimplyBlue SafeStain (Life Technologies) for 1 hour, followed with an overnight wash in deionized water.

Cell Culture

KB (ATCC) and T6-17 cell lines (NIH/3T3 (ATCC) cells engineered to stably express Her2/neu; kindly provided by Dr. Mark Greene, University of Pennsylvania) were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin/ streptomycin at 37°C and 5% CO_2. SK-BR-3 cells (ATCC) were maintained in McCoy's medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin at 37°C and 5% CO_2.

Flow Cytometry

Cells were lifted off their plates with enzyme-free dissociation buffer (PBS base, Millipore) and concentrated to 5×10^5/mL. Serial dilutions of affibody (100 nM - 100 fM, 10-fold dilutions) were incubated with 5×10^4 cells in full media for 30 min at 37°C while shaking to maintain reaction homogeneity. Cells were then held on ice until samples were analyzed on a BD LSR II flow cytometer (5,000 counts/sample, 532 nm laser, 575/26 nm filter).

Computational Analysis

The mean fluorescence intensities (MFIs) for each cell line/affibody pair were simultaneously fit to determine that affibody's apparent dissociation constant ($K_{D,app}$) and receptor copy number (N_R) for each cell line using Equation 4.1 in MATLAB. Because the affinities were weaker than expected for all of the ligands, the data could not be fit to the standard ligand binding model of $[L]/(K_D+[L])$, which requires multiple fully saturated points to fit a reliable dissociation constant. Therefore, data was fit to a more complicated model that does not rely on the excess ligand assumption:

$$MFI = F_0 + (F_{max} - F_0) * \frac{[\mathrm{A}] + [\mathrm{R_i}] + K_{D,app} - \sqrt{([\mathrm{A}] + [\mathrm{R_i}] + K_{D,app})^2 - 4\,[\mathrm{A}]\,[\mathrm{R_i}]}}{2\,[\mathrm{R_T}]} \tag{4.1}$$

$$[\mathrm{R}] = \frac{N_R * N_C}{N_A * V} \tag{4.2}$$

where F_{max} and F_0 scale the fractional saturation curve to the fluorescence data, [A] is the affibody concentration, [$\mathrm{R_i}$] is the concentration of surface receptor that binds to the affibody being fit (Her2/neu, EGFR, or both), [$\mathrm{R_T}$] is the sum of Her2/neu and EGFR ligand concentrations, N_C is the number of cells per well, N_A is Avogadro's number, and V is the volume of media in the well. Parameter confidence intervals were also determined by MATLAB. For plotting, all curves were normalized such that the predicted values spanned the range of [0,1].

The enhancement index (EI) was calculated by dividing the single-positive dissociation constant by the double-positive dissociation constant for each affibody (ER, Equation 4.3) and submitting it to the transformation in Equation 4.4, which ensures the magnitude of EI is greater than 1 while

changing the sign if the ligand is more selective for the single-positive cells.

$$ER = \frac{K_{D,app,\text{T6-17 or KB}}}{K_{D,app,\text{SK-BR-3}}} * \frac{N_{R,\text{SK-BR-3}}}{N_{R,\text{T6-17 or KB}}} \tag{4.3}$$

$$EI = \exp\left(|\ln ER|\right) * \frac{\ln ER}{|\ln ER|} \tag{4.4}$$

The optimal concentration of bispecific affibodies was then determined by subtracting the modeled fractional saturation curve for T6-17 or KB cells from the curve for SK-BR-3 cells for each of the affibodies.

Fluorescence Imaging

All cell lines were incubated with the optimal concentration of each bispecific affibody for 1 hour in full media. Cells were washed 3 times with affibody-free media before being imaged in serum-free DMEM. Imaging was performed with an Olympus IX81 inverted fluorescent microscope with a back-illuminated EMCCD camera (Andor) and a SOLA excitation source (Lumencor). Images of 5-FAM and TAMRA fluorescence were acquired using the filter sets (HQ545/30, HQ620/60, Q570lp) and (HQ710/75, HQ810/90, Q750LP) (Chroma) after focusing with white light. A LUC PLAN FLN 40X objective (NA 0.6) was used for all imaging studies. All images were exposed for 2 seconds and are displayed with the same linear contrast curve.

Chapter 5

Enhancing STEPL Yield through Directed Evolution

Abstract

Sortase-tag expressed protein ligation (STEPL), a robust and flexible system for protein expression and bioconjugation, is plagued by a reduction in protein yield due to unwanted sortase activity during protein expression. This study set out to evolve the STEPL system to reduce *in situ* cleavage by decreasing the calcium sensitivity of the sortase enzyme or by changing the trigger ion to manganese, a less-abundant, biocompatible, and industry-friendly divalent cation. Bacterial, yeast, and bacteriophage evolution systems were explored to produce an assay for sortase activity that contained a genotype-phenotype linkage. Phage display libraries were produced by randomizing the amino acids in the two loops known to bind the native calcium ion with directed, optimized degenerate codons. A number of screens and screening strategies were employed. However, none of the screens were able to produce a viable candidate for an improved STEPL system.

5.1 Introduction

In Chapter 3, it was described in detail how well the STEPL system works once the chimeric protein is in a controlled buffer setting. However, it was only briefly mentioned that STEPL can be unstable when in culture media during expression, due to inherent calcium in the media. In fact, the initial expression studies show that much of the STEPL protein is cleaved before ever reaching the column (Figure 5.1). While this early cleavage does not effect the final product's purity, it does greatly reduce the overall yield.

We postulated that the early cleavage was due to calcium ions present in the growth media and within the cells themselves [243]. This problem could be fixed in one of three ways. First, calcium could be eliminated from the growth media by using M63 minimal media produced in reverse-osmosis water [254]. However, protein expression is difficult in minimal media, requiring a number of potential additives to finally produce a "witch's brew" STEPL expression media. This route could potentially work in *E. coli*, which do not require calcium to grow, but it restricts STEPL's use in eukaryotes and even other prokaryotes. A second strategy would be to alter sortase such that it needs an additional small molecule to be activated. Finally, the sortase catalytic domain could be evolved to alter its relationship with calcium in one of the following ways:

Eliminate Calcium Ion Sensitivity By eliminating the need for calcium entirely, we would also abolish the calcium-driven background cleavage. Instead, sortase activity would be triggered solely by the presence of a triglycine peptide. Thus, no cleavage would occur until the peptide was added to the column. This would greatly increase purity as well.

Weaken Calcium Binding If it were harder for sortase to bind calcium, more would be needed to trigger cleavage. Raising the necessary calcium concentration to well above what would be present in a cell would also avoid cleavage until it is desired on the column.

Alter the Trigger Ion While calcium is generally present at high concentration in biological systems, many other divalent cations are not. By altering sortase to bind one of those instead of calcium, it would be likely to avoid unwanted cleavage due to a trace element's presence.

The evolution option is the most promising because it is the smallest change to the STEPL system that could yield the desired outcome.

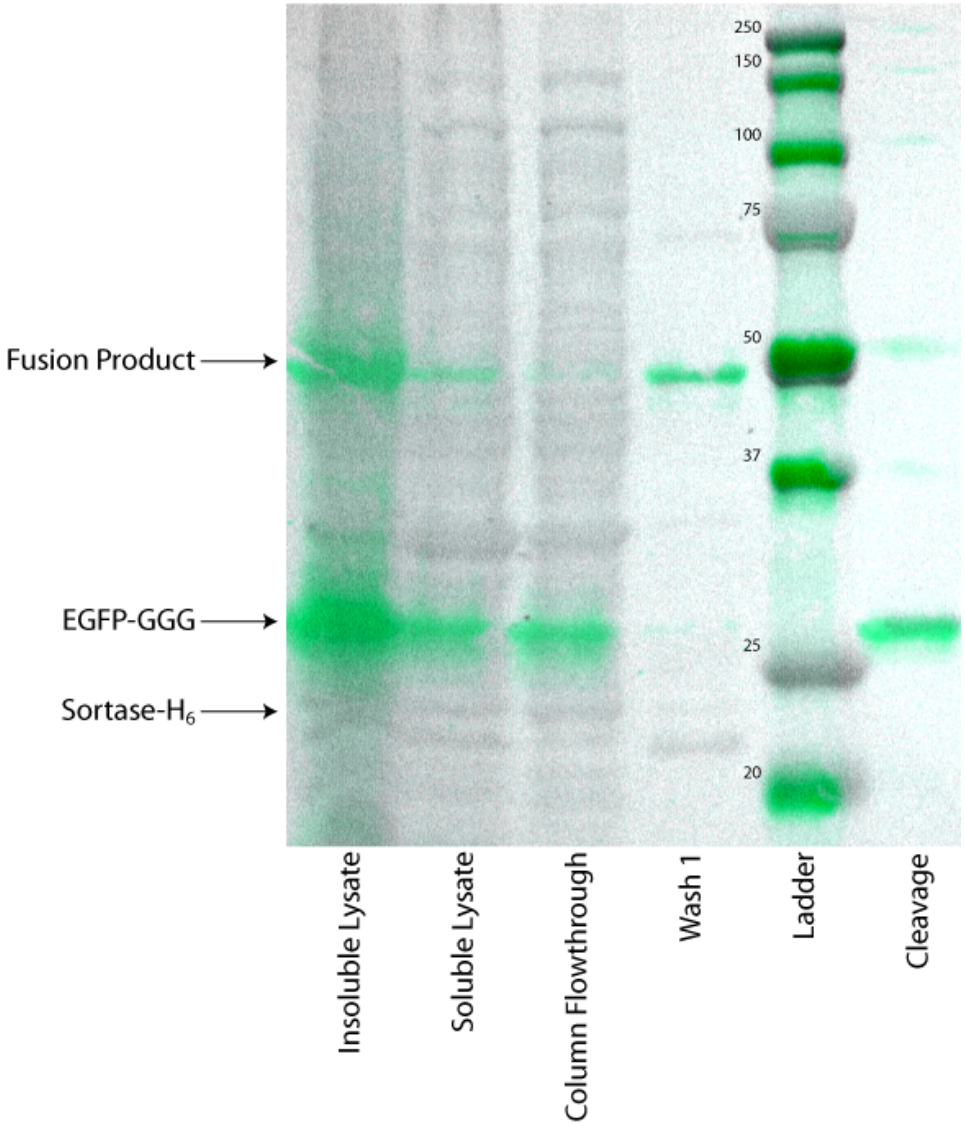


Figure 5.1: Early Cleavage Expression Fractions. A large amount of cleavage product can be seen in the lysate lanes (i.e. EGFP and Sa-SrtAΔ59-His_6), indicating that most of the STEPL chimera is cleaved during expression and therefore cannot be conjugated or recovered.

In this chapter, we will evaluate the feasibility of making these alterations by exploring the evolution systems and library designs that could be used to reduce the unwanted background cleavage during STEPL expression.

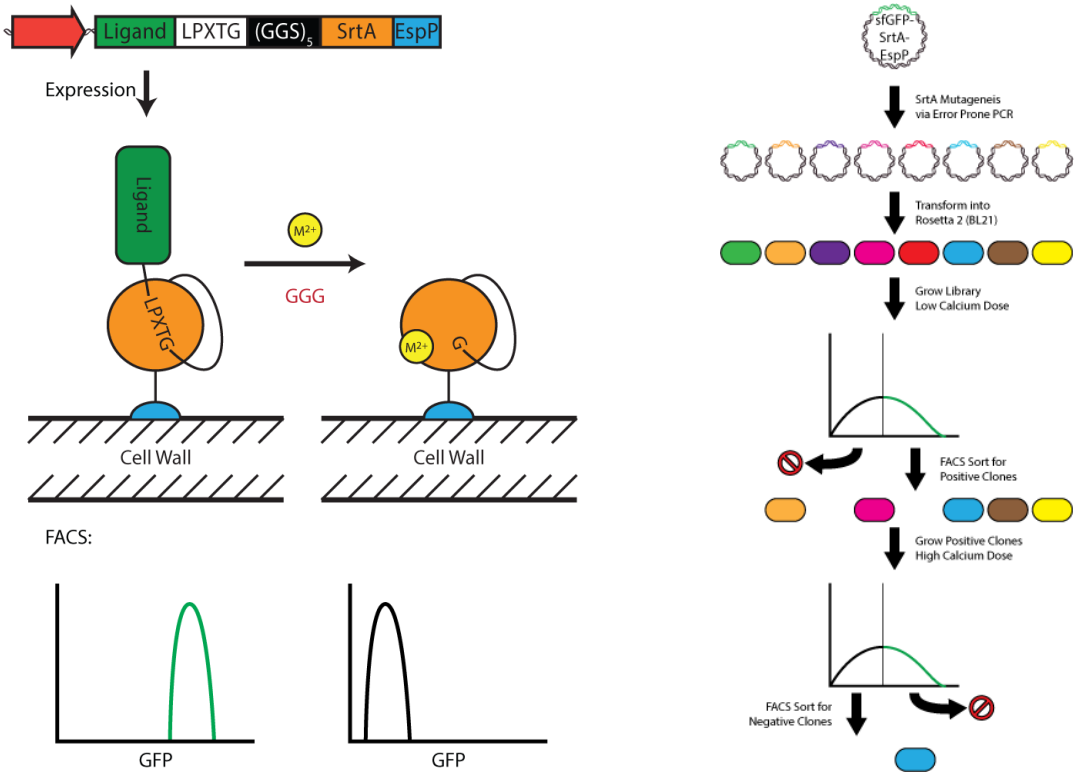
5.2 Selection of Evolution System

The first consideration when evaluating a directed evolution system is the choice of genotype-phenotype linkage, a number of which are described in Section 2.4.2. Background cleavage is occurring within the cell, so a living display technology is ideal because it will most closely replicate those conditions. In addition, a cellular screen will also select for constructs with better expression profiles. This is normally a downfall of display screens, but because STEPL is an expression system, it is a desirable trait in this case. Therefore, *E. coli*, yeast, and phage display systems were considered as potential screens.

5.2.1 *E. coli* Display

E. coli display seemed like the best option to evolve the STEPL system for many reasons. Sortase naturally occurs on the surface of gram-positive bacterial cells, so it is likely that STEPL should work extracellularly as well. In addition, STEPL has already been thoroughly proven in a bacterial system. Also, the entire problem of background cleavage during expression was defined in the context of *E. coli*. It may not exist or have a different cause in other systems. For all these reasons, *E. coli* display was chosen as the first evolution system to be evaluated.

The screening strategy for an *E. coli* display system is depicted in Figure 5.2. In the proposed system, *E. coli* is transformed with a membrane-anchored STEPL system that has a fluorescent ligand. The cells are grown to a high density, pelleted, and resuspended in a negative screen buffer. The negative buffer contains a calcium level deemed to be tolerable and no triglycine. Sequences with background cleavage, such as wild-type, will cleave while incubating in this buffer and lose their fluorescent ligand. The cells are then sorted by fluorescence-activated cell sorting (FACS), which is able to sort individual cells into different tubes based on the cell's fluorescent signal. Cells with high signal will pass the negative screen and be regrown to density. These cells are pelleted and resuspended in a positive reaction buffer, which contains the desired metal ion and Gly_3. Mutants



(A) Molecular Mechanism (B) Screen Design

Figure 5.2: Bacterial Display Scheme. A) STEPL can be adapted for bacterial display by using a fluorescent ligand and a C-terminal membrane anchor. This will give a high fluorescent signal if the construct has not cleaved and a low signal if it has cleaved. B) A library of Sa-SrtAΔ59 mutants is cloned into the surface display system and expressed in *E. coli* Mutants are then exposed to a negative screen and sorted by FACS. Highly fluorescent cells will pass the screen, be amplified, and be incubated in the desired reaction conditions. FACS sorting for non-fluorescent cells will provide the desired genotypes.

that were unable to cleave in the previous buffer but can cleave in the positive buffer now lose their fluorescent ligand. After incubation, these cells are also sorted by FACS, but this time, the cells with low signal will be chosen. Surviving clones are randomly chosen for sequencing and biochemical analysis.

There are a few strategies for trafficking proteins to the outer membrane of *E. coli* (see Figure 2.13), but only one that allows the C-terminal anchor needed for STEPL, the autotransporter. Autotransporters are a class of structural proteins in gram-negative bacteria that are expressed in the periplasm. The C-terminal domain is a transmembrane pore that actively translocates the N-terminal domain out of the periplasm into extracellular space. Due to the inherent modularity of autotransporters and their ability to translocate large globular domains across the outer membrane, they are the display strategy of choice for enzymatic screens [208]. Therefore, pSED/sfGFP (Plasmid 5.1) was created by cloning in frame an N-terminal periplasmic signal sequence, super-folder GFP (sfGFP), the sortase motif LPETG, $(GGS)_5$, Sa-SrtAΔ59, and the autotransporter EspP. Super-folder GFP was used because EGFP is known to export poorly from *E. coli* and sfGFP was designed to survive the *E. coli* secretion pathways [255]. The autotransporter EspP was chosen because it performed well in other studies [208, 209].

To validate pSED/sfGFP, the construct was expressed in *E. coli* overnight. Cells were then pelleted and resuspended into STEPL buffer, where the cleavage reaction could be monitored. Since sortase concentration was a critical parameter while initially characterizing the STEPL system, 1 mL aliquots of the cells were pelleted and resuspened in 0.1 to 1 mL of STEPL buffer to create different resuspension densities and therefore different sortase concentrations. As shown in Table 5.1, even after incubating the reactions at 37°C overnight with high concentrations of $CaCl_2$ and Gly_3 (5 mM each), no sfGFP cleavage was observed.

Because sortase is known to work extracellularly, it was hypothesized that the STEPL construct was trapped in the bacterial cytoplasm or periplasm. To test this, cells expressing pSED/sfGFP were labeled and assayed for surface exposed GFP. Figure 5.3 shows that the cells expressing the surface displayed pSED/sfGFP were able to be stained by α-GFP antibodies as well as Co-NTA resin bound to EGFP. The cytosolically expressed pSTEPL/EGFP could not be stained by the antibody. Thus, the pSED construct must be correctly trafficked to the bacterial surface.

Sortase activity could also be impaired if the enzyme is physically too close to the autotransporter

Table 5.1: pSED/sfGFP Expression Preliminary Data. Different volumes of cells that expressed pSED/sfGFP were pelleted and resuspended in STPEL reaction buffers to determine sortase activity over time. sfGFP fluorescence was estimated by UV transillumination to be nonfluorescent (—), slightly fluorescent (—), fluorescent (++), or highly fluorescent (+++) as compared to cells expressing pSTEPL/EGFP.

			0hr		1hr		2hr		3hr		4hr		5hr		6hr		Overnight	
Sample	$[CaCl_2]$ (mM)	$[Gly_3]$ (mM)	Pellet	Super.	Pellet	Super	Pellet	Super	Pellet	Super	Pellet	Super	Pellet	Super	Pellet	Super	Pellet	Super
1 mL	0	0	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
	5	5	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
0.5 mL	0	0	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
	5	5	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
0.25 mL	0	0	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
	5	5	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
0.1 mL	0	0	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—
	5	5	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—	+++	—

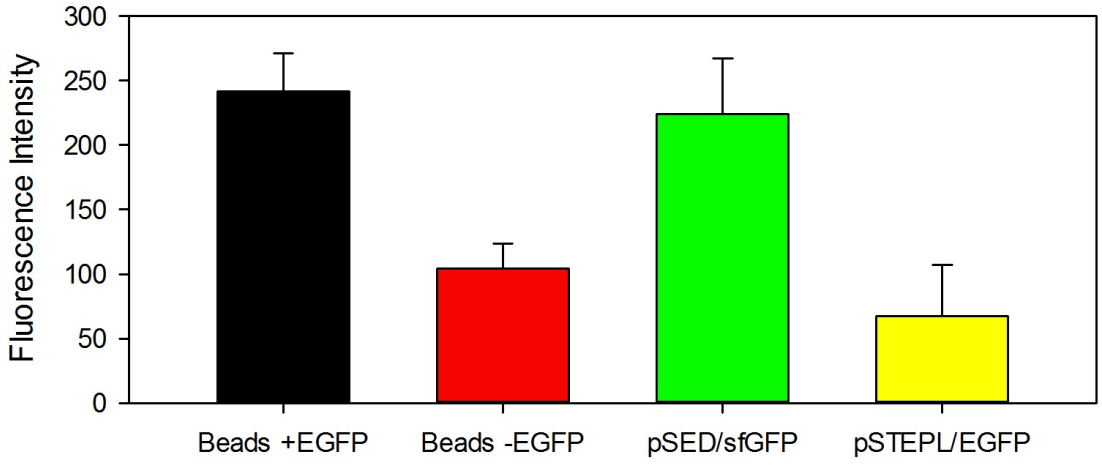


Figure 5.3: GFP Surface Expression. *E. coli* expressing pSED/sfGFP were stained with an α-GFP antibody, indicating surface expression of sfGFP. Cells expressing the internal construct pSTEPL/EGFP as well as Co-NTA beads with and without bound EGFP were used as staining controls.

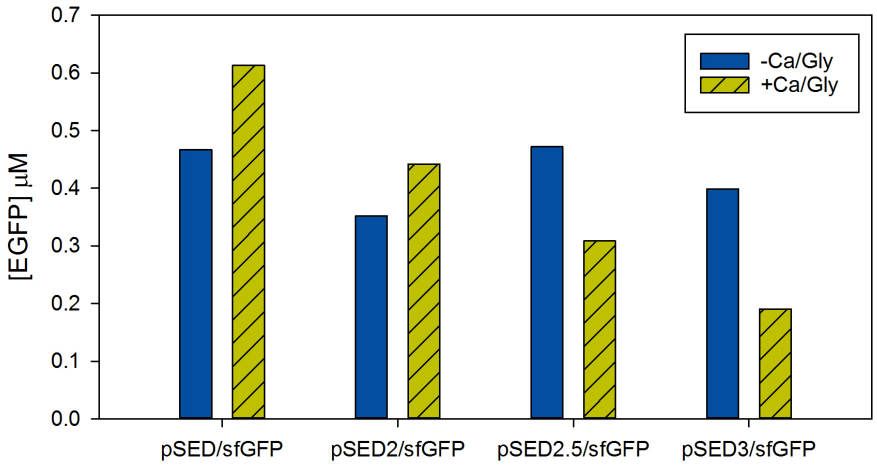


Figure 5.4: Bacterial Display Cleavage. All three generations of pSED constructs were assayed for EGFP release into the supernatant. None show enough cleavage to be used in a directed evolution system.

to refold or find its LPXTG motif. If this is the case, adding a linker between the autotransporter and sortase should allow the enzyme to regain activity. With that in mind, pSED2/sfGFP (Plasmid 5.2) and pSED2.5/sfGFP (Plasmid 5.3) were cloned to insert $(GGS)_5$ and ECFP between Sa-SrtAΔ59 and EspP. The new clones were expressed in *E. coli*, resuspended in STEPL buffer with and without 5mm $CaCl_2$ and Gly_3, and allowed to react for six hours at 37°C. Unfortunately, neither provided a working STEPL display system, as shown in Figure 5.4. Thus, spacing from the bacterial surface was not likely to be the issue.

If sortase is active in the extracellular context naturally, being expressed on the surface, and sufficiently spaced from cell but is still inactive in the display context, it may not be refolding properly after transport through the EspP pore complex. A potential fix for this problem is to use another autotransporter. Therefore, pSED3/sfGFP (Plasmid 5.4) was cloned by replacing EspP with EstA, the second best performing system as per Binder *et al.* [209]. This construct was also expressed in *E. coli*, resuspended in STEPL buffer with and without 5mm $CaCl_2$ and Gly_3, and allowed to react for six hours at 37°C. It still showed no activity (Figure 5.4). Thus, it was presumed that the STEPL system was incompatible with *E. coli* display.

5.2.2 Yeast Display

The second system that was considered was yeast display. The yeast display screen is conceptually identical to the bacterial display screen. The major difference is that the STEPL chimera would be trafficked to the yeast secretory pathway co-translationally, rather than the post-translational translocation through the *E. coli* autotransporter. This is much easier on the enzyme, as evidenced by the prior use of sortase within eukaryotic secretory pathways [256, 257]. However, Dr. Colin Greineder, a collaborator working to adapt STEPL to yeast systems, discovered that Sa-SrtAΔ59 is heavily glycosylated in the secretory pathways of both *P. pastoris* (Figure 5.5) and *S. cerevisiae* (data not shown). These glycogens have the potential to influence the evolution process by interacting with the sortase in ways that would not occur within the cell. Therefore, yeast display was likely not a viable evolutionary technique in this case.

