CYTOSOLIC DELIVERY OF INHIBITORY ANTIBODIES WITH CATIONIC LIPIDS

Hejia Henry Wang

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

_____

Andrew Tsourkas

Professor of Bioengineering

Graduate Group Chairperson

_____

Kim Sharp. Associate Professor of Biochemistry and Biophysics

Dissertation Committee

Rahul Kohli, Assistant Professor of Medicine

E. James Petersson, Associate Professor of Chemistry

Mark Greene, John W. Eckman Professor of Medical Science

Michael Mitchell, Skirkanich Assistant Professor of Innovation and Bioengineering

CYTOSOLIC DELIVERY OF INHIBITORY ANTIBODIES WITH CATIONIC LIPIDS

COPYRIGHT

2019

Hejia Henry Wang

This work is licensed under the

Creative Commons Attribution

NonCommercial-NoDerivatives 4.0 International

License

To view a copy of this license, visit

https://creativecommons.org/licenses/by-nc-sa/4.0/us/

Acknowledgment

I would like to thank my advisor, Andrew Tsourkas, for letting me embark on such a crazy thesis project, helping me troubleshoot when initial experiments inevitably fall short, and giving me the space to become independent and grow as a scientist. I want to thank all past and current members of the Tsourkas lab for helping me bounce ideas off them and for putting up with my grouchiness (at times) in lab. I want to thank my thesis committee – Rahul Kohli, Mark Greene, and James Petersson – for providing me guidance and helping me realize when I'm venturing too far deep into a scientific rabbit hole.

It takes much more than a passing interest in science to motivate someone to get a PhD. I want to thank everyone who helped to foster my passion in science, but especially James Shorter – who let me start working in his lab as a sophomore at Penn – and Zamia Diaz, who taught me so much in the Shorter lab. I want to thank James Petersson again, but for teaching a phenomenal Chem 700 Chemical Biology course, which has had a profound influence on my current scientific interests

I want to thank my parents for pushing me to go into science and medicine as well as supporting me through my time at Penn. I want to also thank all of my friends from medical school and the Biochemistry & Molecular Biophysics graduate group who kept me sane during these years. Finally, I am grateful to my girlfriend, Tanya Kapoor, who has filled my life with so much joy for the last 2 years of my PhD.

ABSTRACT

CYTOSOLIC DELIVERY OF INHIBITORY ANTIBODIES WITH CATIONIC LIPIDS

Hejia Henry Wang

Andrew Tsourkas

Antibodies have become powerful therapeutics and research tools because they can be developed to directly inhibit almost any protein, but their inaility to enter the cytosol limits inhibitory antibodies to extracellular targets. Given that roughly two-thirds of drug targets lie inside of cells and many targets lack binding targets for small-molecule drugs, developing a cytosolic antibody delivery system would dramatically expand the druggable proteome. Cytosolic antibodies also offer unique opportunities to directly inhibit and study intracellular protein function.

Here we demonstrate that immunoglobulin G (IgG) antibodies that are conjugated with anionic polypeptides (ApPs) can be complexed with cationic lipids through electrostatic interactions, enabling close to 90% cytosolic delivery efficiency with only 500 nM IgG. The ApP is fused to a small photoreactive antibody-binding domain (pAbBD) that can be site-specifically photocrosslinked to nearly all off-the-shelf IgGs without perturbing IgG binding affinity. Furthermore, the pAbBD can be functionalized with chemical moieties such as fluorophores at its C-terminus via proximity-based sortase-mediated ligation (PBSL), a chemoenzymatic bioconjugation approach that we have developed.

We show that cytosolically delivered IgGs can inhibit the drug efflux pump multidrug resistance-associated protein 1 (MRP1) and the transcription factor NFκB. This work establishes a new approach for using existing antibody collections to modulate intracellular protein function and provides the foundations for therapeutic cytosolic antibodies.

Table of Contents

Acknowledgment iii

ABSTRACT iv

List of Tables ix

List of Illustrations x

Chapter 1 – Background 1

1.1 Why cytosolic antibody delivery? 1

1.2 Antibodies 2
1.2.1 Immunoglobulin G (IgG) structure 2
1.2.2 IgG function 4
1.2.3 Generating antibodies for research and therapeutic uses 5
1.2.4 Research applications for antibodies 6
1.2.5 scFvs, nanobodies, and antibody-like binding proteins 8
1.2.6 Therapeutic monoclonal antibodies 10
1.2.7 Novel mechanisms of action for therapeutic mAbs – antibody-drug conjugates (ADCs) 11
1.2.8 Novel mechanisms of action for therapeutic mAbs – bispecific antibodies (bsAbs) 12
1.2.9 The frontiers of therapeutic antibodies 13

1.3 Protein Bioconjugation 14
1.3.1 Classical side chain labeling chemistries 14
1.3.2 Glycoconjugation 16
1.3.3 Bioorthogonal Reactions 17
1.3.4 Native and expressed chemical ligation 18
1.3.5 Genetic incorporation of unnatural amino acids 19
1.3.6 Peptide tags and chemoenzymatic bioconjugation 21
1.3.7 Site-specific protein bioconjugation with sortase A (SrtA) 23
1.3.8 Self-labeling protein domains 26
1.3.9 Light activated site-specific conjugation of native IgGs (LASIC) 28

1.4 Cytosolic Protein Delivery 30
1.4.1 Physical delivery methods 30
1.4.2 Physical delivery of inhibitory antibodies 31
1.4.3 Intrabodies 32
1.4.4 Detecting cytosolic protein delivery 33
1.4.5 Cell-penetrating peptides (CPPs) 36
1.4.6 Improving the endosomal escape efficiency of cell-penetrating peptides 38
1.4.7 Activatable cell penetrating peptides (aCPPs) 39
1.4.8 Cytosolic delivery with bacterial toxins 39
1.4.9 Small proteins with intrinsic cytosolic delivery capabilities 40
1.4.10 Antibodies with intrinsic cytosolic delivery capabilities 41

1.4.11 Nanocarriers for cytosolic protein delivery 42
1.4.12 Targeting with nanocarriers 45

1.5 References 47

Chapter 2 – Proximity based sortase-mediated ligation 70

2.1 Abstract 70

2.2 Introduction 70

2.3 Results 72

2.4 Discussion 78

2.5 Supplementary Information 79

2.6 Methods 85
2.6.1 Cloning 85
2.6.2 Protein Expression 86
2.6.3 SpyCatcher-SrtA and SrtA-SpyCatcher Resin Preparation 86
2.6.4 eGFP-LPETG-SpyTag Studies 87
2.6.5 SpyCatcher-SrtA and SrtA-SpyCatcher Release Modeling 90
2.6.6 ScFv Purification and Labeling Studies 91

2.7 References 91

Chapter 3 – Cytosolic delivery of proteins with cell-penetrating peptides.. 95

3.1 Introduction 95

3.2 Results 97
3.2.1 Detecting cytosolic protein delivery with a splitGFP reporter system 97
3.2.2 Cytosolic delivery of pAbBD with CPPs. 99
3.2.3 CPPs fail to cytosolically deliver antibodies 101

3.3 Discussion 103

3.4 Supplementary Figure 104

3.5 Methods 105
3.5.1 Cloning 105
3.5.2 pAbBD expression and purification 107
3.5.3 His6-splitGFP(1-10) expression, purification, and complementation 108
3.5.4 Photocrosslinking pAbBD variants to Rituximab 109
3.5.5 Cell culture 109
3.5.6 Protein delivery 110
3.5.7 Fluorescence microscopy 110
3.5.8 Flow cytometry 110

3.6 References **110**

Chapter 4 – Cytosolic delivery of inhibitory antibodies with cationic lipids **114**

4.1 Abstract **114**

4.2 Introduction **114**

4.3 Results **116**
4.3.1 Stringent detection of cytosolic protein delivery 116
4.3.2 Cytosolic pAbBD delivery 118
4.3.3 Cytosolic IgG delivery 120
4.3.4 A cytosolically delivered IgG can inhibit MRP1, a drug efflux pump 123
4.3.5 Cytosolically delivered IgGs can inhibit the transcription factor NFκB. 125

4.4 Discussion **127**

4.5 Supplementary Information **130**

4.6 Methods **146**
4.6.1 Plasmids and Cloning 146
4.6.2 pAbBD expression and purification 147
4.6.3 His6-splitGFP(1-10) expression, purification, and complementation 147
4.6.4 Photocrosslinking pAbBD variants to IgGs 147
4.6.5 Cell culture 147
4.6.6 splitGFP(1-10) reporter cell line generation 148
4.6.7 Protein delivery 149
4.6.8 Fluorescence microscopy 149
4.6.9 Flow cytometry 149
4.6.10 Calcein efflux assay 149
4.6.11 Doxorubicin and vincristine sensitization 150
4.6.12 RelA (NFκB) immunofluorescence 151
4.6.13 NFκB transcriptional activity 151
4.6.14 Statistical Analysis 152

4.7 References **152**

Chapter 5 – Summary and Future Directions **156**

5.1 Proximity-based sortase-mediated ligation **156**
5.1.1 Summary 156
5.1.2 High purity protein purification with PBSL 157

5.2 Cytosolic delivery of proteins with cell-penetrating peptides **161**
5.2.1 Summary 161
5.2.2 Screening for improved cell-penetrating peptide and endosomolytic peptide pairs 161

5.3 Cytosolic delivery of inhibitory antibodies with cationic lipids **164**
5.3.1 Summary 164

5.3.2 Additional delivery characterization 164
5.3.3 Improving cytosolic delivery 165
5.3.4 Cytosolic delivery of nanobodies and antibody-like binding proteins 166
5.3.5 Antibody-induced protein degradation 170
5.3.6 *In vivo* delivery 172

5.4 Concluding remarks 174

5.5 References 174

List of Tables

Table 2.1 PBSL amino acid sequences. 84

Table 3.1 List of purified proteins used for CPP studies 105

Table 4.1 List of purified proteins used for cytosolic delivery of IgGs. 143

List of Illustrations

Figure 1.1 Schematic of IgG Structure. 2

Figure 1.2 SrtA ligation mechanism. 24

Figure 1.3 Schematic of STEPL. 26

Figure 2.1 PBSL Schematic 72

Figure 2.2 SDS-PAGE of PBSL. 73

Figure 2.3 PBSL Capture Efficiency 73

Figure 2.4 PBSL release yields and predicted ligation efficiency. 74

Figure 2.5 Model of eGFP release. 75

Figure 2.6 PBSL Comparisons. 77

Figure 2.7 PBSL resin stability. 79

Figure 2.8 PBSL eGFP MALDI-TOF. 80

Figure 2.9 PBSL labeling in solution and on resin. 81

Figure 2.10 PBSL CD3 scFv MALDI-TOF. 81

Figure 2.11 PBSL ScFv Quantification 82

Figure 2.12 PBSL kinetics. 83

Figure 3.1 Schematic of splitGFP reporter system. 96

Figure 3.2 SDS-PAGE of pAbBD-S11 photocrosslinking to Rituximab. 97

Figure 3.3 pAbBD-S11 and Ritux-$(pAbBD\text{-}S11)_2$ can complement splitGFP(1-10) in vitro. 98

Figure 3.4 pAbBD-S11 and Ritux-$(pAbBD\text{-}S11)_2$ can complement splitGFP(1-10) in cells. 98

Figure 3.5 Cytosolic pAbBD delivery with the TAT CPP 99

Figure 3.6 Cytosolic pAbBD delivery with the R8 CPP 100

Figure 3.7 Cytosolic pAbBD delivery with dual CPPs. 100

Figure 3.8 Comparisons of different CPP formats for cytosolic pAbBD delivery. 101

Figure 3.9 CPPs fail to cytosolically deliver Rituximab. 102

Figure 3.10 Cytosolic delivery with 80µM pAbBD-CPP 104

Figure 4.1 Schematic of antibody delivery approach. 117

Figure 4.2 Optimizing ApPs for cytosolic pAbBD delivery. 119

Figure 4.3 Optimizing ApPs for cytosolic IgG delivery. 121

Figure 4.4 IgG delivery scope. 122

Figure 4.5 Cytosolic QCRL3 delivery inhibits MRP1. 124

Figure 4.6 Cytosolic anti-RelA IgG delivery inhibits NFκB. 126

Figure 4.7 SDS-PAGE of pAbBD-ApP-S11 and Ritux-(pAbBD-ApP-S11)$_2$. 130

Figure 4.8 Detailed flow cytometry data for optimizing ApPs for pAbBD delivery. 132

Figure 4.9 Detailed flow cytometry data for optimizing ApPs for Ritux-(pAbBD)$_2$ delivery with Lipo 2000. 134

Figure 4.10 Optimizing ApPs for cytosolic IgG delivery with Lipo RNAiMax. 136

Figure 4.11 IgG delivery scope with a 25 residues-long poly-glutamate ApP. 137

Figure 4.12 Detailed flow cytometry data for cytosolic IgG delivery with additional cationic lipids in HEK293T splitGFP(1-10) cells. 138

Figure 4.13 Detailed flow cytometry data for cytosolic IgG delivery with additional cationic lipids in A549 and HT1080 splitGFP(1-10) cells. 140

Figure 4.14 RelA immunofluorescence quantification. 142

Figure 4.15 Cytosolic delivery efficiencies for NFκB experiment IgGs. 142

Figure 5.1 Comparing C-terminal PBSL purification purity and yield against IMAC. 158

Figure 5.2 C-terminal PBSL purification elution kinetics. 159

Figure 5.3 Schematic of N-terminal PBSL purification. 159

Figure 5.4 C-terminal PBSL purification capture kinetics. 160

Figure 5.5 CPP and ELP library generation. 162

Figure 5.6 Cytosolic delivery of αGFPnb, αTaqAffi, and DARPinK27. 167

Figure 5.7 DARPinK27 inhibits KRas-G12S signaling in A549 cells. 168

Chapter 1 – Background

1.1 Why cytosolic antibody delivery?

With the advent of the genomic revolution, increasing portions of the human proteome have been causatively linked to human disease[1–4]. However, the disease-associated proteome overlaps poorly with the druggable proteome – the human proteins that can bind drug-like small molecules – significantly limiting the potential for basic science research on the biological mechanisms underlying disease to translate into therapeutics[1–4]. This is most apparent in oncology, where famous oncogenes such as Myc have been studied for decades without any therapeutic breakthroughs[5].

The poor overlap between the disease-associated and the druggable proteome is primarily attributable to the need for small-molecule therapeutics to bind to target proteins via small pockets that are integral to the protein's biological activity[1,2,6]. These pockets are frequently found in privileged protein classes, such as enzymes or nuclear hormone receptors, that natively bind to small-molecule ligands, but large swaths of the proteome do not contain druggable binding pockets[1,7]. Instead, many protein classes, such as scaffolding proteins and transcription factors, operate through protein-protein interactions (PPIs), which are notoriously difficult to inhibit with small-molecules because such interactions are spread over large surface areas[6–9]. New strategies for identifying small-molecule PPI modulators are constantly being developed[8,9], but proteins with druggable binding pockets are still significantly overrepresented as targets of newly approved therapeutics[10].

Antibodies are attractive candidates for expanding the druggable proteome because they can be developed to tightly and specifically bind to nearly any protein epitope[11]. However, antibodies lack any intrinsic capability to enter the cytosol of cells and are currently limited to membrane-associated or secreted targets. In this thesis, we address this problem by developing

a cytosolic antibody delivery system using cationic lipids, thus providing the foundations for therapeutic cytosolic antibodies.

1.2 Antibodies

Antibodies, also known as immunoglobulins (Ig), are large proteins produced by the humoral immune system to recognize and neutralize pathogens. In humans, there are five antibody classes – IgA, IgD, IgE, IgG, and IgM – that not only differ in structure, but also differ in the types of immune responses they can recruit upon antigen binding[12]. We will primarily focus on discussing IgGs because they are by far the most commonly used antibody class for therapeutics and as research tools. For the rest of the thesis, antibody will be used interchangeably with the IgG antibody class.

1.2.1 Immunoglobulin G (IgG) structure

IgG antibodies are approximately 150 kDa and consist of four polypeptide chains – a pair of identical heavy chains (~50 kDa) and a pair of identical light chains (~25 kDa) – linked to each other by disulfide bonds. There are two classes of light chains – κ and λ light chains –, but they are functionally the same. Heavy chains, however, are divided into four subclasses (in humans) that do result in moderate functional differences[13]. Either light chain class can be paired with any heavy chain subclass, resulting in four IgG isotypes – IgG1, IgG2, IgG3, and IgG4 – that are named based on the heavy chain constituent.

Both the heavy and light chains contain an N-terminal variable domain (VL for light chains and VH for heavy chains) that are joined together to form the antigen-binding variable region. Because IgGs are formed from two heavy and light chain pairs, each IgG molecule is biparatopic with two antigen-binding sites

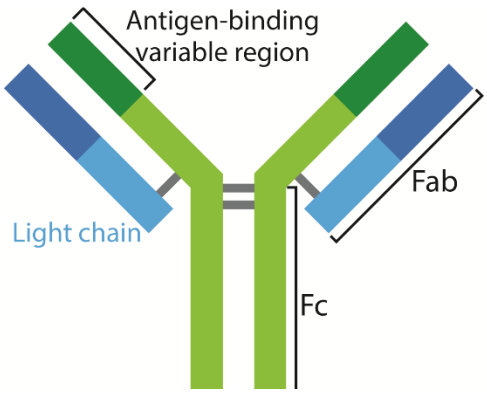

Figure 1.1 Schematic of IgG Structure.

(Fig. 1.1). Each variable region contains three hypervariable complementarity-determining regions (CDRs), which are usually only ~10 amino acids long and form the actual antigen-binding site. To produce antibodies that tightly bind to specific antigens, the immune system can generate and screen an incredible number of variable regions by undergoing V(D)J recombination, somatic hypermutation, and affinity maturation[14]. Because this process can generate IgGs that tightly and specifically bind to almost any epitope, IgGs have become not only powerful therapeutics, but also valuable research tools.

The remainder of the light and heavy chains consist of the constant regions, which, as the name implies, do not change within each light chain class or heavy chain subclass. The light chain constant region contains one constant domain (CL) whereas the heavy chain constant region contains three constant domains (CH1, CH2, CH3) with a hinge region joining CH1 and CH2. IgGs can also be divided into an antigen-binding fragment (Fab), which consists of the VL and CL domain of a light chain paired with the VH and CH1 domain of a heavy chain, as well as the crystallizable fragment (Fc), which consists of the CH2 and CH3 domains of both heavy chains paired together (Fig. 1.1).

The tertiary and quaternary structure of IgGs are held together by an intricate series of disulfide bonds that are key to antibody stability and function. Multiple inter-chain disulfide bonds in the hinge region link both heavy chains together – the precise number depends on the IgG subclass. Each heavy chain is also joined to its partner light chain by one inter-chain disulfide bond. Finally, a buried intra-chain disulfide bond stabilizes each variable and constant domain found within an IgG[15].

All IgGs have a conserved glycosylation site at Asn 297 (N297) of each heavy chain that can be linked to dozens of different glycans, affecting both antibody stability and function. Interactions between N297 glycans and the heavy chain's protein backbone can stabilize the Fc quaternary structure[13]. More importantly, N297 glycans can significantly affect IgG binding to

effector proteins either by inducing conformational changes in the Fc region or via glycan-protein or glycan-glycan interactions[13,16].

1.2.2 IgG function

Following generation, IgGs can directly inhibit their antigen's biological activity, serve as a hub for recruiting various arms of the immune system, or both[12,13]. Because IgGs are large bulky proteins, they can often sterically block their antigens from engaging in key protein-protein or protein-ligand interactions. Examples include antibody-mediated neutralization of bacterial virulence factors[12], the anthrax toxin[17,18], as well as viruses such as HIV[19,20], Ebola[21,22], and influenza[19,23]. A wide array of effector proteins can also bind to the IgG Fc region to induce IgG effector responses. Examples include C1q binding for complement activation and FcγRIIIa binding on natural killer cells to induce antibody-dependent cellular cytotoxicity (ADCC)[12,13].

Once secreted into circulation, IgGs have a remarkably long ~21-day half-life due to their ability to evade degradation by binding to the neonatal Fc receptor (FcRn) only in slightly acidic environments via their Fc region[13,24]. FcRn was first identified as the receptor responsible for maternal-fetal transfer of IgGs across the placenta and the intestine[24–26], but was subsequently found to also protect IgGs from catabolism[27]. Normally, vascular endothelium and monocytes will nonspecifically take up serum proteins via endocytosis into endosomes and subsequently degrade them following endosome maturation into lysosomes. IgGs, however, can bind to FcRn under slightly acidic conditions (pH 6-6.5) within endosomes, allowing them to be sorted into recycling endosomes and ultimately released back into circulation by the FcRn because of the physiological pH at the cell surface[13,24,28]. By strengthening or weakening antibody Fc affinity to the FcRn, antibody half-life can be extended or shortened, respectively, depending on the therapeutic need[24].

1.2.3 Generating antibodies for research and therapeutic uses

Antibodies were first used as therapeutics for infectious diseases in the early 20th century in the form of serum obtained from immunized animals, but they were largely abandoned with the advent of antibiotics due to variable efficacy, the need for intravenous administration, and the potential for severe allergic responses[29]. Although non-human polyclonal antibodies continue to be used in the form of antivenom, variability and toxicity remain a major limitation[29].

In 1975, Milstein and coworkers developed hybridomas, which allowed monoclonal antibodies to be easily isolated[30]. This technique involves isolating B cells from the spleen of an animal immunized with the desired antigen and fusing the B cells with myeloma cells to create immortal hybridomas that can perpetually secrete IgGs. Individual hybridoma cells can then be isolated and cultured to create clonal hybridoma cell lines, each secreting a specific or monoclonal IgG. Panels of monoclonal IgGs can then be screened to identify ones with the desired affinity, specificity, or activity[30–32].

Although hybridomas can be stored and continually cultured to produce IgGs, they can be genetically unstable and their non-human origin can lead to significant immunogenicity in patients[29,33]. Furthermore, because their Fc site binds poorly to many human effector proteins, non-human IgGs have a short half-life and are poor at inducing IgG effector functions in humans[34]. As a result, large scale IgG production has primarily shifted towards recombinant expression in mammalian systems[29,33]. Not only can immunogenicity be decreased by grafting animal-derived CDR or variable region sequences to a human framework IgG, but mammalian production systems also result in greater yields and consistency[29,33–35].

Even though antibody humanization has produced numerous clinically approved therapeutic IgGs[11], the humanization process requires considerable antibody engineering and often results in some loss in antigen binding affinity[35]. As an alternative, immunization can be

performed on transgenic animals containing human IgG genes[36]. Immunization, however, is time consuming with some antigens having poor immunogenicity due to immunological tolerance[31,37].

Bypassing animals and immunization altogether, phage display and other *in vitro* display technologies have also emerged as a prominent method for generating new IgGs. Libraries of VH and VL sequences can be expressed as IgG fragments (discussed in subchapter 1.2.5) on the coat protein of a bacteriophage[31,32,37,38] or the surface of yeast[39], and undergo rounds of selection to identify IgGs with the desired properties. Not only can libraries be constructed from human sequences, but display techniques are also far more high-throughput than traditional immunization.

1.2.4 Research applications for antibodies

Most biological samples are complex mixtures of macromolecules that cannot be easily distinguished from each other. Because IgGs can be developed to recognize a wide range of antigens with great specificity, they grant researchers the ability to quantify, visualize, or isolate almost any biologically relevant macromolecule. Thus, IgGs have become ubiquitous tools across biomedical research. Over decades, large and well validated collections of IgGs have been built containing IgGs that can specifically recognize almost any well-studied protein[40,41]. We will discuss some of their applications briefly.

The most basic application for IgGs is for quantifying the amount of antigens present in a biological sample. This is commonly done via an enzyme-linked immunosorbent assay (ELISA). First, the biological sample containing the antigen of interest is immobilized onto a microplate either directly through non-specific interactions with the plate surface or via a pre-immobilized capture antibody. Afterwards, enzyme-linked antibodies can be used to specifically recognize and quantify the amount of antigen present[31]. Typically, two antibodies will be used for detection – a primary and a secondary antibody. The primary antibody is the one that specifically binds to the antigen whereas the secondary antibody is conjugated to an enzyme used for detection and

specifically recognizes the primary antibody. A two-step dual antibody approach is used rather than directly conjugating the detection enzyme with the primary antibody because conjugating enzymes to an antibody without negatively perturbing the function of either the enzyme or the antibody can be quite difficult (discussed in subchapter 1.3). By using secondary antibodies that only bind to IgGs of a specific species and subclass, enzyme conjugation can be limited to only a handful of secondary antibodies, but almost any antibody can be still used as a primary antibody.

Western blotting, or immunoblotting, uses antibodies to detect the presence of specific proteins and their molecular weights in a biological sample. Proteins from a sample are first separated by molecular weight via electrophoresis and then transferred to a solid support – either a nitrocellulose or polyvinylidene difluoride (PVDF) membrane. The proteins of interest can then be detected by chemiluminescence or fluorescence via labeled primary antibodies or primary antibodies paired with a labeled secondary antibody[31].

Immunofluorescence or immunohistochemistry uses antibodies to detect antigens and their localization in cells or tissues. Cells or tissues are first chemically fixed by crosslinking or precipitation to increase sample stability and to preserve their internal structure. Then, the antigens of interest can be visualized by microscopy via labeled primary antibodies or primary antibodies paired with a labeled secondary antibody[31].

Immunoprecipitation uses antibodies to isolate specific antigens from complex biological mixtures. Antibodies are first incubated with a sample to allow for antibody-antigen binding. Then, a solid-support – typically an agarose resin or magnetic beads – capable of binding to the antibody is added. The solid-support, antibody, and antigen can subsequently be separated from the rest of the sample by centrifugation or magnetic separation[31]. Immunoprecipitation can be used to determine whether and how the antigen interacts with other proteins, RNA, or DNA.

1.2.5 scFvs, nanobodies, and antibody-like binding proteins

As an increasing number of new technologies prominently featuring antibodies have been developed, researchers have also recognized some of their limitations. Most research antibodies are still produced from hybridomas, but hybridomas can be genetically unstable and largely preclude antibody engineering. Full-length IgGs can be recombinantly expressed only in mammalian systems because the numerous disulfide bonds that are crucial to IgG structure do not form properly in the reducing environment of the *E. coli* cytosol[32,42]. Although mammalian expression systems are routinely used in industry, they are expensive and more labor intensive than bacteria systems, which are major barriers for easy adoption by academic researchers[42]. Finally, many applications use antibodies only for their binding ability, rendering the entire IgG Fc region as unnecessary bulk[42,43].

The first antibody alternative developed was the single-chain variable fragment (scFv) in which the variable domain from a heavy chain (VH) is linked to the variable domain of its paired light chain (VL) by a flexible linker to create a monovalent binding protein composed of a single polypeptide chain with no Fc region[44]. Unfortunately, scFvs are often unstable and have lower binding affinity than their parent IgG, requiring significant engineering to be useful[43,45]. Furthermore, scFvs often require proper disulfide bond formation within their VH and VL domains, which can introduce significant hurdles for recombinant scFv expression in *E. coli*[45]. To purify scFvs in any appreciable amount, they often must be either exported to the oxidizing environment of the *E. coli* periplasm[45,46] or expressed in *E. coli* engineered to have increased cytoplasm disulfide bond formation[47]. In spite of these drawbacks, scFvs continue to be popular because their amino acid sequence can be derived from previously established hybridoma stocks and they are widely used with display technologies[38].

Unexpectedly, camelids – which includes camels, alpacas, and llamas – were discovered to have antibodies solely composed of heavy chains[48,49], which antibody engineers quickly seized

upon to generate recombinant single camelid VH domain binding proteins called nanobodies[42,43]. Because nanobodies are extremely stable and express remarkably well in *E. coli*, they have emerged as often a superior alternative to scFvs[49]. Their main limitation is a more laborious development process – libraries of VH domains have to first be obtained from the B-cells of immunized camelids and then screened via phage display[49]. However, large synthetic nanobody libraries have been developed that are beginning to abrogate the need for any immunization[49,50].

Finally, designed antibody-like binding proteins have also emerged as a class of antibody alternatives. These designed proteins are often built upon small protein scaffolds that are easily expressed in *E. coli* and contain regions amenable to inserting recognition residues to create artificial binding pockets. Libraries of the scaffold can then be screened with various display technologies to identify potent binders against a given target[42]. Some of the more prominent protein scaffolds include affibodies, DARPins, affitins, knottins, and monobodies[42]. Two scaffolds that we utilize in this thesis are affibodies and DARPins. Affibodies are derived from the B-domain of *Staphylococcus aureas* protein A and are only ~6kDa, allowing them to be recombinantly expressed or synthesized via solid-phase peptide-synthesis[42,51]. Designed ankyrin repeat proteins (DARPins) are derived from the consensus sequence of human ankyrin repeat proteins and are larger (~14-21kDa) than affibodies, but have a larger target protein binding site[52,53].

Clinically, antibody alternatives suffer from a short half-life due to their small size and the absence of an Fc region that can engage the FcRn receptor[43]. A short half-life can be advantageous when rapid clearance is desired, such as for molecular imaging[54], but is usually prohibitive for therapeutic applications. Strategies for half-life extension include conjugating proteins to polyethylene glycol (PEG) to increase the protein's hydrodynamic size or fusing proteins to domains with a naturally long half-life such as the IgG Fc region or human serum albumin[55].

1.2.6 Therapeutic monoclonal antibodies

The first approved monoclonal IgG therapeutic was in 1985 with muromonab-CD3 (OKT3), a mouse IgG2a antibody used to treat acute rejection in organ transplants. Although OKT3 suffered from many of the limitations of non-human IgGs discussed in subchapter 1.2.3, it illustrated the massive potential for monoclonal IgG therapeutics. Due to the advent of numerous strategies for developing humanized or human IgGs, the number of approved monoclonal antibody (mAb) therapeutics has exploded (>60), resulting in ~$89 billion in revenue in 2016[10,11]. Almost all pharmaceutical and biotechnology companies have developed substantial expertise with mAbs, resulting in hundreds of therapeutic mAbs under clinical development. For some clinical indications, such as oncology, most new therapeutic approvals are now mAbs.

Monoclonal IgGs have emerged as an important class of therapeutics largely because of their ability to overcome many of the shortcomings and limitations faced by small molecule drugs. Small-molecule drug development depends on screening large libraries in a high-throughput manner to first identify promising chemotypes. Then, significant medicinal chemistry efforts must be made to identify a clinical lead that can potently modulate its target's activity while having minimal off-target engagement, sufficient cell permeability, and adequate pharmacokinetic properties[56]. This process can take years and success is not guaranteed.

In contrast, current immunization and display technologies almost guarantee identification of a mAb that can potently bind its target. IgGs bind to targets via much larger surface areas[6,57], resulting in much greater specificity and reducing toxicity due to off-target effects. Therapeutic IgGs must be administered intravenously, but they naturally have a long ~21 day half-life and thus, only require dosing every two to four weeks[13,24]. Furthermore, in addition to inhibiting their target's biological activity, mAbs can recruit various effector arms of the immune system (discussed in 1.2.2), and thus operate through modalities unavailable to small-molecules. For example, rituximab, an anti-CD20 mAb used to treat Non-Hodgkin's lymphoma and chronic

lymphocytic leukemia, kills CD20-positive B cells by simultaneously inducing antiproliferative signaling, complement-dependent cytotoxicity, and ADCC[11,58]. Even if a potent small-molecule CD20 binder could be identified, it would likely only induce antiproliferative signaling and ultimately may not result in a therapeutic benefit. Finally, antibody engineering has allowed for therapeutic antibodies to operate via mechanisms that are not available to endogenous IgGs, which will be discussed below.

1.2.7 Novel mechanisms of action for therapeutic mAbs – antibody-drug conjugates (ADCs)

A key challenge for developing chemotherapeutic agents is maximizing their therapeutic index – that is, identifying dosing regimens where the agent is capable of killing cancer cells without causing undue toxicity to normal tissues. Although oncology often tolerates very narrow therapeutic indexes, some agents remain too toxic for therapeutic use. Antibody-drug conjugates (ADCs) offer a solution by conjugating extremely potent cytotoxic agents to recombinant mAbs in order to grant them specificity towards tumors and increase their therapeutic index.

The first ADC, gemtuzumab ozogamicin, an anti-CD33 mAb conjugated to calicheamicin, was approved for acute myeloid leukemia (AML) in 2000, but was withdrawn due to lack of efficacy compared to standard of care. Significant advances in ADC development, though, led to the approval of brentuximab vedotin in 2011 and trastuzumab emtansine in 2013, which remain in clinical use. Currently, there are dozens of ADC in clinical trials and they have become an important subclass of the mAb therapeutic space[59].

The four key components of ADCs are the payload, the targeting mAb, the linker between the payload and the mAb, and the mAb labeling chemistry. The payload must be extremely potent because ADCs deliver relatively low concentrations of drugs into cancer cells. Auristatins and maytansinoids, both tubulin-disrupting agents, are two of the most commonly used payloads[59]. The targeting mAb must bind to an antigen that is highly expressed on cancer cells to be capable

of delivering large quantities of payload, but poorly expressed or even absent on normal tissue to reduce off-target toxicity. The linker must be stable in circulation to premature release of the payload, but must also be cleavable once internalized in cancer cells[59]. Finally, it is preferable to use site-specific labeling chemistry for conjugating the linker to the mAb because it results in more homogeneous ADC products[60].

1.2.8 Novel mechanisms of action for therapeutic mAbs – bispecific antibodies (bsAbs)

The immense clinical success of chimeric antigen receptor (CAR) T cell therapy[61–63] and checkpoint blockade[64,65] has ushered in a new age of immuno-oncology, which is the utilization of the immune system to treat cancer. One prominent strategy is to reprogram the body's T cells to recognize and attack cancer cells. CAR T cells achieve this by using lentivirus to genetically introduce a synthetic receptor that recognizes a cancer-associated antigen into autologous T cells[63]. Although extremely effective, CAR T cells are autologous and must be specifically manufactured for each patient. Not only does this result in significant logistical and manufacturing challenges, but CAR T cell therapy is also tremendously expensive, which poses a significant challenge for massive adoption of the therapy[63]. There are significant efforts underway to develop allogenic off-the-shelf CAR T cells via genome editing technologies, but their efficacy and safety have yet to been proven[66,67].

Bispecific antibodies (bsAbs) – antibodies that have been engineered to be capable of simultaneously binding two different antigens – offer a promising alternative to T cell-based therapies. By simultaneously binding the CD3 receptor on T cells and a cancer-specific antigen on cancer cells, bsAbs can potentially offer the efficacy of T cell reprogramming, but with the lower cost and off-the-shelf nature of biologics[68]. Although the concept of T cell redirecting bsAbs date back to the 1980s[69], due to challenges in bsAb manufacturing, the first bsAb - catumaxomab (anti-CD3 and anti-epithelial cell adhesion molecule) – wasn't approved until 2009, but it performed poorly commercially and was withdrawn in 2017. Blinatumomab (anti-CD3 and anti-

CD19), the second bsAb approval, however, induced impressive response rates in acute lymphoblastic leukemia (ALL) patients and in conjunction with the rise of immune-oncology, reinvigorated the field[70]. Currently, there over 40 T cell redirecting bsAbs under clinical development[68].

Because bsAbs are engineered antibodies, they can be constructed in almost limitless ways, but bsAbs that are under clinical development broadly fall into three categories. Fragment-based bispecifics are the simplest and involves linking two antibody fragments together to form a bivalent binder[68]. Fragment-based bispecifics can be produced in bacterial expression systems, but are small and do not contain a Fc region, resulting in poor circulation half-life. Symmetric bispecifics uses a conventional mAb scaffold, but fuse it to antibody fragments, usually resulting in a tetravalent binder[68]. Finally, asymmetric bispecifics consist of fragment-based bispecifics linked to a Fc domain or antibodies with two different variable regions to create monovalent binders[68]. Both symmetric and asymmetric bispecifics retain the long half-life of mAbs, but symmetric formats have greater avidity, which can affect potency[68].

1.2.9 The frontiers of therapeutic antibodies

The rapid rise of mAb and engineered antibody therapeutics over the past two decades has illustrated their nearly boundless therapeutic potential. However, there are two major compartments within the body in which small-molecule drugs can access that antibodies cannot – the central nervous system (CNS) and the interior of cells.

The major barrier for developing therapeutics for targets in the CNS is the blood-brain barrier, which only allows privileged molecules to traffic between the blood and the cerebrospinal fluid. IgGs do have some access to the CNS, but their concentration in cerebrospinal fluid is typically 100- to 1,000-fold lower than in circulation[11]. Thus, current antibody therapeutics under clinical development with CNS targets must be extremely potent and dosed at high concentrations to have any therapeutic utility[11,71]. Emerging strategies for improving antibody

delivery into the brain have primarily focused on using bispecific formats to hijack endogenous blood-brain barrier transcytosis pathways. For example, bispecific antibodies that engaged the transferrin receptor were shown to be capable of delivering functional anti-β-Amyloid or anti-β-secretase antibodies into the brains of mice and primates, respectively[72,73]. Although encouraging, identifying the optimal bispecific format and the ideal transcytosis receptor remains an extremely active area of research[11,74].

Antibodies can natively enter the cell via macropinocytosis or internalization of a receptor they are bound to, but they can only access the endosome-lysosome system, which is topologically equivalent to the extracellular space. Given that the majority of drug targets lie within the cytosol, organelles, or nucleus[75], a cytosolic antibody delivery system would dramatically expand the therapeutic utility of antibodies. Current progress on cytosolic protein delivery will be discussed in subchapter 1.4.

1.3 Protein Bioconjugation

Conjugating diverse chemical moieties and macromolecules to proteins has been an integral part of both basic science research as well as therapeutics. For example, conjugating enzymes to antibodies has enabled ELISA, western blotting, and immunohistochemistry. Conjugating fluorophores to antibodies has enabled fluorescence microscopy, flow cytometry, and fluorescence-activated cell sorting (FACS). Conjugating polymers to antibody-like proteins extends their circulation half-life and cytotoxic agents to antibodies enable ADCs. In this subchapter, we will discuss the various strategies that have been developed for protein bioconjugation, particularly ones commonly used for antibodies.

1.3.1 Classical side chain labeling chemistries

Chemically modifying the side chains of naturally occurring amino acids is extremely attractive since it enables easy functionalization of most endogenous proteins without any genetic

engineering. Although side chain selective labeling chemistries exist for most of the non-hydrophobic amino acids[76,77], currently cysteine and lysine are overwhelmingly the most commonly targeted residues. Exploiting the reactivity of natural occurring amino acids, however, comes with three major limitations – 1) Some proteins cannot be labeled because they lack surface cysteine or lysine residues. 2) Lysine-selective chemistries often generate heterogeneous products because proteins often contain multiple surface lysines[60,78–80]. 3) Endogenous proteins have to be purified prior to labeling due to the abundance of cysteines and lysines in most biological samples.

Cysteines have been extensively used as reactive labeling handles because not only are they the most nucleophilic of the canonical amino acids, but they can also be easily oxidized to form disulfide bonds. Furthermore, because they are a relatively rare residue, cysteine can often be genetically introduced into proteins without any native surface cysteines to enable site-selective modifications. The most commonly used cysteine labeling chemistries are disulfide bond conjugations to form S-S bonds and Michael addition with maleimides to form S-C bonds, but it should be noted that both can be reversible[76,78]. Disulfide bonds are reversible under reducing conditions and maleimides can undergo retro-Michael deconjugation, resulting in thiol exchange with cysteines on unlabeled proteins[81,82]. Maleimide reversibility, however, can be addressed by hydrolyzing the maleimide ring following conjugation[59,81,82].

Unfortunately, IgG antibodies are held together by an intricate series of disulfide bonds, which renders cysteine bioconjugation quite tricky. IgG disulfide bonds, however, differ in their susceptibility to reduction[83], enabling semi-selective conjugation to interchain cysteines[84,85]. Selective disulfide reduction, however, is imprecise and results in heterogeneous products, which can be a major challenge for therapeutics such as ADCs[60,82,84,85]. To address this, in 2008, Genentech developed antibodies containing a free reactive cysteine on their heavy chain CH1 domains called THIOMABs[86]. THIOMAB production require a complex reduction followed by re-

oxidation purification strategy, but ultimately resulted in nearly homogeneous ADCs with two drugs per antibody[86].

Lysine side chains contain a very nucleophilic primary amine and can be selectively labeled over thiols with activated esters, sulfonyl chlorides, and isocyanates[76,78]. Due to their abundance, lysines are generally labeled when product heterogeneity is less of a concern. Some proteins, however, contain specific lysines with significantly enhanced reactivity due to their microenvironment[87], making site-selective lysine modification possible with specialized reagents.

Antibodies typically contain dozens of surface lysines, rendering lysine-reactive chemistry inherently imprecise. Labeling must be performed carefully since excessive amounts can often reduce antibody stability or impair antigen binding. Regardless, N-hydroxysuccinimide ester (NHS) chemistry is used extremely often for functionalizing antibodies in research settings, such as with fluorophores for microscopy or flow cytometry. Some early generation ADCs such as gemtuzumab ozogamicin and trastuzumab emtansine also used lysine chemistries, but their therapeutic efficacy were at least partially compromised by labeling hetereogeneity[59].

1.3.2 Glycoconjugation

Bioconjugation reactions targeting glycans are uncommon outside of metabolic glycoengineering[76] due to the heterogeneous glycosylation patterns found in most proteins[85]. Glycoconjugation, however, has been frequently employed for functionalizing antibodies since IgGs are naturally glycosylated at Asn 297 (N297) of each heavy chain. Historically, sodium periodate was used to oxidize antibody glycans to aldehydes, which can then be used in additional biorthogonal labeling reactions. This reaction was used to label antibodies with horseradish peroxidase to produce secondary antibodies used for chemiluminescence detection[88]. However, the need to use high periodate concentrations resulted in heterogeneous glycan oxidation and methionine oxidation[60,85]. Antibody methionine oxidation has been associated with decreased FcRn binding and thus, decrease circulation half-life[60].

Current strategies for antibody glycoconjugation have focused on increasing antibody glycan homogeneity prior to functionalization[60,85]. For example, a mixture of galactosyltransferases and sialyltransferases can be used to incorporate sialic acid onto glycans at the N297 glycosylation site[89]. Sialic acid can then be oxidized under mild conditions with extremely low concentrations of sodium periodate to generate a homogeneous population of aldehyde-functionalized antibodies[89]. A similar enzymatic strategy was also used to install azido-functionalized sialic acids, which can then be used in bioorthogonal labeling reactions[90].

1.3.3 Bioorthogonal Reactions

Labeling reactions are bioorthogonal if they can proceed without interfering with biological processes[76]. For protein bioconjugation, this generally means that canonical amino acids cannot participate in bioorthogonal reactions. In practice, many site-selective protein bioconjugation reactions using bioorthogonal reactions require two steps. First, a reactive handle that is chemically distinguishable from the canonical amino acids is introduced at a desired position within a target protein. Afterwards, a bioorthogonal reaction can be used to site-specifically conjugate functional chemical moieties to the target protein via the reactive handle.

Aldehydes and ketones are electrophiles that can react with α-effect amines such as alkoxylamines or hydrazines to form oxime and hydrazone linkages, respectively, under mild conditions[76,91]. A major advantage of these linkages is that they are very small and are less likely to perturb protein function[91]. It should be noted, however, that these reactions are not strictly bioorthogonal since cells contain high concentrations of endogenous aldehydes and ketones (for example, glucose and pyruvate), but they are effectively bioorthogonal with purified proteins.

Azides are particularly powerful reactive handles that can participate in bioorthogonal reactions with not only purified proteins, but also in the cellular milieu. Azides can participate in 3 main bioorthogonal reactions – the Staudinger ligation, the copper catalyzed azide-alkyne 1,3-dipolar cycloaddition (CuAAC), and the strain-promoted alkyne-azide cycloaddition (SPAAC)[76,80].

The Staudinger ligation was introduced as a bioorthogonal labeling reaction between azides and triarylphosphines in 2000[92]. Although instrumental in making azides a popular reactive handle, it has become less popular due to slow reaction kinetics[76,78]. CuAAC was then developed in 2002[93,94] as a copper(I) catalyzed reaction between azides and alkynes that forms triazoles. Due to its reaction efficiency and speed, CuAAC quickly gained widespread adoption for protein bioconjugation. The only limitation to CuAAC is that copper(I) can be quite toxic to cells[76], which prompted the development of the copper-free SPAAC reaction[95], which uses strained cyclooctynes to react with azides. Although SPAAC reactions were extremely slow compared to CuAAC[95], modern cyclooctynes reagents have become considerably faster[96,97], albeit still slower than CuAAC, and are currently workhorse bioorthogonal conjugation reagents.

Inverse-electron demand Diels-Alder (IEDDA) reactions between strained alkenes and tetrazines are emerging as powerful bioorthogonal reaction[76,80,98,99]. The reaction occurs under physiological conditions, does not require metals, and occurs orders of magnitude faster than CuAAC[98]. The rapidity of IEDDA even under low reactant concentrations and in the presence of serum has made it increasingly popular for use in pre-targeted PET and SPECT imaging[100,101].

1.3.4 Native and expressed chemical ligation

Solid-phase peptide synthesis can be used to synthesize polypeptides conjugated to nearly any chemical matter, but it is rarely successful for peptides significantly longer than 50 amino acids. To address this limitation, native chemical ligation (NCL) was developed in the early 1990s and enabled the ligation of an unprotected peptide with a C-terminal thioester with an unprotected peptide containing an N-terminal cysteine[102]. Originally, NCL always resulted in a cysteine at the ligation site, but the development of auxiliary thiol-derived variants of canonical amino acids has largely abrogated that requirement[103,104]. By chaining NCL reactions together, proteins as large as erythropoiesis protein[105] (166 amino acids) and HIV-1 protease[106] (203 amino acids) have been synthesized.

Although it possible to synthesize proteins up to a moderate size with NCL, it quickly becomes prohibitively cumbersome. Expressed protein ligation (EPL) uses peptide synthesis and recombinant protein expression to greatly extend the length of proteins that can be produced via semi-synthesis[103]. Semisynthetic proteins with N-terminal modifications can be produced using a synthesized peptide with a C-terminal thioester and a recombinantly expressed protein with an N-terminal cysteine[103]. Following the discovery of proteins that self-splice to create C-terminal thioesters[107], EPL was extended to produce semisynthetic proteins with C-terminal modifications [108]. EPL, however, is still subject to the length limitations of peptide synthesis, requires reducing agents to be present, and is generally limited to producing proteins with modifications close to either termini[82].

1.3.5 Genetic incorporation of unnatural amino acids

Genetically incorporating unnatural amino acids (UAAs) into proteins allows for precise control of the placement of reactive handles for bioorthogonal labeling reactions while minimally perturbing protein structure. The genetic code has three redundant stop codons that could in principle be hijacked with an engineered tRNA. The amber stop codon (UAG) is the most amenable to reassignment because it is the least-used stop codon in *E. coli*, *S. cerevisiae*, and humans. Codon reassignment was initially demonstrated by chemically aminoacylating an amber suppressor tRNA engineered to recognize the amber stop codon and using it for *in vitro* translation[109,110]. However, because *in vitro* translation produces such minute amounts of protein, this strategy was not widely adopted for protein bioconjugation.

In 2001, Schultz and coworkers first demonstrated site-specific incorporation of the unnatural amino acid *O*-methyl-L-tyrosine in *E. coli* via the amber suppression approach[111]. To accomplish this feat, an amber suppressor tRNA that could not be recognized by endogenous aminoacyl-tRNA synthetase and an orthogonal aminoacyl-tRNA synthetase that could charge the

suppressor tRNA with the desired unnatural amino acid, but not recognize any endogenous tRNAs had to be simultaneously introduced into *E. coli.*

This was made possible by the discovery that the thermophilic archael bacteria *M. jannaschii* contained a $tRNA^{Tyr}$ and *M*. *jannaschii* tyrosyl-tRNA synthetase (TyrRS) pair that was not only mostly orthogonal to endogenous *E. coli* tRNAs and aminoacyl tRNA-synthetases, but also amenable to re-engineering[112–114]. The *M*. *jannaschii* $tRNA^{Tyr}$ was evolved to be completely orthogonal in *E coli*. by generating a library of tRNA mutants and subjecting it to multiple rounds of negative selection to ensure orthogonality following by positive selection to ensure that it could still be aminoacylated by its *M*. *jannaschii* TyRS pair[115]. The *M*. *jannaschii* TyRS was engineered to only aminoacylate the amber suppressor tRNA with *O*-methyl-L-tyrosine by subjecting a library of *M*. *jannaschii* TyRS active site mutants to multiple rounds of positive selection for functionality and negative selection to ensure *O*-methyl-L-tyrosine selectivity over its native tyrosine[111].

Since the initial report of *O*-methyl-L-tyrosine incorporation, similar tRNA and aminoacyl-tRNA synthetase engineering strategies have been employed to enable the incorporation of more than a hundred different UAAs in *E. coli*[116–118]. For some applications, such as the incorporation of small fluorophores, fluorescent UAA can be directly incorporated into recombinant proteins with amber suppression[116,118]. Other chemical moieties, though, can be easily site-specifically conjugated to proteins via genetically encoded UAAs that contain reactive handles such as azides, strained alkenes, alkynes, and ketones[76,116]. Though the selection is more limited, UAA incorporation is also possible in mammalian cells and even in mice[116–119].

Although tremendously powerful, UAA incorporation does have some drawbacks. Depending on the exact UAA tRNA/tRNA-synthetase pair, UAA incorporation efficiency can vary and allow for unwanted incorporation of canonical amino acids into the protein[117]. Because amber suppressor tRNAs can fail to outcompete endogenous release factors, UAA incorporation can result in truncated recombinant proteins and thus, reduced yields[117,120]. Premature truncation also

makes it extremely challenging to incorporate multiple UAAs in a protein. Extensive optimization of amber suppressor tRNA and orthogonal tRNA synthetase expression can improve some of these parameters[120,121]. Even greater improvements can be achieved with significant engineering of either the *E. coli* genome or the *E. coli* translational machinery. Examples include deleting release factors[122,123], reassigning endogenous amber stop codons[123,124], developing orthogonal ribosomes[125], and relaxing elongation factor quality control[126].

Genetically encoded UAAs have been used in mammalian expression systems to produce antibody drug conjugates with site-specifically labeled payloads. *p*-acetylphenylalanine and azide-containing lysine analogues have been introduced into the heavy chain of IgGs to enable site-specific auristatin conjugation via oxime ligation and CuAAC, respectively[127–129]. Although UAA incorporation in large scale mammalian cultures may require significant optimization[128], it results in homogeneous ADC products that have superior efficacy and pharmacokinetics compared to cysteine-based ADCs.

1.3.6 Peptide tags and chemoenzymatic bioconjugation

Rather than exploiting bioorthogonal chemistry for site-specific labeling, chemoenzymatic bioconjugation approaches utilizes the selectivity of enzyme active sites to specifically recognize peptide tags installed on proteins. Compared to genetic incorporation of UAAs, peptide tags do not require any modification of the expression system, but they are larger in size and are often constrained to modifications at protein termini. In this section, we will briefly introduce some enzymes that are commonly used for chemoenzymatic bioconjugation, but in subchapter 1.3.6, we will extensively discuss the history and applications of sortase A (SrtA) for site-specific protein bioconjugation.