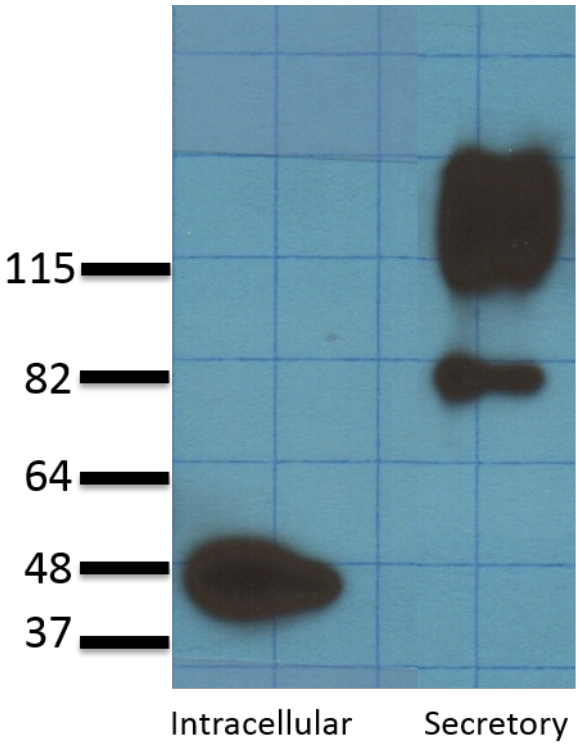


Figure 5.5: Sortase Glycosylation in Yeast. STEPL constructs were produced in *P. pastoris* expression vectors and targeted to the cytosol or the secretory pathway. Cultures were grown of each and analyzed via Western blot. When expressed intercellularly, the STEPL construct runs at the expected molecular weight of 46 kDa. The secreted STEPL construct runs much higher as a result of N-linked glycosylation. Figure courtesy of Dr. Colin Greineder.

5.2.3 Phage Display

Phage display is another viable option for evolving the STEPL system to acquire altered divalent cation sensitivity. Like bacterial display, the constructs would be expressed in the proven *E. coli* system. But Bacteriophage M13 is assembled in the periplasm, so the enzyme would only need to survive one translocation in phage display. This is expected to help mediate the problems that were associated with *E. coli* display.

Figure 5.6 shows a schematic of the phage display screen for STEPL. First, a library of Sa-SrtAΔ59 mutants fused to pIII is transformed into *E. coli*. Those cells are then amplified and infected with CM13 helper phage to induce phage production. STEPL-pIII fusions are randomly incorporated into phage as they are assembled and secreted. The phage are isolated and resuspended into the negative screening buffer. The STEPL "ligand" mediates phage attachment to an antigen coated plate. After incubating the phage in the negative buffer, the wells are washed to remove any phage that cleave during expression or the negative screen. The wells are then filled with positive screening buffer and allowed to incubate again. Supernatant containing phage that cleaved under the desired conditions is used to infect fresh *E. coli* cells, which are then amplified for additional screening and/or plated onto agar plates for clonal sequencing and analysis.

Two different phagemids were constructed to perform the above screen. Both encoded a fusion protein with the PelB periplasmic signal sequence, a ligand, LPETG sortase recognition motif, $(GGS)_5$ linker, Sa-SrtAΔ59, a flexible linker, and the C-terminal portion of M13 pIII. As the "ligand", the plasmid pComb3XM/SBP (Plasmid 5.6) contained the sequence for the streptavidin binding peptide (SBP) [258] and the other, pComb3XM/Avi (Plasmid 5.7), contained the sequence for the biotin acceptor peptide, also known as the Avi-Tag [157]. Both mediate specific binding to streptavidin (SAv), although with very different strengths. SBP has nanomolar affinity for SAv, whereas a biotinylated Avi-Tag has femtomolar affinity, approximately a million times tighter [213, 258]. The Avi-Tag construct is obviously preferred, but both were moved forward in parallel. To validate the assay, both phagemids were mutated to eliminate sortase activity by changing the active cysteine, C180, to an alanine residue.

All four constructs were transformed into *E. coli* cells, which were then infected with CM13 helper phage to induce phage production. The Avi-Tag phage were assembled in an *E. coli* strain

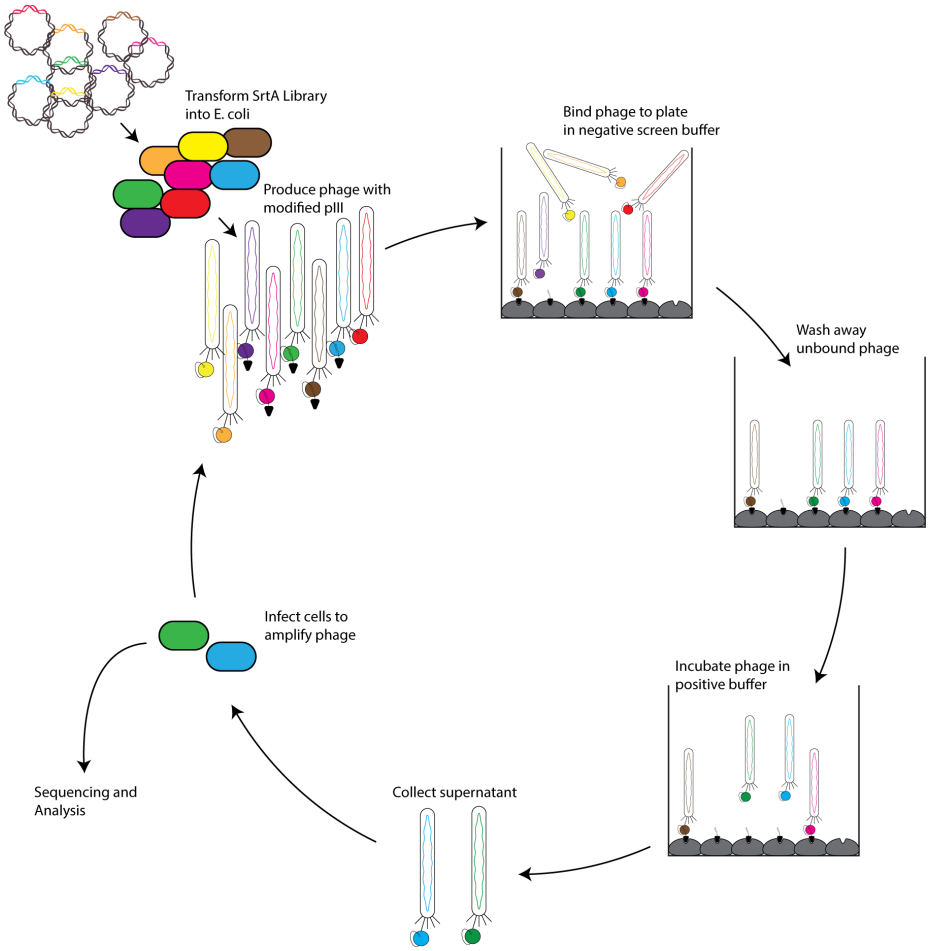


Figure 5.6: Phage Display Screening Cycle. Phagemid libraries were electroporated into *E. coli*, which produced modified phage displaying the modified pIII and a streptavidin-binding ligand. Phage were then bound to wells coated in streptavidin and incubated in the negative buffer. Phage with sortase mutants that cleaved during expression (**yellow**, **orange**, **red**) or during the negative screen (**purple**) were washed away. Next, the remaining phage were incubated in positive buffer. Phage that cleaved during the positive screen (**blue**, **green**) were collected, leaving phage that could not cleave (**brown**, **pink**) bound to the plate. That supernatant then infected *E. coli* cells that could be amplified for another round of phage display or plated for further analysis.

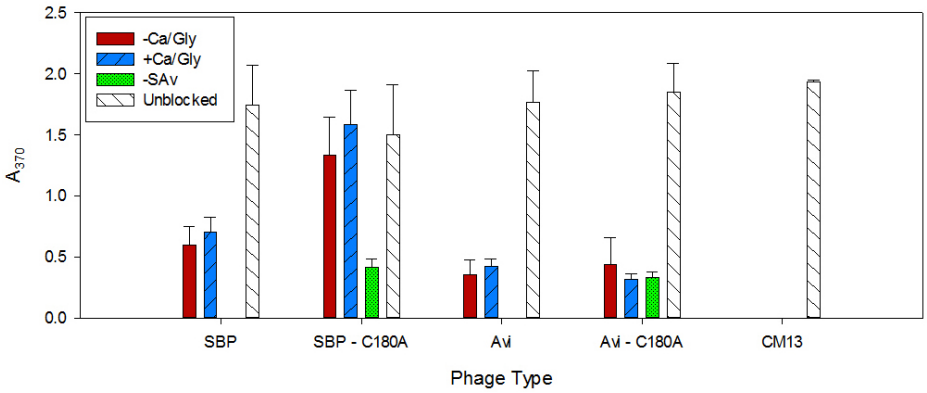


Figure 5.7: Phage Display Validation Assay. SBP and Avi-Tag containing STEPL-pIII phage were incubated in STEPL blocking buffer with and without calcium and triglycine. All phage were also added to unblocked wells as a positive control and inactivated C180A phage were added to blocked wells without streptavidin as a negative control. Stock CM13 phage bound to unblocked wells was a positive control for expression. Avi-Tag constructs did not bind the plate. SBP constructs bound if the sortase enzyme had been inactivated.

that overexpresses biotin ligase so that the STEPL-pIII fusion would contain a biotin conjugation. Isolated phage were split into two pools and incubated in STEPL blocking buffer (STEPL buffer with 2% w/v powdered, biotin-free milk) or STEPL blocking buffer supplemented with 100 μM $CaCl_2$ and 200 μM Gly_3. After four hours, the phage were applied to blocked wells that had been coated with SAv. As a positive control, phage in STEPL buffer without milk were added to unblocked wells, where they could freely adsorb to the surface. Blocked wells with no streptavidin provided negative controls. After washing the wells vigorously, phage remaining in the wells were quantified by phage ELISA. As shown in Figure 5.7, the Avi-Tag phage did not bind to any blocked wells, indicating a lack of biotin conjugation. The PelB signal sequence translocates proteins to the periplasm post-translationally, however this may occur too quickly for biotin ligase to label the Avi-Tag in the cytoplasm. The SBP phage showed robust binding when inactivated, but binding was significantly reduced in the active constructs both with and without calcium and triglycine. Most likely, this is due to the high levels of calcium in the powdered milk, nearly 4 mM in 2% milk [259]. This does not allow for fine control of calcium concentrations, so a number of other blocking buffers were analyzed to determine the best option (Figure 5.8). Milk performed significantly better than the other options. Since the purpose of the screen is to make STEPL insensitive to high levels of calcium, the calcium in the powdered milk is acceptable and was used in all subsequent assays.

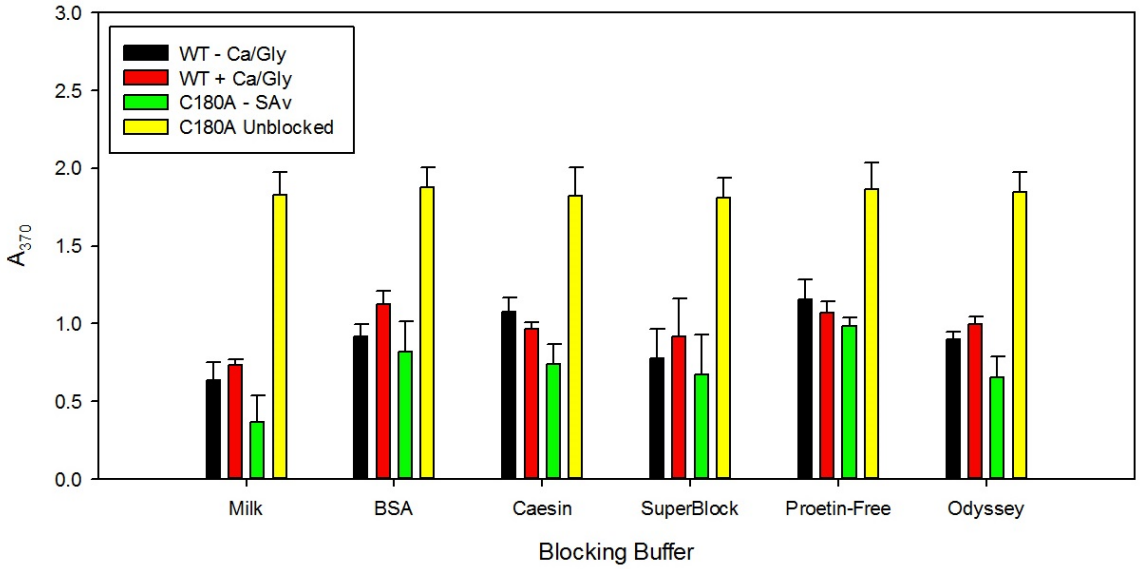


Figure 5.8: Blocking Buffers. Wild-type and C180A SBP-STEPL phage were incubated in various blocking buffers before being bound to streptavidin coated plates, washed, and assayed for binding. Biotin-free 2% milk showed less non-specific binding than any other buffer.

Table 5.2: Trace Elements Approved for Bolus Injection. Metal ions in this table are approved for bolus inection by the FDA. Maximum bolus injection doses are calculated for a 80kg patient with a 5 L blood volume.

Ion	Blood Concentration [260]	Bolus Concentration	Citation
Cr(III)	7 μM	100 μM	[261]
Cu(III)	16 μM	30 mM	[262]
Mn(II)	14 nM	19 mM	[263]
Zn(II)	16 μM	78 mM	[264]

Thus, pComb3XM/SBP can be screened for STEPL activity and served as the wild-type construct for phage display.

5.3 Selection of Metal Ion

One of the evolution options identified in Section 5.1 was changing STEPL's calcium sensitivity to a different metal ion. The ion must be a biocompatible trace element, which greatly reduces the potential field of elements. Those elements were further narrowed to chemicals approved for bolus injection by the U.S. Food and Drug Administration because they will be safe if any of the metal contaminates a STEPL-derived imaging agent and will enable future applications that involve

(A) NNK Library Structure (B) Optimized Library Structure

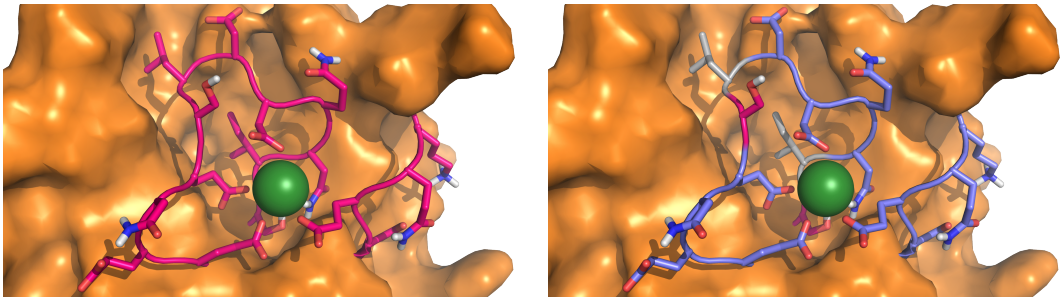


Figure 5.9: Sortase Calcium Binding Loop. Crystal structures of the Sa-SrtAΔ59 calcium binding loop are color coded to show the hydropathy index of each randomized amino acid in the NNK (A) and optimized (B) libraries. **Pink** NNK codons are slightly hydrophilic. **Blue** VRK codons are very hydrophilic. **Gray** codons are hydrophobic.

Table 5.3: Degenrate Codons

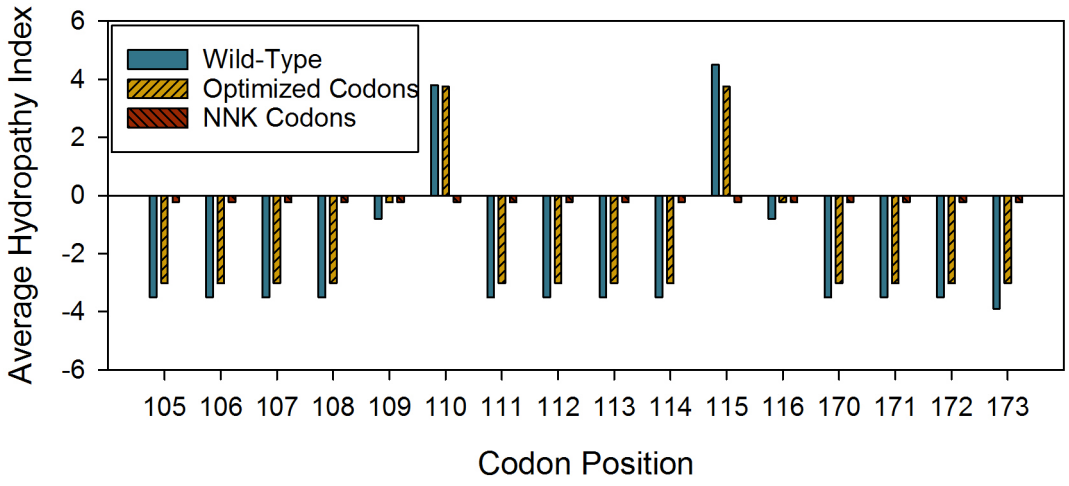
Codon	Hydropathy Index		Possible Amino Acids	Number of Codons
NNK	-0.22	20	ACDEFGHIKLMNPQRSTVWY*	32
VRK	-3.02	9	DEGHKNQRS	12
NTR	3.75	4	ILMV	8

STEPL as a cleaving agent *in vivo* (see Chapter 6). Those metals are described in Table 5.2. Of those metals, chromium was eliminated due to the low bolus dose. Copper was eliminated due to toxicity [141]. Zinc, while having the largest maximum dose, is heavily regulated and eliminated due to potential downstream manufacturing issues. Manganese had a reasonable bolus concentration, the lowest basal concentration in the blood, and is compatible with mass production. Therefore, Mn^{2+} was chosen to replace calcium as the STEPL trigger ion.

5.4 Library Design

Structural biology revealed that Sa-SrtAΔ59 binds its calcium ion through five specific amino acids in the $\beta3/\beta4$ and $\beta6/\beta7$ loops [265]. Therefore, the phage display libraries used in this study focus on the twenty amino acids that make up those loops, as shown in Figure 5.9. If all 20 positions were to be randomized with standard NNK codons, which code for all possible amino acids, the codon and amino acid spaces would be astronomical, 1.2×10^{24} and 1.4×10^{21}, respectively. The vast majority of these libraries will be non-functional because they will not code for valid loops, greatly decreasing the probability of finding the desired protein sequence. By optimizing the codons

(A) Codon Hydropathy



(B) Codon Diversity

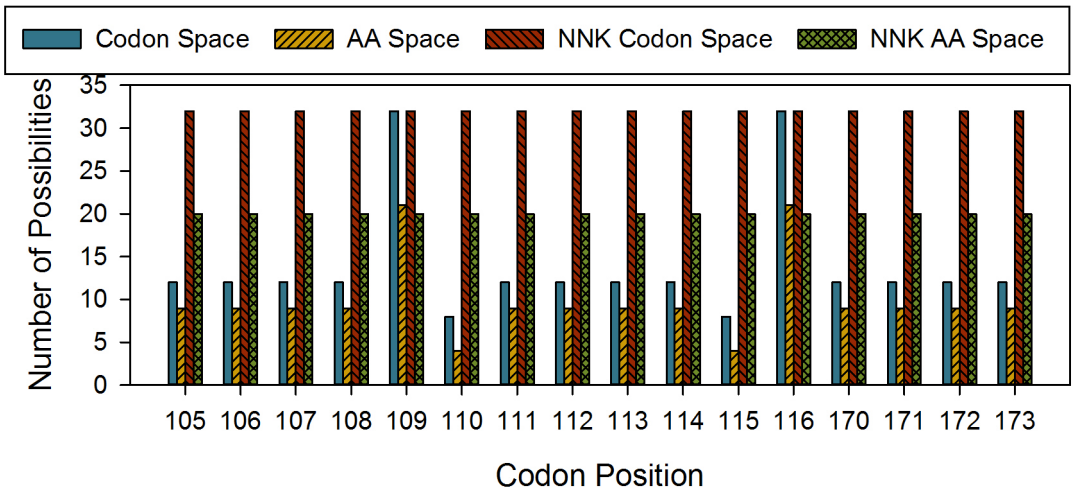


Figure 5.10: Degenerate Codon Properties. (A) Average codon hydropathy index is plotted for each randomized position for both libraries. The optimized library attempts to mirror wild-type residues. (B) Codon and amino acid diversity at each position are plotted for both libraries. As desired, the optimized library has significantly reduced diversities.

in the library, overall amino acid and codon spaces can be reduced and the library skewed toward sequences that are more likely to encode functional sortase clones. Figures 5.9 and 5.10 compare the properties of an optimized library. The optimized library was designed by picking degenerate codons that encode for amino acids with similar overall properties to the wild-type residues (Figure 5.10A). Most of the residues, 16 of 20, should be hydrophilic, polar amino acids because most of the loops are surface exposed and oxygen or nitrogen atoms would be needed to coordinate with calcium or manganese ions. These residues are encoded by VRK, whose possible amino acids are enumerated in Table 5.3. Two residues in the wild-type sequence are hydrophobic. Labeled in gray in Figure 5.9B, these reach into the core of the protein and hold the $\beta3/\beta4$ loop close to the protein. Because these are most likely necessary for a stable structure, those two positions were encoded by NTR, which will translate to a hydrophobic residue. The final two positions are serines in the wild-type enzyme, which are only slightly hydrophilic. It is unclear how these interact with their neighboring residues and the overall structure, so they were encoded by NNK, whose average hydropathy also mimics that of serine. As shown in Figure 5.10B, the optimized library's diversity was much lower than the NNK library, reducing the codon and amino acid spaces to 5.8×10^{17} and 2.0×10^{15}, respectively. These are still larger than the most diverse phage display libraries (approximately 1×10^{11}), but a much higher percentage of the potential library could be translated to phage.

Before a screen could be performed, it was necessary to define a robust method of library creation. The first successful Sa-SrtAΔ59 libraries were created by synthesizing degenerate oligonucleotides and using them to amplify the internal portion of Sa-SrtAΔ59 between the $\beta3/\beta4$ and $\beta6/\beta7$ loops. Those were then woven together with the C-terminal portion of Sa-SrtAΔ59 by PCR, and inserted into the rest of pComb3XM/SBP via restriction-ligation reactions (see Section 5.7 for more detail). Sequencing data from a pool of the resulting plasmids showed that this method produced libraries heavily biased toward the wild-type sequence (Figure 5.11, top and middle chromatograms). This most likely occurred because oligonucleotides that resemble the wild-type sequence will bind better to the region and therefore act as better primers in the amplification reaction. To avoid this, a new strategy was developed whereby the entire region of Sa-SrtAΔ59 between the $\beta3/\beta4$ and $\beta6/\beta7$ loop libraries was constructed from synthetic oligonucleotides. This strategy is depicted in Figure 5.12. Briefly, a series of overlapping oligonucleotides were designed that included the entire region between and including the two libraries. The oligos served as primers for one another in an

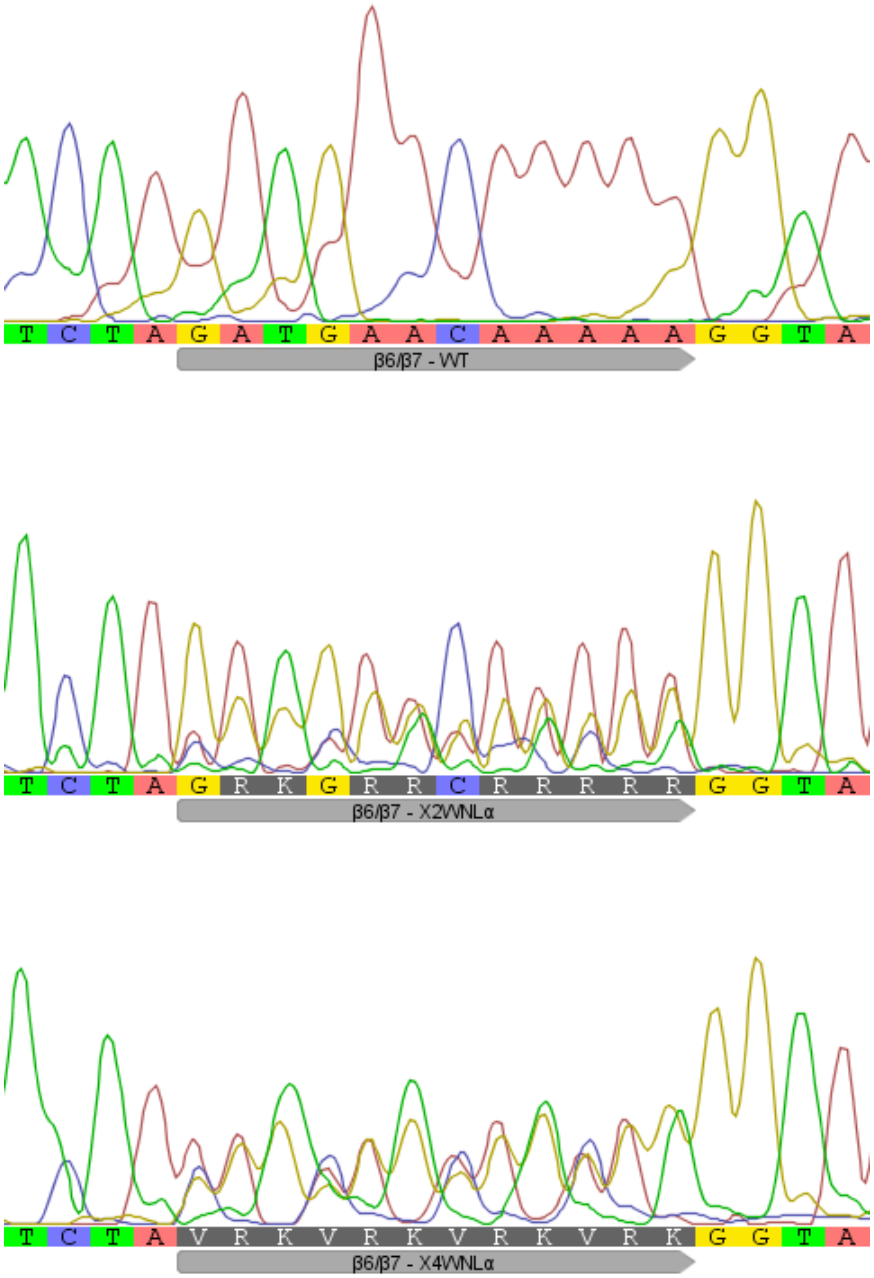


Figure 5.11: Library Quality by Generation Method. Base calls and the chromatagram for the $\beta 6/\beta 7$ loop of the wild-type Sa-SrtAΔ59 sequence (top) and libraries generated by amplifying a Sa-SrtAΔ59 section off of the plasmid with degenerate oligonucleotides (middle) or by weaving together synthetic oligonucleotides (bottom). The amplification sequence shows a heavy bias toward the wild-type sequence, especially for the G/C bases. The woven library shows no such bias. Randomized bases are annotated and should be VRK VRK VRK VRK (V=A/C/G, R=A/G, K=G/T).