Transglutaminases (TGases) are enzymes that normally crosslink proteins together by catalyzing amide bond formation between the primary amine of a lysine and the amide of a glutamine[130]. TGases, however, are relatively promiscuous with regards to their amine-containing

substrate and was first shown in 2006 by Ting and coworkers to be useful for protein bioconjugation[131]. Unfortunately, TGases have a poorly defined glutamine recognition motif, which can make it challenging to perform site-specific protein bioconjugation[132]. Regardless, bacterial TGases have had some utility for producing antibody-drug conjugates. *S. mobaraensis* TGase was found to be capable of modifying IgG heavy chains at Q295, but only following enzymatic deglycosylation of the glycan at N297 to increase Q295 accessibility to the enzyme[133]. Alternatively, *S. mobaraense* TGase was found to have a more restrictive active site that primarily recognizes a LLQG tag, enabling site-specific labeling of engineered antibodies[134].

Sulfatases in eukaryotes and prokaryotes contains a catalytic formylglycine which is posttranslationally generated via cysteine oxidation by the formylglycine generating enzyme (FGE)[135]. *M. tuberculosis* FGE, the most commonly used for bioconjugation, canonically recognizes a CXPXR motif[136], but also tolerates some noncanonical sequences[137]. Thus, by co-expressing FGE with proteins engineered to contain the CXPXR motif, the final purified protein will contain an aldehyde which can subsequently be used as a reactive handle for bioorthogonal labeling reactions[136]. This strategy works not only in *E. coli*[136], but also in mammalian cells[138], which has subsequently enabled the production of antibody-drug conjugates using the hydrazine-Pictet-Spengler (HIPS) ligation[139]. Although much more efficient than other aldehyde bioconjugation chemistries[140], HIPS ligation still requires extremely high protein concentrations and long reaction times[139], which limits the utility of FGE.

Trypsiligase is an engineered version of the trypsin protease that recognizes a YRH motif to enable N-terminal and C-terminal protein modification[141]. For N-terminal modifications, trypsiligase can ligate guanidinophenyl ester (OGp) derivatives to the arginine[141] whereas for C-terminal modifications, trypsiligase can act as a transpeptidase between the YRH motif and a nucleophilic peptide beginning with RH[142]. Although trypsiligase is an extremely fast enzyme, it is

limited by relatively few commercially available OGp derivatives for N-terminal modifications and low ligation efficiencies due to a hydrolysis side reaction for C-terminal modifications[142].

Subtiligase is an engineered version of the *B. subtilis* subtilisin serine protease that can ligate peptides with an α-amine to peptides with a C-terminal ester or thioester[143,144]. It has a broad substrate scope, which can allow the installation of chemical moieties in a native protein sequence context, but can also make ligation efficiency difficult to predict[143,145]. This has been somewhat addressed with modern proteomic techniques have allowed for extensive substrate mapping[146]. Regardless, the need to generate a C-terminal esters or thioesters on target proteins makes C-terminal bioconjugation with subtiligase more laborious to implement when compared to other chemoenzymatic bioconjugation approaches.

Butelase is a cysteine transpeptidase from the plant *C. ternatea* that normally catalyzes peptide cyclization, but also has utility as an extremely fast peptide ligase[147]. Butelase recognizes a D/N-HV motif and can then catalyze amide bond formation between the aspartate/asparagine and almost any other peptide[148]. Because butelase recognizes peptide sequences that are orthogonal to those of sortase A, both transpeptidase can be used together for one-pot double labeling of target proteins[149]. The main drawback for butelase is that it currently cannot be produced recombinantly and must be extracted directly from *C. ternatea*[147,149]. However, the closely related *O. affinis* asparaginyl endopeptidase (OaAEP1) can be recombinantly purified from *E. coli* and initially appears to have the same ligase ability as that of butelase[150].

1.3.7 Site-specific protein bioconjugation with sortase A (SrtA)

Sortase A (SrtA) is a calcium-dependent bacterial transpeptidase that is crucial for anchoring surface proteins onto the cell wall of gram-positive bacteria[151,152]. In the presence of Ca^{2+}, *S. aureas* SrtA binds to its LPXTG recognition motif and attacks the peptide bond between the Thr and the Gly with a catalytic Cys to form a thioester acyl-enzyme intermediate (Fig. 1.2). The thioester acyl-enzyme intermediate is relatively long-lived and is normally resolved by

nucleophilic attack from the oligoglycine of lipid II in bacteria or a peptide for protein bioconjugation (Fig 1.2)[153–155]. Water, though, can also serve also as a nucleophile, albeit at relatively slow rates, resulting in the production of a hydrolysis side-product (Fig. 1.2).

Originally pursued as an antibiotic target, in 2004, Mao and coworkers demonstrated that *S. aureas* SrtA tolerates nucleophilic oligoglycines containing additional chemical substituents

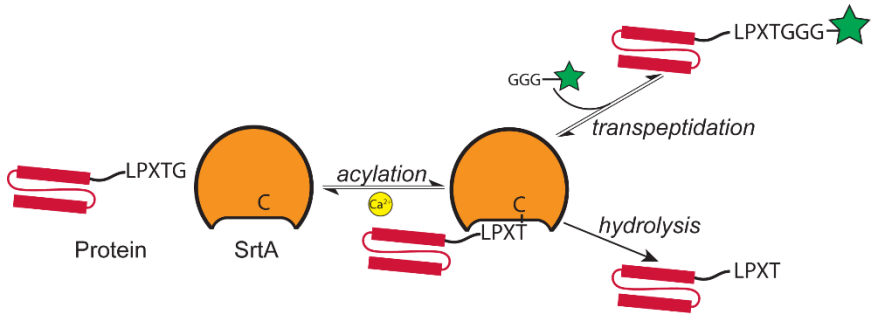

Figure 1.2 SrtA ligation mechanism.

and thus, could be used as a peptide-protein ligase[156]. For sortase-mediated ligation, peptides containing an N-terminal oligoglycine are ligated onto target proteins engineered to contain a C-terminal LPXTG SrtA recognition motif[153,157]. Alternatively, for N-terminal protein modification, SrtA can ligate a peptide containing its LPXTG recognition motif to a protein with an N-terminal glycine[158]. Because SrtA can be recombinantly produced in high amounts[157], the LPXTG SrtA recognition motif is relatively small, and the synthetic ease of producing functionalized short peptides, SrtA mediated-ligation has rapidly become a popular approach for site-specific protein bioconjugation[153]. SrtA has been employed to conjugate proteins to not only small chemical handles[159], but also fluorophores[156,159], nucleic acids[160], lipids[161], polymers[162], glycans[163], and surfaces[162,164].

In addition to the canonical LPXTG recognition motif, naturally occurring and engineered SrtA variants have been also identified that recognize other motifs[153,165]. *S. pyogenes* SrtA can recognize LPXTA motifs in addition to the canonical LPXTG motif and ligate nucleophilic peptides containing an N-terminal alanine, allowing it to be used in conjugation with *S. aureas* SrtA for dual labeling[158]. A *S. aureas* SrtA F40A mutant is capable of recognizing the APATG motif, but has

reduced reaction rates relative to wild-type *S. aureas* SrtA[166]. Finally, yeast display has identified a series of heavily mutated SrtA that were capable of recognizing LAETG and LPEXG motifs[167].

SrtA is also not limited to peptide nucleophiles – under certain conditions, SrtA can tolerate many primary amines as the nucleophile. Class C pilin-specific sortases, for example, normally recognizes the ε-amine of lysines as the nucleophile[152]. Armed with this observation, Chilkoti and coworkers found that *S. aureas* SrtA could also catalyze isopeptide bond formation between a labeled LPETG peptide and the ε-amine of a lysine within a pilin domain sequence[168]. Similar lysine labeling activity was also detected with an engineered version of the *C. diphtheria* SrtA[169]. Finally, a heavily engineered SrtA7M variant was also found to be capable of ligating a variety of amine-containing small molecules to LPETG sequences[170].

Although extremely versatile, SrtA has relatively poor reaction rates due to a mM affinity to its LPXTG recognition motif and a moderate k_{cat}[171], resulting in long reaction times, incomplete reactions, and poor ligation efficiencies[156,160,162]. Using large excesses of the peptide nucleophile can partially address these shortcomings, but can be cost-prohibitive for custom-synthesized peptides and impossible for peptides with low solubility in aqueous buffers. Long reaction times can drive ligation reactions towards completion, but also result in decreased ligation efficiency due to accumulation of the hydrolysis side product[160].

This motivated the development of engineered SrtA variants with superior performance. The first major effort was led by Liu and coworkers, who used yeast display to identify a pentamutant SrtA (SrtA5M) with 120-fold increase in k_{cat}/K_M relative to the native SrtA[171]. Additional screening using a high-throughput FRET assay using a SrtA5M-based library identified an addition three mutations that further increased k_{cat}/K_M by 5-fold relative to SrtA5M[172]. These two variants, with or without additional mutations that abolished Ca^{2+} dependency[173,174], have largely replaced native SrtA for in solution bioconjugation[153].

The Tsourkas lab has taken a different approach to improving SrtA-mediated ligation with the development of Sortase-tag expressed protein ligation (STEPL)[175]. Rather than expressing target proteins with just the LPXTG SrtA recognition motif as in sortase-mediated ligation, in STEPL, the target protein is expressed in series with a LPETG SrtA recognition motif followed by a flexible linker, the native SrtA enzyme, and an affinity tag (Fig. 1.3). Following expression, the target protein-SrtA fusion protein can be immobilized on a resin via the affinity tag and then the SrtA ligation reaction can be performed on resin (Fig. 1.3). Because the LPXTG SrtA recognition motif is tethered to SrtA, its effective concentration is significantly increased, accelerating the ligation reaction. Furthermore, only the labeled target protein is released from the resin following the reaction, consolidating target protein purification, labeling, and separation from the SrtA enzyme to a single process (Fig. 1.3). By using STEPL, proteins have been labeled to fluorophores[175], MRI contrast agents[175], nanoemulsions[176], and hydrogels[177]. Unfortunately, occasionally, fusing SrtA to target proteins can significantly decrease target protein expression, which limits STEPL's utility. We have addressed this limitation by developing proximity-based sortase-mediated ligation (PBSL), which will be discussed in chapter 2.

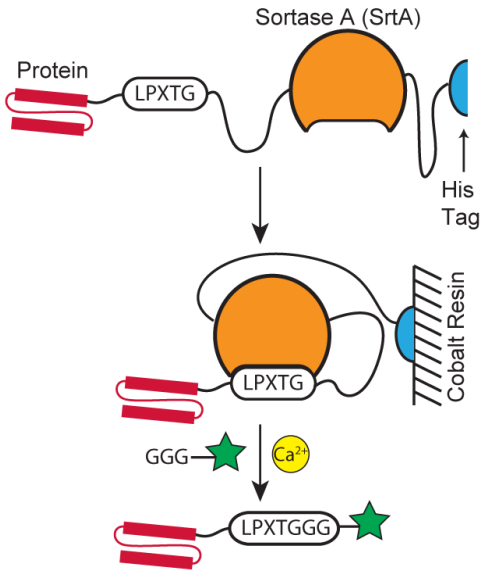

Figure 1.3 Schematic of STEPL.

1.3.8 Self-labeling protein domains

UAA incorporation of bioorthogonal reactive handles and chemoenzymatic bioconjugation approaches are invaluable when one desires to minimally perturb the protein to be labeled. However, they may require significant optimization to achieve high labeling efficiency and may be difficult to adopt in complex environments such as inside cells or in lysates. Self-labeling protein domains are typically engineered enzymes that can covalently link themselves to specific

chemicals or peptides extremely rapidly and efficiently. Although they generally cannot be used in therapeutics and their bulkiness can deleteriously affect protein function, self-labeling protein domains offer the simplicity of genetically encoded protein tags with the power of synthetic chemistry and have become extremely useful research tools.

SNAP tags and CLIP tags are both engineered versions of the human DNA repair enzyme O^6-alkylguanine-DNA alkyltransferase (hAGT). Normally, hAGT transfers an alkyl group from mutagenic O^6-alkylguanine-DNA to its catalytic cysteine to irreversibly form a methionine and restore the nucleobase back to a guanine[178]. hAGT, however, can also tolerate O^6-benzylguanosine with substituted benzyl rings as substrates[179]. Thus, fluorophores or chemical handles can be installed onto the O^6-benzylguanosine benzyl ring and be covalently linked to an engineered version of hAGT called SNAP-tag[179]. By using phage display, a second version of hAGT called CLIP-tag was also produced which recognizes O^2-benzylcytosine derivatives[180]. SNAP-tag and CLIP-tag are orthogonal to each other[180] and have been used extensively as genetic fusions in cells for microscopy.

HaloTag is based on an engineered version of the *Rhodococcus* dehalogenase capable of forming an irreversible covalent bond between its catalytic Asp and a chloroalkane[181]. Thus, similarly to SNAP-tag and CLIP-tag, HaloTag fusions can be easily labeled with fluorophores or chemical handles that are linked to a chloroalkane. Of note, in addition to standard fluorophores, HaloTag can also be coupled with newer generation fluorophores that can be photoactivatable or compatible with *in vivo* imaging[182–184].

The SpyCatcher-SpyTag domain-peptide pair was derived from the CnaB2 domain of the *S. pyogenes* fibronectin binding protein (FbaB), which normally spontaneously forms an intramolecular isopeptide bond between an Asp and a Lys to increase protein stability[185]. Howarth and coworkers split the CnaB2 domain into a domain containing the isopeptide bond-forming Lys as well as a peptide containing the isopeptide bond-forming Asp and found that the halves still

retained the ability to spontaneously form an isopeptide bond. After significant optimization, they arrived at the SpyCatcher-SpyTag domain-peptide pair, which can undergo isopeptide bond formation rapidly and efficiently, even at low concentrations[186,187]. Furthermore, isopeptide bond formation readily occurs under a wide range of conditions, including within cells[186].

1.3.9 Light activated site-specific conjugation of native IgGs (LASIC)

As discussed in this subchapter, there exists a wealth of approaches for site-specifically modifying recombinant antibodies. Recombinant antibody expression, however, is a time consuming process and the amino acid sequence for many commercial antibody collections are not easily attainable by researchers. Without genetic re-engineering, the only site-specific antibody bioconjugation methods available are primarily via glycoconjugation to glycans at N297 (subchapter 1.3.2) or modifying Q295 with transglutaminase following N297 deglycosylation (subchapter 1.3.5), both laborious and relatively inefficient processes.

Because IgG antibodies are crucial for the immune system's response against pathogens, many bacteria have evolved virulence factors that bind to IgGs as part of an immune evasion strategy[188]. Because of their high affinity and specificity towards IgGs, fragments from IgG-binding virulence factors such as protein A from *S. aureas*, protein L from *P. magnus*, and protein G from *Streptococcus* have become indispensable research tools[189]. For example, once immobilized on a solid support, they can be used for purifying antibodies and for isolating antibody-antigen complexes. Although these proteins bind IgGs with high affinity (nM), non-covalent interactions are reversible and can be disrupted in environments with low pH or high concentration of non-specific antibodies[189]. Thus, the Tsourkas lab and others have sought to develop antibody-binding proteins that could be site-specifically and covalently linked to native IgGs.

Protein Z is derived from the B domain of *S. aureas* protein A[190] and only binds to IgG Fc regions[191]. In 2011, Hober and coworkers used solid-phase peptide synthesis to produce protein

Z variants containing the photocrosslinker unnatural amino acid benzoylphenylalanine (BPA)[191]. They identified that BPA replacement of either F5 or H18 did not perturb protein Z Fc binding affinity, but did allow protein Z to be photocrosslinked to IgGs following irradiation with 365nm light[191]. In 2014, the Tsourkas lab demonstrated that not only can BPA be genetically incorporated via amber suppression in recombinantly expressed protein Z[121,192], but protein Z could also be integrated into the STEPL system to enable labeling with fluorophores and reactive handles[193]. Subsequently, the Tsourkas lab identified a panel of BPA incorporation sites that enabled photocrosslinking to human IgG1 (hIgG1), hIgG2, hIgG4, mouse IgG1 (mIgG1), mIgG2a, mIgG2b, mIgG3, rat IgG1 (rIgG1), rIgG2b, rIgG2c, hamster IgG1, and rabbit IgG (rabIgG) antibodies[194]. Unfortunately, IgGs were not fully photocrosslinked, even with a large excess of protein Z[194].

The Tsourkas lab then focused a thermally hyperstable IgG-binding fragment derived from the β1 domain of *Streptococcal* protein G[195] mutated to only be capable of Fc binding[196]. After screening a panel of positions, the Tsourkas lab found that BPA incorporation at A24 resulted in nearly complete photocrosslinking to hIgG1, hIgG2, hIgG3, higG4, mIgG2a, mIgG3, and rIgG2c as well as significant photocrosslinking to mIgG2b and rabbit IgGs[197]. Photocrosslinking efficiency is almost quantitative, did not perturb IgG-antigen binding affinity, and, in conjunction with STEPL, could be used to label IgGs with fluorophores and reactive handles[197]. Mass spectrometry revealed that protein G-A24BPA specifically photocrosslinked to M252 on IgG heavy chains due to BPA's strong photocrosslinking preference towards alkyl sulfides[198–200]. Thus, protein G-A24BPA cannot photocrosslink to IgGs that lack M252 such as mIgG1, rIgG1, rIgG2a, and rIgG2b[197].

In chapters 3 and 4, we will utilize LASIC and protein G-A24BPA to conjugate cell penetrating peptides or anionic polypeptides, respectively, to IgGs to enable cytosolic antibody delivery.

1.4 Cytosolic Protein Delivery

As discussed in subchapter 1.1, developing a cytosolic delivery system for IgG antibodies or antibody-like binding proteins would dramatically expand the druggable proteome, allowing targeting of protein considered largely 'undruggable' by small molecules.

Cytosolic delivery of inhibitory antibodies also offers unique advantages as a research tool over genetic or small-molecule approaches for modulating protein function. siRNA and shRNA are both popular nucleic acid-based approaches for protein knockdown or knockout, but because they do not directly act on target proteins, they perform poorly against proteins with long half-lives[201] and can induce significant cellular compensatory responses[202]. Small-molecule inhibitors and modulators can avoid these limitations, but even for proteins druggable by small-molecules, identifying potent compounds and then validating their selectivity can be challenging[203]. In contrast, inhibitory antibodies directly bind to target proteins, act on fast time scales, and can be generated far more easily than small-molecules. Finally, cytosolic inhibitory antibodies also offer the opportunity to discriminate between and modulate the activity of proteins with specific post-translational modifications or certain isoforms of a protein, which are generally not possible with traditional approaches.

Because of its potential profound therapeutic and basic science research implications, efficient cytosolic protein delivery systems have been pursued for more than two decades. In this subchapter, we will discuss the current progress on approaches that have been employed for cytosolic protein delivery.

1.4.1 Physical delivery methods

Since the plasma membrane is the primary barrier for macromolecule entry into the cytosol, physical delivery methods take the most straightforward approach by temporarily disrupting the plasma membrane without killing the cell.

The oldest cytosolic macromolecule delivery method, first reported in 1911[204], is microinjection, which uses micromanipulators to physically pierce cells with glass micropipettes and inject macromolecules into the cell. Microinjection requires significant user skill, is extremely low throughput, and thus, is currently rarely used outside of pronuclear injection for creating transgenic mice.

The other physical delivery method still frequently used is electroporation, first reported in 1982, which uses short high voltage electrical pulses to induce pores to transiently form in the plasma membrane[205]. Although electroporation protocols have improved significantly from its first inception and can be extremely effective[206,207], it requires expensive equipment, can result in significant toxicity, and has only moderate throughput.

New physical delivery methods such as microfluidic electroporation, microfluidic constriction, localized heating, nanoscale electroporation, nanoneedles, and targeted cavitation are under active development, but their ultimate utility remain to be demonstrated[208]. Even if newer physical delivery technologies replace electroporation, they cannot be adopted for therapeutic purposes other than *ex vivo* manipulation of cells. Thus, effective biology-inspired or carrier-mediated approaches must be developed to unlock the potential of cytosolic protein therapeutics.

1.4.2 Physical delivery of inhibitory antibodies

Although currently rarely used, historical studies using microinjection to cytosolically deliver antibodies have not only demonstrated the feasibility of, but also the power of cytosolic antibody-dependent inhibition of proteins. Following the development of hybridomas in 1975, one of the first steps when studying a new protein was to develop a panel of IgGs against the protein and epitope map them. Afterwards, antibodies against the protein of interest can be microinjected into cells and then one can assess for perturbation of phenotypes predicted to be associated with the protein. This strategy was used to demonstrate in 1982 that the tumor suppressor p53

regulated DNA synthesis[209] and in 1985 that the Ras oncogene was crucial for cell division in cancer cells[210]. Since, microinjecting inhibitory antibodies have helped to unveil the biological roles of the heat shock protein hsp70[211], the Jun and Fos transcription factors[212], the cyclin proteins[213,214], the signaling protein phosphatidylinositol 3-kinase (PI3-K)[215], the regulatory protein Raf kinase inhibitory protein (RKIP)[216], and many others.

Physical antibody delivery has also demonstrated that cytosolic IgGs can interact with the cytosolic antibody receptor tripartite motif 21 (TRIM21) via their Fc region to stimulate innate immune signaling[217,218] and degrade antibody-bound proteins[207]. Normally, the TRIM21 pathway plays a crucial role in disabling and degrading antibody-bound viruses and bacteria that are internalized by cells[219]. However, Schuh and coworkers demonstrated that the TRIM21 pathway could be hijacked to rapidly degrade almost any cytosolic or nuclear protein with antibodies cytosolically delivered by microinjection or electroporation[207].

Finally, physical antibody delivery studies have revealed that IgG antibodies have a surprisingly long half-life in the cytosol. Tracking cytosolic IgG concentrations by radioactivity, autoradiography, or western blotting have all revealed a cytosolic IgG half-life of approximately 24 hours[207,220,221]. Initially, this may seem perplexing given that IgGs cannot fold properly in the reducing environment of the cytosol, but once folded, the majority of IgG disulfide bonds are buried and become significantly more resistant to reduction.

1.4.3 Intrabodies

Intrabodies are intracellularly expressed antibodies that can modulate intracellular protein function. Initially pursued as a strategy for treating HIV with gene therapy[222,223], intrabodies have now emerged as important research tools after major advances in scFv engineering and the advent of small antibody-like binding proteins[224,225]. Initial efforts at using intrabodies primarily relied upon expressing scFvs or Fabs, which are functional in the endoplasmic reticulum (ER)[222], but express poorly in the reducing environment of the cytosol due to poor disulfide bond

formation[45,226]. Accordingly, a multitude of strategies have emerged for increasing scFv stability[225], such as identifying scFv framework sequences that are stable in the cytosol[227,228] or generating scFvs that do not contain disulfide bonds[229,230]. Currently, nanobodies and antibody-like binding proteins (discussed in subchapter 1.2.5) have become the most commonly used intrabody frameworks due to their high stability and the spread of expertise in *in vitro* display technologies.

Well-validated inhibitory intrabodies have been generated against dozens of targets[224,225] and have proved to be particularly useful tools for drug discovery. Rabbitts and coworkers used an anti-Ras inhibitory intrabody to counter-screen a fragment-based small-molecule library and identify a compound capable of inhibiting Ras-effector protein interactions[231]. O'Bryan and coworkers generated an anti-H/KRas monobody that demonstrated H/Kras can be inhibited by blocking their dimerization[232]. Finally, Koide and coworkers developed a monobody that inhibited WDR5, a component of the mixed lineage leukemia 1 (MLL1) methyltransferase complex, to validate WDR5 as a viable drug target for acute myeloid and acute lymphoblastic leukemia[233].

Although inhibitory intrabodies are clearly powerful research tools, they require considerable expertise in *in vitro* display technologies to implement, which can be a major barrier for easy adoption by researchers. Therapeutic intrabodies are possible, but they would require gene therapy, which introduces significant risks associated with insertional mutagenesis. Thus, there is a major need for a system capable of therapeutically delivering inhibitory proteins into the cytosol.

1.4.4 Detecting cytosolic protein delivery

It is crucial to use reporter systems with a low false positive rate when developing cytosolic protein delivery technologies. Most delivery technologies will first induce cargo proteins to bind to the outer surface of the plasma membrane, then be taken up into the cell via some form of endocytosis, and then finally induce endosome escape for protein delivery into the cytosol.

Because the outer plasma membrane and the endosome-lysosome system is topologically equivalent to the extracellular space, ideal reporter systems will only report on cargo delivery to the cytosol[234]. In this section, *total cellular uptake* refers to cargo protein that remain stuck in the endosome-lysosome system in addition to cargo successfully delivered to the cytosol. *Total cellular association* refers to cargo protein bound to the outside surface of the plasma membrane in addition to the total cellular uptake.

The most straightforward method for detecting cytosolic protein delivery is to label cargo proteins with a fluorophore and then following delivery, look for fluorescence in cells via microscopy or flow cytometry. Although simple, fluorescent labeling of cargo proteins only reports the total cellular association of cargo proteins. This method can be improved to report on total cellular uptake by stringently washing cells with trypsin[235], heparin[236], or trypan blue[236] to remove any cargo proteins bound to the outer surface of the plasma membrane[234]. In theory, cargo proteins stuck in endosomes can be distinguished from cytosolically delivered proteins by their punctate fluorescence, but in practice, it is extremely difficult. Fluorescence microscopy must be done on live cells since even mild fixation conditions can cause cargos trapped in the endosome to redistribute into the cytosol[237]. Furthermore, confocal fluorescence microscopy must be used to eliminate diffuse out-of-plane fluorescence, but even then, it can be difficult to distinguish between aggregated cargo proteins and endo-lysosomal cargo proteins[234]. Finally, fluorescently labeled cargo proteins that are degraded in lysosomes may release the fluorescent dye into the cytosol, resulting in diffuse cytosolic fluorescence without any cytosolic protein delivery. Due to the multiple potential sources of false positives associated with tracking fluorescently labeled cargo proteins, any study that claims cytosolic protein delivery using only fluorescently labeled cargo proteins must be treated skeptically.