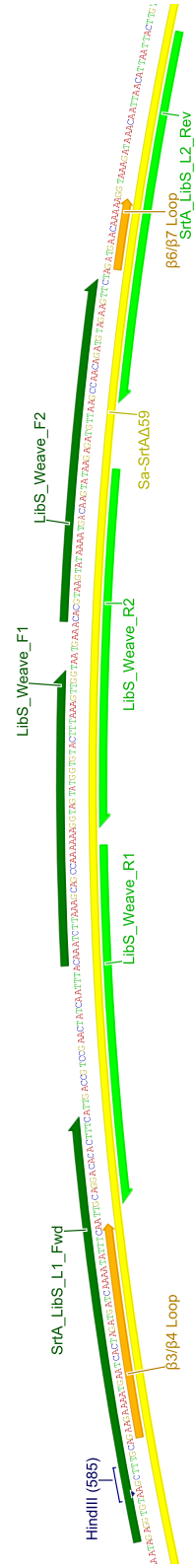


Figure 5.12: Isothermal Library Weaving Scheme. Four single-stranded DNA oligonucleotides were designed to bridge the sequence between the two library-generating oligos. Isothermal assembly was used to fill in missing bases and ligate the structure together. Note that there is no synthetic nucleotides opposing the randomized sections (the β_3/β_4 and β_6/β_7 loops), so there can be no GC or wild-type sequence bias.

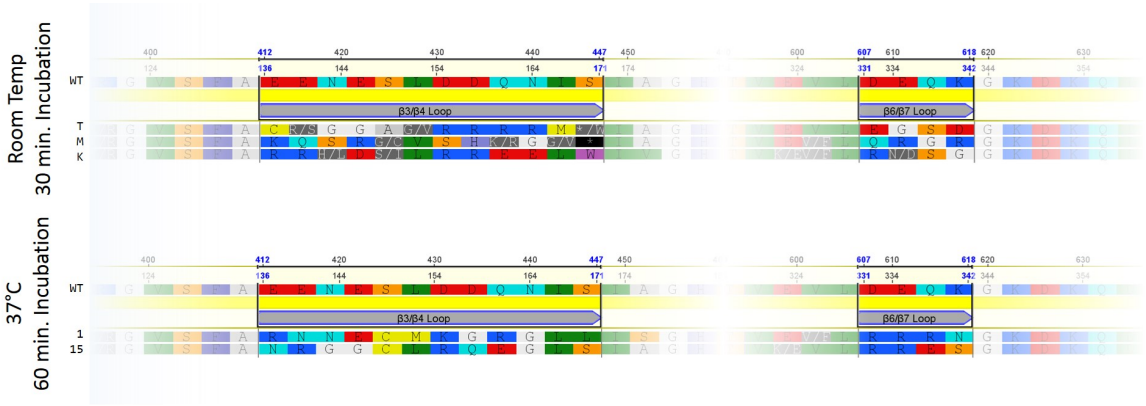


Figure 5.13: Recovered Sequences from Library X4Gβ.

isothermal weaving reaction, which filled in the missing bases and ligated the strands together. The most important aspect of this is that no oligonucleotides bound to the randomized regions of the degenerate oligonucleotides. The complement strand was thus created by a DNA polymerase that has no bias toward any sequence, wild-type or other. This library generation method allowed for a truly random library, as shown by the bottom chromatogram in Figure 5.11.

5.5 Screening Results

5.5.1 Phage Library X4Gβ

The first library to be screened was created by isothermal weaving and also combined with the rest of the pComb3XM/SBP vector by isothermal assembly, creating a library of 5.4×10^6 unique phage. Although this was a small library, it was large enough to screen. Resulting phage were subjected to a negative screen of 2 hours in 4 mM Ca^{2+} and a positive screen of 4mM Ca^{2+} and 200 μM Gly_3 for either 1 hour at 37°C or 30 minutes at room temperature. After 5 rounds of screening (successive, with no mutagenesis in between rounds), 24 colonies from each group were sent to sequencing. Of those 48 colonies, 5 sequences were recovered, shown in Figure 5.13. These five sequences were cloned into pSTEPL/EGFP (Plasmid 3.2) and assayed for observable cleavage. The results are tabulated in Table 5.4. All of the clones either did not cleave under any conditions or cleaved no matter the condition. Note that colony M, which contained a stop codon in the library sequences, was not assayed.

Table 5.4: Phage Display Preliminary Data. Sequences recovered from phage display were cloned into the pSRTA/EGFP vector. The resulting proteins were expressed, bound to cobalt resin, and their expression was measured qualitatively for preliminary studies. EGFP fluorescence was estimated by UV transillumination to be nonfluorescent (—), slightly fluorescent (+), fluorescent (++), or highly fluorescent (+++) as compared to the wild-type control. Fluorescence was observed for the column flowthroughs, resin before cleavage, and after 4 hours (24 hours) at the sample's positive screen temperature under the following cleavage conditions: Neg (0mM $CaCl_2$, 0mM Gly_3, 0mM $MnCl_2$), Base (100μM $CaCl_2$, 100μM Gly_3, 0mM $MnCl_2$), ↑Ca (5mM $CaCl_2$, 100μM Gly_3, 0mM $MnCl_2$), ↑Ca, Mn (5mM $CaCl_2$, 100μM Gly_3, 5mM $MnCl_2$), ↑Ca, -Gly (5mM $CaCl_2$, 0mM Gly_3, 0mM $MnCl_2$), Ca, ↑Mn (5mM $CaCl_2$, 200μM Gly_3, 80mM $MnCl_2$). Clones that were subjected to additional mutagenesis and screening are bolded.

		Negative Screen Conditions					Positive Screen Conditions					EGFP Cleavage Conditions							
Library	Colony	[$CaCl_2$] (μM)	[$MnCl_2$] (μM)	[Gly_3] (μM)	Temp (°C)	Hours	[$CaCl_2$] (μM)	[$MnCl_2$] (μM)	[Gly_3] (μM)	Temp (°C)	Hours	Flowthrough	Beads	Neg	Base	↑Ca	↑Ca, Mn	↑Ca, -Gly	Ca, ↑Mn
X4Gβ	WT											—	++	—	++				
	RT5.1	4,000	0	0	37	2	4,000	0	200	RT	0.5	—	+	—	—				
	RT5.15	4,000	0	0	37	2	4,000	0	200	RT	0.5	—	+	—	—				
	375.K	4,000	0	0	37	2	4,000	0	200	37	1	—	+	+	+				
	375.T	4,000	0	0	37	2	4,000	0	200	37	1	—	+	—	—				
X4WNLε	WT											—	++	—	++	++	++	+	
	GM4.B	4,000	0	0	37	2	4,000	5,000	200	37	4	+++	—						
	GM4.E	4,000	0	0	37	2	4,000	5,000	200	37	4	+++	—						
	GM5.S	4,000	0	0	37	2	4,000	5,000	200	37	4	—	+++	—(—)	—(—)	—(—)	—(—)		
	GM5.T	4,000	0	0	37	2	4,000	5,000	200	37	4	+++	—						
	GM5.DD	4,000	0	0	37	2	4,000	5,000	200	37	4	—	+	—(—)	—(—)	—(—)	—(—)		
	G4.11	4,000	0	0	37	2	4,000	0	200	37	4	++	++	—(—)	—(—)	—(—)		—(—)	
	G5.18	4,000	0	0	37	2	4,000	0	200	37	4	++	+	—(—)	—(—)	—(—)		—(—)	
	G5.22	4,000	0	0	37	2	4,000	0	200	37	4	++	+++	—(—)	—(—)	—(+++)		—(+++)	
	G5.24	4,000	0	0	37	2	4,000	0	200	37	4	+	—						
	G5.30	4,000	0	0	37	2	4,000	0	200	37	4	++	+++	—(—)	—(—)	—(—)		—(—)	
XH4WSLα	WT											—	++	—	++	++	++		
	G5.G1	4,000	0	0	37	2	4,000	0	200	RT	1	—	++	—(—)	—(—)	—(—)			
	G5.H1	4,000	0	0	37	2	4,000	0	200	RT	1	—	+++	—(—)	—(—)	—(—)			
	G5.E2	4,000	0	0	37	2	4,000	0	200	RT	1	+	+++	—(—)	—(—)	—(—)			
	G5.G3	4,000	0	0	37	2	4,000	0	200	RT	1	+++	—						
	GM5.B5	4,000	0	0	37	2	4,000	5,000	200	RT	1	—	+	—(—)	—(—)	—(—)			
	G4a.A7	4,000	0	200	37	2	10,000	0	200	RT	1	+	+	—(—)	—(—)	—(—)			
	G4a.C7	4,000	0	200	37	2	10,000	0	200	RT	1	—	+	—(—)	—(—)	—(—)			
	G4a.G7	4,000	0	200	37	2	10,000	0	200	RT	1	—	++	—(—)	—(—)	—(—)			
	G4a.A9	4,000	0	200	37	2	10,000	0	200	RT	1	+++	—						
	G4a.D9	4,000	0	200	37	2	10,000	0	200	RT	1	—	++	—(—)	—(—)	—(—)			
	GM4a.C10	4,000	0	200	37	2	4,000	5,000	200	RT	1	—	+++	—(—)	—(—)		—(—)		
	GM4a.D10	4,000	0	200	37	2	4,000	5,000	200	RT	1	+	+	—(—)	—(—)		—(—)		
	GM4a.A11	4,000	0	200	37	2	4,000	5,000	200	RT	1	+	+++	—(—)	—(—)		—(—)		
	GM4a.B11	4,000	0	200	37	2	4,000	5,000	200	RT	1	—	+++	—(—)	—(—)		—(—)		
	GM4a.G11	4,000	0	200	37	2	4,000	5,000	200	RT	1	—	+++	—(—)	—(—)		—(—)		
	GM4a.E12	4,000	0	200	37	2	4,000	5,000	200	RT	1	++	+	—(—)	—(—)		—(—)		
X4WEALβ	WT											—	++	—	++				++
	B.F3	4,000	0	200	RT	1.5	4,000	80,000	200	RT	0.5	—	+++	—(—)	—(—)				—(—)
X5ALα	WT											—	++	—	++				++
	F.C11	4,000	0	200	RT	1.5	4,000	80,000	200	RT	1.5	—	+++	—(—)	—(—)				—(—)

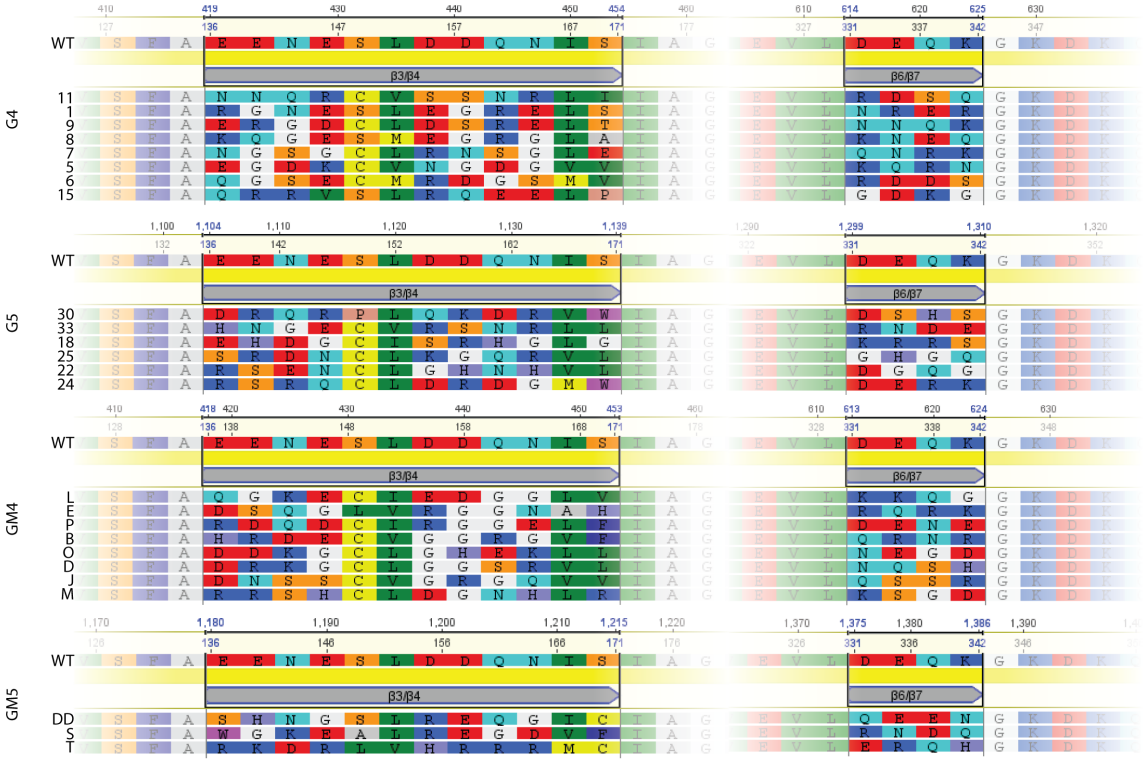


Figure 5.14: Recovered Sequences from Library X4WNLε.

The vast majority of the clones submitted to sequencing in this and the previous library did not contain a coding sequence for Sa-SrtAΔ59. The plasmids contained a deletion between the $(GGS)_5$ linker in STEPL and the flexible linker at the N-terminus of pIII. These sequences are highly similar and GC rich, so this was likely due to a recombination event, which is possible during isothermal assembly. Therefore, the remaining libraries were placed into the pComb3XM/SBP vector by ligation.

5.5.2 Phage Library X4WNLε

With no success in the initial trial of tuning sortase's calcium sensitivity, the next library was also screened for manganese-triggered activity. This library of 4.0×10^6 unique phage was also split into two screens. All went through a 2 hour negative screen in 4 mM Ca^{2+} and positive screens for 4 hours at 37°C with either 4mM Ca^{2+} and 200 μM Gly_3 or 4mM Ca^{2+}, 200 μM Gly_3, and 10 mM $MnCl_2$. 16 colonies in each group were sequenced after both rounds 4 and 5, resulting in 26 unique

potential sequences (Figure 5.14). Of these, 5 sequences from each screen were chosen to be cloned into pSTEPL/EGFP and assayed. Table 5.4 again shows that none of the sequences performed as expected within 4 hours.

One clone that was screened for reduced calcium sensitivity, however, did show cleavage only under high-calcium conditions at 24 hours, G5.22. That clone was therefore subjected to random mutagenesis along the entire length of the Sa-SrtAΔ59 sequence to create a phage library of 2.5×10^6 unique clones. Those were screened for activity at room temperature in 1 hour. After 5 rounds, 16 clones were sent to sequencing. None contained DNA that could be sequenced. It might not have been possible to increase the activity of G5.22, so it was no longer pursued.

A common feature of the clones resulting from the calcium screens is a cysteine at position 109. That implies that it plays an important role in how these phage are passing the screen. Since the bacterial cytoplasm (where the assayed proteins are produced) is a reducing environment, but the periplasm (where the phage are produced) is oxidizing, the pSTEPL/EGFP mutants containing the conserved cysteine may not be able to make a critical disulfide. They were therefore expressed in a special *E. coli* strain that has an oxidizing cytoplasm. Unfortunately, all colonies again showed negative results (data not shown), so a reduced thiol was not the problem.

5.5.3 Phage Library XH4WSLα

After having no duplicate sequences in Library X4WNLε, the screens were made harder to pass, with the hope that this would further narrow the field and allow a consensus sequence to emerge. A library of 6.0×10^7 clones was screened under four different conditions:

1. A negative screen of 4mM Ca^{2+} for 2 hours at 37°C followed by a positive screen of 4mM Ca^{2+} and 200 μM Gly_3 for 1 hour at room temperature.

2. A negative screen of 4mM Ca^{2+} for 2 hours at 37°C followed by a positive screen of 4mM Ca^{2+}, 5mM $MnCl_2$, and 200 μM Gly_3 for 1 hour at room temperature.

3. A negative screen of 4mM Ca^{2+} and 200 μM Gly_3 for 2 hours at 37°C followed by a positive screen of 10mM Ca^{2+} and 200 μM Gly_3 for 1 hour at room temperature.

4. A negative screen of 4mM Ca^{2+} and 200 μM Gly_3 for 2 hours at 37°C followed by a positive screen of 4mM Ca^{2+}, 5mM $MnCl_2$, and 200 μM Gly_3 for 1 hour at room temperature.

Figure 5.15: Recovered Sequences from Library XH4WSLα.

Table 5.5: Sortase Activity after Acid Shock. pSTEPL/EGFP proteins were expressed, bound to cobalt resin, and their ability to release EGFP into the supernatant was measured qualitatively by UV transillumination to be nonfluorescent (—), slightly fluorescent (+), fluorescent (++), or highly fluorescent (+++) after 4 hours at 37°C.

Acid Conditons		Cleavage Conditions	
pH	Time	0 μM $CaCl_2$, 0 μM Gly_3	100 μM $CaCl_2$, 200 μM Gly_3
4	10 min	—	+++
7	N/A	—	+++

After 5 rounds of screening conditions 1 and 2, and 4 rounds of screening conditions 3 and 4, 24 clones of each were sent to sequencing. 17 viable sequences were returned (Figure 5.15). Again these sequences had no duplicates or consensus. Still, all 17 were cloned into pSTEPL/EGFP and assayed for activity in the positive and negative buffers. None showed conditional activity (Table 5.4).

5.5.4 Phage Library X4WEALβ

To try and improve the display results, a new screening method was employed, as shown in Figure 5.16. It was hypothesized that the SBP-streptavidin interaction was not strong enough to hold for the long negative and positive incubations in the coated wells, so the new screen was designed to minimize the time that phage spent bound to the plate. Briefly, libraries are transformed into *E. coli*, which

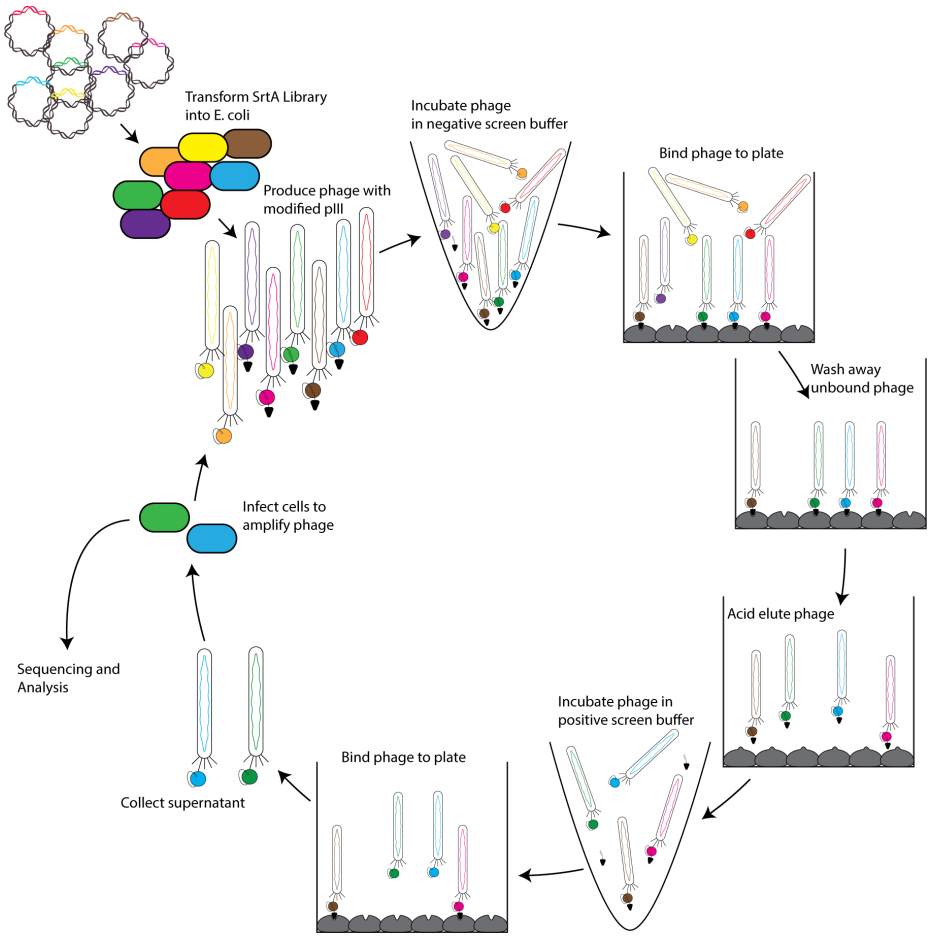


Figure 5.16: Altered Phage Display Screening Cycle. Phagemid libraries were electroporated into *E. coli*, which produced modified phage displaying the modified pIII and a streptavidin-binding ligand. Phage were then incubated in the negative buffer before being bound to wells coated in streptavidin. Phage with sortase mutants that cleaved during expression (**yellow**, **orange**, **red**) or during the negative screen (**purple**) could not bind and were washed away. Next, the remaining phage acid eluted off of the plate before being incubated in positive buffer. These phage were then applied to a different streptavidin-coated well. Phage that cleaved during the positive screen (**blue**, **green**) were collected, leaving phage that could not cleave (**brown**, **pink**) bound to the plate. That supernatant then infected *E. coli* cells that could be amplified for another round of phage display or plated for further analysis.

are induced to produce phage that randomly include the mutant STEPL-pIII chimeras as well as the encoding plasmid. The phage are incubated in the negative selection buffer in a polypropylene tube. With 30 minutes remaining in the incubation, the phage are transferred to a streptavidin coated plate and allowed to bind. Then, unbound phage are washed away with vigorous pipetting, which removes any mutants that cleave during expression or in the negative screening buffer. Uncleaved phage that remain bound to the plate are eluted in pH 4 acetate buffer for 10 minutes. These phage are then neutralized with sodium hydroxide and incubated in positive screening buffer in another polypropylene tube. Again, phage are transferred to a streptavidin-coated well with 30 minutes remaining in the positive incubation. Supernatant, which contains STEPL mutants that were inactive during expression and the negative screen but gained activity during the positive screen, is collected and used to infect fresh bacterial cells. The bacteria are then used to produce phage for another screen or grown on plates for sequencing and analysis. STEPL's ability to maintain activity after the acid elution was determined by assaying wild-type pSRTA/EGFP for EGFP cleavage after a 10 minute pH 4 acid shock (Table 5.5).