Even some techniques that try to separate and distinguish between cytosolic and endo-lysosomal cargo proteins can be flawed. Ultracentrifugation can be used to separate the

endosome fraction of cellular lysates from the cytosolic fraction, but it is technically very challenging and rarely achieves complete separation. Cargo proteins can be conjugated to pH-sensitive fluorophores that are non-fluorescent in the acidic environment of endosomes or lysosomes, but the fluorophore may still be weakly fluorescent in some endosomes and free fluorophores can still be released into the cytosol following lysosomal degradation of the cargo protein. Given that many cytosolic protein delivery processes achieve an extremely low (<<10%) endosomal escape efficiency[238,239], even a small contaminant signal from non-cytosolic cargo proteins can swamp out the cytosolic signal and significantly complicate data interpretation.

Toxic enzymes that induce cell death such as caspase 3[240] or RNAse[241] are frequently used as cargo proteins so cell death can be used as a proxy for cytosolic delivery efficiency. These systems should be approached cautiously because it can be difficult to differentiate between toxicity due to successful cytosolic delivery from toxicity due to the delivery system[234].

Several stringent reporter-gene assays have been developed for detecting cytosolic delivery. The splice-correction assay uses reporter cells containing a luciferase gene interrupted by an intron that can be spliced out by cytosolic RNA delivery, resulting in turn-on luciferase expression[242]. Delivery of genome-editing proteins such as Cre recombinase, TALENS, or Cas9 can be used to turn on expression of fluorescent proteins[243,244]. Finally, cargo proteins can be conjugated to dexamethasone, which upon cytosolic delivery, can activate a synthetic glucocorticoid receptor that induces nuclear translocation of a transcription factor that activates reporter gene expression[245]. These reporter-gene assays generally have low false positive rates and are extremely sensitive due to transcriptional amplification of the cytosolic delivery signal[234]. However, these assays have poor dynamic range and their signals plateau once cytosolic cargo concentrations rise above a low cutoff concentration. For example, genome-editing reporter systems only require a few molecules to reach the nucleus to fully turn on. Thus, reporter-gene assays may be misleading if the goal is to achieve high cytosolic concentrations of the cargo.

Reporter systems that rely on protein cargos to interact with cytosolic enzymes can be extremely reliable because not only do they have a low false-positive rate, but they also have a large dynamic range[234]. Plückthun and coworkers labeled cargo proteins with an avi-tag, a substrate for the *E. coli* biotin ligase (BirA), so they could be biotinylated upon cytosolic delivery into BirA-expressing cells, which can be subsequently detected by western blotting[238]. The main downside to the biotinylation reporter system is the low-throughput nature of western blotting. Kritzer and coworkers developed the chloroalkane penetration assay (CAPA), which uses chloroalkane-labeled cargos and HaloTag-expressing reporter cells[246]. Upon cytosolic delivery, cargos will covalently link to HaloTag via the chloroalkane, which will subsequently block HaloTag from interacting with chloroalkane-labeled fluorophores, resulting in a fluorescent signal inversely proportional to the extent of cargo delivery. The main limitations to CAPA is its low sensitivity, the need to conjugate chloroalkanes to cargos, and the potential for cargo-degradation related artifacts.

Finally, protein complementation-based reporter systems have also emerged as reliable methods for assessing cytosolic protein delivery[234]. The most prominent is the splitGFP reporter system, which relies upon a self-reassembling splitGFP that is asymmetrically split between a large splitGFP(1-10) half and a short 15 residues-long splitGFP S11 half[247]. Cargo proteins are expressed with the S11 reporter peptide and upon cytosolic delivery into reporter cells expressing splitGFP(1-10), induces splitGFP complementation and turn-on fluorescence[242,248–250]. The splitGFP reporter system has a low false positive rate, has moderate throughput, and requires no extraneous modification of the cargo protein.

1.4.5 Cell-penetrating peptides (CPPs)

Cell penetrating peptides (CPPs) are short, poly-cationic peptides that can induce cellular uptake of not only the CPP, but also cargos conjugated to them. The first CPPs were identified almost 30 years ago as fragments of naturally occurring proteins such as penetratin from the

Drosophilia transcription factor Antennapedia[251] and Tat from the HIV protein trans-activator of transcription (Tat)[252]. Sugiura and coworkers then found that simply poly-arginine peptides could also translocate into cells[253]. These initial observations resulted in immense excitement for the potential of CPPs resulting in the development of well over a hundred purported CPP sequences[254].

Even though there has been an enormous amount of research into CPPs, evaluating CPP efficacy can be extremely difficult. CPPs were initially believed to directly cross the plasma membrane since their cytosolic uptake appeared to be unaffected by low temperature (4°C) and ATP-depletion[251–253], but Lebleu and coworkers demonstrated in 2003 that these observations were due to artifacts associated with chemical fixation[237]. Thus, any conclusions about CPPs drawn from fluorescence microscopy of fixed cells must be discarded. Even though CPP studies since 2003 have largely avoided fixation, most studies still relied upon tracking fluorescently labeled CPPs, which tend to give rise to false-positives as discussed in subchapter 1.4.4.

Due to the widespread use of cytosolic delivery detection methods that are prone to false-positives and the tendency for researchers to focus on different CPPs, the mechanisms by which CPPs enter the cytosol remain largely elusive[234,255]. The most accepted mechanism for CPP-mediated delivery of large cargos is that the CPP first binds to the outer surface of the plasma membrane through either electrostatic interactions between the CPP and negatively charged surface proteoglycans[234,256–258] or direct binding of the CPP to surface receptors[259]. Afterwards, the CPP as well as any cargo is taken up by the cell via some form of endocytosis or macropinocytosis[234,255,257]. Finally, the CPP induces disruption of the endosome to induce endosomal escape and cytosolic delivery for the cargo. The endosomal escape step is the least understood due to difficulties associated with studying endosome dynamics within living cells[234,257,260], but is believed to be extremely inefficient for CPPs[243] and is the bottleneck for CPP-mediated cytosolic delivery of large cargos.

Although the enthusiasm for them has largely waned, improved CPPs continue to be developed by using more stringent reporter systems[242,246,261,262]. Unfortunately, many of these studies only track delivery when the CPPs are conjugated to very small cargos and it is not clear how effective newer CPPs are at delivering larger proteins.

1.4.6 Improving the endosomal escape efficiency of cell-penetrating peptides

In a seminal paper in 2004, Dowdy and coworkers demonstrated using a stringent Cre recombinase reporter system that the limiting factor for CPP-mediated cytosolic delivery of large protein cargos is the endosome escape step[243]. Endosomal escape efficiency and thus cytosolic delivery was improved for the Tat CPP when it was conjugated it to the endosomolytic peptide HA2[243]. HA2 is derived from the N-terminus of the influenza hemagglutinin protein, which undergoes pH-dependent conformational changes as the lumen of a maturing endosome acidifies to facilitate endosome escape for the influenza virus[263]. At a slightly acidic pH, protonation of HA2 side chains increases its hydrophobicity, allowing it to insert and then disrupt the endosome membrane[264]. The HA2 endosomolytic peptide has since been used to improve cytosolic delivery efficiencies for a supercharged GFP[265], the tumor suppressor p53[266] and siRNA[267].

Given the potential of endosomolytic peptides to solve the endosome escape problem, numerous endosomolytic peptides have now been developed. We will briefly discuss a few notable examples. Pellois and coworkers found that dimerized Tat CPP (dfTAT) exhibited profoundly improved endosomal escape efficiency, allowing for cytosolic delivery of transcription factors that activated a luciferase reporter gene[261]. Liu and coworkers screened a panel of antimicrobial peptides and found that aurein 1.2 disrupted endosomes to improve delivery of the Cre recombinase[268]. Finally, Futaki and coworkers developed L17E by modifying a membrane-lytic peptide to only disrupt membranes in low pH environments and used it to deliver Cre recombinase and the toxin saporin[269]. Although it is clear that endosomolytic peptides improve CPP effectiveness, endosomal escape efficiency remains low, though, since high peptide or

concentrations must be used for detectable delivery. Furthermore, most studies have still only delivered cargo proteins capable of amplifying their effects such as enzymes or transcription factors. In chapter 3, we will discuss out attempts at improving the endosomal escape efficiency for CPPs to enable delivery of small proteins and IgG antibodies.

1.4.7 Activatable cell penetrating peptides (aCPPs)

Even if the cytosolic delivery efficiency for CPPs can be dramatically improved for large protein cargos, CPPs lack any intrinsic targeting capability, limiting the potential targets for therapeutic cytosolic antibodies. Tsien and coworkers have developed aCPPs, which utilize the elevated expression of specific extracellular proteases under pathological conditions to engineer tissue specificity into CPPs[270]. Since CPPs are largely poly-cationic, they can be linked to a poly-anionic sequence to mask their capacity to induce cellular uptake[270]. By designing the linker sequence to be protease cleavable, aCPPs are only functional in tissues with high expression of the protease[270–273]. Matrix metalloproteinases 2 and 9 (MMP-2/9) are commonly used for aCPP activation in therapeutic applications for oncology because not only are they heavily secreted in tumors, but their substrate sequence and specificity are well known[274–276]. Currently, aCPPs have demonstrated utility in tumor imaging[270] as well as fluorescence-guided surgery[271].

1.4.8 Cytosolic delivery with bacterial toxins

Many bacteria naturally secrete toxins that can penetrate into the cytosol of cells. The most studied bacterial toxins are of the AB-toxin class such as the anthrax and diphtheria toxins[277]. AB-toxins consist of an enzymatic A domain that causes toxicity and a delivery B domain that is responsible for cell uptake and translocation of the A domain across endosome membranes into the cytosol[277]. Due to their natural modularity, AB-toxins have been exploited to delivery cargo proteins such as DARPins, monobodies, and affibodies into the cytosol[238,278,279].

Although there has been some success, toxin-mediated delivery of protein cargos faces multiple major limitations. First, for translocation, A domains must be unfolded to be translocated across endosome membranes and then refold in the cytosol[277,280]. This limits cargo delivery to proteins that can also be unfolded and then spontaneously refold in the cytosol to regain functionality[278]. Second, cell or tissue-specific delivery using toxin-mediated delivery requires significant engineering to simultaneously delete the B domain's innate receptor binding activity and to introduce a new binding moiety[281]. Finally, bacterial toxins can be extremely immunogenic and it is not clear whether immunogenicity can be decreased without compromising activity[282,283].

1.4.9 Small proteins with intrinsic cytosolic delivery capabilities

Because CPP internalization into cells is primarily due to their poly-cationic nature, many groups have grafted positively charged amino acids onto solvent exposed area of small proteins to grant them the ability to enter the cytosol. Schepartz and coworkers developed small α-helical proteins with arginines displayed on one face and found that they could be taken upon by cells via endocytosis and escape early endosomes[284–286]. Barbas and coworkers used tandem zinc finger domains engineered to have a high positive charge to deliver protein cargos into the cytosol[287]. Finally, Liu and coworkers developed supercharged green fluorescent proteins (GFPs) with a net charge up to +48 and demonstrated that they were capable of delivering Cre recombinase into cells[288,289].

Initial studies with these small engineered proteins have been promising, but it is not clear how they escape the endosome, why they appear to be more effective than CPPs, and what kinds of cargo they can efficiently deliver[234]. It may be extremely challenging to significantly reshape the surface of therapeutic proteins without perturbing their activity and there is no obvious route for building in cell-type or tissue-selectivity with this approach.

1.4.10 Antibodies with intrinsic cytosolic delivery capabilities

Under pathological conditions, such as in autoimmune diseases, humans can develop autoantibodies against cytosolic or nuclear antigens. Anti-DNA autoantibodies are commonly found in systemic lupus erythematosus and in rare cases, are capable of entering the cytosol of cells[290,291]. Weisbart and coworkers identified an anti-DNA 3E10 antibody that appear to be taken up by cells through the nucleoside salvage pathway by the equilibrative nucleoside transporter ENT2[292]. They later found that a 3E10 scFv was capable of sensitizing cancer cells to radiation and DNA-damaging agents via inhibition of DNA repair[293].

Kim and coworkers identified an anti-DNA m3D8 monoclonal antibody from a mouse model of autoimmune disease[294], found that it could be spontaneously taken up into cells by caveolae-mediated endocytosis, and ultimately spontaneously enter the cytosol[291]. They mapped m3D8's cell-penetrating capability to its VL domain and then grafted m3D8's VL CDR sequences to a human IgG1 framework to generate cytotransmabs – humanized IgGs containing a fixed VL domain that grants it intrinsic cell-penetrating capability[295]. Using quantitative western blotting and a modified splitGFP reporter system, Kim and coworkers estimated a cellular uptake efficiency of ~0.5% and an endosomal escape efficiency of 1 to 4%[296]. Cytotransmabs are believed to first bind to heparin-sulfate proteoglycans to induce cellular uptake and then, in the acidic environment of maturing endosomes, become capable of binding to and disrupting endosome membranes for endosome escape[297]. To demonstrate the utility of cytotransmabs, Kim and coworkers used yeast display to identify a VH domain capable of inhibiting KRas and transplanted the VH domain to the cytotransmabs framework to create an anti-KRas cytotransmab called RT11[298]. RT11 was capable of inhibiting KRas signaling moderately at high concentrations in cell culture, but inhibited tumor growth in murine models rather modestly even at extremely high concentrations[298].

Although Kim and coworkers have demonstrated with RT11 that cytotransmabs could potentially be powerful therapeutic cytosolic antbodies[298], cytotransmabs come with several major drawbacks. First, its delivery capability is found within the VL domain so cytotransmabs can only bind to target proteins via its VH domain. This imposes significant constraints when generating new inhibitory cytotransmabs and cytotransmabs cannot be easily adapted from pre-existing scFv or antibody sequences. Second, RT11 inhibited KRas rather modestly so cytotransmabs likely need additional improvements in cytosolic delivery efficiency to achieve any significant therapeutic benefits in humans. Finally, Kim and coworkers engineered tissue specificity into RT11 by conjugating it to the RGD10 cyclic peptide, which moderately improved tumor accumulation. Bispecific antibodies must be constructed, however, to use protein-based targeting ligands for cytotransmabs, which can be technically challenging. Given these limitations, it remains to be seen whether cytotransmabs will be adopted by the cytosolic protein delivery community.

1.4.11 Nanocarriers for cytosolic protein delivery

Similarly to nonviral nucleic acid delivery, in nanocarrier-mediated protein delivery, cargo proteins are complexed with delivery lipid or polymer formulations that can induce endocytic uptake and then destabilization of the endosome membrane to allow for cytoplasmic release of cargo proteins[206,234]. This takes advantage of advances in non-viral nucleic acid delivery that have led to delivery formulations much better at endosomal escape than CPPs[299,300]. For example, Alnylam has recently received FDA approval for lipid nanocarrier-mediated delivery of a siRNA (Patisiran) into hepatocytes for hereditary transthyretin amyloidosis[300]. Depending on the formulation, nanocarriers can induce endosome escape via membrane fusion, using osmotic pressure, via nanocarrier swelling, or by directly destabilizing the membrane[301].

Although many nanocarrier systems for protein delivery have been developed, they should be evaluated cautiously – most were developed by tracking fluorescently labeled protein

cargos, which is extremely prone to false positives (see subchapter 1.4.4). In fact, evaluation of 4 commercially available protein transfection systems for IgG delivery using a very stringent Cre-recombinase based cytoplasmic delivery reporter system revealed that none were capable of delivery to > 6% of cells, indicating that these systems suffer from poor efficacy, poor reproducibility, or likely both[302]. We will, however, review some of the more well-validated approaches for cytosolic protein delivery.

One of the major challenges for nanocarrier-mediated protein delivery is poor protein encapsulation efficiency. Classic methods such as reverse phase evaporation, detergent removal, thin film hydration, or freeze-thaw cycling result in low protein loading due to poor protein stability during the encapsulation process[303,304]. Nucleic acids are highly negative charged due to their polyphosphate backbone, allowing them to be efficiently encapsulated via electrostatic interactions with cationic or ionizable delivery formulations[305,306]. Similarly, Behr and coworkers found that proteins that are natively negatively charged – such as β-galactosidase (β-Gal) – can be efficiently complexed with the cationic lipospermine dioctadecylglycylspermine (DOGS) under low salt conditions[307]. Cheng and coworkers also found that negatively charged bovine serum albumin (BSA) and β-Gal could be complexed with positively charged fluoroamphiphiles[308]. Using these principles, some IgGs with a negative net charge have been reported to be delivered into cells with nanocarriers, though it is unclear what their cytosolic delivery efficiencies were[307,309].

Given that protein encapsulation efficiency increases when cargo proteins can interact with the delivery lipids[304,307], many groups have sought to directly engineer cargo proteins to improve protein loading. Kataoka and coworkers decreased the net charge of cargo proteins by reversibly conjugating citraconic or *cis*-aconitic amides to the protein's surface lysines, allowing them to be complexed with cationic delivery polymers[310,311]. These modifications are reversible at endosomal pH, allowing for the release of unmodified proteins into the cytosol. Using this protein

charge-conversion method, they were able to demonstrate successful cytosolic delivery of cytochrome c[311] and IgGs[310]. Charge-conversion, however, must be carefully monitored to prevent excessive cargo protein modification[312]. Rotello and coworkers found that β-Gal could complex with and be cytosolically delivered by cationic gold nanoparticles[313]. They extended this approach to other proteins that are not natively negatively charged by fusing them to an oligo-glutamate sequence and complexing them with arginine-functionalized gold nanoparticles[314]. Anderson and coworkers conjugated oligonucleotides to horseradish peroxidase as well as neutravidin and demonstrated that they could be complexed with potent lipidoids designed for siRNA delivery[315]. Finally, Zuber and coworkers bound His-tagged proteins to a Ni^{2+}-functionalized polyanionic polymer, which could then be complexed with cationic pyridylthiourea-grafted polyethylenimine to enable cytosolic protein delivery[316].

The most exciting progress on nanocarrier-mediated protein delivery has been with genome editing proteins. Because genome editing protein delivery applications use gene activation reporter assays, the resulting methods tend to have low false-positive rates. In a pioneering study, Liu and coworkers discovered that a negatively supercharged GFP (-30GFP) could be fused to Cas9, Cre recombinase, and TALENs to enable complexation with commercially available cationic lipids for genome editing[244]. This approach was used to treat autosomal dominant hearing loss via gene-disruption following local injections of Cas9-RNA-lipid complexes into the cochlea of neonatal mice[317]. The complexation strategy was also compatible with bio-reducible delivery formulations and enabled the delivery of Cre recombinase into mouse brains following a local injection[318]. More recently, anionic human proteins such as prothymosin alpha (ProTα) were found to be even more effective than -30GFP at cytosolic protein delivery following complexation with cationic lipids[319].

Because the Cas9 protein is similar in size to IgGs, these delivery principles may be also applicable to cytosolic antibody delivery. However, since inhibitory antibodies must be

stoichiometric relative to their target to achieve adequate target inhibition, they will require greater delivery efficiency than genome editing proteins. Furthermore, all of the genome editing protein delivery systems described thus far have only been demonstrated to have *in vivo* efficacy following local injection[244,317,318,320,321], suggesting that they form protein-lipid complexes that are unstable in serum or have poor tissue-targeting capability.

Making therapeutic cytoplasmically delivered IgGs a reality requires 1) a reproducible method for efficiently encapsulating a large variety of IgGs into stable nanocarrier formulations, 2) stringent testing of formulations using a cell reporter assay that only detects cytoplasmic protein delivery, and 3) demonstration that cytoplasmically delivered IgGs can inhibit therapeutically relevant targets in both cell culture and model organisms. In chapter 4, we will discuss our strategy for complexing IgG antibodies with cationic lipids to enable efficient cytosolic delivery in living cells.

1.4.12 Targeting with nanocarriers

Following successful protein encapsulation or complexation into a cytosolic delivery formulation, the nanocarrier must then be capable of delivering its cargo protein to the desired tissue. If the target tissue or tumor is easily accessible, local delivery can be an attractive option since the entire nanocarrier dose reaches the target and it abrogates the need to optimize nanocarriers for serum half-life. Unfortunately, local delivery cannot target metastases and it may be difficult to access primary tumors deep within the body. For systematic nanocarrier administration, targeting is typically accomplished via passive or active approaches, but can also be strongly affected by the protein corona that forms on the nanocarrier surface following administration[322]. Currently, most targeting studies have focused on nanocarrier-mediated delivery to solid tumors, which we will give a brief overview on.

Passive targeting typically refers to utilization of the enhanced permeability and retention (EPR) effect. First described in 1986 by Maeda and coworkers[323], EPR refers to the passive

accumulation of small nanocarriers in tumors due to the tumors' leaky vasculature and poor lymphatic drainage. Although the EPR effect has been validated in tumor xenograft models[322,324], it is unclear how translatable these findings are to tumors in humans[325]. Variable endothelial leakiness, perfusion, and interstitial pressure in tumors often result in heterogeneous nanoparticle extravasation[322]. A recent meta-analysis of nanoparticle delivery to solid tumors via calculates that with only passive targeting, ~0.6% of the injected dose (ID) actually reaches the target tumor[325].

Active targeting usually refers to functionalizing nanocarriers with targeting ligands that bind to targets associated with a desired tissue to increase nanocarrier retention. Common ligands for tumor targeting includes small molecules like folate[326], peptides that bind to integrins[327], and antibodies or antibody-like binding proteins that bind to membrane proteins upregulated in cancer cells[328]. The design of active targeting nanocarriers can be tricky, requiring significant optimization of ligand identity, conjugation approach, and ligand density[322,328]. There is considerable debate as to whether active targeting even improves nanocarrier accumulation in tumors[322,328,329]. Kirpotin and coworkers provided evidence that ligand targeting increased nanoparticle internalization rates at tumors, but not accumulation[330]. Similarly, modeling analysis suggests that targeting fails to improve accumulation for nanocarriers larger than 50 nm[329]. A recent meta-analysis, however, suggests that active targeting does improves ID from 0.6% to 0.9%[325]. Since active targeting definitely does improve cellular uptake of nanocarriers[330], it has been used extensively to selectively induce targeted cells to uptake macromolecules or cytotoxic drugs[322,328]. The limiting factor for actively targeting nanocarriers to deliver payloads into the cytosol is its endosomal escape efficiency.

Upon intravenous administration, serum proteins rapidly adsorb onto the surface of nanocarriers, creating the protein corona[331]. The exact composition of the protein corona can negate the capabilities of targeting ligands[332], extend or reduce circulation time[333], or even

determine nanocarrier tissue specificity. For the latter, one prominent example is that Alnylam's Patisiran achieves hepatocyte targeting due to adsorption of apolipoprotein E (apoE) onto their lipid nanoparticle surface[334]. Hepatocyte gene silencing was completely abolished when patisiran was injected into ApoE$^{-/-}$ mice, but was restored with co-administration of synthetic ApoE along with patisiran in the transgenic mice.

For therapeutic cytosolic protein delivery into solid tumors, protein payloads must be designed while factoring in that for many delivery formulations, < 1% of the injected nanocarrier dose will accumulate in the tumor[325]. The proportion of injected cargo protein that ultimately enters the cytosol of cancer cells will be even lower given the endosomal escape bottleneck. Thus, tumors must be exquisitely sensitive to inhibition of the therapeutic cytosolic protein's target or, alternatively, the therapeutic protein cargo must be catalytic in nature. In subchapter 5.3.5, we will discuss strategies for making inhibitory antibodies catalytic through antibody-induced protein degradation.

1.5 References

1. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
2. Patel, M. N., Halling-Brown, M. D., Tym, J. E., Workman, P. & Al-Lazikani, B. Objective assessment of cancer genes for drug discovery. *Nat. Rev. Drug Discov.* **12**, 35–50 (2013).
3. Finan, C. *et al.* The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166 (2017).
4. Oprea, T. I. *et al.* Unexplored therapeutic opportunities in the human genome. *Nat. Rev. Drug Discov.* **17**, 317–332 (2018).
5. Whitfield, J. R., Beaulieu, M.-E. & Soucek, L. Strategies to Inhibit Myc and Their Clinical Applicability. *Front. Cell Dev. Biol.* **5**, 10 (2017).
6. Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71–75 (2007).
7. Mapp, A. K., Pricer, R. & Sturlis, S. Targeting transcription is no longer a quixotic quest. *Nat. Chem. Biol.* **11**, 891–894 (2015).
8. Cesa, L. C., Mapp, A. K. & Gestwicki, J. E. Direct and Propagated Effects of Small

Molecules on Protein–Protein Interaction Networks. *Front. Bioeng. Biotechnol.* **3**, 119 (2015).

9. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001–1009 (2007).
10. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
11. Carter, P. J. & Lazar, G. A. Next generation antibody drugs: pursuit of the 'high-hanging fruit'. *Nat. Rev. Drug Discov.* **17**, 197–223 (2018).
12. Lu, L. L., Suscovich, T. J., Fortune, S. M. & Alter, G. Beyond binding: Antibody effector functions in infectious diseases. *Nat. Rev. Immunol.* **18**, 46–61 (2018).
13. Vidarsson, G., Dekkers, G. & Rispens, T. IgG subclasses and allotypes: from structure to effector functions. *Front. Immunol.* **5**, 520 (2014).
14. Alberts, B. *et al.* B Cells and Immunoglobulins. in *Moleulcar Biology of the Cell* 1315–1323 (Garland Science, 2015).
15. Liu, H. & May, K. Disulfide bond structures of IgG molecules: Structural variations, chemical modifications and possible impacts to stability and biological function. *MAbs* **4**, 17–23 (2012).
16. Jennewein, M. F. & Alter, G. The Immunoregulatory Roles of Antibody Glycosylation. *Trends Immunol.* **38**, 358–372 (2017).
17. Radjainia, M., Hyun, J.-K., Leysath, C. E., Leppla, S. H. & Mitra, A. K. Anthrax toxin-neutralizing antibody reconfigures the protective antigen heptamer into a supercomplex. *Proc. Natl. Acad. Sci.* **107**, 14070–14074 (2010).
18. Wild, M. A. *et al.* Human antibodies from immunized donors are protective against anthrax toxin in vivo. *Nat. Biotechnol.* **21**, 1305–1306 (2003).
19. Corti, D. & Lanzavecchia, A. *Broadly Neutralizing Antiviral Antibodies*. *Annual Review of Immunology* **31**, (2013).
20. Sok, D. & Burton, D. R. Recent progress in broadly neutralizing antibodies to HIV. *Nat. Immunol.* **19**, 1179–1188 (2018).
21. Misasi, J. *et al.* Structural and molecular basis for Ebola virus neutralization by protective human antibodies. *Science* **351**, 1343–1346 (2016).
22. Corti, D. *et al.* Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. *Science* **351**, 1339–1342 (2016).
23. Ekiert, D. C. *et al.* Antibody Recognition of a Highly Conserved Influenza Virus Epitope. *Science* **324**, 246–251 (2009).
24. Roopenian, D. C. & Akilesh, S. FcRn: The neonatal Fc receptor comes of age. *Nat. Rev.*

Immunol. **7**, 715–725 (2007).

25. Israel, E. J., Patel, V. K., Taylor, S. F., Marshak-Rothstein, A. & Simister, N. E. Requirement for a beta 2-microglobulin-associated Fc receptor for acquisition of maternal IgG by fetal and neonatal mice. *J. Immunol.* **154**, 6246–51 (1995).
26. Simister, N. E. & Mostov, K. E. An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184–187 (1989).
27. Junghans, R. P. P. & Anderson, C. L. The protection receptor for IgG catabolism is the beta2-microglobulin-containing neonatal intestinal transport receptor. *Proc. Natl. Acad. Sci.* **93**, 5512–5516 (1996).
28. Raghavan, M., Bonagura, V. R., Morrison, S. L. & Bjorkman, P. J. Analysis of the pH Dependence of the Neonatal Fc Receptor/Immunoglobulin G Interaction Using Antibody and Receptor Variants. *Biochemistry* **34**, 14649–14657 (1995).
29. Casadevall, A. Passive Antibody Therapies : Progress and Continuing Challenges. *Clin. Immunol.* **93**, 5–15 (1999).
30. Köhler, G. & Milstein, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
31. Saeed, A. F. U. H., Wang, R., Ling, S. & Wang, S. Antibody engineering for pursuing a healthier future. *Front. Microbiol.* **8**, 1–28 (2017).
32. Liu, J. K. H. The history of monoclonal antibody development - Progress, remaining challenges and future innovations. *Ann. Med. Surg.* **3**, 113–116 (2014).
33. Kunert, R. & Reinhart, D. Advances in recombinant antibody manufacturing. *Appl. Microbiol. Biotechnol.* **100**, 3451–3461 (2016).
34. Buss, N. A. P. S., Henderson, S. J., McFarlane, M., Shenton, J. M. & De Haan, L. Monoclonal antibody therapeutics: History and future. *Curr. Opin. Pharmacol.* **12**, 615–622 (2012).
35. Winter, G. & Harris, W. J. Humanized antibodies. *Trends Pharmacol. Sci.* **14**, 139–43 (1993).
36. Lonberg, N. Human antibodies from transgenic animals. *Nat. Biotechnol.* **23**, 1117–1125 (2005).
37. Bradbury, A. R. M., Sidhu, S., Dübel, S. & McCafferty, J. Beyond natural antibodies: The power of in vitro display technologies. *Nat. Biotechnol.* **29**, 245–254 (2011).
38. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552–554 (1990).
39. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).

40. Venkataraman, A. *et al.* A toolbox of immunoprecipitation-grade monoclonal antibodies to human transcription factors. *Nat. Methods* **15**, 330–338 (2018).
41. Marcon, E. *et al.* Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat. Methods* **12**, 725–731 (2015).
42. Yu, X., Yang, Y.-P., Dikici, E., Deo, S. K. & Daunert, S. Beyond Antibodies as Binding Partners: The Role of Antibody Mimetics in Bioanalysis. *Annu. Rev. Anal. Chem.* **10**, 293–320 (2017).
43. Holliger, P. & Hudson, P. J. Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* **23**, 1126–1136 (2005).
44. Ward, E. S. *et al.* Generating binding activities from Escherichia coli by expression of a repertoire of immunoglobulin variable domains. *Nature* **341**, 544–546 (1989).
45. Wörn, A. & Plückthun, A. Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989–1010 (2001).
46. Skerra, A. & Plückthun, A. Assembly of a functional immunoglobulin Fv fragment in Escherichia coli. *Science* **240**, 1038–1041 (1988).
47. Proba, K., Ge, L. & Plückthun, A. Functional antibody single-chain fragments from the cytoplasm of Escherichia coli: influence of thioredoxin reductase (TrxB). *Gene* **159**, 203–207 (1995).
48. Hamer-Casterman Atarchouch,T, C. *et al.* Naturally occurring antibodies devoid of light chains. *Nature* **363**, 446–448 (1998).
49. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. *Annu. Rev. Biochem.* **82**, 775–797 (2013).
50. Moutel, S. *et al.* NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. *Elife* **5**, 1–31 (2016).
51. Nord, K. *et al.* Coinbinatorial Libraries of an a-Helical Bacterial Receptor Doinain. *Nat. Biotechnol.* **15**, 772–777 (1997).
52. Plückthun, A. Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).
53. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).
54. James, M. L. & Gambhir, S. S. A Molecular Imaging Primer: Modalities, Imaging Agents, and Applications. *Physiol. Rev.* **92**, 897–965 (2012).
55. Strohl, W. R. Fusion Proteins for Half-Life Extension of Biologics as a Strategy to Make Biobetters. *BioDrugs* **29**, 215–239 (2015).

56. Hoelder, S., Clarke, P. A. & Workman, P. Discovery of small molecule cancer drugs: Successes, challenges and opportunities. *Mol. Oncol.* **6**, 155–176 (2012).
57. Ramaraj, T., Angel, T., Dratz, E. A., Jesaitis, A. J. & Mumey, B. Antigen-antibody interface properties: Composition, residue interactions, and features of 53 non-redundant structures. *Biochim. Biophys. Acta - Proteins Proteomics* **1824**, 520–532 (2012).
58. Maloney, D. G. Anti-CD20 Antibody Therapy for B-Cell Lymphomas. *N. Engl. J. Med.* **366**, 2008–2016 (2012).
59. Beck, A., Goetsch, L., Dumontet, C. & Corvaïa, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16**, 315–337 (2017).
60. Chudasama, V., Maruani, A. & Caddick, S. Recent advances in the construction of antibody-drug conjugates. *Nat. Chem.* **8**, 114–119 (2016).
61. Maude, S. L. *et al.* Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N. Engl. J. Med.* **378**, 439–448 (2018).
62. Schuster, S. J. *et al.* Tisagenlecleucel in Adult Relapsed or Refractory Diffuse Large B-Cell Lymphoma. *N. Engl. J. Med.* **380**, 45–56 (2018).
63. June, C. H. & Sadelain, M. Chimeric Antigen Receptor Therapy. *N. Engl. J. Med.* **379**, 64–73 (2018).
64. Garon, E. B. *et al.* Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
65. Hodi, F. S. *et al.* Improved Survival with Ipilimumab in Patients with Metastatic Melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
66. Qasim, W. *et al.* Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci. Transl. Med.* **9**, eaaj2013 (2017).
67. Sadelain, M., Themeli, M., Rivie, I., Rivière, I. & Sadelain, M. New cell sources for T cell engineering and adoptive immunotherapy. *Cell Stem Cell* **16**, 357–366 (2015).
68. Labrijn, A. F., Janmaat, M. L., Reichert, J. M. & Parren, P. W. H. I. H. I. Bispecific antibodies: a mechanistic review of the pipeline. *Nat. Rev. Drug Discov.* (2019). doi:10.1038/s41573-019-0028-1
69. Staerz, U. D., Kanagawa, O. & Bevan, M. J. Hybrid antibodies can target sites for attack by T cells. *Nature* **314**, 628–631 (1985).
70. Kantarjian, H. *et al.* Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia. *N. Engl. J. Med.* **376**, 836–847 (2017).
71. Sevigny, J. *et al.* The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* **537**, 50–56 (2016).
72. Yu, Y. J. *et al.* Therapeutic bispecific antibodies cross the blood-brain barrier in nonhuman

primates. *Sci. Transl. Med.* **6**, 261ra154-261ra154 (2014).

73. Niewoehner, J. *et al.* Increased Brain Penetration and Potency of a Therapeutic Antibody Using a Monovalent Molecular Shuttle. *Neuron* **81**, 49–60 (2014).
74. Zuchero, Y. J. Y. *et al.* Discovery of Novel Blood-Brain Barrier Targets to Enhance Brain Uptake of Therapeutic Antibodies. *Neuron* **89**, 70–82 (2016).
75. Wang, X., Wang, R., Zhang, Y. & Zhang, H. Evolutionary survey of druggable protein targets with respect to their subcellular localizations. *Genome Biol. Evol.* **5**, 1291–1297 (2013).
76. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).
77. Means, G. E. & Feeney, R. E. Chemical Modifications of Proteins: History and Applications. *Bioconjug. Chem.* **1**, 2–12 (1990).
78. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**, 1–14 (2014).
79. Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **7**, 876–884 (2011).
80. Hoyt, E. A., Cal, P. M. S. D., Oliveira, B. L. & Bernardes, G. J. L. Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **3**, 147–171 (2019).
81. Shen, B. Q. *et al.* Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nat. Biotechnol.* **30**, 184–189 (2012).
82. Krall, N., Da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113 (2016).
83. Liu, H., Chumsae, C., Gaza-Bulseco, G., Hurkmans, K. & Radziejewski, C. H. Ranking the susceptibility of disulfide bonds in human IgG1 antibodies by reduction, differential alkylation, and LC-MS analysis. *Anal. Chem.* **82**, 5219–5226 (2010).
84. Sun, M. M. C. *et al.* Reduction-alkylation strategies for the modification of specific monoclonal antibody bisulfides. *Bioconjug. Chem.* **16**, 1282–1290 (2005).
85. Agarwal, P. & Bertozzi, C. R. Site-specific antibody-drug conjugates: The nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug. Chem.* **26**, 176–192 (2015).
86. Junutula, J. R. *et al.* Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* **26**, 925–932 (2008).
87. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
88. Nakane, P. K. & Kawaoi, A. Peroxidase-labeled antibody. A new method of conjugation. *J.*

Histochem. Cytochem. **22**, 1084–1091 (1974).

89. Zhou, Q. *et al.* Site-Specific Antibody–Drug Conjugation through Glycoengineering. *Bioconjug. Chem.* **25**, 510–520 (2014).
90. Li, X., Fang, T. & Boons, G. Preparation of Well-Defined Antibody-Drug Conjugates through Glycan Remodeling and Strain-Promoted Azide-Alkyne Cycloadditions. *Angew. Chem. Int. Ed.* **53**, 7179–7182 (2014).
91. Kölmel, D. K. & Kool, E. T. Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem. Rev.* **117**, 10358–10376 (2017).
92. Saxon, E. & Bertozzi, C. R. Cell Surface Engineering by a Modified Staudinger Reaction. *Science* **287**, 2007–2010 (2000).
93. Rostovtsev, V. V, Green, L. G., Fokin, V. V & Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
94. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **67**, 3057–3064 (2002).
95. Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* **126**, 15046–15047 (2004).
96. Baskin, J. M. *et al.* Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci.* **104**, 16793–16797 (2007).
97. Ning, X., Guo, J., Wolfert, M. A. & Boons, G.-J. Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-Free and Fast Huisgen Cycloadditions. *Angew. Chem. Int. Ed.* **47**, 2253–2255 (2008).
98. Oliveira, B. L., Guo, Z. & Bernardes, G. J. L. Inverse electron demand Diels-Alder reactions in chemical biology. *Chem. Soc. Rev.* **46**, 4895–4950 (2017).
99. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine Ligation : Fast Bioconjugation Based on Tetrazine Ligation : Fast Bioconjugation Based on Inverse-Electron-Demand. *J. Am. Chem. Soc.* **130**, 13518–13519 (2008).
100. Zeglis, B. M. *et al.* A Pretargeted PET Imaging Strategy Based on Bioorthogonal Diels-Alder Click Chemistry. *J. Nucl. Med.* **54**, 1389–1396 (2013).
101. Rossin, R. *et al.* In vivo chemistry for pretargeted tumor imaging in live mice. *Angew. Chem. Int. Ed.* **49**, 3375–3378 (2010).
102. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

103. Muir, T. W. Semisynthesis of Proteins by Expressed Protein Ligation. *Annu. Rev. Biochem.* **72**, 249–289 (2003).
104. Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nat. Rev. Chem.* **2**, (2018).
105. Kochendoerfer, G. G. Design and Chemical Synthesis of a Homogeneous Polymer-Modified Erythropoiesis Protein. *Science* **299**, 884–887 (2003).
106. Torbeev, V. Y. & Kent, S. B. H. Convergent chemical synthesis and crystal structure of a 203 amino acid 'covalent dimer' HIV-1 protease enzyme molecule. *Angew. Chem. Int. Ed.* **46**, 1667–1670 (2007).
107. Shah, N. H. & Muir, T. W. Inteins: Nature's gift to protein chemists. *Chem. Sci.* **5**, 446–461 (2014).
108. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95**, 6705–6710 (1998).
109. Bain, J. D., Glabe, C. G., Dix, T. A., Chamberlin, A. R. & Diala, E. S. Biosynthetic Site-Specific Incorporation of a Non-Natural Amino Acid into a Polypeptide. *J. Am. Chem. Soc.* **111**, 8013–8014 (1989).
110. Noren, C., Anthony-Cahill, S., Griffith, M. & Schultz, P. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
111. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **292**, 498–500 (2001).
112. Steer, B. A. & Schimmel, P. Major Anticodon-binding Region Missing from an Archaebacterial tRNA Synthetase. *J. Biol. Chem.* **274**, 35601–35606 (1999).
113. Jakubowski, H. & Goldman, E. Editing of errors in selection of amino acids for protein synthesis. *Microbiol. Rev.* **56**, 412–29 (1992).
114. Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. A New Functional Suppressor tRNA/Aminoacyl−tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins. *J. Am. Chem. Soc.* **122**, 5010–5011 (2000).
115. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883–890 (2001).
116. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
117. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
118. Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal Labeling of Proteins. *Chem. Rev.* **114**, 4764–4806 (2014).
119. Ernst, R. J. *et al.* Genetic code expansion in the mouse brain. *Nat. Chem. Biol.* **12**, 776–778

(2016).
120. Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014).
121. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *J. Mol. Biol.* **395**, 361–374 (2010).
122. Johnson, D. B. F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779–786 (2011).
123. Lajoie, M. J. *et al.* Genomically Recoded Organisms Expand Biological Functions. *Science* **342**, 357–360 (2013).
124. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* **569**, 514–518 (2019).
125. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25**, 770–777 (2007).
126. Park, H.-S. *et al.* Expanding the Genetic Code of Escherichia coli with Phosphoserine. *Science* **333**, 1151–1154 (2011).
127. Axup, J. Y. *et al.* Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc. Natl. Acad. Sci.* **109**, 16101–16106 (2012).
128. Tian, F. *et al.* A general approach to site-specific antibody drug conjugates. *Proc. Natl. Acad. Sci.* **111**, 1766–1771 (2014).
129. Vanbrunt, M. P. *et al.* Genetically Encoded Azide Containing Amino Acid in Mammalian Cells Enables Site-Specific Antibody-Drug Conjugates Using Click Cycloaddition Chemistry. *Bioconjug. Chem.* **26**, 2249–2260 (2015).
130. Lorand, L. & Graham, R. M. Transglutaminases: Crosslinking enzymes with pleiotropic functions. *Nat. Rev. Mol. Cell Biol.* **4**, 140–156 (2003).
131. Lin, C. W. & Ting, A. Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J. Am. Chem. Soc.* **128**, 4542–4543 (2006).
132. Fontana, A., Spolaore, B., Mero, A. & Veronese, F. M. Site-specific modification and PEGylation of pharmaceutical proteins mediated by transglutaminase. *Adv. Drug Deliv. Rev.* **60**, 13–28 (2008).
133. Jeger, S. *et al.* Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. *Angew. Chem. Int. Ed.* **49**, 9995–9997 (2010).
134. Strop, P. *et al.* Location matters: Site of conjugation modulates stability and pharmacokinetics of antibody drug conjugates. *Chem. Biol.* **20**, 161–167 (2013).

135. Dierks, T. *et al.* Multiple sulfatase deficiency is caused by mutations in the gene encoding the human Cα-formylglycine generating enzyme. *Cell* **113**, 435–444 (2003).
136. Carrico, I. S., Carlson, B. L. & Bertozzi, C. R. Introducing genetically encoded aldehydes into proteins. *Nat. Chem. Biol.* **3**, 321–322 (2007).
137. Rush, J. S. & Bertozzi, C. R. New Aldehyde Tag Sequences Identified by Screening Formylglycine Generating Enzymes in Vitro and in Vivo. *J. Am. Chem. Soc.* **130**, 12240–12241 (2008).
138. Wu, P. *et al.* Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proc. Natl. Acad. Sci.* **106**, 3000–3005 (2009).
139. 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. *Bioconjug. Chem.* **25**, 1331–1341 (2014).
140. Agarwal, P. *et al.* Hydrazino-pictet-spengler ligation as a biocompatible method for the generation of stable protein conjugates. *Bioconjug. Chem.* **24**, 846–851 (2013).
141. Liebscher, S. *et al.* N-terminal protein modification by substrate-activated reverse proteolysis. *Angew. Chem. Int. Ed.* **53**, 3024–3028 (2014).
142. Liebscher, S. *et al.* Derivatization of antibody fab fragments: A designer enzyme for native protein modification. *ChemBioChem* **15**, 1096–1100 (2014).
143. Jackson, D. Y. *et al.* A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243–247 (1994).
144. Henager, S. H. *et al.* Enzyme-catalyzed expressed protein ligation. *Nat. Methods* **13**, 925–927 (2016).
145. Braisted, A. C., Kevin Judice, J. & Wells, J. A. Synthesis of proteins by subtiligase. in *Methods in Enzymology* **289**, 298–313 (1997).
146. Weeks, A. M. & Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol.* **14**, 50–57 (2018).
147. Nguyen, G. K. T. *et al.* Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–738 (2014).
148. Nguyen, G. K. T. *et al.* Butelase-mediated cyclization and ligation of peptides and proteins. *Nat. Protoc.* **11**, 1977–1988 (2016).
149. Harmand, T. J. *et al.* One-Pot Dual Labeling of IgG 1 and Preparation of C-to-C Fusion Proteins Through a Combination of Sortase A and Butelase 1. *Bioconjug. Chem.* **29**, 3245–3249 (2018).
150. Yang, R. *et al.* Engineering a Catalytically Efficient Recombinant Protein Ligase. *J. Am.*

Chem. Soc. **139**, 5351–5358 (2017).

151. Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763 (1999).
152. Hendrickx, A. P. A., Budzik, J. M., Oh, S. Y. & Schneewind, O. Architects at the bacterial surface-sortases and the assembly of pili with isopeptide bonds. *Nat. Rev. Microbiol.* **9**, 166–176 (2011).
153. Pishesha, N., Ingram, J. R. & Ploegh, H. L. Sortase A: A Model for Transpeptidation and Its Biological Applications. *Annu. Rev. Cell Dev. Biol.* **34**, 163–188 (2018).
154. Ton-that, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc. Natl. Acad. Sci.* **96**, 12424–12429 (1999).
155. Jacobitz, A. W., Kattke, M. D., Wereszczynski, J. & Clubb, R. T. *Sortase Transpeptidases: Structural Biology and Catalytic Mechanism*. *Advances in Protein Chemistry and Structural Biology* **109**, (Elsevier Inc., 2017).
156. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
157. Schmidt, M., Toplak, A., Quaedflieg, P. J. & Nuijens, T. Enzyme-mediated ligation technologies for peptides and proteins. *Curr. Opin. Chem. Biol.* **38**, 1–7 (2017).
158. Antos, J. M. *et al.* Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–10801 (2009).
159. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: A versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708 (2007).
160. Pritz, S. *et al.* Synthesis of Biologically Active Peptide Nucleic Acid-Peptide Conjugates by Sortase-Mediated Ligation Stephan. *J. Org. Chem.* **72**, 3909–3912 (2007).
161. Antos, J. M., Miller, G. M., Grotenbreg, G. M. & Ploegh, H. L. Lipid Modification of Proteins through Sortase-Catalyzed Transpeptidation. *J. Am. Chem. Soc.* **130**, 16338–16343 (2008).
162. Parthasarathy, R., Subramanian, S. & Boder, E. T. Sortase A as a novel molecular ‘stapler’ for sequence-specific protein conjugation. *Bioconjug. Chem.* **18**, 469–476 (2007).
163. Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V. K. & Roy, R. P. Peptide-sugar ligation catalyzed by transpeptidase sortase: A facile approach to neoglycoconjugate synthesis. *J. Am. Chem. Soc.* **130**, 2132–2133 (2008).
164. Chan, L. *et al.* Covalent attachment of proteins to solid supports and surfaces via sortase-mediated ligation. *PLoS One* **2**, e1164 (2007).
165. Antos, J. M., Truttmann, M. C. & Ploegh, H. L. Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118 (2016).

166. Piotukh, K. *et al.* Directed evolution of sortase A mutants with altered substrate selectivity profiles. *J. Am. Chem. Soc.* **133**, 17536–17539 (2011).
167. Dorr, B. M., Ham, H. O., An, C., Chaikof, E. L. & Liu, D. R. Reprogramming the specificity of sortase enzymes. *Proc. Natl. Acad. Sci.* **111**, 13343–13348 (2014).
168. Bellucci, J. J., Bhattacharyya, J. & Chilkoti, A. A noncanonical function of sortase enables site-specific conjugation of small molecules to lysine residues in proteins. *Angew. Chem. Int. Ed.* **54**, 441–445 (2015).
169. McConnell, S. A. *et al.* Protein Labeling via a Specific Lysine-Isopeptide Bond Using the Pilin Polymerizing Sortase from Corynebacterium diphtheriae. *J. Am. Chem. Soc.* **140**, 8420–8423 (2018).
170. Glasgow, J. E., Salit, M. L. & Cochran, J. R. In Vivo Site-Specific Protein Tagging with Diverse Amines Using an Engineered Sortase Variant. *J. Am. Chem. Soc.* **138**, 7496–7499 (2016).
171. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399–11404 (2011).
172. Chen, L. *et al.* Improved variants of SrtA for site-specific conjugation on antibodies and proteins with high efficiency. *Sci. Rep.* **6**, 1–12 (2016).
173. Hirakawa, H., Ishikawa, S. & Nagamune, T. Design of Ca 2+-independent Staphylococcus aureus sortase A mutants. *Biotechnol. Bioeng.* **109**, 2955–2961 (2012).
174. Hirakawa, H., Ishikawa, S. & Nagamune, T. Ca^{2+}-independent sortase-A exhibits high selective protein ligation activity in the cytoplasm of Escherichia coli. *Biotechnol. J.* **10**, 1487–1492 (2015).
175. 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. *Anal. Chem.* **85**, 11090–11097 (2013).
176. Amirshaghaghi, A. *et al.* Site-Specific Labeling of Cyanine and Porphyrin Dye-Stabilized Nanoemulsions with Affibodies for Cellular Targeting. *J. Am. Chem. Soc.* **140**, 13550–13553 (2018).
177. Shadish, J. A., Benuska, G. M. & DeForest, C. A. Bioactive site-specifically modified proteins for 4D patterning of gel1. Shadish, J. A., Benuska, G. M. & DeForest, C. A. Bioactive site-specifically modified proteins for 4D patterning of gel biomaterials. Nat. Mater. (2019). doi:10.1038/s41563-019-0367-. *Nat. Mater.* (2019). doi:10.1038/s41563-019-0367-7
178. Pegg, A. E. Repair of O6-alkylguanine by alkyltransferases. *Mutat. Res. Mutat. Res.* **462**, 83–100 (2000).
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. Gautier, A. *et al.* An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15**, 128–136 (2008).
181. Los, G. V *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
182. Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
183. Grimm, J. B. *et al.* Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* **13**, 985–988 (2016).
184. Grimm, J. B. *et al.* A general method to fine-tune fluorophores for live-cell and in vivo imaging. *Nat. Methods* **14**, 987–994 (2017).
185. Hagan, R. M. *et al.* NMR spectroscopic and theoretical analysis of a spontaneously formed lys-asp isopeptide bond. *Angew. Chem. Int. Ed.* **49**, 8421–8425 (2010).
186. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E690-7 (2012).
187. Li, L., Fierer, J. O., Rapoport, T. A. & Howarth, M. Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* **426**, 309–317 (2014).
188. Falugi, F., Kim, H. K. & Missiakas, D. M. Role of Protein A in the Evasion of Host Adaptive Immune Responses. *MBio* **4**, 1–9 (2013).
189. Choe, W., Durgannavar, T. A. & Chung, S. J. Fc-binding ligands of immunoglobulin G: An overview of high affinity proteins and peptides. *Materials (Basel).* **9**, (2016).
190. Nilsson, B. *et al.* A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.* **1**, 107–13 (2007).
191. Konrad, A., Eriksson Karlström, A. & Hober, S. Covalent immunoglobulin labeling through a photoactivable synthetic Z domain. *Bioconjug. Chem.* **22**, 2395–2403 (2011).
192. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci.* **99**, 11020–11024 (2002).
193. Hui, J. Z. *et al.* Facile method for the site-specific, covalent attachment of full-length IgG onto nanoparticles. *Small* **10**, 3354–3363 (2014).
194. Hui, J. Z. & Tsourkas, A. Optimization of Photoactive Protein Z for Fast and Efficient Site-Specific Conjugation of Native IgG. *Bioconjug. Chem.* **25**, 1709–1719 (2014).
195. Malakauskas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5**, 470–5 (1998).