In addition, it was theorized that the optimized $\beta3/\beta4$ and $\beta6/\beta7$ loop library may have been too optimal, resulting in almost entirely functional proteins because loops are very tolerant of mutations between similar amino acids. Therefore, another layer of diversity was added by running error-prone PCR reactions on top of the randomized loops. The inclusion of random mutations should allow for even harsher screening conditions because residues that more directly affect the enzymatic mechanism could be altered.

A library of 1.0×10^7 mutants was screened under four conditions:

1. A negative screen of 4mM Ca^{2+} for 1.5 hours at room temperature followed by a positive screen of 4mM Ca^{2+} and 200 μM Gly_3 for 4 hours at room temperature.

2. A negative screen of 4mM Ca^{2+} for 1.5 hours at room temperature followed by a positive screen of 4mM Ca^{2+} and 200 μM Gly_3 for 30 minutes at room temperature.

3. A negative screen of 4mM Ca^{2+} and 200 μM Gly_3 for 1.5 hours at room temperature followed by a positive screen of 4mM Ca^{2+}, 80mM $MnCl_2$, and 200 μM Gly_3 for 4 hours at room temperature.

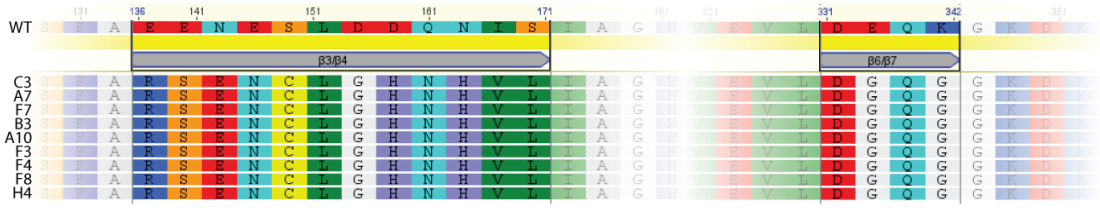


Figure 5.17: Recovered Sequences from Library X4WEALβ.

4. A negative screen of 4mM Ca^{2+} and 200 μM Gly_3 for 1.5 hours at room temperature followed by a positive screen of 4mM Ca^{2+}, 80mM $MnCl_2$, and 200 μM Gly_3 for 30 minutes at room temperature.

The manganese screens had a much larger concentration of $MnCl_2$ to try and pull out sequences with even a low affinity for manganese, which could then be increased in subsequent libraries.

After five days, 16 clones from each condition were sent for sequencing. No viable sequences resulted from conditions 1 or 2, but 9 identical sequences resulted from conditions 3 and 4 (Figure 5.17). However, Table 5.4 again shows that the sequence had no activity in cells with oxidizing or reducing cytoplasms.

5.5.5 Phage Library X5ALα/X4WSLβ

Finally, a pooled library of Sa-SrtAΔ59 with only random mutagenesis and Sa-SrtAΔ59 with only the optimized loop library was screened under the following conditons:

1. A negative screen of 4mM Ca^{2+} for 1.5 hours at room temperature followed by a positive screen of 4mM Ca^{2+} and 200 μM Gly_3 for 4 hours at room temperature.

2. A negative screen of 4mM Ca^{2+} and 200 μM Gly_3 for 1.5 hours at room temperature followed by a positive screen of 4mM Ca^{2+}, 80mM $MnCl_2$, and 200 μM Gly_3 for 4 hours at room temperature.

Beginning with 3.7×10^6 sequences, 16 clones from each condition were sequenced. One good sequence was returned, from the manganese screen, containing a single point mutation, Q64R (Figure 5.18). The clone showed no activity (Table 5.4).

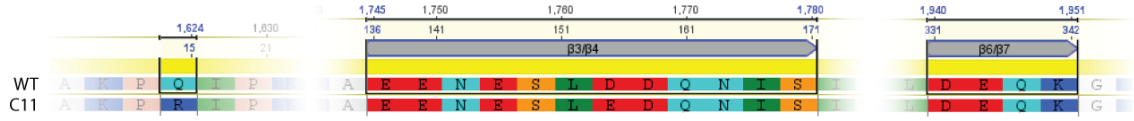


Figure 5.18: Recovered Sequences from Library X5ALα/X4WSLβ.

5.6 Conclusions

Unfortunately, this study was unable to produce STEPL mutants with a lower calcium sensitivity or a manganese sensitivity. Bacterial display was not a viable strategy because the sortase enzyme could not survive translocations across both *E. coli* membranes. Yeast display was not attempted due to heavy glycosylation of the sortase in the endoplasmic reticulum. Phage display produced a construct with an active sortase and an assay that physically held or released the phage based on enzymatic activity. However, despite the use of both randomly generated and specifically optimized libraries under a number of different screening conditions, no clones were obtained that were also functional in the original STEPL assay.

5.7 Materials & Methods

Bacterial Display

Cloning

To construct pSED/sfGFP (see Plasmid Map 5.1), the coding sequence for EspP and the OmpA signal were obtained from the NCBI nucleotide database (accession no. FJ875095) and synthesized by IDT. sfGFP was amplified off of the instructional plasmid pGlo (Bio-Rad). The sequences for LPETG, $(GGS)_5$, and Sa-SrtAΔ59 were amplified together off of pSTEPL. These three amplicons were then woven together via PCR and cloned into pRSET-A via restriction-ligation. pSED2/sfGFP (Plasmid Map 5.2) and pSED2/sfGFP (Plasmid Map 5.3) were cloned by first inserting an EagI restriction site between the coding sequences for Sa-SrtAΔ59 and EspP in pSED/sfGFP. Oligos containing $(GGS)_5$ and appropriate overhangs were ligated into pSED/sfGFP that had been digested with EagI to create pSED2/sfGFP. pSED2.5/sfGFP was then constructed by amplifying ECFP off of pECFP and cloning it into the same restriction site. pSED3/sfGFP was created by ordering the

entire expression cassette from IDT and cloning it into pRSET-A. The EstA sequence was obtained from the NCBI nucleotide database (accession no. AF005091.1). Insert directionality and all final vectors were verified by sequencing.

Cleavage Assay

First, single colonies or stabs of glycerol stocks were grown for 24-hours in 10 mL of ZYP-5052 (Ameresco) supplemented with 100 μg/mL Amp. 1 mL aliquots were pelleted at 6,000g for 5 minutes and then resuspended in the appropriate amount of STEPL buffer with or without 5 mM $CaCl_2$ and 5 mM Gly_3 (Sigma). Cultures were then shook at 37°C overnight. For the first six hours, cells were pelleted and observed with UV light once per hour.

Immunohistochemistry

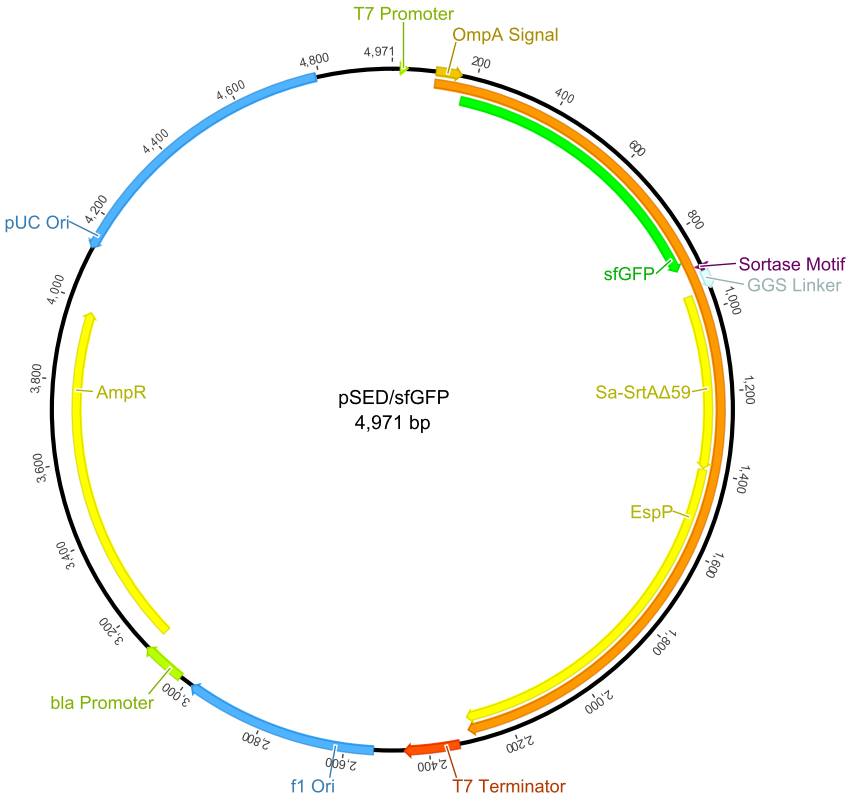
Both pSED/sfGFP and pSTEPL/EGFP expression cells were grown for 24-hours in 10 mL of ZYP-5052 (Ameresco) supplemented with 100 μg/mL Amp. For half of the pSTEPL/EGFP stock, the chimeric protein was purified and bound to Talon resin (Clontech) as previously described (Section 3.9). Cells and beads were pelleted at 8,600g for 3 minutes and washed three times with 1 mL ice-cold STEPL buffer. Samples were then resuspeneded in 1 mL ice-cold STEPL buffer with Rabbitα-GFP (GE) antibody and shook for 2 hours at 4°C. After another three washes, samples were resuspended in 1 mL ice-cold STEPL buffer with α-Rabbit secondary antibody conjugated to a 700nm fluorophore (LiCor) and shook for 1 hour at 4°C. Samples were again washed 3 times and then resuspended in 100 μL ice-cold STEPL buffer. These were then transferred to a 96-well plate that was imaged and quantified by the LiCor Odyssey fluorescence scanner.

Phage Display

Phagemid Cloning

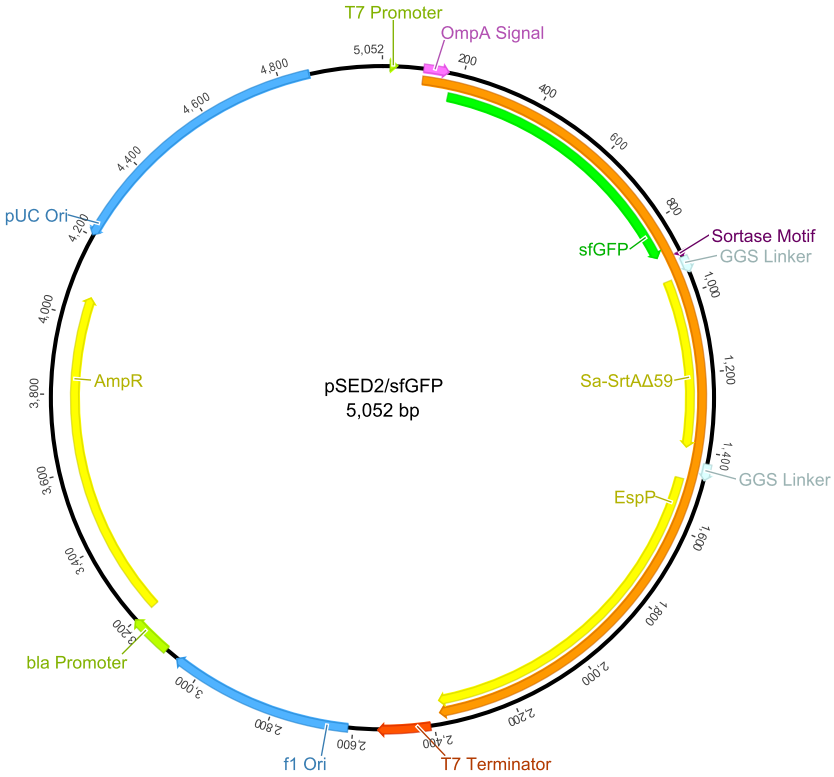
All of the phagemids were derived from pComb3X (generously gifted by Dr. Donald Siegel). Before any other cloning, the amber stop codon in pComb3X's linker region was deleted by site-directed mutagenesis (Quikchange II Kit, Agilent Genomics), creating the vector pComb3XM (see Plasmid Map 5.5). The mutation was validated by sequencing.

Plasmid Map 5.1: pSED/sfGFP



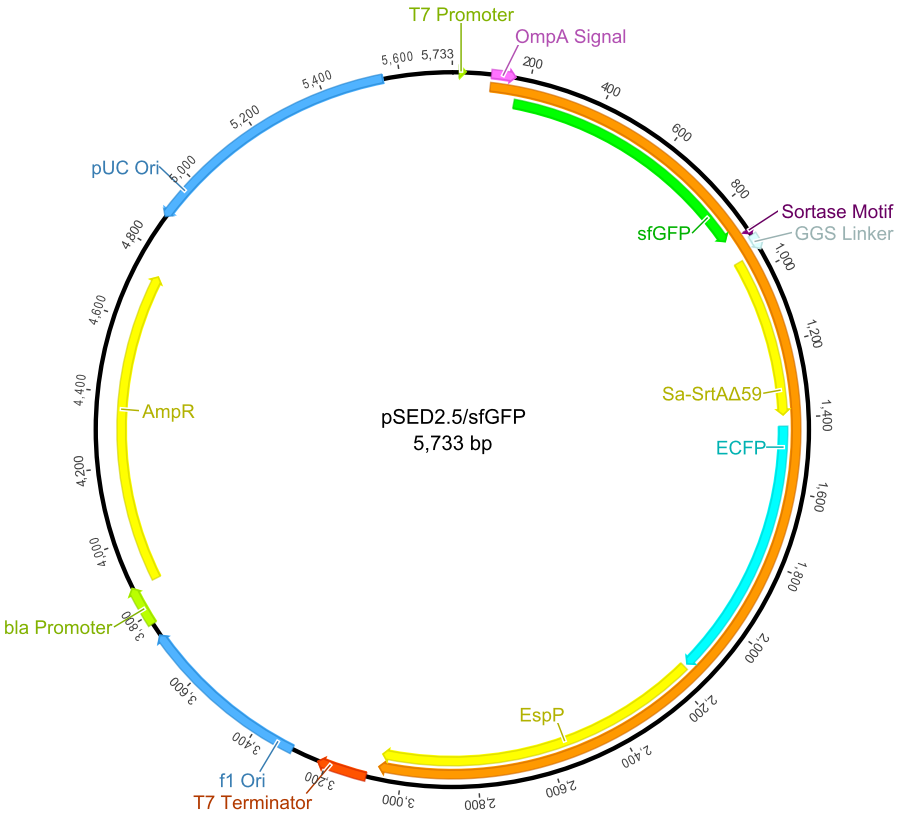
Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-2382
OmpA Signal	103-165
sfGFP	172-888
Sortase Motif	895-909
$(GGS)_5$ Linker	910-954
Sa-SrtAΔ59	955-1395
EspP β Domain	1396-2301
T7 Terminator	2330-2495
f1 Origin	2530-2985
bla Promoter	3017-3121
Amp^R	3116-3976
pUC Origin	4794-4121

Plasmid Map 5.2: pSED2/sfGFP



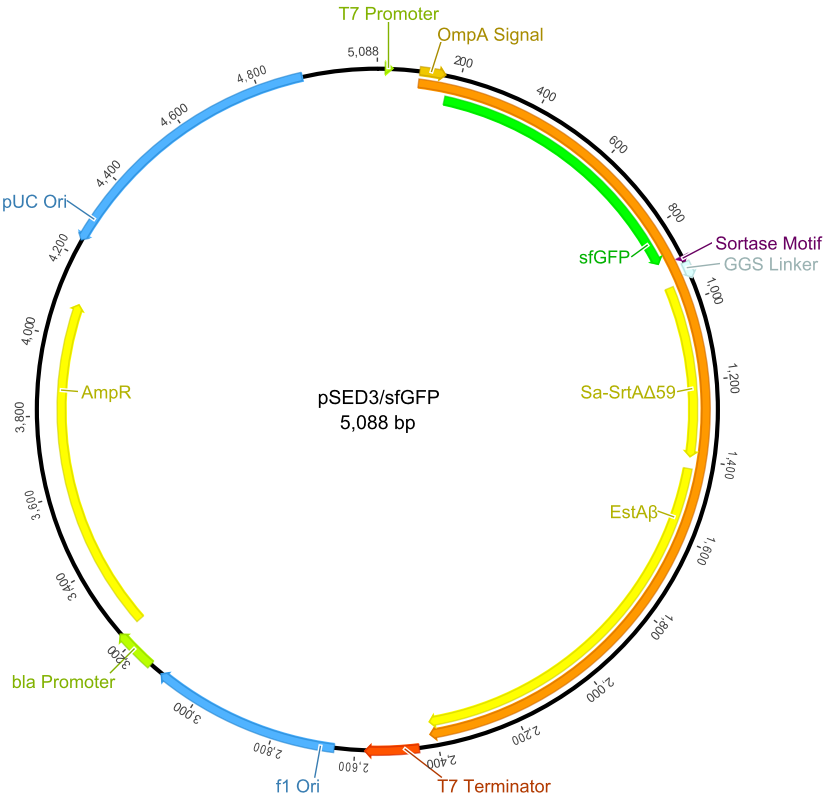
Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-2382
OmpA Signal	103-165
sfGFP	172-888
Sortase Motif	895-909
$(GGS)_5$ Linker	910-954
Sa-SrtAΔ59	955-1395
$(GGS)_5$ Linker	1426-1470
EspP β Domain	1477-2382
T7 Terminator	2411-2540
f1 Origin	2611-3066
bla Promoter	3098-3202
Amp^R	3197-4057
pUC Origin	4875-4202

Plasmid Map 5.3: pSED2.5/sfGFP



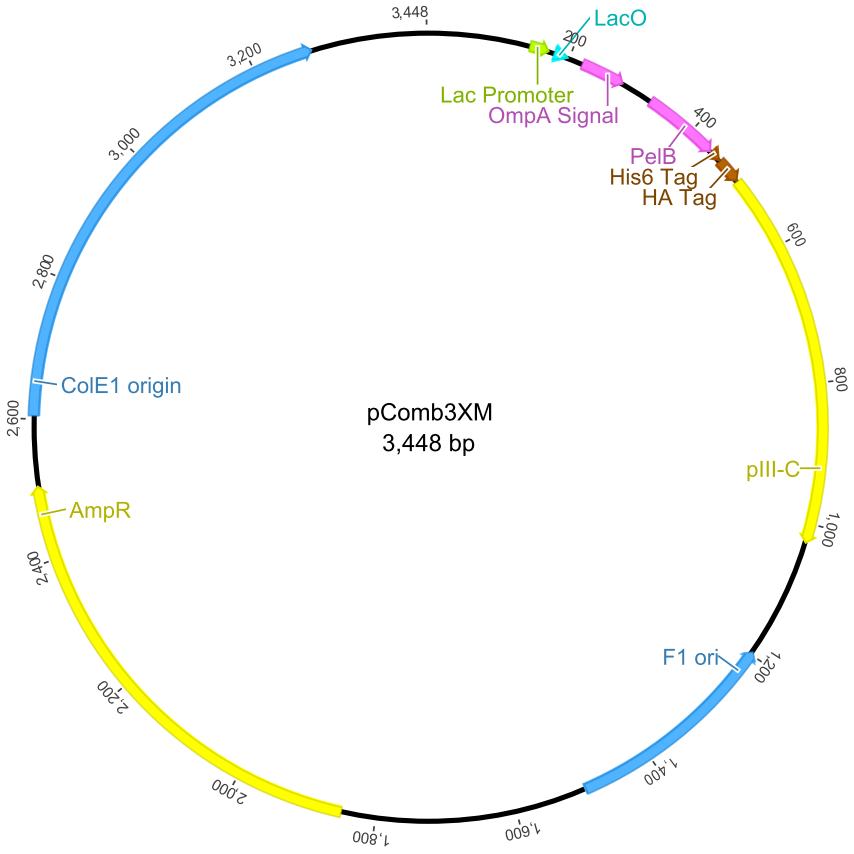
Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-2382
OmpA Signal	103-165
sfGFP	172-888
Sortase Motif	895-909
$(GGS)_5$ Linker	910-954
Sa-SrtAΔ59	955-1395
ECFP	1426-2151
EspP β Domain	2158-3063
T7 Terminator	3092-3221
f1 Origin	3292-3747
bla Promoter	3779-3883
Amp^R	3878-4738
pUC Origin	5556-4883

Plasmid Map 5.4: pSED3/sfGFP



Feature	*Location*
T7 Promoter	20-39
Chimeric ORF	100-2415
OmpA Signal	103-165
sfGFP	172-888
Sortase Motif	895-909
$(GGS)_5$ Linker	910-954
Sa-SrtAΔ59	955-1395
EstA β Domain	1426-2415
T7 Terminator	2447-2576
f1 Origin	2647-3102
bla Promoter	3134-3121
Amp^R	3116-3976
pUC Origin	4794-4121

Plasmid Map 5.5: pComb3XM



Feature	*Location*
Lac Promoter	145-172
LacO	179-201
OmpA Signal Sequence	222-284
M13 pIII-C	494-1027
F1 Ori	1501-1195
Amp^R	1848-2507
ColE1 Ori	2605-3287

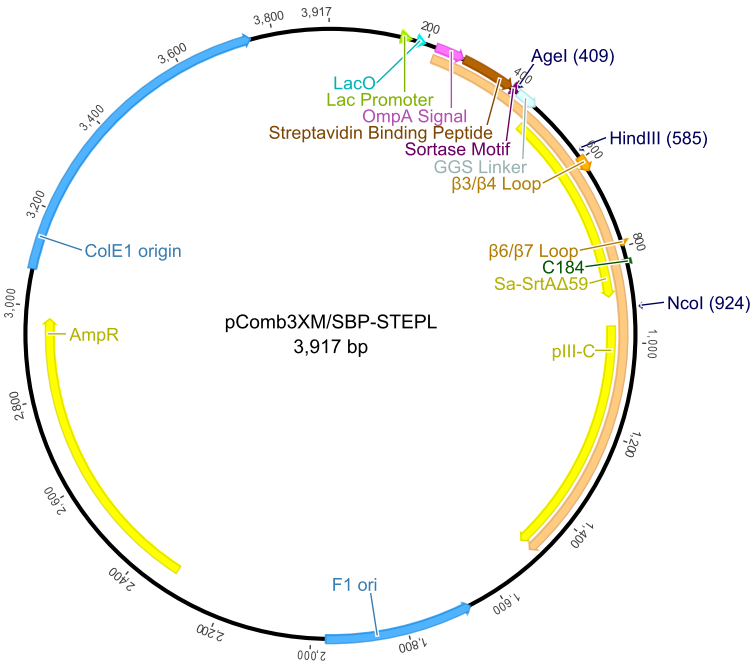
The streptavidin binding peptide (SBP) sequence (MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP) and the Avi-Tag sequence (GLNDIFEAQKIEWHE) were codon-optimized and synthesized by IDT. The j5 program was used to design PCR primers to amplify the pComb3XM backbone and sortase A active site with the linker and LPETG motif from pSTEPL [266]. The linear PCR fragments were combined through isothermal assembly [267] to create pComb3XM/SBP-STEPL and pComb3XM/Avi-STEPL (see Plasmid Maps 5.6 and 5.7). Gibson reactions were transformed into Top10 competent cells (Life Technologies). Resultant vectors were validated by sequencing. Site-directed mutagenesis (Quikchange II Kit, Agilent Genomics) was used to introduce the inactivating C180A mutation into both vectors, creating pComb3XM/SBP-C180A and pComb3XM/Avi-C180A (Plasmid Maps not shown). Resultant vectors were validated by Sanger sequencing.

To create a phagemid that linked the STEPL system to pIII via a helical linker, the α-helical linker sequence (GSG$(EAAAK)_4$GSG) plus 20bp upstream/22bp downstream overlaps was synthesized by IDT as a series of single-stranded oligonucleotides. The pComb3XM/SBP vector was amplified to create a linear product that excluded the native glycine-serine linker but included the upstream and downstream overlaps. The PCR product was then digested with DpnI (New England Biolabs), purified on a 0.7% low-melt agarose gel (Lonza), and re-amplified by PCR with the PhireII-Hotstart Enzyme (Thermo Fisher). 1 pmol of backbone was then combined with 10 μM of each oligo and 2X Gibson Assembly Master Mix (New England Biolabs) and incubated at 50°C for 1 hour before being transformed into Top10 cells (Life Technologies). Resultant vectors (see Plasmid Map 5.8) were validated by sequencing.

Helper Phage Preparation

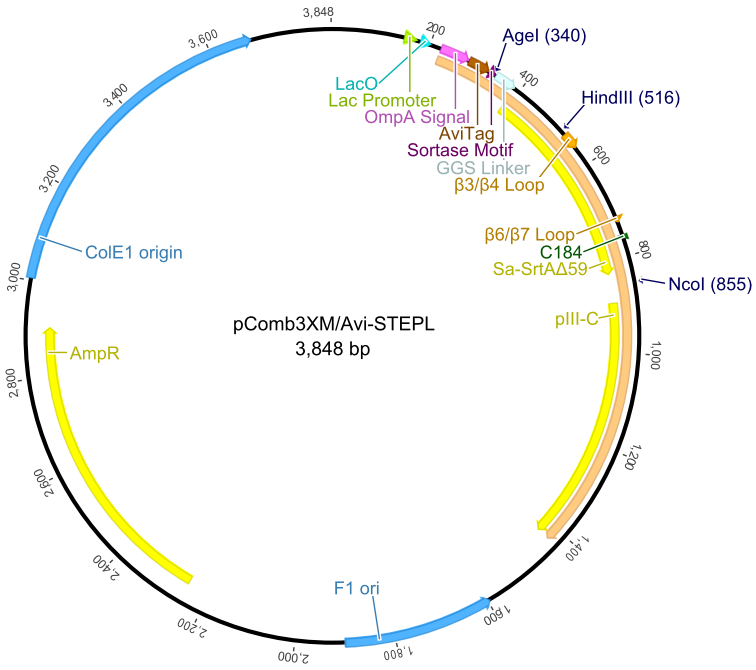
Overnight cultures of XL1-Blue cells (Agilent) were grown in SB Media (Gentrox) containing 10 μg/mL tetracycline. The following day, 10 μL of the culture was added to 10 mL of SB Media with tetracycline and shook at 37°C for 1 hour. Then 10 μL of CM13 helper phage stock (Antibody Design Labs) was added to the culture and allowed to infect at room temperature for 15 minutes before being shook for an additional 2 hours at 37°C. The culture was then added to 500 mL of SB Media containing 70 μg/mL kanamycin before being shook overnight at 37°C. The next day, the culture was split into 50 mL aliquots and bacterial cells were removed by pelleting at 3,000g

Plasmid Map 5.6: pComb3XM/SBP-STEPL



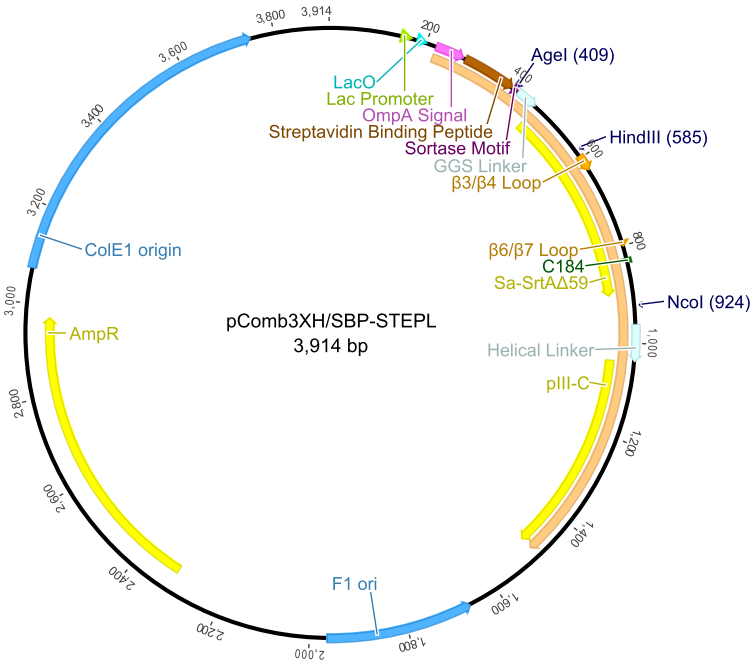
Feature	*Location*
Lac Promoter	145-172
LacO	179-201
Chimeric ORF	222-1496
OmpA Signal Sequence	222-284
Streptavidin Binding Peptide	285-398
Sortase Motif	399-413
$(GGS)_5$ Linker	414-458
Sa-SrtAΔ59	459-899
β_3/β_4 Loop	594-629
β_6/β_7 Loop	789-800
C184 Catalytic Residue	831-833
M13 pIII-C	963-1496
F1 Ori	1970-1664
Amp^R	2317-2976
ColE1 Ori	3074-3756

Plasmid Map 5.7: pComb3XM/Avi-STEPL



Feature	*Location*
Lac Promoter	145-172
LacO	179-201
Chimeric ORF	222-1496
OmpA Signal Sequence	222-284
Avi-Tag®	285-329
Sortase Motif	330-344
$(GGS)_5$ Linker	345-389
Sa-SrtAΔ59	390-830
β_3/β_4 Loop	525-560
β_6/β_7 Loop	720-731
C184 Catalytic Residue	762-764
M13 pIII-C	894-1427
F1 Ori	1901-1595
Amp^R	2248-2907
ColE1 Ori	3005-3687

Plasmid Map 5.8: pComb3XH/SBP-STEPL



Feature	*Location*
Lac Promoter	145-172
LacO	179-201
Chimeric ORF	222-1496
OmpA Signal Sequence	222-284
Streptavidin Binding Peptide	285-398
Sortase Motif	399-413
$(GGS)_5$ Linker	414-458
Sa-SrtAΔ59	459-899
β_3/β_4 Loop	594-629
β_6/β_7 Loop	789-800
C184 Catalytic Residue	831-833
α-helical Linker	963-1040
M13 pIII-C	1041-1493
F1 Ori	1967-1661
Amp^R	2314-2973
ColE1 Ori	3071-3753

for 15 minutes. Supernatants were then incubated at 70°C for 20 miuntes before being centrifuged again at 3,000g for 15 minutes. Supernatants were transferred to new tubes and stored at 4°C until needed.

Preparation of Streptavidin plates

10 μg of purified streptavidin (Fisher, diluted to 4 mg/mL in PBS, pH 5) was added to 96-well, clear, flat bottom, half area, high bind plates (Corning). Plates were sealed and incubated for 2 hours at 37°C. The PBS solution was removed and wells were then blocked by filling them with STEPL buffer with 2% w/v biotin-free powdered milk (LabScientific) and incubating for at least 1 hour at 37°C or room temperature.

Phage Assay Validation

First, the vectors pComb3XM/SBP-STEPL, pComb3XM/Avi-STEPL, pComb3XM/SBP-C180A, and pComb3XM/Avi-C180A were transformed into Biotin XCell™ F' electrocompetent cells (Lucigen). Transformed colonies were used to inoculate 10 mL overnight cultures in SB Media (50 μg/mL carbenicillin, 10 μg/mL tetracycline) at 37°C. The next day, the cultures were expanded to 50 mL of SB Media (50 μg/mL carbenicillin, 10 μg/mL tetracycline) and grown to an OD_{600} of 1 by shaking at 37°C. At that point 2 mL of helper phage were added to each culture, and allowed to infect stationary at room temperature for 15 minutes. Cultures were then expanded to 200 mL of SB Media (50 μg/mL carbenicillin, 10 μg/mL tetracycline) and shook for an additional 2 hours at 37°C. Finally, kanamycin was added to a concentration of 70 μg/mL and cultures were shook overnight at 25°C.

The following day, bacterial cells were removed from the overnight cultures by pelleting at 3,000g for 15 minutes. Sodium chloride (Fisher, 3% w/v) and poly(ethylene glycol) (Sigma, MW 8000, 4% w/v) were dissolved into the supernatants by shaking vigorously at 37°C for 5 minutes before being incubated on ice for 30 minutes. Phage were then pelleted by centrifugation at 15,000g for 15 minutes. Pellets were resuspended in 2 mL of STEPL Buffer with 2% w/v bovine serium albumin by gentle rocking for 30 minutes at room temperature. Impurities were removed by centrifugation at 10,000g for 5 minutes at 4°C. Supernatants were finally passed through a steile filter to obtain purified phage.

Table 5.6: Phage Library Nomenclature. In order to keep track of the libraries that were created, a common nomenclature was employed. All libraries follow the scheme [Vector(s)][Generation][Cloning][Attempt]. The vector, generation, and cloning portions are described in the table. The attempt portion is a sequential Greek letter used to separate libraries created with the same cloning plan.

Vectors

Symbol	Vector
H	pComb3X**H**
X	pComb3**X**M

Library Generation

Symbol	Technique
1	Kunkel Mutagenesis
2	Site-Directed Mutagenesis
3	Degenerate Oligo PCR
4	Gibson Oligo Weaving
5	Error-Prone PCR

Cloning Techniques

Symbol	Technique
A	**A**geI/NcoI Digest
E	**E**rror-prone PCR Amplifcation
G	**G**ibson Assembly
H	**H**indIII Digest
L	**L**igation
M	**M**luI Digest
N	HindIII/**N**coI Digest
S	HindIII/NcoI Digest, **S**hort primers
W	PCR **W**eaving

Phage were then added in a 4:1 ratio to STEPL Buffer containing 10% w/v biotin-free powdered milk (LabScientific, final 2% w/v) and incubated at room temperature for 1 hour. Triglycine (Sigma) and $CaCl_2$ (Sigma) were added with final concentrations of 5 mM and 100 μM, respectively, in half of the sample and phage were allowed to cleave for 1 hour at 37°C. Phage were then added to a streptavidin-coated plate (see above) and incubated for 4 hours at 37°C. Then, wells were washed six times with TBS + 0.05% Tween-20 by vigorously pipetting and incubating with wash buffer for 4 minutes before shaking out the solution. Anti-M13 antibody conjugated to horseradish peroxidase (Abcam) was diluted to working concentrations in STEPL buffer with 2% w/v biotin-free powdered milk and incubated in the wells for 1 hour at 37°C. Wells were again washed six times with TBS + 0.05% Tween-20 by vigorously pipetting and incubating with wash buffer for 4 minutes before shaking out the solution. TMB substrate solution (Pierce) was added to the wells and allowed to react for 30 minutes before measuring the absorbance at 370 nm on a Tecan Infinate 200 Plate Reader. Positive controls were produced by incubating phage in unblocked wells. A control for streptavidin binding was produced by incubating C180A samples in blocked wells that had not been coated in streptavidin. All samples were run in triplicate.

Phage Library Cloning

Phage Libraries were cloned in a number of different ways. The component techniques are summarized in Table 5.6 and described below.

Library Generation: Method 1 – Kunkel Mutagenesis

First, pComb3XM/SBP-STEPL was electroporated into electrocompetent XL1-Blue cells (Agilent Genomics, 2500 V, 25 μF, 200 Ω, 2 mm cuvette). 6 colonies were picked in the morning and used to inoculate 200 mL of 2xYT Media with 100 μg/mL carbenicillin and 12.5 μg/mL tetracycline. The culture was shook at 37°C until it reached an OD_{600} of 0.5. 5 mL of helper phage (see above) was added to the culture and it was allowed to infect for 15 minutes at room temperature before being shook for an additional hour at 37°C. Cells were then pelleted by centrifugation at 3,000g for 15 min and resuspended in 200 mL of fresh 2xYT Media with 100 μg/mL carbenicillin, 12.5 μg/mL tetracycline, and 0.25 μg/mL uridine. The culture was then shook for 24 hours at 25°C. The following day, bacterial cells were removed from the overnight cultures by pelleting at 3,000g for 15 minutes. Sodium chloride (Fisher, 1.5M) and poly(ethylene glycol) (Sigma, MW 8000, 24% w/v) were added to the supernatant and the phage were allowed to precipitate at room temperature for 15 minutes. Phage were then pelleted by centrifugation at 15,000g for 10 minutes at 15°C. Tubes were inverted for 10 minutes to drain excess supernatant before phage were resuspended in 6 mL of PBS. Uracil-doped-single-stranded DNA (ssDNA) was then obtained with a QiaPrep® Spin M13 Kit (Qiagen).

4.4 pmol of 5'-phosphorylated mutagenic oligonucleotides (IDT) were annealed to 8 μg of ssDNA in buffer TM (50 M Tris−HCl, 10 mM $MgCl_2$, pH 7.5) by heating to 90°C for 2 minutes and then decreasing the temperature by 1°Cminute until it reached 25°C. The DNA was then supplemented with ATP (0.8 mM final concentration, New England Biolabs), DTT (5 mM final concentration, Fisher), dNTP mix (0.8 mM final concentration, each, Fisher), T4 ligase (20,000 units, New England Biolabs), and T7 polymerase (30 units, New England Biolabs). The reaction was run for 3 hours at 20°C followed by 20 minutes at 75°C to inactivate the enzymes. The QiaQuik PCR Cleanup Kit (Qiagen) was used to recover double-stranded DNA, which was then electroporated into 200 μL of electrocompetent XL1-Blue cells (Agilent Genomics, 2500 V, 25 μF, 200 Ω, 2 mm cuvette).

Library Generation: Method 2 – Site-Directed Mutagenesis

Methylated pComb3XM/SBP-STEPL was obtained via the QIAquick Spin miniprep Kit (Qiagen) from overnight cultures of transformed Top10 cells. 10 PCR reactions were run, each consisting

of 1X Phire Green Reaction Buffer (Thermo Fisher), 200 μM dNTPs, 2% v/v DMSO, 100 ng of pComb3XM/SBP-STEPL, 0.5 μM of each forward strand mutagenic oligonucleotide, and 1 unit Phire Hot Start II polymerase (Thermo Fisher). After 50 cycles of PCR, according to the product protocol, 20 units of the restriction enzyme DpnI (New Englnad Biolabs) was added to each reaction and incubated at 37°C for 2 hours to degrade the template DNA. Reactions were then concentrated and purified using the QIAquick PCR Cleanup Kit. DNA was then electroporated into 200 μL of electrocompetent XL1-Blue cells (Agilent Genomics, 2500 V, 25 μF, 200 Ω, 2 mm cuvette).

Library Generation: Method 3 – Degenerate Oligonucleotide PCR

The central portion of Sa-SrtAΔ59 was amplified from 50 ng of pComb3XM/SBP-STEPL using the sense-strand mutagenic oligonucleotides for the β_3/β_4 loop and the antisense-strand mutagenic oligonucleotide for the β_6/β_7 loop. After 50 cycles of PCR, according to the Phire Hot Start II polymerase protocol, product was purified from the template by agarose gel electrophoresis. The gel band was then cut out and purified with a QIAquick Gel Extraction Kit (Qiagen).

Library Generation: Method 4 – Gibson Oligonucleotide Assembly

A series of overlapping oligonucleotides were designed with melting temperatures near 50°C (see Figure 5.12), which were produced by IDT. Oligos were diluted to 10 μM and mixed together. The mixture was heated to 95°C for 5 minutes and then cooled 1°C/minute until reaching 25°C. 1 μL of the annealed oligos was then mixed with 9 μL of deionized water and 10 μL of Gibson Assembly Master Mix (New England Biolabs). The reaction was held at 60°C for 1 hour. After, the entire Gibson reaction was used as template DNA in a PCR reaction, using the primers Weave_Lib_Fwd and Weave_Lib_Rev (see Figure 5.19). After 50 cycles of PCR, according to the Phire Hot Start II polymerase protocol, product was purified from the template by agarose gel electrophoresis. The gel band was then cut out and purified with a QIAquick Gel Extraction Kit (Qiagen).

Library Generation: Method 5 – Error-Prone PCR

Random mutations were added to the wild-type Sa-SrtAΔ59 sequence by amplifying the entire sequence with an error-prone polymerase (Clontech Diversify Kit). 50 ng of pComb3XM/SBP-STEPL was amplified by the TITANIUM *Taq* Polymerase using the Weave_N_Fwd and Weave_C_Rev_Short

Figure 5.19: Library Cloning Primers.

primers (see Figure 5.19) with the addition of 640 μM $MnSO_4$ to further reduce the *Taq* fidelity to 3.5 mutations/kb (roughly 2 mutations per Sa-SrtAΔ59 sequence). The PCR product was then diluted 1:1000 and used as the template in a second round of error-prone PCR, with the same conditions (roughly 4 mutations per Sa-SrtAΔ59 sequence). All of the second round product was used in a high-fidelity PCR reaction (Phire Polymerase, Thermo Fisher) to amplify the library to the necessary quantities.

Cloning Technique E

For libraries with an E in the cloning strategy, the entire Sa-SrtAΔ59 gene was assembled by other methods (generally PCR weaving). The Sa-SrtAΔ59 fragment was then amplified by error-prone PCR in the same way as Library Generation Method 5 above.

Cloning Technique G

For libraries with a G in the cloning strategy, backbone vectors and libraries were amplified by PCR. Products were then subjected to DpnI digestion and agarose gel purification, which removed all template DNA. Purified fragments were again amplified by PCR to obtain at least 1 μg of backbone DNA. Backbone and library amplicons were combined in a 1:3 ratio and added to 2X Gibson Assembly Master Mix (New England Biolabs, 20 μL total reaction per 1 pmol of backbone DNA) and incubated at 50°C for 1-4 hours.

Phage Library Screening

Phage Screening: Libraries X4Gβ, X4WNLε, and XH4WSLα

6 μg of each library DNA was electroporated into 800 μL of electrocompetent XL1-Blue cells (Agilent Genomics, 2500 V, 25 μF, 200 Ω, 2 mm cuvette). Transformants were diluted into 5 mL of SOC Media (Corning) and allowed to recover by shaking for 1 hour at 37°C. Then, 8 mL of pre-warmed SB Media was added to the cells along with carbenicillin (20 μg/mL final concentration) and tetracycline (10 μg/mL final concentration) and shook for an hour at 37°C. Carbenicillin was then added to raise the concentration to 50 μg/mL and the cells were shook for another hour at 37°C. Next, 2 mL of CM13 helper phage (see above for preparation) was added and cells were incubated stationary at

room temperature for 15 minutes. Infected cultures were added to 185 mL of pre-warmed SB Media (50 μg/mL carbenicillin, 10 μg/mL tetracycline) and shook at 37°C for 2 hours. Finally, kanamycin was added to a final concentration of 70 μg/mL and cultures were shook overnight at 37°C.

For each day of screening, phage were harvested as described above (Phage Assay Validation). Phage were added to STEPL Buffer containing 10% w/v biotin-free powdered milk (LabScientific) and appropriate solutes to produce the negative screening conditions described in Section 5.5. Phage were then added to a streptavidin-coated plate (see above) and incubated for the amount of time indicated in Section 5.5 at 37°C. Then, wells were washed six times with TBS + 0.05% Tween-20 by vigorously pipetting and incubating for 4 minutes before shaking out the wash buffer. STEPL Buffer containing 2% w/v biotin-free powdered milk and appropriate solutes to produce the positive screening conditions described in Section 5.5 was then added to the wells, which were then incubated as described in Section 5.5. Supernatants were gently transferred to 2 mL of XL1-Blue cells (OD_{600} = 1) and allowed to infect for 15 minutes at room temperature. 2 μL of infected culture was used for titering.

6 mL of pre-warmed SB Media was added to the infected cells along with carbenicillin (20 μg/mL final concentration) and tetracycline (10 μg/mL final concentration) and shook for an hour at 37°C. Carbenicillin was then added to raise the concentration to 50 μg/mL and the cells were shook for another hour at 37°C. Next, 1 mL of CM13 helper phage was added and cells were incubated stationary at room temperature for 15 minutes. Infected cultures were added to 91 mL of pre-warmed SB Media (50 μg/mL carbenicillin, 10 μg/mL tetracycline) and shook at 37°C for 2 hours. Kanamycin was added to a concentration of 70 μg/mL and cultures were shook overnight at 37°C.

Phage Screening: Libraries G5.22-5ALα, X4WEALβ, X5ALα, and X4WSLβ

3 μg of each library DNA was electroporated into 300 μL of electrocompetent XL1-Blue cells (Agilent Genomics, 2500 V, 25 μF, 200 Ω, 2 mm cuvette). Transformants were diluted into 5 mL of SOC Media (Corning) and allowed to recover by shaking for 1 hour at 37°C. Then, 8 mL of pre-warmed SB Media (Gentrox) was added to the cells along with carbenicillin (20 μg/mL final concentration) and tetracycline (10 μg/mL final concentration) and shook for an hour at 37°C. Carbenicillin was then added to raise the concentration to 50 μg/mL and the cells were shook for another hour at 37°C. Next, 2 mL of CM13 helper phage (see above for preparation) was added and cells were

incubated stationary at room temperature for 15 minutes. Infected cultures were added to 185 mL of pre-warmed SB Media (50 μg/mL carbenicillin, 10 μg/mL tetracycline) and shook at 37°C for 2 hours. Finally, kanamycin was added to a concentration of 70 μg/mL and cultures were shook overnight at 37°C.

For each day of screening, phage were harvested as described above. Phage were added to STEPL Buffer containing 10% w/v biotin-free powdered milk (LabScientific) and appropriate solutes to produce the negative screening conditions described in Section 5.5 and incubated at room temperature for 1 hour. Phage were then added to a streptavidin-coated plate (see above) and incubated for 30 minutes at room temperature. Then, wells were washed six times with TBS + 0.05% Tween-20 by vigorously pipetting and immediately shaking out the wash buffer. Remaining phage were then eluted with acetate buffer (pH 4) for 10 minutes. Eluate was transferred to a microcentrifuge tube containing 3.5% v/v 0.5 M sodium hydroxide. 5X STEPL Buffer containing 10% w/v biotin-free powdered milk and appropriate solutes to produce the positive screening conditions described in Section 5.5 was then added to the neutralized phage, which were then incubated at room temperature for the amount of time conditions described in Section 5.5. With 30 minutes remaining in the incubation time, phage were added to a streptavidin-coated plate (see above) and incubated at room temperature. Supernatants were gently transferred to 2 mL of XL1-Blue cells ($OD_{600} = 1$) and allowed to infect for 15 minutes at room temperature. 2 μL of infected culture was used for titering. Finally, phage were amplified as described above.

Cloning of Library Results into pSTEPL/EGFP

The library regions of interesting clones and the pSTEPL/EGFP backbone were amplified by PCR and agarose gel purified. Products were then re-amplified to obtain at least 1 pmol of pSTEPL/EGFP backbone and 3 pmol library per reaction. The fragments were incubated in 1X Gibson Assembly Master Mix (New England Biolabs) for 1 hour at 50°C. 1:3 dilutions of the isothermal assembly reactions were then transformed into One-Shot Top10 cells (Life Technologies). Plasmids were verified by sequencing before being transformed into Rosetta2 BL21 DE3 cells (EMD Millipore) or the disulfide-forming SHuffle T7 Express cells (New England Biolabs).

Preliminary Clone Characterization

25mL starter cultures were grown overnight in LB-Ampicillin. These were added to 225mL of LB and grown to an OD_{600} of 0.8-1 before induction with 0.5mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Cultures were allowed to express for 24hrs at 25°C. Cells were then harvested by centrifugation (6000 RPM, 15min) and resuspended in 10mL of lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 1mg/mL Lysozyme, 1 EDTA-free cOmplete Mini protease inhibitor tablet (Roche), pH 7.5). Lysates were incubated at room temperature for 30min under gentle agitation before freezing overnight at -80°C. Samples were then thawed and incubated for 30min with DNAse I (Roche) under gentle agitation. Lysates were then sonicated and separated by centrifugation (10,000 rpm, 30min). 8mL of clarified lysate was incubated for 1 hr with 0.5mL Talon resin (Clontech, equilibrated in lysis buffer). The lysate and beads were then added to a column and beads were washed with 5mL STEPL buffer. Washed beads were resuspended in a total of 1mL of STEPL buffer containing the indicated amounts of $CaCl_2$, triglycine, and $MnCl_2$ (Sigma-Aldrich) and aliquoted into 1.5mL microcentrifuge tubes. Samples were shaken at 1,000 rpm, the indicated temperature, and protected from light. At each timepoint, samples were spun down at 3,000 rpm for 5min. GFP fluorescence was then observed with a UV transilluminator.

Chapter 6

Overall Conclusions and Future Directions

6.1 Thesis Discussion

6.1.1 Design and Optimization of Sortase-Tag Expressed Protein Ligation

The ability to attach a payload to a targeting ligand is a fundamental requirement in the field of molecular imaging. While this is simple to do if both the ligand and payload are small molecules, the problem becomes much more complex when either is a protein. Many chemical approaches to bioconjugation have been developed. However, these techniques are often not efficient, stoichiometric, or site specific. They may also exclude large numbers of potential ligands, such as those with disulfide bonds. Enzymatic bioconjugation techniques ameliorate some of these issues, but are also limited by constraining useable payloads to structures chemically similar to their natural substrates, requiring large fusions that alter the properties of the targeting ligand, or necessitating difficult downstream purifications. Keeping all of these considerations in mind, we set out to create a novel bioconjugation technology that combines the small size and flexibility of chemical techniques with the efficiency and specificity of enzymatic approaches.