196. Chung, B. H. *et al.* Photoactivable antibody binding protein: Site-selective and covalent coupling of antibody. *Anal. Chem.* **81**, 936–942 (2009).
197. Hui, J. Z., Tamsen, S., Song, Y. & Tsourkas, A. LASIC: Light Activated Site-Specific Conjugation of Native IgGs. *Bioconjug. Chem.* **26**, 1456–1460 (2015).
198. Dormán, G. & Prestwich, G. D. Benzophenone Photophores in Biochemistry. *Biochemistry* **33**, 5661–5673 (1994).
199. Wittelsberger, A., Thomas, B. E., Mierke, D. F. & Rosenblatt, M. Methionine acts as a 'magnet' in photoaffinity crosslinking experiments. *FEBS Lett.* **580**, 1872–1876 (2006).
200. Lancia, J. K. *et al.* Sequence context and crosslinking mechanism affect the efficiency of in vivo capture of a protein-protein interaction. *Biopolymers* **101**, 391–397 (2014).
201. Smoak, E. M., Stein, P., Schultz, R. M., Lampson, M. A. & Black, B. E. Long-Term Retention of CENP-A Nucleosomes in Mammalian Oocytes Underpins Transgenerational Inheritance of Centromere Identity. *Curr. Biol.* **26**, 1110–1116 (2016).
202. Rossi, A. *et al.* Genetic compensation induced by deleterious mutations but not gene knockdowns. *Nature* **524**, 230–233 (2015).
203. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nat. Chem. Biol.* **11**, 536–541 (2015).
204. Barber, M. A. A Technic for the Inoculation of Bacteria and Other Substances Into Living Cells. *J. Infect. Dis.* **8**, 348–360 (1911).
205. Neumann, E., Schaefer-Ridder, M., Wang, Y. & Hofschneider, P. H. Gene transfer into mouse lyoma cells by electroporation in high electric fields. *EMBO J.* **1**, 841–845 (1982).
206. Marschall, A. L. J., Frenzel, A., Schirrmann, T., Schüngel, M. & Dübel, S. Targeting antibodies to the cytoplasm. *MAbs* **3**, 3–16 (2011).
207. Clift, D. *et al.* A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell* **171**, 1692-1706.e18 (2017).
208. Stewart, M. P. *et al.* In vitro and ex vivo strategies for intracellular delivery. *Nature* **538**, 183–192 (2016).
209. Mercer, W. E., Nelson, D., DeLeo, A. B., Old, L. J. & Baserga, R. Microinjection of monoclonal antibody to protein p53 inhibits serum-induced DNA synthesis in 3T3 cells. *Proc. Natl. Acad. Sci.* **79**, 6309–6312 (1982).
210. Mulcahy, L. S., Smith, M. R. & Stacey, D. W. Requirement for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* **313**, 241–243 (1985).
211. Riabowol, K., Mizzen, L. & Welch, W. Heat shock is lethal to fibroblasts microinjected with antibodies against hsp70. *Science* **242**, 433–436 (1988).
212. Kovary, K. & Bravo, R. The jun and fos protein families are both required for cell cycle

progression in fibroblasts. *Mol. Cell. Biol.* **11**, 4466–72 (1991).

213. Baldin, V., Lukas, J., Marcote, M. J., Pagano, M. & Draetta, G. Cyclin D1 is a nuclear protein required for cell cycle progression in G1. *Genes Dev.* **7**, 812–821 (1993).
214. Ohtsubo, M., Theodoras, A. M., Schumacher, J., Roberts, J. M. & Pagano, M. Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. *Mol. Cell. Biol.* **15**, 2612–2624 (1995).
215. Roche, S., Koegl, M. & Courtneidge, S. A. The phosphatidylinositol 3-kinase alpha is required for DNA synthesis induced by some, but not all, growth factors. *Proc. Natl. Acad. Sci.* **91**, 9185–9189 (1994).
216. Yeung, K. *et al.* Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature* **401**, 173–177 (1999).
217. McEwan, W. A. *et al.* Intracellular antibody-bound pathogens stimulate immune signaling via the Fc receptor TRIM21. *Nat. Immunol.* **14**, 327–336 (2013).
218. Fletcher, A. J., Mallery, D. L., Watkinson, R. E., Dickson, C. F. & James, L. C. Sequential ubiquitination and deubiquitination enzymes synchronize the dual sensor and effector functions of TRIM21. *Proc. Natl. Acad. Sci.* **112**, 10014–10019 (2015).
219. Rhodes, D. A. & Isenberg, D. A. TRIM21 and the Function of Antibodies inside Cells. *Trends Immunol.* **38**, 916–926 (2017).
220. Zavortink, M., Thacher, T. & Rechsteiner, M. Degradation of proteins microinjected into cultured mammalian cells. *J. Cell. Physiol.* **100**, 175–185 (1979).
221. Yamaizumi, M., Uchida, T., Mekada, E. & Okada, Y. Antibodies introduced into living cells by red cell ghosts are functionally stable in the cytoplasm of the cells. *Cell* **18**, 1009–1014 (1979).
222. Marasco, W. A., Haseltine, W. A. & Chen, S. Y. Design, intracellular expression, and activity of a human anti-human immunodeficiency virus type 1 gp120 single-chain antibody. *Proc. Natl. Acad. Sci.* **90**, 7889–7893 (1993).
223. Rondon, I. J. & Marasco, W. A. Intracellular antibodies (intrabodies) for gene therapy of infectious diseases. *Annu. Rev. Microbiol.* **51**, 257–283 (1997).
224. Böldicke, T. Single domain antibodies for the knockdown of cytosolic and nuclear proteins. *Protein Sci.* **26**, 925–945 (2017).
225. Marschall, A. L. J., Dübel, S. & Böldicke, T. Specific in vivo knockdown of protein function by intrabodies. *MAbs* **7**, 1010–1035 (2015).
226. Biocca, S., Ruberti, F., Tafani, M., Pierandrei-Amaldi, P. & Cattaneo, A. Redox state of single chain Fv fragments targeted to the endoplasmic reticulum, cytosol and mitochondria. *Bio/technology* **13**, 1110–5 (1995).

227. Tanaka, T., Lobato, M. N. & Rabbitts, T. H. Single domain intracellular antibodies: A minimal fragment for direct in vivo selection of antigen-specific intrabodies. *J. Mol. Biol.* **331**, 1109–1120 (2003).
228. Visintin, M. *et al.* The intracellular antibody capture technology (IACT): towards a consensus sequence for intracellular antibodies. *J. Mol. Biol.* **317**, 73–83 (2002).
229. Proba, K., Wörn, A., Honegger, A. & Plückthun, A. Antibody scFv fragments without disulfide bonds, made by molecular evolution 1 1Edited by I. A. Wilson. *J. Mol. Biol.* **275**, 245–253 (1998).
230. Colby, D. W. *et al.* Potent inhibition of huntingtin aggregation and cytotoxicity by a disulfide bond-free single-domain intracellular antibody. *Proc. Natl. Acad. Sci.* **101**, 17616–17621 (2004).
231. Quevedo, C. E. *et al.* Small molecule inhibitors of RAS-effector protein interactions derived using an intracellular antibody fragment. *Nat. Commun.* **9**, (2018).
232. Spencer-Smith, R. *et al.* Inhibition of RAS function through targeting an allosteric regulatory site. *Nat. Chem. Biol.* **13**, 62–68 (2017).
233. Gupta, A. *et al.* Facile target validation in an animal model with intracellularly expressed monobodies. *Nat. Chem. Biol.* **14**, 895–900 (2018).
234. Deprey, K., Becker, L., Kritzer, J. & Plückthun, A. Trapped! A Critical Evaluation of Methods for Measuring Total Cellular Uptake versus Cytosolic Localization. *Bioconjug. Chem.* **30**, 1006–1027 (2019).
235. Mueller, J., Kretzschmar, I., Volkmer, R. & Boisguerin, P. Comparison of Cellular Uptake Using 22 CPPs in 4 Different Cell Lines. *Bioconjug. Chem.* **19**, 2363–2374 (2008).
236. Illien, F. *et al.* Quantitative fluorescence spectroscopy and flow cytometry analyses of cell-penetrating peptides internalization pathways: Optimization, pitfalls, comparison with mass spectrometry quantification. *Sci. Rep.* **6**, 1–13 (2016).
237. Richard, J. P. *et al.* Cell-penetrating peptides: A reevaluation of the mechanism of cellular uptake. *J. Biol. Chem.* **278**, 585–590 (2003).
238. Verdurmen, W. P. R., Mazlami, M. & Plückthun, A. A quantitative comparison of cytosolic delivery via different protein uptake systems. *Sci. Rep.* **7**, 1–13 (2017).
239. Du, S., Liew, S. S., Li, L. & Yao, S. Q. Bypassing Endocytosis: Direct Cytosolic Delivery of Proteins. *J. Am. Chem. Soc.* **140**, 15986–15996 (2018).
240. Postupalenko, V. *et al.* Intracellular delivery of functionally active proteins using self-assembling pyridylthiourea-polyethylenimine. *J. Control. Release* **178**, 86–94 (2014).
241. Ressler, V. T., Mix, K. A. & Raines, R. T. Esterification Delivers a Functional Enzyme into a Human Cell. *ACS Chem. Biol.* **14**, 599–602 (2019).

242. Kauffman, W. B., Guha, S. & Wimley, W. C. Synthetic molecular evolution of hybrid cell penetrating peptides. *Nat. Commun.* **9**, (2018).
243. Wadia, J. S., Stan, R. V. & Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **10**, 310–315 (2004).
244. Zuris, J. A. *et al.* Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73–80 (2015).
245. Yu, P., Liu, B. & Kodadek, T. A high-throughput assay for assessing the cell permeability of combinatorial libraries. *Nat. Biotechnol.* **23**, 746–751 (2005).
246. Peraro, L. *et al.* Cell Penetration Profiling Using the Chloroalkane Penetration Assay. *J. Am. Chem. Soc.* **140**, 11360–11369 (2018).
247. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).
248. Milech, N. *et al.* GFP-complementation assay to detect functional CPP and protein delivery into living cells. *Sci. Rep.* **5**, 1–11 (2015).
249. Schmidt, S. *et al.* Detecting Cytosolic Peptide Delivery with the GFP Complementation Assay in the Low Micromolar Range. *Angew. Chem. Int. Ed.* **54**, 15105–15108 (2015).
250. Lönn, P. *et al.* Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **6**, 1–9 (2016).
251. Derossi, D., Joliot, A. H., Chassaing, G. & Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* **269**, 10444–50 (1994).
252. Vivès, E., Brodin, P. & Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biol. Chem.* **272**, 16010–16017 (1997).
253. Futaki, S. *et al.* Arginine-rich Peptides. *J. Biol. Chem.* **276**, 5836–5840 (2001).
254. Milletti, F. Cell-penetrating peptides: Classes, origin, and current landscape. *Drug Discov. Today* **17**, 850–860 (2012).
255. Bechara, C. & Sagan, S. Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.* **587**, 1693–1702 (2013).
256. Nakase, I. *et al.* Interaction of arginine-rich peptides with membrane-associated proteoglycans is crucial for induction of actin organization and macropinocytosis. *Biochemistry* **46**, 492–501 (2007).
257. Peraro, L. & Kritzer, J. A. Emerging Methods and Design Principles for Cell-Penetrant

Peptides. *Angew. Chem. Int. Ed.* **57**, 11868–11881 (2018).

258. Richard, J. P. *et al.* Cellular uptake of unconjugated TAT peptide involves clathrin-dependent endocytosis and heparan sulfate receptors. *J. Biol. Chem.* **280**, 15300–15306 (2005).
259. Kawaguchi, Y. *et al.* Syndecan-4 Is a Receptor for Clathrin-Mediated Endocytosis of Arginine-Rich Cell-Penetrating Peptides. *Bioconjug. Chem.* **27**, 1119–1130 (2016).
260. Brock, D. J., Kondow-Mcconaghy, H. M., Hager, E. C. & Pellois, J. P. Endosomal Escape and Cytosolic Penetration of Macromolecules Mediated by Synthetic Delivery Agents. *Bioconjug. Chem.* **30**, 293–304 (2019).
261. Erazo-Oliveras, A. *et al.* Protein delivery into live cells by incubation with an endosomolytic agent. *Nat. Methods* **11**, 861–867 (2014).
262. Schmidt, S. *et al.* Identification of Short Hydrophobic Cell-Penetrating Peptides for Cytosolic Peptide Delivery by Rational Design. *Bioconjug. Chem.* **28**, 382–389 (2017).
263. Han, X., Bushweller, J. H., Cafiso, D. S. & Tamm, L. K. Membrane structure and fusion-triggering conformational change of the fusion domain from influenza hemagglutinin. *Nat. Struct. Biol.* **8**, 715–720 (2001).
264. Skehel, J. J., Cross, K., Steinhauer, D. & Wiley, D. C. Influenza fusion peptides. *Biochem. Soc. Trans.* **29**, 623–626 (2001).
265. McNaughton, B. R., Cronican, J. J., Thompson, D. B. & Liu, D. R. Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc. Natl. Acad. Sci.* **106**, 6111–6116 (2009).
266. Michiue, H. *et al.* The NH 2 Terminus of Influenza Virus Hemagglutinin-2 Subunit Peptides Enhances the Antitumor Potency of Polyarginine-mediated p53 Protein Transduction. *J. Biol. Chem.* **280**, 8285–8289 (2005).
267. Lundberg, P., El-Andaloussi, S., Sütlü, T., Johansson, H. & Langel, Ü. Delivery of short interfering RNA using endosomolytic cell-penetrating peptides. *FASEB J.* **21**, 2664–2671 (2007).
268. Li, M. *et al.* Discovery and Characterization of a Peptide That Enhances Endosomal Escape of Delivered Proteins in Vitro and in Vivo. *J. Am. Chem. Soc.* **137**, 14084–14093 (2015).
269. Akishiba, M. *et al.* Cytosolic antibody delivery by lipid-sensitive endosomolytic peptide. *Nat. Chem.* **9**, 751–761 (2017).
270. Jiang, T. *et al.* Tumor imaging by means of proteolytic activation of cell-penetrating peptides. *Proc. Natl. Acad. Sci.* **101**, 17867–17872 (2004).
271. Nguyen, Q. T. *et al.* Surgery with molecular fluorescence imaging using activatable cell-penetrating peptides decreases residual cancer and improves survival. *Proc. Natl. Acad.*

Sci. **107**, 4317–4322 (2010).

272. Olson, E. S. *et al.* Activatable cell penetrating peptides linked to nanoparticles as dual probes for in vivo fluorescence and MR imaging of proteases. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4311–6 (2010).
273. Savariar, E. N. *et al.* Real-time in Vivo molecular detection of primary tumors and metastases with ratiometric activatable cell-penetrating peptides. *Cancer Res.* **73**, 855–864 (2013).
274. Kridel, S. J. *et al.* Substrate Hydrolysis by Matrix Metalloproteinase-9. *J. Biol. Chem.* **276**, 20572–20578 (2001).
275. Turk, B. E., Huang, L. L., Piro, E. T. & Cantley, L. C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol.* **19**, 661–667 (2001).
276. Overall, C. M. & Kleifeld, O. Validating matrix metalloproteinases as drug targets and anti-targets for cancer therapy. *Nat. Rev. Cancer* **6**, 227–239 (2006).
277. Falnes, P. & Sandvig, K. Penetration of protein toxins into cells. *Curr. Opin. Cell Biol.* **12**, 407–413 (2000).
278. Verdurmen, W. P. R., Luginbühl, M., Honegger, A. & Plückthun, A. Efficient cell-specific uptake of binding proteins into the cytoplasm through engineered modular transport systems. *J. Control. Release* **200**, 13–22 (2015).
279. Liao, X., Rabideau, A. E. & Pentelute, B. L. Delivery of antibody mimics into mammalian cells via anthrax toxin protective antigen. *ChemBioChem* **15**, 2458–2466 (2014).
280. Young, J. A. T. & Collier, R. J. Anthrax Toxin: Receptor Binding, Internalization, Pore Formation, and Translocation. *Annu. Rev. Biochem.* **76**, 243–265 (2007).
281. Mechaly, A., McCluskey, A. J. & John Collier, R. Changing the receptor specificity of anthrax toxin. *MBio* **3**, 1–6 (2012).
282. Bruce, V. J. & McNaughton, B. R. Inside Job: Methods for Delivering Proteins to the Interior of Mammalian Cells. *Cell Chem. Biol.* **24**, 924–934 (2017).
283. Barth, H., Aktories, K., Popoff, M. R. & Stiles, B. G. Binary bacterial toxins: biochemistry, biology, and applications of common Clostridium and Bacillus proteins. *Microbiol. Mol. Biol. Rev.* **68**, 373–402, table of contents (2004).
284. Smith, B. A. *et al.* Minimally cationic cell-permeable miniature proteins via α-helical arginine display. *J. Am. Chem. Soc.* **130**, 2948–2949 (2008).
285. Appelbaum, J. S. *et al.* Arginine topology controls escape of minimally cationic proteins from early endosomes to the cytoplasm. *Chem. Biol.* **19**, 819–830 (2012).
286. Larochelle, J. R., Cobb, G. B., Steinauer, A., Rhoades, E. & Schepartz, A. Fluorescence correlation spectroscopy reveals highly efficient cytosolic delivery of certain penta-arg

proteins and stapled peptides. *J. Am. Chem. Soc.* **137**, 2536–2541 (2015).

287. Gaj, T., Liu, J., Anderson, K. E., Sirk, S. J. & Barbas, C. F. Protein Delivery Using Cys 2 – His 2 Zinc-Finger Domains. *ACS Chem. Biol.* **9**, 1662–1667 (2014).
288. Lawrence, M. S., Phillips, K. J. & Liu, D. R. Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–10112 (2007).
289. Cronican, J. J. *et al.* A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo. *Chem. Biol.* **18**, 833–838 (2011).
290. Zack, D. J., Stempniak, M., Wong, A. L., Taylor, C. & Weisbart, R. H. Mechanisms of cellular penetration and nuclear localization of an anti-double strand DNA autoantibody. *J. Immunol.* **157**, 2082–8 (1996).
291. Jang, J. Y. *et al.* A nucleic acid-hydrolyzing antibody penetrates into cells via caveolae-mediated endocytosis, localizes in the cytosol and exhibits cytotoxicity. *Cell. Mol. Life Sci.* **66**, 1985–1997 (2009).
292. Hansen, J. E. *et al.* Intranuclear protein transduction through a nucleoside salvage pathway. *J. Biol. Chem.* **282**, 20790–20793 (2007).
293. J.E., H. *et al.* Targeting cancer with a lupus autoantibody. *Sci. Transl. Med.* **4**, 1–9 (2012).
294. Kim, Y. R. *et al.* Heavy and light chain variable single domains of an anti-DNA binding antibody hydrolyze both double- and single-stranded DNAs without sequence specificity. *J. Biol. Chem.* **281**, 15287–15295 (2006).
295. Choi, D. K. *et al.* A general strategy for generating intact, full-length IgG antibodies that penetrate into the cytosol of living cells. *MAbs* **6**, 1402–1414 (2014).
296. Kim, J. S. *et al.* Quantitative assessment of cellular uptake and cytosolic access of antibody in living cells by an enhanced split GFP complementation assay. *Biochem. Biophys. Res. Commun.* **467**, 771–777 (2015).
297. Kim, J.-S. S. *et al.* Endosomal acidic pH-induced conformational changes of a cytosol-penetrating antibody mediate endosomal escape. *J. Control. Release* **235**, 165–175 (2016).
298. Shin, S.-M. M. *et al.* Antibody targeting intracellular oncogenic Ras mutants exerts anti-tumour effects after systemic administration. *Nat. Commun.* **8**, 15090 (2017).
299. Yin, H., Kauffman, K. J. & Anderson, D. G. Delivery technologies for genome editing. *Nat. Rev. Drug Discov.* **16**, 387–399 (2017).
300. Adams, D. *et al.* Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *N. Engl. J. Med.* **379**, 11–21 (2018).
301. Smith, S. A., Selby, L. I., Johnston, A. P. R. & Such, G. K. The Endosomal Escape of Nanoparticles: Toward More Efficient Cellular Delivery. *Bioconjug. Chem.* **30**, 263–272 (2019).

302. Marschall, A. L. J. *et al.* Delivery of antibodies to the cytosol: Debunking the myths. *MAbs* **6**, 943–956 (2014).
303. Xu, X., Costa, A. & Burgess, D. J. Protein encapsulation in unilamellar liposomes: High encapsulation efficiency and a novel technique to assess lipid-protein interaction. *Pharm. Res.* **29**, 1919–1931 (2012).
304. Colletier, J. P., Chaize, B., Winterhalter, M. & Fournier, D. Protein encapsulation in liposomes: Efficiency depends on interactions between protein and phospholipid bilayer. *BMC Biotechnol.* **2**, 1–8 (2002).
305. Kanasty, R., Dorkin, J. R., Vegas, A. & Anderson, D. Delivery materials for siRNA therapeutics. *Nat. Mater.* **12**, 967–977 (2013).
306. Hajj, K. A. & Whitehead, K. A. Tools for translation: Non-viral materials for therapeutic mRNA delivery. *Nat. Rev. Mater.* **2**, 1–17 (2017).
307. Dalkara, D., Zuber, G. & Behr, J. P. Intracytoplasmic delivery of anionic proteins. *Mol. Ther.* **9**, 964–969 (2004).
308. Zhang, Z. *et al.* The fluorination effect of fluoroamphiphiles in cytosolic protein delivery. *Nat. Commun.* **9**, (2018).
309. Courtete, J. *et al.* Suppression of cervical carcinoma cell growth by intracytoplasmic codelivery of anti-oncoprotein E6 antibody and small interfering RNA. *Mol. Cancer Ther.* **6**, 1728–1735 (2007).
310. Lee, Y. *et al.* Efficient delivery of bioactive antibodies into the cytoplasm of living cells by charge-conversional polyion complex micelles. *Angew. Chem. Int. Ed.* **49**, 2552–2555 (2010).
311. Lee, Y. *et al.* Charge-conversional polyionic complex micelles-efficient nanocarriers for protein delivery into cytoplasm. *Angew. Chem. Int. Ed.* **48**, 5309–5312 (2009).
312. Kim, A. *et al.* Intracellular Delivery of Charge-Converted Monoclonal Antibodies by Combinatorial Design of Block/Homo Polyion Complex Micelles. *Biomacromolecules* **17**, 446–453 (2016).
313. Ghosh, P. *et al.* Intracellular Delivery of a Membrane-Impermeable Enzyme in Active Form Using Functionalized Gold Nanoparticles. *J. Am. Chem. Soc.* **132**, 2642–2645 (2010).
314. Mout, R. *et al.* General Strategy for Direct Cytosolic Protein Delivery via Protein-Nanoparticle Co-engineering. *ACS Nano* **11**, 6416–6421 (2017).
315. Eltoukhy, A. A. *et al.* Nucleic acid-mediated intracellular protein delivery by lipid-like nanoparticles. *Biomaterials* **35**, 6454–6461 (2014).
316. Postupalenko, V. *et al.* Protein Delivery System Containing a Nickel-Immobilized Polymer for Multimerization of Affinity-Purified His-Tagged Proteins Enhances Cytosolic Transfer.

Angew. Chem. Int. Ed. **54**, 10583–10586 (2015).