We were able to do this by utilizing the transpeptidation activity of *S. aureus* Sortase A. In the presence of Ca^{2+}, the enzyme recognizes the pentapeptide sequence LPXTG and then catalyzes

the replacement of the glycine residue with any other polypeptide that has an N-terminal glycine. Nearly any molecule, biological or synthetic, can flank the LPXTG peptide or follow the N-terminal glycine of the second molecule, yielding an efficient, site-specific transformation unprecedented in its flexibility. Our novel technique, Sortase-Tag Expressed Protein Ligation (STEPL), harnessed the power of the sortase enzyme by expressing its active domain as part of a larger chimeric protein. The STEPL chimera consisted of, from N- to C-terminus, the protein to be conjugated, the sortase recognition motif LPETG, a flexible linker sequence, the sortase active domain, and a hexahistidine tag. The construct was then expressed in the *E. coli* cytoplasm and purified by metal affinity chromatography. However, rather than eluting the chimeric protein off of the metal resin, the bound beads were incubated with $CaCl_2$ and a short, synthetic peptide. This activated the sortase domain, cleaving the target protein off of the column via conjugation of the peptide. In theory, linking protein cleavage to bioconjugation should result in complete bioconjugation of all target protein that is released from the beads, eliminating the need for large excesses of synthetic peptide and greatly simplifying downstream workups.

In practice, the system works in much the same way. The major difference is due to a peptide-independent hydrolytic shunt in the sortase mechanism. To quantify the effect of this side reaction, a simple test system was constructed using the reporter protein EGFP as the target protein to be conjugated and triglycine as the synthetic peptide. We were able to monitor STEPL activity under a number of initial conditions by measuring the EGFP concentration in the supernatant, which would only occur if the enzyme actively cleaved the protein off of the beads. First, we used this assay to determine the optimal $CaCl_2$ concentration. While the hydrolytic shunt is active in the absence of peptide, calcium is necessary even for this side reaction. By titering $CaCl_2$ in the presence and absence of 25 μM triglycine, we calculated a ratio of ligated to unligated free EGFP as a function of calcium concentration. In less than 100 μM $CaCl_2$, the system reported a high percentage of conjugated product. This ratio quickly dropped off at higher concentrations, so 100 μM was determined to be the optimal calcium concentration for the STEPL system.

After that, we used the EGFP-STEPL reporter construct to analyze the effects of reaction time, reaction temperature, and peptide concentration. These parameters are not independent of one another, so a mass-action kinetics model was developed that used the Arrhenius definition of a rate constant to capture temperature-dependence. The model was fit to wet-lab quantification of EGFP

release under varying conditions to determine the pre-exponential constants and activation energies of both the glycine-dependent and glycine-independent pathways of the sortase mechanism. The calculated parameters had reasonable magnitudes and model-predicted EGFP release fit the data well. The model was finally used to predict the yield of conjugated protein, the percentage of released protein that would be conjugated, and the percentage of peptide consumed by the STEPL reaction. Conjugation percentage was heavily dependent on initial peptide concentration and weakly temperature dependent, with more peptide and higher temperatures leading to more pure results. A ratio of initial peptide to initial chimera that gave >95% conjugated protein could be found for all temperatures and times greater than 30 minutes. For a 6 hour reaction, only a two-fold excess of peptide was needed for 95% purity at 37°C. Peptide utilization was heavily dependent on both temperature and initial peptide concentrations. The peptide dependence was the inverse of conjugation purity, with peptide utilization increasing with decreasing initial peptide concentrations. Although peptide utilization of >90% were possible, this only occurred within 6 hours at 37°C. Finally, overall yield increased with temperature and peptide concentration. Thus, 37°C was determined to be the optimal temperature for all outcomes. Interestingly, peptide utilization and conjugate purity were found to be competing interests. Therefore, STEPL reactions can be optimized for peptide utilization if downstream applications facilitate easy separation of conjugated and unconjugated protein, such as subsequent surface functionalization, or for high purity if further purifications would be difficult, for instance if the protein will be used directly for imaging or therapy.

This novel system for protein purification and bioconjugation offers many advantages over other chemical and enzymatic ligation technologies by combining their advantages and avoiding their shortcomings. Like chemical crosslinkers, but unlike most enzymes, STEPL can conjugate a broad selection of bio-orthogonal functionality to a protein while adding minimal bulk. However, being an enzymatic reaction, it is more efficient, site-specific, and quantitative, unlike most chemical conjugations. Because STEPL consists of a single chimeric protein, it eliminates the need for subsequent purification of the sortase enzyme from the ligation products. What makes STEPL truly unique in the field of bioconjugation, however, is that it links protein purification to bioconjugation. In doing this, it can produce a high-purity, quantitatively labeled product or utilize most of an expensive reagent. Both approaches reduce the time, cost, and complexity of bioconjugation, making it a significant advance in the field with applications throughout academia and industry. STEPL has

the potential to aid in the production of protein-drug conjugates, PEGylated biologics, biophysical probes, and of course molecular imaging agents, among many other applications.

6.1.2 Applications of the STEPL System

Though promising, the STEPL system is only useful if it eases or enables the production of useful conjugated proteins. With that in mind, we tested STEPL's ability to create ligands for interesting downstream applications. For this thesis, STEPL was used to express the affibodies Z_{Her2} and Z_{EGFR}, small proteins that bind tightly and specifically to the extracellular domains of the Her2/neu and EGFR receptors, respectively. Affibodies were chosen because they express easily in the *E. coli* cytoplasm and are very stable after purification, allowing us to focus on the contributions of the STEPL system without adding complications derived from the targeting ligand itself.

Fluorescence Imaging

STEPL's first use was the generation of targeted fluorescence contrast agents. The Z_{Her2} affibody was cloned into pSTEPL and expressed in bacterial cells. A synthetic peptide containing the near-IR fluorophore HiLyteFluor750 was used to cleave the affibody off of the column. After simple dialysis of the column eluate into Tris-buffered saline to remove excess peptide, the sample was analyzed by SDS-PAGE, which showed a single, fluorescent band that corresponded to the size of Z_{Her2}. No unconjugated affibody could be visualized on the gel, indicating that STEPL did indeed produce a high-purity protein conjugate as predicted by the model. The conjugated affibody was then applied to Her2/neu positive and negative cell lines. Fluorescence imaging revealed that the fluorescent Z_{Her2} successful labeled only the Her2/neu positive cells, which could also be greatly reduced by competitive inhibition with unlabeled Z_{Her2}. In addition to microscopy, affibody labeling of the Her2/neu positive cell line was quantified with a fluorescence scanner, essentially creating an in-cell western reagent. These results were then duplicated by cleaving an Z_{EGFR}-STEPL construct with a synthetic peptide containing a different fluorophore, 5-FAM, a fluorescein derivative. Again, the reagent specifically labeled an EGFR-positive cell line, as determined by florescence microscopy. Thus, STEPL quickly and efficiently produced a fluorescence contrast agent with all of the expected properties and a simpler downstream purification. The only workup step, dialysis to remove excess labeling reagent, would be necessary in any system.

Nanoparticle Synthesis

The Z_{Her2}-STPEL chimera was also cleaved with a peptide containing a lysine residue that was modified to have an azide functional group, which is capable of participating in the azide-alkyne cycloaddition click reaction. While the label used for fluorescence imaging could be incorporated anywhere that does not interfere with binding, the C-terminal location of the azide functionalization has been proven to significantly improve the final performance of a targeted nanoparticle [2]. The Z_{Her2}-azide conjugate was then reacted with superparamagnetic iron oxide (SPIO) nanoparticles that displayed a strained cyclooctyne derivative, which efficiently formed covalent bonds with the azide moiety on the affibody. The resulting targeted SPIO particle was able to decrease the T_2^* relaxivity of Her2/neu positive cells, while inducing no change in the relaxivities of Her2/neu negative cells or those in subjected to competitive inhibition. These changes in T_2^* were echoed in MR images, which showed significant darkening of the Her2/neu positive cell lines. STEPL was able to quickly and efficiently create a targeted nanoparticle with all of its expected properties intact.

Modular Generation of Bispecific Ligands

The Z_{Her2} and Z_{EGFR} chimeras were each cleaved with three different peptides: one containing a fluorophore and a DBCO, one containing a fluorophore and a DBCO separated from the peptide by a flexible $(PEG)_4$ linker, and the azide peptide described above. DBCO-conjugated affibodies were then incubated with azide-conjugated affibodies to create a set of six dimeric affibodies Z_{Her2}-Z_{Her2}, Z_{Her2}-$(PEG)_4$-Z_{Her2}, Z_{Her2}-Z_{EGFR}, Z_{Her2}-$(PEG)_4$-Z_{EGFR}, Z_{EGFR}-Z_{EGFR}, and Z_{EGFR}-$(PEG)_4$-Z_{EGFR}. Dimerization was confirmed by SDS-PAGE. Affibody titers were then incubated with a Her2/neu-positive cell line, an EGFR-positive cell line, and a double-positive cell line. By measuring the mean fluorescence intensity of each binding reaction, we were able to calculate apparent dissociation constants for each of the affibody-cell line pairs and compare the effects of homodimerization against heterodimerization and using a long spacer against a short spacer. The calculated dissociation constants were then used to determine the concentration of heterodimers that will most enhance selection for the double-positive cell line, 10 nM. That concentration of each affibody was finally incubated with cells to obtain fluorescence images that supported our hypothesis that the bispecific ligand would show enhanced specificity for the double-positive cell line.

This was a simple illustration of what promises to be a much more powerful technology. Using STEPL and click chemistry as a method for generating modular bispecifics expands the field from specific IgG heavy-light chain interactions to any two ligands that can be expressed in bacterial cells. The inherent flexibility in STEPL would allow for the mixing of different classes of targeting ligands and easy assaying of different spacer moieties.

6.1.3 Directed Evolution of the STEPL System

The STEPL system has proven to be very useful, but it could always be better. One of STEPL's shortcomings was that there was always a reduction of protein yield due to calcium-induced cleavage within the expression cell. To remedy this, we proposed evolving the sortase enzyme to either require much higher calcium concentrations for activity or to require a different divalent metal ion entirely. We then attempted bacterial and phage display screens that would enable directed evolution of the sortase active domain in the STEPL context.

An *E. coli* display system was designed that used an autotransporter protein to facilitate transport of super-folder GFP (sfGFP), the LPETG recognition motif, a flexible linker, and the sortase active domain across the outer membrane and hold the chimera on the cell surface. This seemed to be the most promising option because sortase is naturally on the surface of gram-positive bacteria and the construct would be expressed in *E. coli*, where all the previous STEPL work had already been done. However, the system had no detectable sortase activity even in saturating levels of triglycine and calcium. Immunohistochemistry assays showed that sfGFP had indeed been exported to the surface of the cell, so the system was altered by introducing a second flexible linker to increase the distance between the cell surface and STEPL domains. This had no effect. Neither did separating the STEPL domains from the surface by including a fully functional ECFP domain or using a different autotransporter. Most likely, the sortase active domain did not re-fold correctly after being denatured during transport to the cell surface.

A yeast display system was briefly considered. However, data showed that the sortase domain was heavily glycosylated in the yeast endoplasmic reticulum. This could potentially have large and unknown effects on the evolutionary process, so yeast display was not pursued.

A phage display system was designed that linked the STEPL domains to the phage surface by the coat protein pIII. The STEPL "ligand" bound the structure to streptavidin either by directly

including the streptavidin binding peptide (SBP) or the biotin acceptor peptide, which is biotinylated *in vivo*. An active sortase domain would cleave off the ligand and release the phage from a streptavidin-coated plate, an easy physical separation that is much less subjective than sorting cells based on fluorescence. When assayed for activity, only the SBP-bearing phage bound to the plate. The phage displaying the biotin acceptor peptide probably did not spend enough time in the cytosolic compartment to be biotinylated, so that construct was discarded. The SBP-STEPL phage, however, showed robust activity in the phage system. Due to high levels of calcium in the milk-based blocking buffer, the only phage that did not cleave were those with the C180A mutation that eliminates sortase activity. Phage were assayed in a number of other blocking buffers, but none prevented non-specific binding as well as the 2% milk. Because all of the screening buffers would have contained high calcium concentrations in order to weaken its binding affinity for sortase, the biotin-free 2% milk was chosen as the blocking buffer. Thus, a phage display screening assay was validated and it was chosen as the method of directed evolution for the STEPL system.

Now that a screen had been established, a library to screen could be designed. The $\beta3/\beta4$ and $\beta6/\beta7$ loops constitute the calcium binding pocket in sortase, so mutations were focused on the twenty residues that make up those two loops. The potential library size of randomizing all twenty positions to all twenty amino acids was astronomically large, so an optimized library was designed. In this library, the available amino acids for each position would mimic the hydrophobicity of the wild-type residue. This design would ensure that the $\beta3/\beta4$ loop was still anchored to the core of the protein by two hydrophobic residues. It would also provide an abundance of residues capable of coordinating a metal ion. By restricting the codons of 18 of the 20 positions, the potential library size was reduced by a factor of 10^6, yielding a library that is still too large to sample most of the sequences but is enriched for sequences more likely to work. While cloning the library, it was observed that the standard method of using degenerate oligonucleotides as PCR primers resulted in biasing the library toward the wild-type sequence, especially for the GC bases. To avoid this effect, libraries were instead produced as a system of overlapping oligonucleotides that were combined by isothermal assembly. Thus, complimentary library regions were created by a polymerase without any bias toward the wild-type bases.

With a screen and a library, evolution could begin. The first evolution experiment consisted of five rounds of screening the phage and resulted in only five sequenceable plasmids, each unique.

These were cloned into the EGFP-STEPL construct used to validate the system and assayed in the same way. Unfortunately, all of the clones either cleaved under all conditions or did not cleave under any conditions.

Many of the sequenced plasmids had undergone recombination to delete the sortase sequence entirely, so the cloning methods were altered to add stability and the screen was run again. After another five rounds of screening, 26 sequences were returned, again all unique. Ten clones were chosen to be assayed. None were modulated by the addition of glycine or manganese, the ion chosen as a calcium replacement during the 4 hour reaction. When left overnight, one of the clones did cleave only in the presence of high calcium. However, this was independent of glycine. This was subjected to random mutagenesis to create a second generation screen, which returned no sequences.

The next library was screened under much harsher conditions to try and obtain a consensus sequence. After five rounds of screening, seventeen sequences were obtained, all unique yet again. All of the clones were assayed. None could be modulated by high calcium or manganese.

At this point, the screening method was altered to reduce non-specific dissociation from the streptavidin-coated plates. This screen included a 20-fold increase in manganese concentration in hopes of obtaining even a weakly dependent clone. After five rounds of screening, nine identical sequences were obtained. However, the sequence yielded inactive sortase under all conditions.

Finally, a library of only random mutations to the entire sortase sequence was screened. After five rounds, one sequence was obtained. This also was inactive under all conditions.

A constant bias for a cysteine at position 109, which was allowed to sample all twenty amino acids, indicated that the screens were indeed selecting for something. Unfortunately, they did not select for conditional sortase activity as intended. One theory is that the screens selected for calcium- or manganese-mediated streptavidin binding. Many affinity ligands are evolved using a similar randomization of two loops, so this was a distinct possibility. However, the SBP domain should have still held inactive phage to the plate, which was the most common phenotype assayed. Another possibility is that the screens selected for a metal-modulated SBP-binding domain, which would explain why they no longer bound the streptavidin in the positive screening buffers. It is also possible that this bound conformation was stabilized by a disulfide bond, explaining the consensus cysteine found in most of the sequences.

6.2 Future Directions

6.2.1 Extending STEPL to Other Expression Systems

An obvious next step in the development of STEPL is to extend it to other prokaryotic and eukaryotic expression systems. While the *E. coli* cytoplasm is a convenient and scalable place to express proteins, it is severely limited in the types of proteins that can be expressed. scFvs for example, have much more complex structures than affibodies and often contain multiple disulfide bonds, which cannot form in the reducing environment of the cytoplasm. When overexpressed proteins are poorly folded, they can precipitate *in vivo*, forming occlusion bodies that complicate harvesting proteins [268]. There are a few strategies for avoiding these precipitates, but the problem is usually best solved by changing expression systems entirely.

The simplest change is to express the STEPL chimera in the bacterial periplasm, which is an oxidizing environment that has been reported to be a favorable compartment for scFv expression [269]. This is accomplished by fusing the OmpA or PelB signaling peptide to the N-terminus of the chimera, before the targeting ligand. The peptide facilitates transport across the inner membrane before being proteolyzed, resulting in the original STEPL chimera. During phage display, the STEPL-pIII protein was transported to the periplasm via the OmpA signal peptide. Phage displaying the sortase enzyme showed activity, which is a good indication that periplasmically expressed STEPL chimeras will also remain functional. We have already begun working on expressing scFv-STEPL chimeras in the periplasm. The STEPL chimera has been cloned into the pET20(+) expression plasmid, which directs periplasmic expression via the PelB signal. Initial experiments resulted in poor yield of scFv fusion proteins, so the project has been delegated to the experts at the Wistar Institute's Protein Expression Facility.

When expressing complicated proteins, however, the best options are eukaryotic. Yeast, insect, and mammalian cells all have a secretory system capable of folding and secreting complex proteins with a myriad of post-translational modifications [270]. Yeast are particularly well suited for protein expression because they can be grown to high densities and secrete the protein for easy isolation and continuous manufacturing [271]. Although they grow slowly compared to bacterial cells, they outpace other eukaryotic systems by a considerable amount. Currently, a collaborator in the Muzykantov group at Penn is working to optimize expression of an scFv-STEPL fusion in the *P. pastoris* yeast

strain. By optimizing STEPL in eukaryotic expression systems, its utility would grow enormously as it could then be capable of expressing and conjugating full-length antibodies and enzyme-derived therapeutic proteins.

6.2.2 Improved Directed Evolution Strategies

While the directed evolution attempts described in Chapter 5 were unable to produce sortase enzymes with reduced background cleavage, this will still be necessary to accelerate adoption of the technology throughout the scientific community. To that end, there are a few potential options for improving the directed evolution strategy, depending on what the problem with the current system actually is.

In Section 5.5.4 we altered the screening protocol in an attempt to solve one potential problem, that SBP-streptavidin binding affinity may be too low, resulting in random release of inactive phage. Another solution to that problem is to change the affinity ligand from SBP to one with a stronger interaction with its target. We are already in the process of doing this, cloning a phagemid that displays a hexahistidine tag rather than SBP. The tight binding of His_6 with immobilized metal resins is well documented [272, 273, 274, 275]. Even in this thesis, we have shown that a cobalt resin is capable of retaining protein for a full 24 hours. Using a His-tag for immobilization may also reduce the potential to evolve a binding pocket for the affinity ligand because it is much shorter than SBP and lacks any sort of structure necessary for protein-protein interactions.

Another way to avoid evolving an affinity ligand binding pocket is to use larger ligands, such as monobodies or nanobodies. Even if a modulated, tight binder were to emerge, it would have a much lower chance of binding the ligand in a configuration that impairs binding to an immobilized target. To further ensure this, several copies of the ligand could be strung together one after another. Not only would that greatly increase the overall avidity to the plate, it would also leave extra ligands that are capable of binding the plate even if one of them was sequestered by an evolved binding loop, making it much more difficult for that phage to pass the screen.

Phage display itself may be too complicated of a system for the delicate evolution of an enzyme. If that is the case, a delicate screening system such as ribosome display may be more appropriate. Ribosome display is entirely *in vitro*, so it removes many possible complications, such as STEPL-phage interactions, the oxidizing environment of the periplasm, and transport across the inner membrane. It would also allow much larger libraries that could increase the possibility of recovering clones with the desired phenotype.

6.2.3 Triggered Cleavage with Dual-Sortase Constructs

Creating a manganese-triggered sortase would allow for more than just higher protein yields. It would also enable STEPL proteins with two independently active sortase domains. By creating a fusion protein that consists of a targeting ligand, Sa-SrtAΔ59$_{Mn}$, the sortase motif LPEGT, a flexible linker, Sa-SrtAΔ59, and a His$_6$ tag, it would be possible to use the Ca-triggered Sa-SrtAΔ59 to perform the standard STEPL conjugation and purification and later use the Mn-triggered Sa-SrtAΔ59 to cleave that label off of the protein. This would be especially useful in nuclear imaging. The radiotoxicity of PET and SPECT agents forces them to either use nuclides with short half-lives, agents that can be excreted quickly, or low doses of contrast agent [1, 276]. With the dual-STEPL system, a nuclide could be conjugated to the targeting ligand, which is then injected and allowed to equilibrate for contrast. After an image is obtained, Mn is injected to trigger the cleavage of the nuclide from the targeting ligand, allowing the toxin to be quickly excreted. Because the time that the targeting protein spends in the patient would be decoupled from the length of time that the radiolabel would be present, higher activities of more stable nuclei could be used. For example, the major nuclide used for PET, ^{18}F, has a half-life of 110 minutes. This short half-life allows high activities to be used for imaging, while limiting the radiation dose to the patient. It also requires an on-site cyclotron to produce the contrast agents, which greatly increases the expense of PET imaging and limits it to only a handful of clinics. With the dual-STEPL approach, it would be possible to produce the contrast agents off-site and in bulk with a different positron-emitter like ^{125}I, which has a half-life of 60 days, because the patient dose could be mitigated by cleavage rather than a short life-span.

6.3 Concluding Remarks

In this thesis, we have documented the design, optimization, and initial applications of a novel protein system capable of condensing protein purification and conjugation into a single enzymatic step. By modeling the underlying chemical principles of the STEPL system, we proved that it is one of the most efficient bioconjugation systems, capable of producing yields of more than 95% purity while at the same time reducing the amount of conjugate needed to achieve that purity. We then showed examples of STEPL's inherent flexibility and modularity by making a panel of affibody-

peptide conjugates that were used as monomeric and dimeric fluorescence imaging agents as well as targeting ligands for magnetic nanoparticles. As proteins continue to become a larger portion of the pharmaceutical portfolio, technologies like STEPL will become foundational to the future of medicine and biotechnology in general. This will be even more likely once STEPL has been adapted for use in eukaryotes and evolved to increase protein yields. The incredible flexibility of the sortase enzyme coupled with STEPL's cost-, material-, and time-efficient single-protein design give STEPL the potential for widespread use in academic and industrial settings.

References

[1] James, M. L. & Gambhir, S. S. A molecular imaging primer: modalities, imaging agents, and applications. *Physiological reviews* **92**, 897–965 (2012).

[2] Elias, D. R., Cheng, Z. & Tsourkas, A. An intein-mediated site-specific click conjugation strategy for improved tumor targeting of nanoparticle systems. *Small (Weinheim an der Bergstrasse, Germany)* **6**, 2460–8 (2010).

[3] Awada, A. & Aftimos, P. G. Targeted therapies of solid cancers: new options, new challenges. *Current opinion in oncology* **25**, 296–304 (2013).

[4] Bertos, N. & Park, M. Breast cancerone term, many entities? *The Journal of clinical investigation* **121**, 3789–3796 (2011).

[5] Harris, R. & Kinsinger, L. S. Less is more: Not "going the distance" and why. *Journal of the National Cancer Institute* **103**, 1726–1728 (2011).

[6] Welch, H. G. & Black, W. C. Overdiagnosis in cancer. *Journal of the National Cancer Institute* **102**, 605–613 (2010).

[7] Esserman, L., Shieh, Y. & Thompson, I. Rethinking Screening for Breast Cancer and Prostate Cancer. *JAMA: the Journal of the American Medical Association* **302**, 1685–1692 (2014).

[8] Gur, D. & Sumkin, J. H. Screening for Early Detection of Breast Cancer: Overdiagnosis versus Suboptimal Patient Management. *Radiology* **268**, 327–328 (2013).

[9] Klotz, L. Cancer overdiagnosis and overtreatment. *Current Opinion in Urology* **22**, 203–209 (2012).

[10] Elmore, J. G. & Fletcher, S. W. Overdiagnosis in breast cancer screening: Time to tackle an underappreciated harm. *Annals of Internal Medicine* **156**, 536–537 (2012).

[11] Esserman, L. J., Thompson, I. M., Jr & Reid, B. Overdiagnosis and overtreatment in cancer: An opportunity for improvement. *JAMA: the Journal of the American Medical Association* **310**, 797–798 (2013).

[12] Brody, J. E. Retesting Breast Cancer Axioms. *The New York Times*. November 11, 2014. Page D5.

[13] Grady, D. Look for Cancer, and Find It. *The New York Times*. April 8, 2014. Page D6.

[14] Aschwanden, C. The Case Against Early Cancer Detection. *FiveThirtyEight*. November 24, 2014.

[15] Miller, A. B. *et al.* Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ (Clinical research ed.)* **348**, g366 (2014).

[16] Kolata, G. & Ageim. Vast Study Casts Doubts on Value of Mammograms. *The New York Times* (2008). February 12, 2014. Page A1.

[17] Maitournam, a. & Simon, R. On the efficiency of targeted clinical trials. *Statistics in Medicine* **24**, 329–339 (2005).

[18] Mankoff, D. a., Pryma, D. a. & Clark, A. S. Molecular imaging biomarkers for oncology clinical trials. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine* **55**, 525–8 (2014).

[19] Hayashi. Analyzing global trends of biomarker use in drug interventional clinical studies. *Drug Discoveries & Therapeutics* **6**, 102–107 (2012).

[20] Goulart, B. H. L. *et al.* Trends in the use and role of biomarkers in phase I oncology trials. *Clinical Cancer Research* **13**, 6719–6726 (2007).

[21] Zelmer, a. & Ward, T. H. Noninvasive fluorescence imaging of small animals. *Journal of Microscopy* **252**, 8–15 (2013).

[22] Müller, J., Wunder, A. & Licha, K. Optical Imaging. In Schober, O. & Riemann, B. (eds.) *Molecular Imaging in Oncology*, vol. 187 of *Recent Results in Cancer Research*, 221–246 (Springer, Berlin, 2013).

[23] Mccann, T. E. *et al.* Molecular imaging of tumor invasion and metastases: The role of MRI. *NMR in Biomedicine* **24**, 561–568 (2011).

[24] Smith, G., Carroll, L. & Aboagye, E. O. New frontiers in the design and synthesis of imaging probes for PET oncology: current challenges and future directions. *Molecular imaging and biology* **14**, 653–66 (2012).

[25] Speck, U. Contrast Agents: X-ray Contrast Agents and Molecular Imaging A Contradiction? In Semmler, W. & Schwaiger, M. (eds.) *Molecular Imaging I*, vol. 185/1 of *Handbook of Experimental Pharmacology*, 167–175 (Springer Berlin Heidelberg, 2008).

[26] Hyvelin, J.-m. *et al.* Use of Ultrasound Contrast Agent Microbubbles in Preclinical Research. *Investigative Radiology* **48**, 570–583 (2013).

[27] Stokes, G. G. On the Change of Refrangibility of Light. *Philosophical Transactions of the Royal Society of London* **142**, 463–562 (1852).

[28] Marescaux, J. & Diana, M. Next step in minimally invasive surgery: hybrid image-guided surgery. *Journal of Pediatric Surgery* **50**, 30–36 (2014).

[29] Kovar, J. L., Simpson, M. a., Schutz-Geschwender, A. & Olive, D. M. A systematic approach to the development of fluorescent contrast agents for optical imaging of mouse cancer models. *Analytical biochemistry* **367**, 1–12 (2007).

[30] Caravan, P. Resonance Imaging (MRI) Contrast Agents : Design and Mechanism of Action. *Accounts of Chemical Research* **42**, 851–862 (2009).

[31] Koenig, S. H., Baglin, C. M. & Brown, R. D. Magnetic field dependence of solvent proton relaxation in aqueous solutions of Fe3+ complexes. *Magnetic Resonance in Medicine* **2**, 283–288 (1985).

[32] Oghabian, M. & Farahbakhsh, N. Potential use of nanoparticle based contrast agents in MRI: a molecular imaging perspective. *Journal of biomedical ...* **6**, 203–213 (2010).

[33] Penfield, J. G. & Reilly, R. F. What nephrologists need to know about gadolinium. *Nature clinical practice. Nephrology* **3**, 654–68 (2007).

[34] Bogdanov Jr, A., Matuszewski, L., Bremer, C., Petrovsky, A. & Weissleder, R. Oligomerization of Paramagnetic Substrates Result in Signal Amplification and can be Used for MR Imaging of Molecular Targets. *Molecular Imaging* **1**, 16–23 (2002).

[35] Cheng, Z., Thorek, D. L. J. & Tsourkas, A. Gadolinium-conjugated dendrimer nanoclusters as a tumor-targeted T1 magnetic resonance imaging contrast agent. *Angewandte Chemie (International ed. in English)* **49**, 346–50 (2010).

[36] Thorek, D. L. J., Chen, A. K., Czupryna, J. & Tsourkas, A. Superparamagnetic iron oxide nanoparticle probes for molecular imaging. *Annals of biomedical engineering* **34**, 23–38 (2006).

[37] Elias, A., Crayton, S. H., Warden-Rothman, R. & Tsourkas, A. Quantitative Comparison of Tumor Delivery for Multiple Targeted Nanoparticles Simultaneously by Multiplex ICP-MS. *Scientific reports* **4**, 5840 (2014).

[38] Elias, D. R., Poloukhtine, A., Popik, V. & Tsourkas, A. Effect of ligand density, receptor density, and nanoparticle size on cell targeting. *Nanomedicine : nanotechnology, biology, and medicine* **9**, 194–201 (2013).

[39] Hui, J. Z. *et al.* Facile method for the site-specific, covalent attachment of full-length IgG onto nanoparticles. *Small (Weinheim an der Bergstrasse, Germany)* **10**, 3354–63 (2014).

[40] Poelzing, S., Smoot, A. F. & Veeraraghavan, R. Novel x-ray attenuation mechanism: role of interatomic distance. *Medical physics* **35**, 4386–4395 (2008).

[41] Maeda, H., Wu, J., Sawa, T., Matsumura, Y. & Hori, K. Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review. *Journal of Controlled Release* **65**, 271–284 (2000).

[42] Konerding, M., Fait, E. & Gaumann, A. 3D microvascular architecture of pre-cancerous lesions and invasive carcinomas of the colon. *British journal of cancer* **84**, 1354–1362 (2001).

[43] Moghimi, S., Hunter, A. & Murray, J. Long-circulating and target-specific nanoparticles: theory to practice. *Pharmacological reviews* **53**, 283–318 (2001).

[44] Vonarbourg, A., Passirani, C., Saulnier, P. & Benoit, J.-P. Parameters influencing the stealthiness of colloidal drug delivery systems. *Biomaterials* **27**, 4356–73 (2006).

[45] Lemarchand, C., Gref, R. & Couvreur, P. Polysaccharide-decorated nanoparticles. *European journal of pharmaceutics and biopharmaceutics : official journal of Arbeitsgemeinschaft für Pharmazeutische Verfahrenstechnik e.V* **58**, 327–41 (2004).

[46] Lemarchand, C. *et al.* Influence of polysaccharide coating on the interactions of nanoparticles with biological systems. *Biomaterials* **27**, 108–18 (2006).

[47] Byrne, J. D., Betancourt, T. & Brannon-Peppas, L. Active targeting schemes for nanoparticle systems in cancer therapeutics. *Advanced drug delivery reviews* **60**, 1615–26 (2008).

[48] Jiang, W., Kim, B. Y. S., Rutka, J. T. & Chan, W. C. W. Nanoparticle-mediated cellular response is size-dependent. *Nature nanotechnology* **3**, 145–50 (2008).

[49] Daldrup-Link, H. E. *et al.* MR Imaging of Tumor Associated Macrophages with Clinically-Applicable Iron Oxide Nanoparticles. *Clinical cancer research : an official journal of the American Association for Cancer Research* 5695–5704 (2011).

[50] Thorek, D., Elias, D. R. & Tsourkas, A. Comparative analysis of nanoparticle-antibody conjugations: carbodiimide versus click chemistry. *Molecular Imaging* **8**, 221–229 (2009).

[51] Madan, J. *et al.* Enhanced noscapine delivery using estrogen-receptor-targeted nanoparticles for breast cancer therapy. *Anti-cancer drugs* **25**, 704–16 (2014).

[52] Banerjee, N. *et al.* Estrone-3-sulphate, a potential novel ligand for targeting breast cancers. *PloS one* **8**, e64069 (2013).

[53] Joner, M. *et al.* Site-specific targeting of nanoparticle prednisolone reduces in-stent restenosis in a rabbit model of established atheroma. *Arteriosclerosis, thrombosis, and vascular biology* **28**, 1960–6 (2008).

[54] Lee, C. *et al.* Novel chondroitin sulfate-binding cationic liposomes loaded with cisplatin efficiently suppress the local growth and liver metastasis of tumor cells in vivo. *Cancer research* **62**, 4282–4288 (2002).

[55] McCluskey, A. J., Olive, A. J., Starnbach, M. N. & Collier, R. J. Targeting HER2-positive cancer cells with receptor-redirected anthrax protective antigen. *Molecular oncology* **7**, 440–51 (2013).

[56] Salter, H. & Holland, R. Biomarkers: Refining diagnosis and expediting drug development - reality, aspiration and the role of open innovation. *Journal of Internal Medicine* **276**, 215–228 (2014).

[57] Forner, A. & Bruix, J. Biomarkers for early diagnosis of hepatocellular carcinoma. *Lancet Oncology* **13**, 750–751 (2012).

[58] Gramont, A. D. *et al.* Pragmatic issues in biomarker evaluation for targeted therapies in cancer. *Nature Publishing Group* 1–16 (2014).

[59] Bailey, A. M. *et al.* Implementation of Biomarker-Driven Cancer Therapy: Existing Tools and Remaining Gaps. *Discovery Medicine* **92**, 1–11 (2014).

[60] Wilson, A. M. *et al.* Beta2-Microglobulin as a biomarker in peripheral arterial disease: Proteomic profiling and clinical studies. *Circulation* **116**, 1396–1403 (2007).

[61] Lempiäinen, A., Hotakainen, K., Blomqvist, C., Alfthan, H. & Stenman, U. H. k. Hyperglycosylated human chorionic gonadotropin in serum of testicular cancer patients. *Clinical Chemistry* **58**, 1123–1129 (2012).

[62] Menczer, J., Ben-Shem, E., Golan, A. & Levy, T. The Significance of Normal Pretreatment Levels of CA125 (¡35 U/mL) in Epithelial Ovarian Carcinoma. *Rambam Maimonides Medical Journal* **6**, e0005 (2015).

[63] Ruibal, A. *et al.* CA15.3 Serum Concentrations in Older Women with Infiltrating Ductal Carcinomas of the Breast. *International Journal of Molecular Sciences* **15**, 19870–19876 (2014).

[64] Kim, J. *et al.* Postoperative Carcinoembryonic Antigen as a Complementary Tumor Marker of Carbohydrate Antigen 19-9 in Pancreatic Ductal Adenocarcinoma. *Journal of Korean Medical Science* **30**, 259–263 (2015).

[65] Jeong, Y. *et al.* Postoperative Radiotherapy for Gallbladder Cancer. *Anticancer Research* **34**, 5621–5629 (2014).

[66] Macuks, R., Baidekalna, I. & Donina, S. An ovarian cancer malignancy risk index composed of HE4, CA125, ultrasonographic score, and menopausal status: Use in differentiation of ovarian cancers and benign lesions. *Tumor Biology* **33**, 1811–1817 (2012).

[67] Wang, Y.-h. *et al.* Chromogranin A as a Marker for Diagnosis, Treatment, and Survival in Patients With Gastroenteropancreatic Neuroendocrine Neoplasm. *Medicine* **93**, e247 (2014).

[68] Cherenfant, J. *et al.* Comparison of tumor markers for predicting outcomes after resection of nonfunctioning pancreatic neuroendocrine tumors. *Surgery* **156**, 1504–10; discussion 1510–1 (2014).

[69] Moore, R. G. *et al.* Serum levels of the ovarian cancer biomarker HE4 are decreased in pregnancy and increase with age. *American Journal of Obstetrics and Gynecology* **206**, 349.e1–349.e7 (2012).

[70] Karagiannis, P., Fittall, M. & Karagiannis, S. N. Evaluating biomarkers in melanoma. *Frontiers in Oncology* **4**, 1–11 (2014).

[71] Toss, A. *et al.* Ovarian cancer: Can proteomics give new insights for therapy and diagnosis? *International Journal of Molecular Sciences* **14**, 8271–8290 (2013).

[72] Tsao, A. S. *et al.* Elevated PDGFRB gene copy number gain is prognostic for improved survival outcomes in resected malignant pleural mesothelioma. *Annals of Diagnostic Pathology* **18**, 140–145 (2014).

[73] Carpenter, K. L. H. *et al.* Systemic, Local, and Imaging Biomarkers of Brain Injury: More Needed, and Better Use of Those Already Established? *Frontiers in Neurology* **6**, 1–20 (2015).

[74] Rinaldi, S. *et al.* Thyroid-stimulating hormone, thyroglobulin, and thyroid hormones and risk of differentiated thyroid carcinoma: The EPIC study. *Journal of the National Cancer Institute* **106** (2014).

[75] Duffy, M. J., McGowan, P. M., Harbeck, N., Thomssen, C. & Schmitt, M. uPA and PAI-1 as biomarkers in breast cancer: validated for clinical use in level-of-evidence-1 studies. *Breast Cancer Research* **16**, 428 (2014).

[76] Agrawal, a., Gutteridge, E., Gee, J. M. W., Nicholson, R. I. & Robertson, J. F. R. Overview of tyrosine kinase inhibitors in clinical breast cancer. *Endocrine-related cancer* **12 Suppl 1**, S135–44 (2005).

[77] Flynn, J. F., Wong, C. & Wu, J. M. Anti-EGFR Therapy: Mechanism and Advances in Clinical Efficacy in Breast Cancer. *Journal of oncology* **2009**, 526963 (2009).

[78] Milane, L., Duan, Z. & Amiji, M. Development of EGFR-targeted polymer blend nanocarriers for combination paclitaxel/lonidamine delivery to treat multi-drug resistance in human breast and ovarian tumor cells. *Molecular pharmaceutics* **8**, 185–203 (2010).

[79] Medina, O. P. *et al.* Optimizing tumor targeting of the lipophilic EGFR-binding radiotracer SKI 243 using a liposomal nanoparticle delivery system. *Journal of controlled release : official journal of the Controlled Release Society* **149**, 292–8 (2011).

[80] Lin, S.-X. *et al.* Molecular therapy of breast cancer: progress and future directions. *Nature Reviews Endocrinology* **6**, 485–493 (2010).

[81] Santin, A. D., Bellone, S., Roman, J. J., McKenney, J. K. & Pecorelli, S. Trastuzumab treatment in patients with advanced or recurrent endometrial carcinoma overexpressing HER2/neu. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* **102**, 128–31 (2008).

[82] Izumi, Y., Xu, L., di Tomaso, E., Fukumura, D. & Jain, R. Tumour biology: herceptin acts as an anti-angiogenic cocktail. *Nature* **416**, 279–280 (2002).

[83] Conlin, A. K. *et al.* Phase II trial of weekly nanoparticle albumin-bound paclitaxel with carboplatin and trastuzumab as first-line therapy for women with HER2-overexpressing metastatic breast cancer. *Clinical breast cancer* **10**, 281–7 (2010).

[84] Ito, A. *et al.* Magnetite nanoparticle-loaded anti-HER2 immunoliposomes for combination of antibody therapy with hyperthermia. *Cancer letters* **212**, 167–75 (2004).

[85] Tan, M. & Yu, D. Molecular Mechanisms of ErbB2-Mediated Breast Cancer Chemoresistance. *Advances in Experimental Medicine and Biology* **608**, 119–129 (2007).

[86] Weng, X. *et al.* Carbon nanotubes as a protein toxin transporter for selective HER2-positive breast cancer cell destruction. *Molecular bioSystems* **5**, 1224–31 (2009).

[87] Saul, J. M., Annapragada, A. V. & Bellamkonda, R. V. A dual-ligand approach for enhancing targeting selectivity of therapeutic nanocarriers. *Journal of controlled release : official journal of the Controlled Release Society* **114**, 277–87 (2006).

[88] May, C., Sapra, P. & Gerber, H.-P. Advances in bispecific biotherapeutics for the treatment of cancer. *Biochemical pharmacology* **84**, 1105–12 (2012).

[89] Yan, Y. *et al.* A new 18F-labeled BBN-RGD peptide heterodimer with a symmetric linker for prostate cancer imaging. *Amino Acids* **41**, 439–447 (2011).

[90] Nie, Y. *et al.* Dual-targeted polyplexes: one step towards a synthetic virus for cancer gene therapy. *Journal of controlled release : official journal of the Controlled Release Society* **152**, 127–34 (2011).

[91] Lu, D. *et al.* A fully human recombinant IgG-like bispecific antibody to both the epidermal growth factor receptor and the insulin-like growth factor receptor for enhanced antitumor activity. *The Journal of biological chemistry* **280**, 19665–72 (2005).

[92] Meng, S. *et al.* Enhanced antitumor effect of novel dual-targeted paclitaxel liposomes. *Nanotechnology* **21**, 415103 (2010).

[93] Willmann, J., Lutz, A. & Paulmurugan, R. Dual-targeted Contrast Agent for US Assessment of Tumor Angiogenesis in Vivo. *Radiology* **248** (2008).

[94] Yazaki, P. J. *et al.* A series of anti-CEA/anti-DOTA bispecific antibody formats evaluated for pre-targeting: comparison of tumor uptake and blood clearance. *Protein engineering, design & selection : PEDS* **26**, 187–93 (2013).

[95] Watts, R. J. & Dennis, M. S. Bispecific antibodies for delivery into the brain. *Current opinion in chemical biology* **17**, 393–9 (2013).

[96] Yu, Y. J. *et al.* Boosting brain uptake of a therapeutic antibody by reducing its affinity for a transcytosis target. *Science translational medicine* **3**, 84ra44 (2011).

[97] Sanford, M. Blinatumomab: First Global Approval. *Drugs* 321–327 (2015).

[98] Bollard, C. Blinatumomab Knocks Out Residual Disease. *Cancer Discovery* **5**, 103–103 (2015).

[99] Zhao, H., Cui, K., Muschenborn, A. & Wong, S. T. Progress of engineered antibody-targeted molecular imaging for solid tumors. *Molecular Medicine Reports* **1**, 131–134 (2008).

[100] Petering, J., McManamny, P. & Honeyman, J. Antibody therapeutics - the evolving patent landscape. *New biotechnology* **28**, 538–44 (2011).

[101] Milstein, C. The hybridoma revolution: an offshoot of basic research. *BioEssays* **21**, 966–973 (1999).

[102] Carson, K. R. *et al.* Monoclonal antibody-associated progressive multifocal leucoencephalopathy in patients treated with rituximab, natalizumab, and efalizumab: a Review from the Research on Adverse Drug Events and Reports (RADAR) Project. *The Lancet. Oncology* **10**, 816–24 (2009).

[103] Kashmiri, S. V. S., De Pascalis, R., Gonzales, N. R. & Schlom, J. SDR grafting–a new approach to antibody humanization. *Methods (San Diego, Calif.)* **36**, 25–34 (2005).

[104] Hawe, A., Hulse, W. L., Jiskoot, W. & Forbes, R. T. Taylor dispersion analysis compared to dynamic light scattering for the size analysis of therapeutic peptides and proteins and their aggregates. *Pharmaceutical Research* **28**, 2302–2310 (2011).

[105] Lobo, E. D., Hansen, R. J. & Balthasar, J. P. Antibody pharmacokinetics and pharmacodynamics. *Journal of pharmaceutical sciences* **93**, 2645–68 (2004).

[106] Siontorou, C. G. Nanobodies as novel agents for disease diagnosis and therapy. *International journal of nanomedicine* **8**, 4215–4227 (2013).

[107] Eigenbrot, C., Ultsch, M., Dubnovitsky, A., Abrahmsén, L. & Härd, T. Structural basis for high-affinity HER2 receptor binding by an engineered protein. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 15039–15044 (2010).

[108] Feldwisch, J. & Tolmachev, V. Engineering of Affibody Molecules for Therapy and Diagnostics. In Voynov, V. & Caravella, J. A. (eds.) *Therapeutic Proteins*, vol. 899 of *Methods in Molecular Biology*, 103–126 (Humana Press, Totowa, NJ, 2012).

[109] Nilsson, B. *et al.* A synthetic IgG-binding domain based on staphylococcal protein a. *Protein Engineering, Design and Selection* **1**, 107–113 (1987).

[110] Arora, P., Oas, T. G. & Myers, J. K. Fast and faster: a designed variant of the B-domain of protein A folds in 3 microsec. *Protein science : a publication of the Protein Society* **13**, 847–853 (2004).

[111] Lin, X., Xie, J. & Chen, X. Protein-based tumor molecular imaging probes. *Amino acids* **41**, 1013–36 (2011).

[112] Friedman, M. *et al.* Engineering and characterization of a bispecific HER2 x EGFR-binding affibody molecule. *Biotechnology and applied biochemistry* **54**, 121–31 (2009).

[113] Orlova, A. *et al.* Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer research* **66**, 4339–48 (2006).

[114] Qi, S. *et al.* Evaluation of Four Affibody-Based Near-Infrared Fluorescent Probes for Optical Imaging of Epidermal Growth Factor Receptor Positive Tumors. *Bioconjugate chemistry* **23**, 1149–1156 (2012).

[115] Vargo, K. B., Zaki, A. A., Warden-Rothman, R., Tsourkas, A. & Hammer, D. A. Superparamagnetic Iron Oxide Nanoparticle Micelles Stabilized by Recombinant Oleosin for Targeted Magnetic Resonance Imaging. *Small* **EPub**, 1–5 (2014).

[116] Gebauer, M. & Skerra, A. Engineered protein scaffolds as next-generation antibody therapeutics. *Current opinion in chemical biology* **13**, 245–55 (2009).

[117] Sha, F. *et al.* Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 14924–9 (2013).

[118] Schönfeld, D. *et al.* An engineered lipocalin specific for CTLA-4 reveals a combining site with structural and conformational features similar to antibodies. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 8198–8203 (2009).

[119] Kohl, A. *et al.* Allosteric inhibition of aminoglycoside phosphotransferase by a designed ankyrin repeat protein. *Structure* **13**, 1131–1141 (2005).

[120] Daly, N. L. *et al.* Structural insights into the role of the cyclic backbone in a squash trypsin inhibitor. *The Journal of biological chemistry* **288**, 0–18 (2013).

[121] Koide, A. & Koide, S. Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. In Arndt, K. M. & Müller, K. M. (eds.) *Methods in molecular biology (Clifton, N.J.)*, vol. 352 of *Methods in Molecular Biology*, 95–109 (Humana Press, Totowa, NJ, 2007).

[122] Skerra, A. Alternative binding proteins: Anticalins - Harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. *FEBS Journal* **275**, 2677–2683 (2008).

[123] Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPins: A new generation of protein therapeutics. *Drug Discovery Today* **13**, 695–701 (2008).

[124] Kolmar, H. Alternative binding proteins: Biological activity and therapeutic potential of cystine-knot miniproteins. *FEBS Journal* **275**, 2684–2690 (2008).

[125] Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic labeling of proteins: techniques and approaches. *Bioconjugate chemistry* **24**, 1277–94 (2013).

[126] Rana, S., Yeh, Y.-C. & Rotello, V. M. Engineering the nanoparticle-protein interface: applications and possibilities. *Current opinion in chemical biology* **14**, 828–34 (2010).

[127] Sapsford, K. E. *et al.* Functionalizing Nanoparticles with Biological Molecules: Developing Chemistries that Facilitate Nanotechnology. *Chemical reviews* **113**, 1904–2074 (2013).

[128] Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nature chemical biology* **7**, 876–84 (2011).

[129] Kocbek, P., Obermajer, N., Cegnar, M., Kos, J. & Kristl, J. Targeting cancer cells using PLGA nanoparticles surface modified with monoclonal antibody. *Journal of Controlled Release* **120**, 18–26 (2007).

[130] Smith, M. H. The amino acid composition of proteins. *Journal of Theoretical Biology* **13**, 261–282 (1966).

[131] Lowe, A. B. Thiol-ene click reactions and recent applications in polymer and materials synthesis. *Polymer Chemistry* **1**, 17 (2010).

[132] Hayworth, D. Overview of Crosslinking and Protein Modification: Accessed 9 Feb 2015. *http://www.piercenet.com/method/overview-crosslinking-protein-modification.*

[133] Fonge, H. *et al.* Preliminary in vivo evaluation of a novel 99mTc-Labeled HYNIC-cys-annexin A5 as an apoptosis imaging agent. *Bioorganic and Medicinal Chemistry Letters* **18**, 3794–3798 (2008).

[134] Rodrigues, M. L. *et al.* Engineering Fab' fragments for efficient F(ab)2 formation in Escherichia coli and for improved in vivo stability. *Journal of immunology* **151**, 6954–6961 (1993).

[135] Doronina, S. O. *et al.* Development of potent monoclonal antibody auristatin conjugates for cancer therapy. *Nature biotechnology* **21**, 778–784 (2003).