317. Gao, X. *et al.* Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217–221 (2018).
318. Wang, M. *et al.* Efficient delivery of genome-editing proteins using bioreducible lipid nanoparticles. *Proc. Natl. Acad. Sci.* **113**, 2868–2873 (2016).
319. Kim, Y. B., Zhao, K. T., Thompson, D. B. & Liu, D. R. An anionic human protein mediates cationic liposome delivery of genome editing proteins into mammalian cells. *Nat. Commun.* **10**, 2905 (2019).
320. Lee, K. *et al.* Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat. Biomed. Eng.* **1**, 889–901 (2017).
321. Sun, W. *et al.* Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing. *Angew. Chem. Int. Ed.* **54**, 12029–12033 (2015).
322. Rosenblum, D., Joshi, N., Tao, W., Karp, J. M. & Peer, D. Progress and challenges towards targeted delivery of cancer therapeutics. *Nat. Commun.* **9**, (2018).
323. Matsumura, Y. & Maeda, H. A new concept for macromolecular therapeutics in cancer chemotherapy: mechanism of tumoritropic accumulation of proteins and the antitumor agent smancs. *Cancer Res.* **46**, 6387–92 (1986).
324. Peer, D. *et al.* Nanocarriers as an emerging platform for cancer therapy. *Nat. Nanotechnol.* **2**, 751–760 (2007).
325. Wilhelm, S. *et al.* Analysis of nanoparticle delivery to tumours. *Nat. Rev. Mater.* **1**, 16014 (2016).
326. Parker, N. *et al.* Folate receptor expression in carcinomas and normal tissues determined by a quantitative radioligand binding assay. *Anal. Biochem.* **338**, 284–293 (2005).
327. Desgrosellier, J. S. & Cheresh, D. A. Integrins in cancer: biological implications and therapeutic opportunities. *Nat. Rev. Cancer* **10**, 9–22 (2010).
328. Cheng, Z., Al Zaki, A., Hui, J. Z., Muzykantov, V. R. & Tsourkas, A. Multifunctional Nanoparticles: Cost Versus Benefit of Adding Targeting and Imaging Capabilities. *Science* **338**, 903–910 (2012).
329. Schmidt, M. M. & Wittrup, K. D. A modeling analysis of the effects of molecular size and binding affinity on tumor targeting. *Mol. Cancer Ther.* **8**, 2861–2871 (2009).
330. Kirpotin, D. B. *et al.* Antibody targeting of long-circulating lipidic nanoparticles does not increase tumor localization but does increase internalization in animal models. *Cancer Res.* **66**, 6732–6740 (2006).
331. Tenzer, S. *et al.* Rapid formation of plasma protein corona critically affects nanoparticle pathophysiology. *Nat. Nanotechnol.* **8**, 772–781 (2013).

332. Salvati, A. *et al.* Transferrin-functionalized nanoparticles lose their targeting capabilities when a biomolecule corona adsorbs on the surface. *Nat. Nanotechnol.* **8**, 137–143 (2013).
333. Bertrand, N. *et al.* Mechanistic understanding of in vivo protein corona formation on polymeric nanoparticles and impact on pharmacokinetics. *Nat. Commun.* **8**, (2017).
334. Akinc, A. *et al.* Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Mol. Ther.* **18**, 1357–1364 (2010).

Chapter 2 – Proximity based sortase-mediated ligation

This chapter is adapted from our previously published work with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

2.1 Abstract

Protein bioconjugation has been a crucial tool for studying biological processes and developing therapeutics. Sortase A (SrtA), a bacterial transpeptidase, has become widely used for its ability to site-specifically label proteins with diverse functional moieties, but a significant limitation is its poor reaction kinetics. In this work, we address this by developing proximity-based sortase-mediated ligation (PBSL), which improves the ligation efficiency to over 95% by linking the target protein to SrtA using the SpyTag–SpyCatcher peptide–protein pair. By expressing the target protein with SpyTag C-terminal to the SrtA recognition motif, it can be covalently captured by an immobilized SpyCatcher–SrtA fusion protein during purification. Following the ligation reaction, SpyTag is cleaved off, rendering PBSL traceless, and only the labeled protein is released, simplifying target protein purification and labeling to a single step.

2.2 Introduction

The ability to modify proteins with diverse chemical groups has been important for both understanding protein biology as well as engineering protein therapeutics[1–3]. Site-specific bioconjugation techniques are especially valuable because they produce samples that are more homogeneously labeled, which is important for many applications such as the production of antibody-drug conjugates[4,5]. Sortase A (SrtA), a bacterial transpeptidase, has become widely used for installing site-specific modifications because it only requires target proteins to have a short LPXTG recognition motif (for S. aureas SrtA)[6,7]. In the presence of Ca^{2+}, SrtA binds to its recognition motif and its catalytic Cys residue cleaves between the Thr and Gly residues to form a thioester acyl-enzyme intermediate, which is resolved most commonly by nucleophilic attack from a peptide with a N-terminal oligo-glycine causing the release of the ligated protein[8,9]. Owing to the

synthetic accessibility of peptides linked to chemical groups (for example, imaging agents and drugs) and SrtA's broad nucleophile scope, SrtA has been used to conjugate proteins to not only peptides but also other proteins[10], lipids[11], sugars[12], nucleic acids[13], polymers[14], and surfaces[15]. Additionally, noncanonical SrtA labeling sites have been identified and can be exploited to label proteins at multiple sites in an orthogonal manner[16].

One major limitation of SrtA, however, is its low binding affinity for its recognition motif (K_m >7.3 mM) and poor reaction kinetics, which result in long reaction times and poor yields[10,14,17,18]. Using a large molar excess of both SrtA and the nucleophilic peptide can improve these parameters[19] but can be cost-prohibitive for custom-synthesized peptides. Directed evolution has been used to engineer faster SrtA variants[17], but these also have reduced ligation efficiency owing to increased rates of the hydrolysis side-reaction[20,21]. We have previously addressed this limitation by developing sortase-tag expressed protein ligation (STEPL), which expresses the target protein in series with both the SrtA recognition motif and SrtA[22]. By linking the target protein to SrtA, its effective concentration in the subsequent intramolecular ligation reaction is significantly increased, accelerating ligation rates by overcoming SrtA's low recognition-motif-binding affinity. However, we have found not only that do some proteins express poorly once linked to SrtA but also that protein–SrtA fusions that do express can in rare instances be catalytically inactive, which limits STEPL's generality.

In this work, we describe a new ligation strategy, proximity-based sortase-mediated ligation (PBSL), in which the target protein is expressed separately from SrtA but still allows for an intramolecular ligation reaction. The target protein is expressed with a C-terminal LPETG SrtA recognition motif linked to SpyTag through a flexible $(GGS)_5$ linker. SpyTag is a short peptide (13 aa) that upon incubation with SpyCatcher, its partner protein, will rapidly form an irreversible isopeptide bond[23,24]. Thus, a SrtA and SpyCatcher fusion protein immobilized on an affinity resin can first capture the target protein before initiating the ligation reaction with the addition of Ca^{2+}

and a peptide with a N-terminal GGG (Fig. 2.1). By separating SrtA and target protein expression, PBSL retains the advantages of STEPL without any of the drawbacks associated with expression. Furthermore, since only labeled protein is released from the resin, PBSL simplifies purification, ligation, and enzyme removal to a single step. Significant efforts have been made just to increase the ease of removing SrtA following ligation[25]. Both SpyCatcher–SrtA and SrtA–SpyCatcher were tested to determine whether fusion-protein orientation affected PBSL effectiveness.

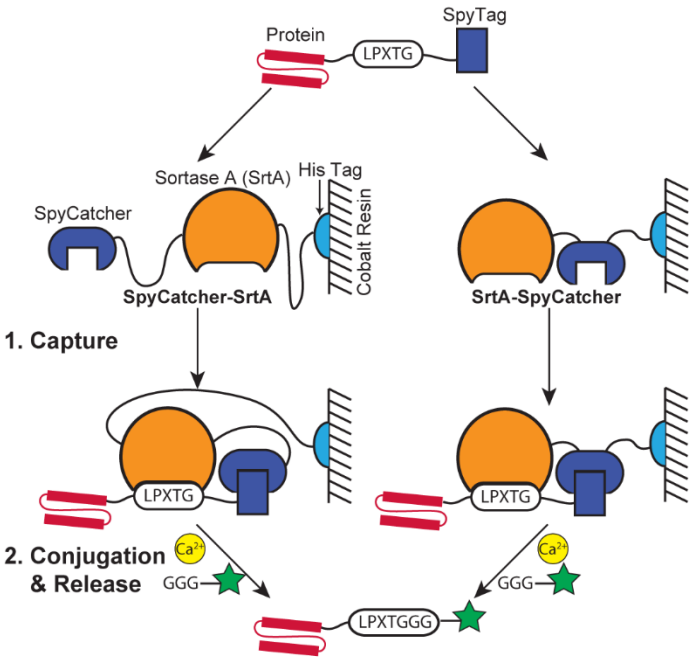

Figure 2.1 PBSL Schematic

The target protein is expressed with a C-terminal LPXTG SrtA recognition motif linked to SpyTag. It is captured during purification by a SpyCatcher-SrtA fusion protein immobilized on cobalt resin. Addition of Ca^{2+} and a peptide with a N-terminal GGG triggers ligation and release of the labeled protein. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

2.3 Results

We used eGFP to characterize PBSL because its concentration can be tracked by measuring absorbance at 488 nm (ε = 56,000$m^{-1}cm^{-1}$)[26]. As confirmed by SDS-PAGE, both SpyCatcher–SrtA and SrtA–SpyCatcher resins captured expressed eGFP–LPETG–SpyTag from clarified lysates (Fig 2.2, Lanes 1 and 2), resulting in eGFP–LPETG–SpyTag covalently linked to the SpyCatcher and SrtA fusion protein on the resin (Fig 2.2, Lanes 3 and 4). Nearly all captured

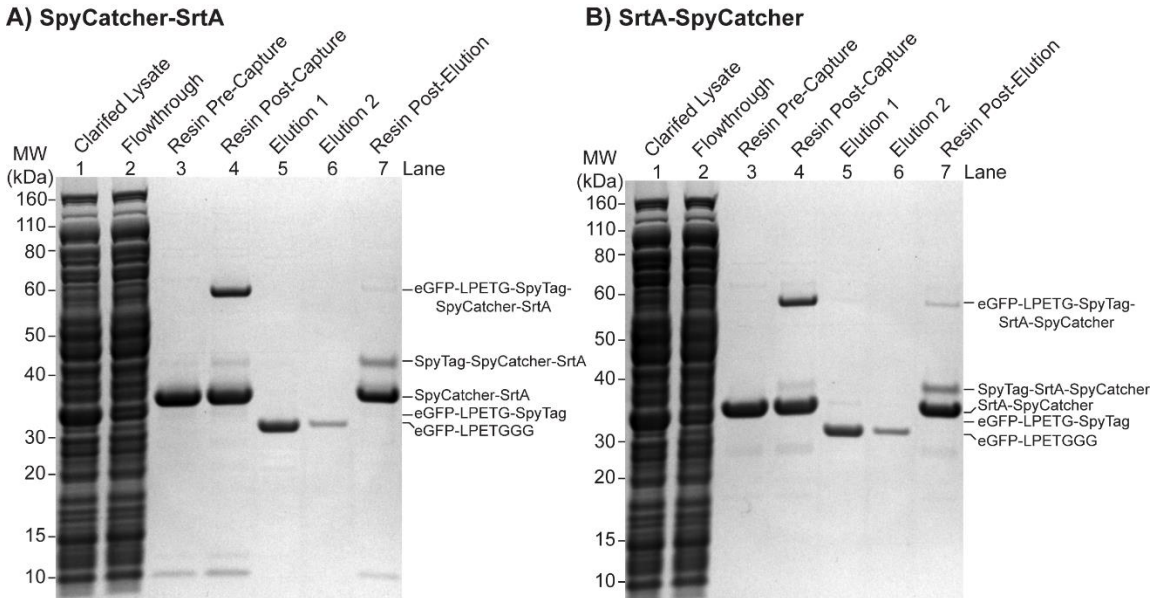

Figure 2.2 SDS-PAGE of PBSL

Following a 30 min incubation, most of the eGFP-LPETG-SpyTag clarified lysates was captured with either **a)** SpyCatcher-SrtA or **b)** SrtA-SpyCatcher resin. Release of almost all captured protein is achieved with a 3h ligation reaction with 2mM GGG in reaction buffer (PBS + 50μM $CaCl_2$, pH 7.4) at 37°C. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

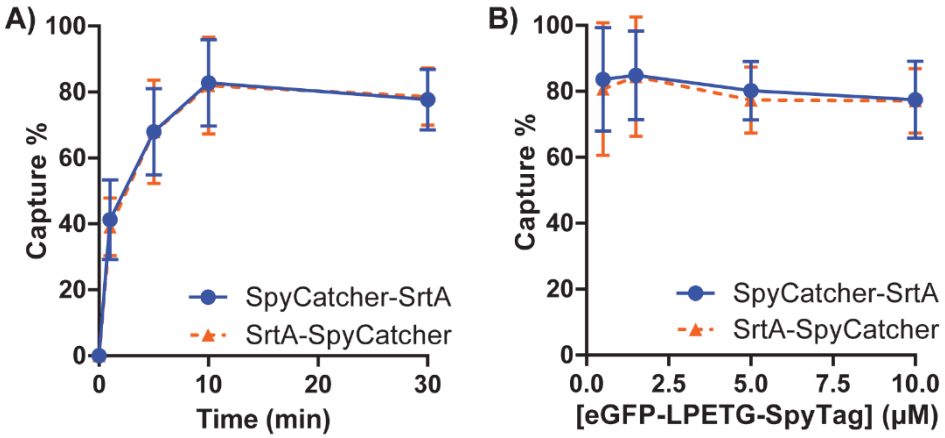

Figure 2.3 PBSL Capture Efficiency.

a) 10μM eGFP-LPETG-SpyTag clarified lysates were incubated with either SpyCatcher-SrtA or SrtA-SpyCatcher resin for varying times. Approximately 80% capture occurred within 10 min for both resins. **b)** Varying concentrations of eGFP-LPETG-SpyTag clarified lysates were incubated with either resins for 30 min. ~80% maximal capture was retained for lysate concentrations as low as 0.5μM. Captured [eGFP-LPETG-SpyTag] were determined by incubating the resin with stripping buffer (PBS + 200mM imidazole, pH 7.4) and measuring eGFP absorbance at 488nm. Data are mean ± SD; n=3. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

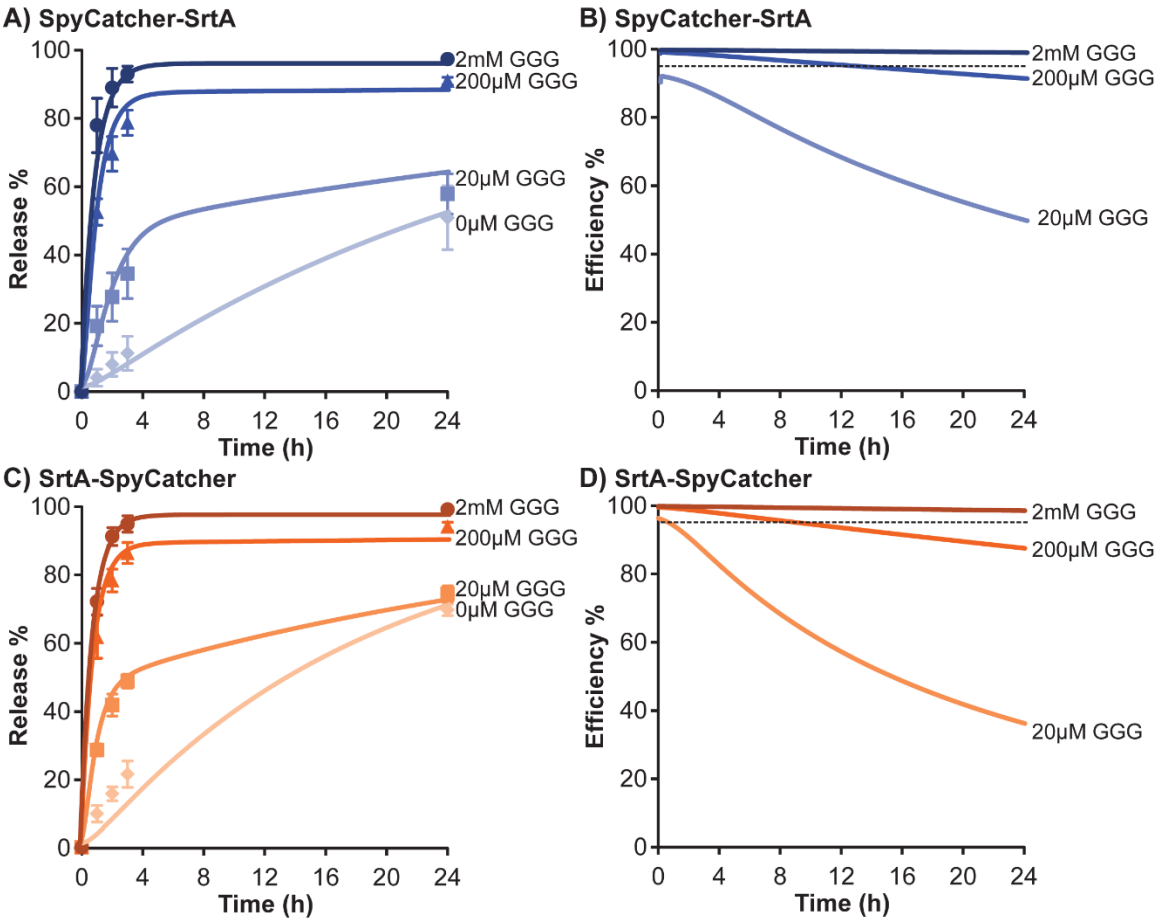

Figure 2.4 PBSL release yields and predicted ligation efficiency.

Yields for eGFP release from **a)** SpyCatcher-SrtA or c) SrtA-SpyCatcher resin was monitored over time following incubation with reaction buffer and varying [GGG] (symbols). $[eGFP]_{released}$ was determined by measuring absorbance at 488nm whereas $[eGFP]_{captured}$ was determined as the sum of $[eGFP]_{released}$ and $[eGFP]_{stripped.}$ Data are mean ± SD; n=3. Release yields were fitted to the model in Fig. 2.5 and superimposed (lines) onto the data. Ligation efficiency for **b)** SpyCatcher-SrtA and d) SrtA-SpyCatcher and varying [GGG] was then calculated. The dashed line indicates 95% ligation efficiency. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

eGFP was released following a 3 h ligation reaction at 37°C with 2 mM triglycine (GGG) (Fig 2.2, Lanes 4–7).

To quantify capture efficiency, we incubated 10 μM eGFP–LPETG–SpyTag clarified lysates with 4 molar equivalents of either SpyCatcher–SrtA or SrtA–SpyCatcher resin for varying times. A capture efficiency of about 80% was achieved with both resins within 10 min, consistent

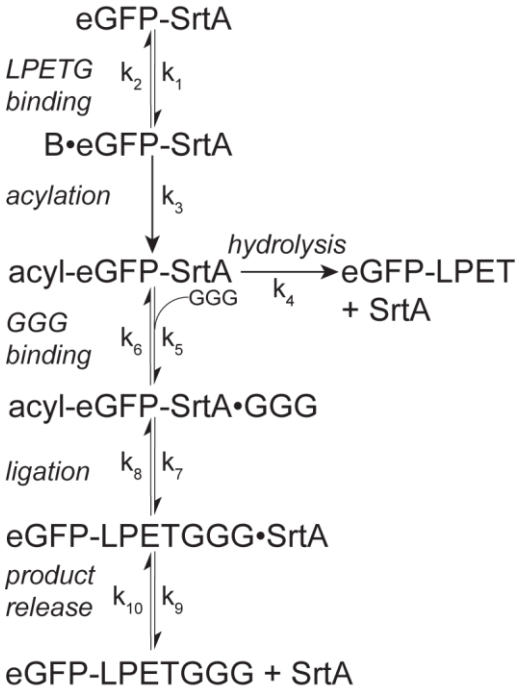

Figure 2.5 Model of eGFP release.

An initial intramolecular acylation reaction is followed by either a reversible, GGG-dependent ligation reaction or an irreversible hydrolysis side-reaction. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

with previously reported SpyTag–SpyCatcher reaction rates (Fig. 2.3a)[23,24]. The capture efficiency was maintained even with 0.5 μM eGFP–LPETG–SpyTag lysates, demonstrating that PBSL is compatible with poorly expressed proteins as well (Fig. 2.3b). SpyTag–SpyCatcher has a binding constant, K_d of 0.2 μM[23,24], suggesting that target proteins expressed at even lower concentrations may still be efficiently captured. Finally, both resins remained stable when stored at 4°C for up to 2 months with no decrease in capture efficiency (Supplementary Information, Figure 2.7a).

Next, we monitored eGFP release by performing the ligation reaction with approximately 5 μM eGFP–LPETG–SpyTag captured on either SpyCatcher–SrtA or SrtA–Spy-Catcher resins with varying concentrations of GGG. As expected, higher GGG concentrations increased release rates for both resins (Fig. 2.4a and c, symbols). Interestingly, the hydrolysis side-reaction, which is the only reaction when 0 μM GGG is used, was slightly faster for SrtA–SpyCatcher. To determine ligation efficiency, the release data was fit to a model in which eGFP–SrtA first undergoes an intramolecular acylation reaction followed by either a reversible ligation reaction or an irreversible hydrolysis side-reaction reaction (Fig. 2.5)[27,28]. This model is similar to the ping-pong bi-bi hydrolytic shunt kinetic mechanism proposed for SrtA alone[18], except that the acylation reaction is intramolecular. As shown in Figure 2.4, the model provides an excellent fit to the observed data (Fig. 2.4a and c, lines).

The predicted ligation efficiency for both resins is excellent with 200 μM and 2 mM GGG but poor for 20 μM GGG. SpyCatcher–SrtA has a slightly higher predicted efficiency owing to its lower base hydrolysis rate (Fig. 2.4b and d). The poor 20 μM GGG ligation efficiency is likely due to a greater relative contribution from the reverse ligation reaction compared to that of the forward reaction. Furthermore, since the hydrolysis reaction is irreversible, ligation efficiency is predicted to slowly decrease over time. Thus, optimal reaction times require a compromise between maximizing product release and product purity. Although both issues could potentially be addressed by incrementally removing any released protein, adopting a flow-based reaction system[29], or using a nucleophile incompatible with the reverse ligation reaction[30], PBSL still results in excellent ligation yields and efficiency. When peptide cost is not a concern, a 3 h reaction at 37°C with 2 mM peptide results in approximately 95% release for SpyCatcher–SrtA and approximately 96% release for SrtA–SpyCatcher with a >99% predicted ligation efficiency for both. When using 200 μM GGG, which is more economical, a 4 h reaction at 37°C is recommended, which offers a balance between high release (ca. 87% for SpyCatcher–SrtA and ca. 89% for SrtA–SpyCatcher) and high predicted ligation efficiency (ca. 98% for SpyCatcher–SrtA and ca. 97% for SrtA–SpyCatcher). The actual ligation efficiency, however, may be higher as no hydrolysis product could be detected by MALDI-TOF when eGFP was labeled with 200 μM GGG-TAMRA for 4 h at 37°C (Supplementary Information, Fig. 2.8). Finally, both resins remained stable with no decrease in eGFP release (Supplementary Information, Fig. 2.7b).

To directly compare labeling yields for PBSL against STEPL and traditional sortase ligation, we attempted to label an anti-CD3 ScFv, which is commonly used for therapeutics and in immunology research, with a fluorescent GGG-TAMRA peptide. For traditional sortase ligation, ScFv–LPETG–His was incubated with increasing molar ratios of SrtA (up to 100:1 SrtA/ScFv) and 200 μM GGG-TAMRA. By using SDS-PAGE, successful ligation was detected by a downward shift in the ScFv band, which also fluoresced under UV (Fig. 2.6, Lanes 3–9). Even with an enormous excess of SrtA, significant amounts of the ScFv remain unlabeled, consistent with

previous reports of SrtA ligation efficiency[10]. The ScFv was incompatible with STEPL and no labeled protein could be purified (Fig. 2.6, Lane 10) owing to poor expression once linked to SrtA. Finally, with PBSL, ScFv–LPETG–SpyTag captured on SpyCatcher–SrtA resin was stripped and labeled in solution with GGG-TAMRA under the same conditions as that of traditional sortase ligation. The reaction was performed in solution to ensure that the same starting amount of unlabeled ScFv was used for both PBSL and traditional SrtA ligation. There are no significant differences in reactivity between PBSL in solution and on resin (Supplementary Information, Fig. 2.9). As expected, the ligation reaction resulted in ScFv release from SpyCatcher–SrtA and the formation of SpyTag–SpyCatcher–SrtA as well as a fluorescent ScFv-LPETGGG-TAMRA product (Fig. 2.6, Lanes 1–2). Consistent with the eGFP studies, little to no ScFv remained captured following the ligation (Fig. 2.6, Lane 2) in contrast with the large amounts of unlabeled ScFv remaining with traditional sortase ligation. The labeled ScFv product obtained from PBSL on resin

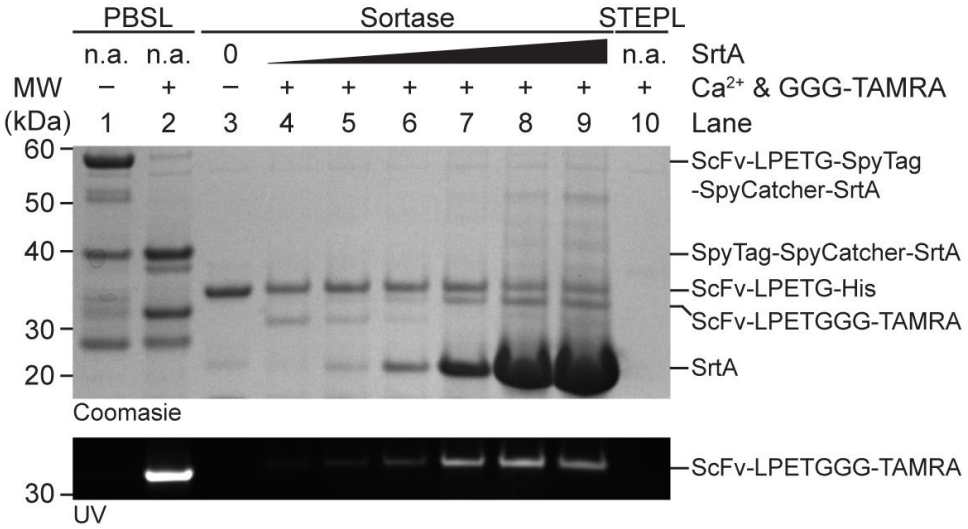

Figure 2.6 PBSL Comparisons.