[136] Hamblett, K. J. *et al.* Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate Effects of Drug Loading on the Antitumor Activity of a Monoclonal Antibody Drug Conjugate. *Clinical cancer research : an official journal of the American Association for Cancer Research* **10**, 7063–7070 (2004).

[137] McDonagh, C. F. *et al.* Engineered antibody-drug conjugates with defined sites and stoichiometries of drug attachment. *Protein Engineering, Design and Selection* **19**, 299–307 (2006).

[138] Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click chemistry: Diveverse chemical funtion from a few good reactions. *Angewandte Chemie International Edition* **40**, 2004–2012 (2001).

[139] Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angewandte Chemie - International Edition* **41**, 2596–2599 (2002).

[140] Taylor, M. S., Zalatan, D. N., Lerchner, A. M. & Jacobsen, E. N. Highly enantioselective conjugate additions to alpha,beta-unsaturated ketones catalyzed by a (salen)Al complex. *Journal of the American Chemical Society* **127**, 1313–1317 (2005).

[141] Gaetke, L. M. & Chow, C. K. Copper toxicity, oxidative stress, and antioxidant nutrients. *Toxicology* **189**, 147–163 (2003).

[142] Agard, N. J., Prescher, J. a. & Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide - Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *Journal of the American Chemical Society* **126**, 15046–15047 (2004).

[143] Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 16793–16797 (2007).

[144] Ning, X., Guo, J., Wolfert, M. a. & Boons, G. J. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. *Angewandte Chemie - International Edition* **47**, 2253–2255 (2008).

[145] Wood, D. W. & Camarero, J. a. Intein Applications: From Protein Purification and Labeling to Metabolic Control Methods. *The Journal of biological chemistry* **289**, 14512–14519 (2014).

[146] Borra, R., Dong, D., Elnagar, A. Y., Woldemariam, G. a. & Camarero, J. a. In-cell fluorescence activation and labeling of proteins mediated by FRET-quenched split inteins. *Journal of the American Chemical Society* **134**, 6344–53 (2012).

[147] Chong, S. *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271–81 (1997).

[148] Clancy, K. W., Melvin, J. a. & McCafferty, D. G. Sortase transpeptidases: insights into mechanism, substrate specificity, and inhibition. *Biopolymers* **94**, 385–96 (2010).

[149] Popp, M. W.-L. & Ploegh, H. L. Making and breaking peptide bonds: protein engineering using sortase. *Angewandte Chemie (International ed. in English)* **50**, 5024–32 (2011).

[150] Proft, T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnology letters* **32**, 1–10 (2010).

[151] Ton-That, H. & Mazmanian, S. Anchoring of Surface Proteins to the Cell Wall of Staphylococcus aureus SORTASE CATALYZED IN VITRO TRANSPEPTIDATION REACTION USING LPXTG. *Journal of Biological ...* **275**, 9876–9881 (2000).

[152] Paterson, B. M. *et al.* Enzyme-mediated site-specific bioconjugation of metal complexes to proteins: sortase-mediated coupling of copper-64 to a single-chain antibody. *Angewandte Chemie (International ed. in English)* **53**, 6115–9 (2014).

[153] Kobashigawa, Y., Kumeta, H., Ogura, K. & Inagaki, F. Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. *Journal of biomolecular NMR* **43**, 145–50 (2009).

[154] Wu, S. & Proft, T. The use of sortase-mediated ligation for the immobilisation of bacterial adhesins onto fluorescence-labelled microspheres: a novel approach to analyse bacterial adhesion to host cells. *Biotechnology letters* **32**, 1713–8 (2010).

[155] Li, Y.-M. *et al.* Irreversible site-specific hydrazinolysis of proteins by use of sortase. *Angewandte Chemie (International ed. in English)* **53**, 2198–202 (2014).

[156] Bellucci, J. J., Amiram, M., Bhattacharyya, J., McCafferty, D. & Chilkoti, A. Three-in-One Chromatography-Free Purification, Tag Removal, and Site-Specific Modification of Recombinant Fusion Proteins Using Sortase A and Elastin-like Polypeptides. *Angewandte Chemie (International ed. in English)* 1–7 (2013).

[157] Chen, I., Howarth, M., Lin, W. & Ting, A. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nature methods* **2**, 99–104 (2005).

[158] Howarth, M. *et al.* Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nature methods* **5**, 397–9 (2008).

[159] Liang, M. *et al.* Multimodality nuclear and fluorescence tumor imaging in mice using a streptavidin nanoparticle. *Bioconjugate chemistry* **21**, 1385–8 (2010).

[160] Perrault, S. D. & Chan, W. C. W. In vivo assembly of nanoparticle components to improve targeted cancer imaging. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11194–9 (2010).

[161] Barat, B. & Wu, A. M. Metabolic biotinylation of recombinant antibody by biotin ligase retained in the endoplasmic reticulum. *Biomolecular engineering* **24**, 283–91 (2007).

[162] Slavoff, S. a., Chen, I., Choi, Y.-A. & Ting, A. Y. Expanding the substrate tolerance of biotin ligase through exploration of enzymes from diverse species. *Journal of the American Chemical Society* **130**, 1160–2 (2008).

[163] Strop, P. Versatility of microbial transglutaminase. *Bioconjugate chemistry* **25**, 855–62 (2014).

[164] Dennler, P. *et al.* Transglutaminase-based chemo-enzymatic conjugation approach yields homogeneous antibody-drug conjugates. *Bioconjugate chemistry* **25**, 569–78 (2014).

[165] Mero, A., Schiavon, M., Veronese, F. M. & Pasut, G. A new method to increase selectivity of transglutaminase mediated PEGylation of salmon calcitonin and human growth hormone. *Journal of controlled release : official journal of the Controlled Release Society* **154**, 27–34 (2011).

[166] Lin, C.-W. & Ting, A. Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *Journal of the American Chemical Society* **128**, 4542–3 (2006).

[167] Sung, K., Kamiya, N., Kawata, N., Kamiya, S. & Goto, M. Functional glass surface displaying a glutamyl donor substrate for transglutaminase-mediated protein immobilization. *Biotechnology journal* **5**, 456–62 (2010).

[168] Kitaoka, M. *et al.* Transglutaminase-mediated synthesis of a DNA-(enzyme)n probe for highly sensitive DNA detection. *Chemistry (Weinheim an der Bergstrasse, Germany)* **17**, 5387–92 (2011).

[169] Hudak, J. E. *et al.* Synthesis of heterobifunctional protein fusions using copper-free click chemistry and the aldehyde tag. *Angewandte Chemie (International ed. in English)* **51**, 4161–5 (2012).

[170] Rush, J. S. & Bertozzi, C. R. New aldehyde tag sequences identified by screening formylglycine generating enzymes in vitro and in vivo. *Journal of the American Chemical Society* **130**, 12240–12241 (2008).

[171] Rabuka, D., Rush, J. S., DeHart, G. W., Wu, P. & Bertozzi, C. R. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nature protocols* **7**, 1052–67 (2012).

[172] Smith, E. L. *et al.* Chemoenzymatic Fc glycosylation via engineered aldehyde tags. *Bioconjugate chemistry* **25**, 788–95 (2014).

[173] Drake, P. M. *et al.* Aldehyde tag coupled with HIPS chemistry enables the production of ADCs conjugated site-specifically to different antibody regions with distinct in vivo efficacy and PK outcomes. *Bioconjugate Chemistry* **25**, 1331–1341 (2014).

[174] Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D. & Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 46–51 (2013).

[175] Li, X., Fang, T. & Boons, G.-J. Preparation of Well-Defined Antibody-Drug Conjugates through Glycan Remodeling and Strain-Promoted Azide-Alkyne Cycloadditions. *Angewandte Chemie* **126**, 7307–7310 (2014).

[176] Zhou, Q. *et al.* Site-specific antibody-drug conjugation through glycoengineering. *Bioconjugate chemistry* **25**, 510–20 (2014).

[177] Ta, H. T., Peter, K. & Hagemeyer, C. E. Enzymatic antibody tagging: toward a universal biocompatible targeting tool. *Trends in cardiovascular medicine* **22**, 105–11 (2012).

[178] Boeggeman, E. *et al.* Site specific conjugation of fluoroprobes to the remodeled Fc N-glycans of monoclonal antibodies using mutant glycosyltransferases: application for cell surface antigen detection. *Bioconjugate chemistry* **20**, 1228–36 (2009).

[179] Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules. *in vivo, Nat. Biotechnol.* **21**, 86–89 (2003).

[180] Zhao, Z. W. *et al.* Spatial organization of RNA polymerase II inside a mammalian cell nucleus revealed by reflected light-sheet superresolution microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 681–6 (2014).

[181] Calebiro, D. *et al.* Single-molecule analysis of fluorescently labeled G-protein-coupled receptors reveals complexes with distinct dynamics and organization. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 743–8 (2013).

[182] Gautier, A. *et al.* An engineered protein tag for multiprotein labeling in living cells. *Chemistry & biology* **15**, 128–36 (2008).

[183] Maurel, D. *et al.* Cell-surface protein-protein interaction analysis with time-resolved FRET and snap-tag technologies: application to GPCR oligomerization. *Nature methods* **5**, 561–567 (2008).

[184] Landgraf, D., Okumus, B., Chien, P., Baker, T. a. & Paulsson, J. Segregation of molecules at cell division reveals native protein localization. *Nature Methods* (2012). 480–482.

[185] Sadelain, M., Brentjens, R. & Rivière, I. The promise and potential pitfalls of chimeric antigen receptors. *Current opinion in immunology* **21**, 215–23 (2009).

[186] Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nature biotechnology* **27**, 753–9 (2009).

[187] Wriggers, W., Chakravarty, S. & Jennings, P. a. Control of protein functional dynamics by peptide linkers. *Biopolymers* **80**, 736–46 (2005).

[188] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1 (2010). `http://www.pymol.org`.

[189] Simons, K., Bonneau, R., Ruczinski, I. & Baker, D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* **Suppl 3**, 171–176 (1999).

[190] Huang, C. C., Meng, E. C., Morris, J. H., Pettersen, E. F. & Ferrin, T. E. Enhancing UCSF Chimera through web services. *Nucleic Acids Research* **42**, 1–7 (2014).

[191] Han, J. *et al.* Structure-Based Rational Design of a Toll-like Receptor 4 (TLR4) Decoy Receptor with High Binding Affinity for a Target Protein. *PLoS ONE* **7**, e30929 (2012).

[192] Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* **487**, 545–74 (2011).

[193] Richter, F. *et al.* Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *Journal of the American Chemical Society* **134**, 16197–16206 (2012).

[194] Stauffer, C. E. & Etson, D. The effect on subtilisin activity of oxidizing a methionine residue. *Journal of Biological Chemistry* **244**, 5333–5338 (1969).

[195] Estell, D. a., Graycar, T. P. & Wells, J. a. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *Journal of Biological Chemistry* **260**, 6518–6521 (1985).

[196] Barberis, S., Quiroga, E., Barcia, C. & Liggieri, C. Effect of laundry detergent formulation on the performance of alkaline phytoproteases. *Electronic Journal of Biotechnology* **16**, 1–8 (2013).

[197] Cooper, a. a., Chen, Y. J., Lindorfer, M. a. & Stevens, T. H. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *The EMBO journal* **12**, 2575–2583 (1993).

[198] Chong, S. *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271–281 (1997).

[199] Reitman, Z. J. *et al.* Enzyme redesign guided by cancer-derived IDH1 mutations. *Nature Chemical Biology* **8**, 887–889 (2012).

[200] Eigen, M. *et al.* Geometry Is of the Code ? Genetic Provides tRNA Statistical an Answer. *Science* **244**, 673–679 (1989).

[201] Wals, K. & Ovaa, H. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Frontiers in chemistry* **2**, 15 (2014).

[202] Hui, J. & Tsourkas, A. Optimization of Photoactive Protein Z for Fast and Efficient Site-Specific Conjugation of Native IgG. *Bioconjugate chemistry* **25**, 1709–1719 (2014).

[203] Zhang, W. H., Otting, G. & Jackson, C. J. Protein engineering with unnatural amino acids. *Current opinion in structural biology* **23**, 581–7 (2013).

[204] Pantoja, R., Rodriguez, E. a., Dibas, M. I., Dougherty, D. a. & Lester, H. a. Single-molecule imaging of a fluorescent unnatural amino acid incorporated into nicotinic receptors. *Biophysical Journal* **96**, 226–237 (2009).

[205] Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *Journal of the American Chemical Society* **133**, 2124–2127 (2011). PMID21271704.

[206] Pless, S. a. *et al.* A cation-π interaction at a phenylalanine residue in the glycine receptor binding site is conserved for different agonists. *Molecular pharmacology* **79**, 742–748 (2011).

[207] Grönwall, C. *et al.* Selection and characterization of Affibody ligands binding to Alzheimer amyloid beta peptides. *Journal of biotechnology* **128**, 162–83 (2007).

[208] van Bloois, E., Winter, R. T., Kolmar, H. & Fraaije, M. W. Decorating microbes: surface display of proteins on Escherichia coli. *Trends in Biotechnology* **29**, 79–86 (2010).

[209] Binder, U., Matschiner, G., Theobald, I. & Skerra, A. High-throughput sorting of an Anticalin library via EspP-mediated functional display on the Escherichia coli cell surface. *Journal of molecular biology* **400**, 783–802 (2010).

[210] Gai, S. A. & Wittrup, K. D. Yeast surface display for protein engineering and characterization. *Current opinion in structural biology* **17**, 467–73 (2007).

[211] Meehl, M. a. & Stadheim, T. a. Biopharmaceutical discovery and production in yeast. *Current opinion in biotechnology* **30C**, 120–127 (2014).

[212] Shaheen, H. H. *et al.* A dual-mode surface display system for the maturation and production of monoclonal antibodies in glyco-engineered Pichia pastoris. *PloS one* **8**, e70190 (2013).

[213] Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nature protocols* **1**, 755–68 (2006).

[214] Kierny, M. R., Cunningham, T. D. & Kay, B. K. Detection of biomarkers using recombinant antibodies coupled to nanostructured platforms. *Nano Reviews* **3**, 1–24 (2012).

[215] Barbas, C. F., Burton, D. R., Scott, J. K. & Silverman, G. J. *Phage Display: A Laboratory Manual*. No. p. 532 in Phage Display: A Laboratory Manual (Cold Spring Harbor Laboratory Press, 2004).

[216] Barbas, C. F., Kang, a. S., Lerner, R. a. & Benkovic, S. J. Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 7978–82 (1991).

[217] Siegel, D. L. & Silberstein, L. E. Expression and characterization of recombinant anti-Rh(D) antibodies on filamentous phage: a model system for isolating human red blood cell antibodies by repertoire cloning. *Blood* **83**, 2334–44 (1994).

[218] Pansri, P., Jaruseranee, N., Rangnoi, K., Kristensen, P. & Yamabhai, M. A compact phage display human scFv library for selection of antibodies to a wide variety of antigens. *BMC biotechnology* **9**, 6 (2009).

[219] Huang, R., Fang, P. & Kay, B. K. Improvements to the Kunkel mutagenesis protocol for constructing primary and secondary phage-display libraries. *Methods (San Diego, Calif.)* **58**, 10–7 (2012).

[220] Askoxylakis, V. *et al.* Challenges in optimizing a prostate carcinoma binding peptide, identified through the phage display technology. *Molecules (Basel, Switzerland)* **16**, 1559–78 (2011).

[221] Sun, N., Funke, S. A. & Willbold, D. Mirror image phage display - Generating stable therapeutically and diagnostically active peptides with biotechnological means. *Journal of Biotechnology* **161**, 121–125 (2012).

[222] Griffiths, a. D. *et al.* Isolation of high affinity human antibodies directly from large synthetic repertoires. *The EMBO journal* **13**, 3245–3260 (1994).

[223] Bábíčková, J., Tóthová, u., Boor, P. & Celec, P. In vivo phage display–a discovery tool in molecular biomedicine. *Biotechnology advances* **31**, 1247–59 (2013).

[224] Arap, W., Pasqualini, R. & Ruoslahti, E. Cancer Treatment by Targeted Drug Delivery to Tumor Vasculature in a Mouse Model. *Science* **279**, 377–380 (1998).

[225] Pasqualini, R. & Ruoslahti, E. Organ targeting In vivo using phage display peptide libraries. *Nature* **380**, 364–366 (1996).

[226] Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).

[227] Plückthun, A. Ribosome Display: A Perspective. In Douthwaite, J. A. & Jackson, R. H. (eds.) *Ribosome Display and Related Technologies SE - 1*, vol. 805 of *Methods in Molecular Biology*, 3–28 (Springer New York, 2012).

[228] Blumberg, D. D. *Guide to Molecular Cloning Techniques*, vol. 152 of *Methods in Enzymology* (Elsevier, 1987).

[229] Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. USA* **100**, 9727–9732 (2003).

[230] Firnberg, E. & Ostermeier, M. PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS ONE* **7**, 1–10 (2012).

[231] Reetz, M. T. & Wu, S. Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions. *Chemical communications (Cambridge, England)* 5499–5501 (2008).

[232] Tang, L. *et al.* Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques* **52**, 149–158 (2012).

[233] Mena, M. a. & Daugherty, P. S. Automated design of degenerate codon libraries. *Protein Engineering, Design and Selection* **18**, 559–561 (2005).

[234] Steffens, D. L. & Williams, J. G. K. Efficient site-directed saturation mutagenesis using degenerate oligonucleotides. *Journal of Biomolecular Techniques* **18**, 147–149 (2007).

[235] Patrick, W. M. & Firth, A. E. Strategies and computational tools for improving randomized protein libraries. *Biomolecular engineering* **22**, 105–12 (2005).

[236] Warden-Rothman, R., Caturegli, I., Popik, V. & Tsourkas, A. Sortase-Tag Expressed Protein Ligation: Combining Protein Purification and Site-Specific Bioconjugation into a Single Step. *Analytical chemistry* **85**, 11090–11097 (2013).

[237] Spirig, T., Weiner, E. M. & Clubb, R. T. Sortase enzymes in Gram-positive bacteria. *Molecular microbiology* **82**, 1044–1059 (2011).

[238] Mao, H. A self-cleavable sortase fusion for one-step purification of free recombinant proteins. *Protein expression and purification* **37**, 253–63 (2004).

[239] Zong, Y., Bice, T. W., Ton-That, H., Schneewind, O. & Narayana, S. V. L. Crystal structures of Staphylococcus aureus sortase A and its substrate complex. *The Journal of biological chemistry* **279**, 31383–9 (2004).

[240] Junutula, J. R. *et al.* Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotechnology* **26**, 925–932 (2008).

[241] Steinhagen, M., Zunker, K., Nordsieck, K. & Beck-Sickinger, A. G. Large scale modification of biomolecules using immobilized sortase A from Staphylococcus aureus. *Bioorganic & medicinal chemistry* **21**, 3504–10 (2013).

[242] Fong, B. a., Wu, W.-Y. & Wood, D. W. The potential role of self-cleaving purification tags in commercial-scale processes. *Trends in biotechnology* **28**, 272–9 (2010).

[243] Norris, V., Grant, S. & Freestone, P. Calcium signalling in bacteria. *Journal of bacteriology* **178**, 3677–3682 (1996).

[244] Zhou, G. Q. & Zhong, W. Z. Diffusion-Controlled Reactions of Enzymes. *European Journal of Biochemistry* **387**, 383–387 (1982).

[245] Alberty, R. & Hammes, G. Application of the theory of diffusion-controlled reactions to enzyme kinetics. *The Journal of Physical Chemistry* **62**, 154–159 (1958).

[246] McRae, S. R., Brown, C. L. & Bushell, G. R. Rapid purification of EGFP, EYFP, and ECFP with high yield and purity. *Protein expression and purification* **41**, 121–7 (2005).

[247] Offterdinger, M., Georget, V., Girod, A. & Bastiaens, P. I. H. Imaging phosphorylation dynamics of the epidermal growth factor receptor. *The Journal of biological chemistry* **279**, 36972–81 (2004).

[248] Lallana, E., Sousa-Herves, A., Fernandez-Trillo, F., Riguera, R. & Fernandez-Megia, E. Click chemistry for drug delivery nanosystems. *Pharmaceutical research* **29**, 1–34 (2012).

[249] Ridgway, J. B., Presta, L. G. & Carter, P. 'Knobs-into-holes' engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein engineering* **9**, 617–621 (1996).

[250] Orcutt, K. D. *et al.* A modular IgG-scFv bispecific antibody topology. *Protein Engineering, Design and Selection* **23**, 221–228 (2010).

[251] Gavrilyuk, J. I. *et al.* An efficient chemical approach to bispecific antibodies and antibodies of high valency. *Bioorganic and Medicinal Chemistry Letters* **19**, 3716–3720 (2009).

[252] Wagner, K. *et al.* Bispecific antibody generated with sortase and click chemistry has broad antiinfluenza virus activity. *Proceedings of the National Academy of Sciences* **111**, 16820–16825 (2014).

[253] Watanabe, H. *et al.* Biomimetic engineering of modular bispecific antibodies for biomolecule immobilization. *Langmuir* **27**, 9656–9661 (2011).

[254] Neidhardt, F., Bloch, P. & Smith, D. Culture medium for enterobacteria. *Journal of bacteriology* **119**, 736–747 (1974).

[255] Fisher, A. C. & DeLisa, M. P. Laboratory evolution of fast-folding green fluorescent protein using secretory pathway quality control. *PloS one* **3**, e2351 (2008).

[256] Strijbis, K., Spooner, E. & Ploegh, H. L. Protein ligation in living cells using sortase. *Traffic (Copenhagen, Denmark)* (2012).

[257] Dorr, B., Ham, H., An, C., Chaikof, E. & Liu, D. Reprogramming the specificity of sortase enzymes. *Proceedings of the National Academy of Sciences* **111**, 13343–13348 (2014).

[258] Keefe, a. D., Wilson, D. S., Seelig, B. & Szostak, J. W. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein expression and purification* **23**, 440–6 (2001).

[259] Basic Report 01212 , Milk , dry , whole , without added vitamin D. Tech. Rep., United States Department of Agriculture (2015).

[260] Versieck, J. & Cornelis, R. Normal levels of trace elements in human blood plasma or serum. *Analytica Chimica Acta* **116**, 217–254 (1980).

[261] Hospira Inc. Chromium 4mcg/mL [package insert] (2004).

[262] Hospira Inc. Copper 0.4 mg/mL [package insert] (2009).

[263] Hospira Inc. Manganese 0.1 mg/mL [package insert] (2009).

[264] Hospira Inc. Zinc 1 mg/mL [package insert] (2004).

[265] Naik, M. T. *et al.* Staphylococcus aureus Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. *The Journal of biological chemistry* **281**, 1817–26 (2006).

[266] Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA assembly design automation software. *ACS synthetic biology* **1**, 14–21 (2012).

[267] Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* **6**, 343–5 (2009).

[268] Schlapschy, M., Grimm, S. & Skerra, A. A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in Escherichia coli. *Protein Engineering, Design and Selection* **19**, 385–390 (2006).

[269] Miller, K. D., Weaver-Feldhaus, J., Gray, S. A., Siegel, R. W. & Feldhaus, M. J. Production, purification, and characterization of human scFv antibodies expressed in Saccharomyces cerevisiae, Pichia pastoris, and Escherichia coli. *Protein expression and purification* **42**, 255–67 (2005).

[270] Geisse, S., Gram, H., Kleuser, B. & Kocher, H. P. Eukaryotic expression systems: a comparison. *Protein expression and purification* **8**, 271–82 (1996).

[271] Digan, M. *et al.* Continuous production of a novel lysozyme via secretion from the yeast, Pichia pastoris. *Biotechnology* **7**, 160–164 (1989).

[272] Pasquinelli, R. S., Shepherd, R. E., Koepsel, R. R., Zhao, A. & Ataai, M. M. Design of affinity tags for one-step protein purification from immobilized zinc columns. *Biotechnology Progress* **16**, 86–91 (2000).

[273] Knecht, S., Ricklin, D., Eberle, A. N. & Ernst, B. Oligohis-tags: Mechanisms of binding to Ni2+-NTA surfaces. *Journal of Molecular Recognition* **22**, 270–279 (2009).

[274] Gaberc-Porekar, V. & Menart, V. Perspectives of immobilized-metal affinity chromatography. *Journal of Biochemical and Biophysical Methods* **49**, 335–360 (2001).

[275] Ueda, E., Gout, P. & Morganti, L. Current and prospective applications of metal ionprotein binding. *Journal of Chromatography A* **988**, 1–23 (2003).

[276] Agdeppa, E. D. & Spilker, M. E. A review of imaging agent development. *The AAPS journal* **11**, 286–299 (2009).