SDS-PAGE comparing PBSL (Lanes 1-2) with traditional sortase ligation (Lanes 3-9) and STEPL (Lane 10). Following ligation with either PBSL in solution or sortase, a fluorescent band corresponding to the ScFv-LPETGGG-TAMRA product appears. PBSL results in significantly more fluorescently labeled product than either SrtA or STEPL. Traditional sortase ligation was carried out with SrtA:ScFv molar ratios ranging from 1:1 to 100:1. The ScFv-LPETGGG-TAMRA product for traditional sortase ligation runs slightly higher than that of PBSL due to a longer linker between the ScFv and the LPETG SrtA recognition motif. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

was of high purity and without any hydrolysis side-products as detected by MALDI-TOF (Supplementary Information, Fig. 2.10). Quantification of TAMRA fluorescence shows that PBSL labeled about 2.5-fold more ScFv than traditional sortase ligation (Fig. 2.6, UV image; Supplementary Information, Fig. 2.11), demonstrating the superiority of PBSL. Finally, kinetic comparisons between PBSL and traditional sortase ligation show that PBSL results in shorter reaction times by increasing the effective concentration of the target protein by more than 100-fold (Supplementary Information, Fig. 2.12).

2.4 Discussion

In conclusion, we have demonstrated that PBSL is a new SrtA ligation strategy with shorter reaction times and greater efficiency than that of WT SrtA alone. This method only requires an additional 28 amino acids (2.5 kDa) to be expressed C-terminal to the SrtA recognition motif, which should minimally perturb target protein expression and folding. Furthermore, PBSL is traceless since both SpyTag and the linker remain with SpyCatcher following protein release. Although the resin cannot be regenerated, SpyCatcher and SrtA fusion proteins can be easily purified in large quantities from E. coli (ca. 80–100 mg L^{-1} culture), allowing for milligram-scale purification and labeling of target proteins (ca. 20–25 mg labeled protein per liter of SpyCatcher and SrtA culture). By separating target protein and SrtA expression, PBSL is compatible with target proteins that require expression environments containing high concentrations of Ca^{2+} such as yeast or mammalian system. Finally, we anticipate that using SpyTag and SpyCatcher to link substrate proteins to labeling enzymes could be applied to other enzymatic bioconjugation strategies.

2.5 Supplementary Information

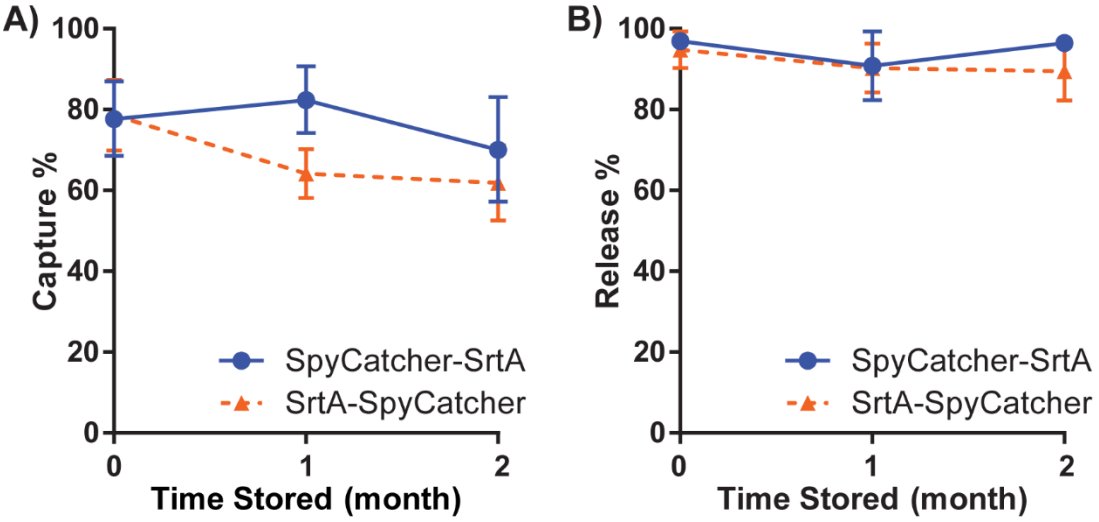

Figure 2.7 PBSL resin stability.

SpyCatcher-SrtA and SrtA-SpyCatcher resin stability at 4°C. **a)** Capture % of stored resins was determined by incubating 10µM eGFP-LPETG-SpyTag clarified lysates with 4 molar equivalents of resin for 30min at room temperature. Captured protein was then stripped from the resin with PBS + 200mM imidazole, pH 7.4 and eGFP concentration determined by measuring its absorbance. **b)** Release % of stored resins was determined by monitoring eGFP release from either resins following a 3h 37°C incubation with PBS + 50µM $CaCl_2$ + 2mM GGG, pH 7.4. Total captured eGFP-LPETG-SpyTag was determined by summing the released eGFP concentration with the eGFP concentration remaining on the resin. eGFP concentrations were determined by measuring eGFP absorbance either directly from elutions ($[eGFP]_{released}$) or following stripping of the resin with PBS + 200mM imidazole, pH 7.4 ($[eGFP]_{stripped}$). No significant decreases in either capture or release % were observed out to 2 months. Data are mean ± SD; n=3. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

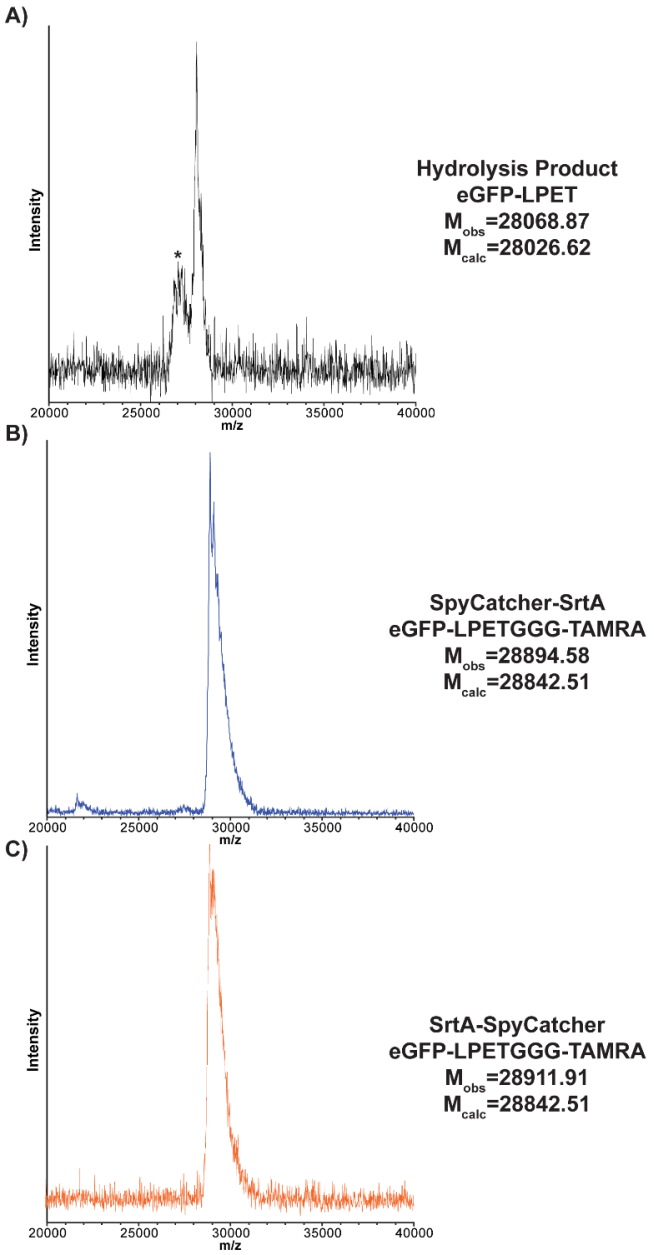

Figure 2.8 PBSL eGFP MALDI-TOF.

MALDI-TOF of eGFP-LPET hydrolysis product or eGFP-LPETGGG-TAMRA products. **a)** To obtain the eGFP-LPET hydrolysis product, eGFP-LPETG-SpyTag captured on SpyCatcher-SrtA resin was incubated in PBS with 50μM $CaCl_2$ at 37°C for 48h. The secondary peak (*) is a contaminant protein irrelevant to the hydrolysis product. **b-c)** eGFP-LPETG-SpyTag captured on either SpyCatcher-SrtA or SrtA-SpyCatcher was labeled with 200μM GGG-TAMRA in PBS with 50μM $CaCl_2$ at 37°C for 4h. No eGFP-LPET peak was detected in either labeled products, suggesting that minimal hydrolysis occurs under the labeling conditions. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

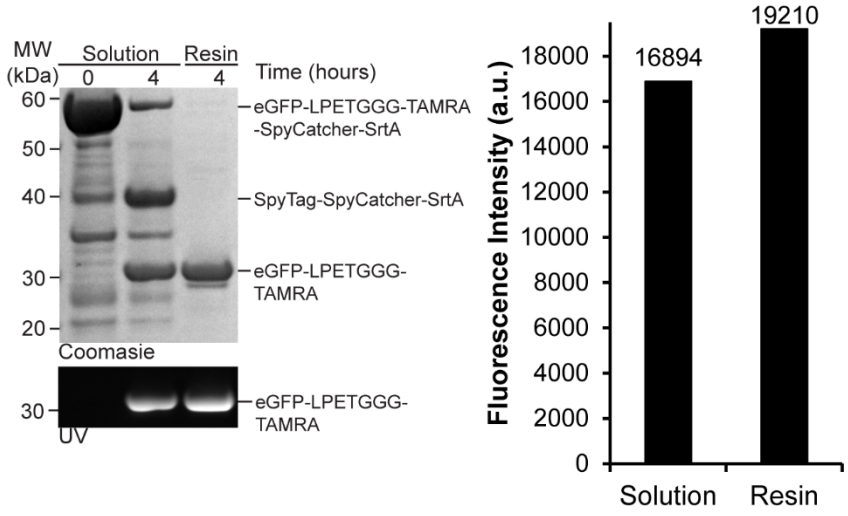

Figure 2.9 PBSL labeling in solution and on resin.

Comparison of PBSL labeling in solution versus on resin. An excess of eGFP-LPETG-SpyTag clarified lysates were bound to SpyCatcher-SrtA resin. The resin was then divided in two halves. One half was kept on the resin while the eGFP-LPETG-SpyTag-SpyCatcher-SrtA protein was stripped off from the other half. Both were then labeled in PBS + 50µM $CaCl_2$ + 200µM GGG-TAMRA at 37°C for 4h. Quantification of the fluorescently labeled product shows that PBSL in solution and on resin results in similar ligation yields. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

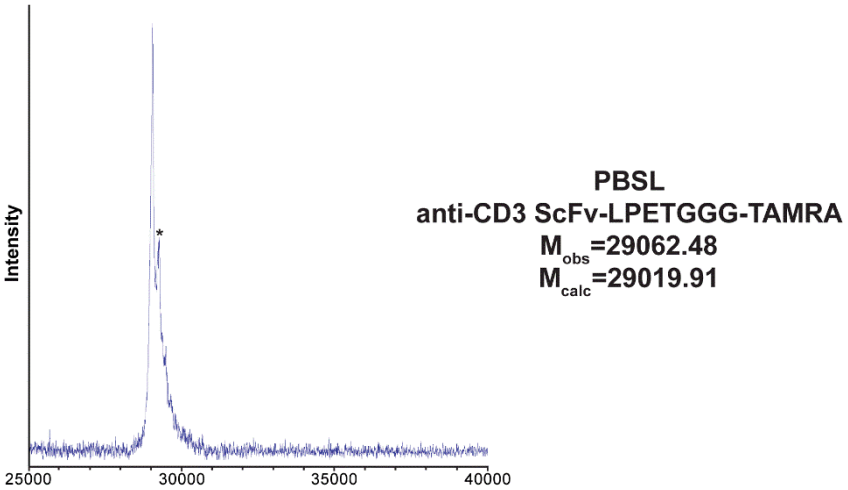

Figure 2.10 PBSL CD3 scFv MALDI-TOF.

MALDI-TOF of PBSL anti-CD3 ScFv-LPETGGG-TAMRA products. No hydrolysis side product was detected (M_{calc}=28606.38). The secondary peak (*) is an adduct from the sinapic acid matrix used in sample preparation. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

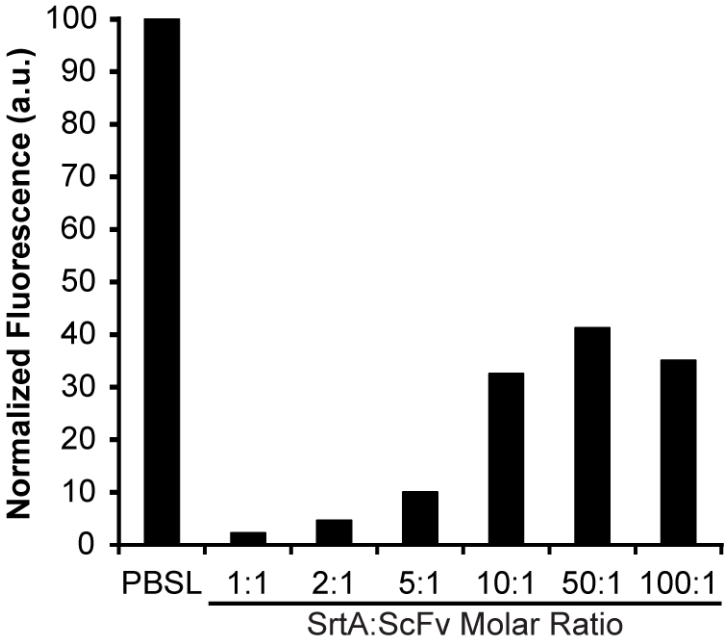

Figure 2.11 PBSL ScFv Quantification

Quantification of ScFv labeling with GGG-TAMRA in Figure 2.6. The fluorescence intensity was normalized to the initial amount of ScFv to be labeled. For PBSL, that was the intensity of the ScFv-SpyTag-SpyCatcher-SrtA (Lane 1 of Fig. 2.6) Coomasie band normalized by its molecular weight. For sortase labeling, that was the intensity of the ScFv-His (Lane 3 of Fig. 2.6) Coomasie band normalized by its molecular weight. Even at extremely high concentrations of SrtA (50-100:1 SrtA:ScFv molar ratio), sortase only results in ~40% of the labeled product when compared to that of PBSL. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

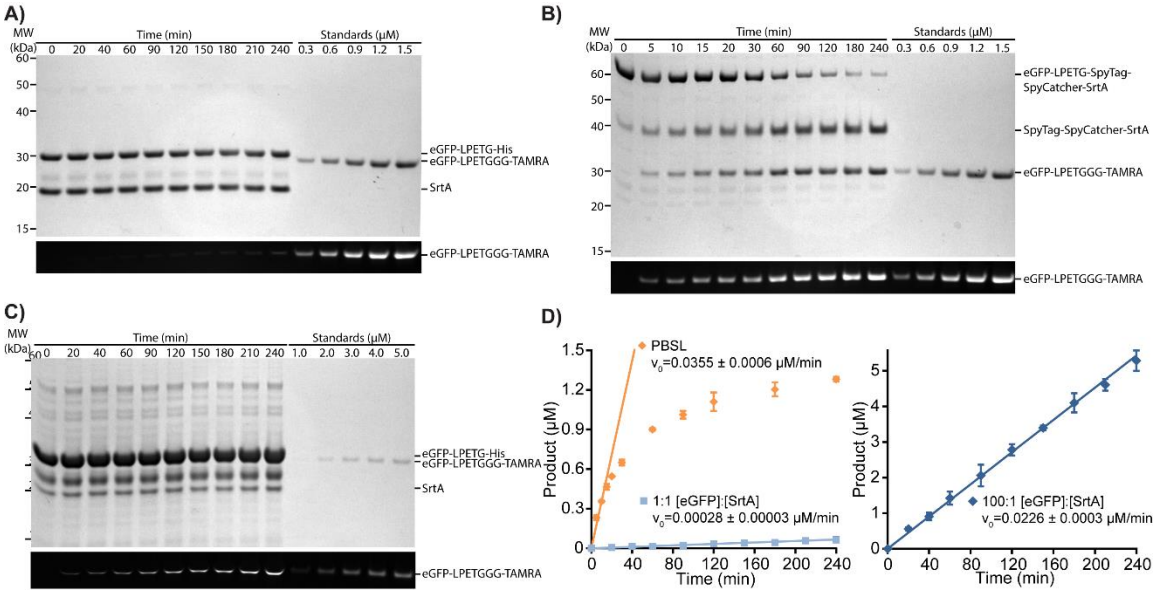

Figure 2.12 PBSL kinetics.

Representative SDS-PAGE gels visualizing the sortase ligation reaction at the indicated time points in PBS + 50μM $CaCl_2$ + 200μM GGG-TAMRA at 37°C with **a)** 1.5μM eGFP-LPETG-His and 1.5μM SrtA (1:1 [eGFP]:[SrtA] ratio), **b)** 1.5μM eGFP-LPETG-SpyTag-SpyCatcher-SrtA in solution, as well as **c)** 150μM eGFP-LPETG-His and 1.5μM SrtA (100:1 [eGFP]:[SrtA] ratio). **d)** Quantification (symbols) shows that within 15min, PBSL results in enough product formation such that the reverse reaction becomes significant and product formation become nonlinear. For 1:1 and 100:1 [eGFP]:[SrtA] reactions, however, product formation remains linear throughout the 4h reaction time. Linear regression for each reaction (lines) shows an excellent fit and reveals that the initial velocity (v_0) for PBSL is 0.0355 ± 0.0006 μM/min, for 1:1 [eGFP]:[SrtA] is 0.00028 ± 0.00003 μM/min, and for 100:1 [eGFP]:[SrtA] is 0.0226 ± 0.0003 μM/min. Since PBSL v_0 is greater than 100:1 [eGFP]:[SrtA] v_0, PBSL increases the effective concentration of the target protein by more than 100-fold. Note that 150μM eGFP-LPETG-His is equivalent to ~4.5mg/ml protein, concentrations that are unrealistic for most protein labeling reactions. Data are mean ± SD; n=3. Reprinted with permission from Wang, H.H., et al. Angew. Chem. Int. Ed. 56, 5349-5352 (2017).

Table 2.1 PBSL amino acid sequences.

Name	Amino Acid Sequence
SpyCatcher-SrtA	MSGSGDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGGGSGGSSGGGSGGSGGSGGSGGSQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVEVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVKHHHHHH
SrtA-SpyCatcher	MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVEVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVKGGSGGSGGSGGSGGSGSGDSATHIKFSKRDEDGKELAGATMELRDSSGKTISTWISDGQVKDFYLYPGKYTFVETAAPDGYEVATAITFTVNEQGQVTVNGHHHHHH
eGFP-LPETG-SpyTag	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKLEGGSGGSLPETGGGSGGSGGSGGSGGSAHIVMVDAYKPTK
eGFP-LPETG-His	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKLEGGSGGSLPETGGGSGGSGGSGGSGGSHHHHHH
CD3ScFv-LPETG	MDIKLQQSGAELARPGASVKMSCKTSGYTFTRYTMHWVKQRPGQGLEWIGYINPSRGYTNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYYCARYYDDHYCLDYWGQGTTLTVSSVEGGSGGSGGSGGSGGVDDIQLTQSPAIMSASPGEKVTMTCRASSSVSYMNWYQQKSGTSPKRWIYDTSKVASGVPYRFSGSGSGTSYSLTISSMEAEDAATYYCQQWSSNPLTFGAGTKLELKLEGGSGGSGGSGGSGGSGGSGGSLPETGGGSGGSGGSGGSGGSHHHHHH
CD3ScFv-LPETG-SrtA (pSTEPL)	MDIKLQQSGAELARPGASVKMSCKTSGYTFTRYTMHWVKQRPGQGLEWIGYINPSRGYTNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYYCARYYDDHYCLDYWGQGTTLTVSSVEGGSGGSGGSGGSGGVDDIQLTQSPAIMSASPGEKVTMTCRASSSVSYMNWYQQKSGTSPKRWIYDTSKVASGVPYRFSGSGSGTSYSLTISSMEAEDAATYYCQQWSSNPLTFGAGTKLELKLEGGSGGSGGSGGSGGSGGSGGSLPETGGGSGGSGGSGGSGGSQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVEVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVKHHHHHH
CD3ScFv-LPETG-SpyTag	MDIKLQQSGAELARPGASVKMSCKTSGYTFTRYTMHWVKQRPGQGLEWIGYINPSRGYTNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYYCARYYDDHYCLDYWGQGTTLTVSSVEGGSGGSGGSGGSGGVDDIQLTQSPAIMSASPGEKVTMTCRASSSVSYMNWYQQKSGTSPKRWIYDTSKVASGVPYRFSGSGSGTSYSLTISSMEAEDAATYYCQQWSSNPLTFGAGTKLELKLEGGSGGSGGSGGSGGSGGSLPETGGGSGGSGGSGGSGGSAHIVMVDAYKPTK
SrtA	MQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLDDQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDVEVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEVKHHHHHH

2.6 Methods

2.6.1 Cloning

For SpyCatcher-SrtA and SrtA-SpyCatcher, their respective DNA sequences were synthesized with 15bp homologous to either the NdeI & NcoI-HF restriction sites (SrtA-SpyCatcher) or the NdeI & HindIII-HF restriction sites (SpyCatcher-SrtA) of the pRSET-A expression vector (Invitrogen).

For eGFP-LPETG-SpyTag and eGFP-LPETG-His, pSTEPL eGFP[22] was used as a starting point. The sequence coding for SrtA was removed from pSTEPL eGFP by double digestion with XhoI & NcoI-HF restriction enzymes (New England Biolabs) and subsequent gel extraction. For eGFP-LPETG-SpyTag, DNA sequences for the LPETG Sortase motif, a $(GGS)_5$ linker sequence, and SpyTag were synthesized with 15bp homologous to both restriction sites. For eGFP-LPETG-His. DNA sequences for the LPETG Sortase motif, a $(GGS)_5$ linker sequence, and His tag were synthesized with 15bp homologous to both restriction sites as well.

For anti-CD3 ScFv-LPETG-SrtA, DNA sequences for the ScFv was synthesized with 15bp homologous to the NdeI & XhoI restriction sites of pSTEPL[22]. For ScFv-LPETG and ScFv-LPETG-SpyTag, DNA sequences were synthesized with 15bp homologous to the NdeI & NcoI restriction sites of pRSET-A.

For sortase A (SrtA), its DNA sequence was synthesized with 15bp homologous to the NdeI & NcoI-HF restriction sites of pRSET-A.

All synthesized DNA sequences were from Integrated DNA Technologies and codon optimized for E. coli. expression. Synthesized DNA sequences were inserted into vectors double digested with the corresponding restriction enzymes using In-Fusion HD Cloning (Clontech). Cloning was verified by Sanger sequencing. For full amino acid sequences, see Supplementary Table 2.1.

2.6.2 Protein Expression

pRSET SpyCatcher-SrtA, pRSET SrtA-SpyCatcher, pRSET eGFP-LPETG-SpyTag, and pRSET SrtA were transformed into T7 Express Crystal Competent E. coli. (New England Biolabs). ScFv vectors were transformed into Origami B(DE3) Competent Cells (EMD Millipore).

Starter cultures were then grown in 3 mL LB + 100µg/mL ampicillin at 37°C in an open air shaker overnight. Starter cultures were added at a 1:1000 dilution to Autoinduction Media LB Broth Base Including Trace Elements (Formedium) with 0.6% UltraPure glycerol (Invitrogen) and 100µg/mL ampicillin.

pRSET eGFP-LPETG-SpyTag was expressed at 25°C in an open air shaker for 48h. pRSET SrtA-SpyCatcher and pRSET SpyCatcher-SrtA were expressed at 25°C in an open air shaker for 54h. pRSET SrtA was expressed at 37°C in an open air shaker for 48h. All ScFv vectors were expressed at 25°C in an open air shaker for 72h.

Following expression, cultures were pelleted by centrifugation (5,500g for 15 min at 4°C) and stored at -20°C.

2.6.3 SpyCatcher-SrtA and SrtA-SpyCatcher Resin Preparation

For lysis, cell pellets overexpressing either SpyCatcher-SrtA or SrtA-SpyCatcher were resuspended in 8mL (per 2g pellet) B-PER (Thermo Fisher Scientific) with 200 µg/mL lysozyme, 4µg/mL DNAseI, and 1 cOmplete Mini EDTA-free Protease Inhibitor Tablet (Roche). They were then rotated at room temperature for 30 min followed by 1 freeze-thaw cycle (-80°C freeze, 37°C water bath thaw). The lysates were then clarified by centrifugation at 16,000g for 20min at room temperature. Clarified lysates were incubated with cobalt resin (TALON Metal Affinity Resin, Clontech, 1.2mL per 100mL expression culture) for 30 min at room temperature for binding. Following binding, the resin was washed with (12mL per 1.2mL resin) PBS + 10mM imidazole, pH

7.4 three times followed by (12mL per 1.2mL resin) PBS once. The resin was stored as a 50% slurry in PBS at 4°C.

To determine the amount of SpyCatcher-SrtA or SrtA-SpyCatcher bound to the resin, 50µL of the corresponding resin (100µL of a 50% slurry) was stripped of all protein by incubation with 250µL stripping buffer (PBS + 200mM imidazole, pH 7.4) for 30 min at room temperature. Protein concentration was then determined using the Pierce BCA Protein Assay (Thermo Fisher Scientific).

2.6.4 eGFP-LPETG-SpyTag Studies

eGFP-LPETG-SpyTag clarified lysates were prepared in the same manner as SpyCatcher-SrtA & SrtA-SpyCatcher resin preparation. The absorbance spectra of eGFP-LPETG-SpyTag clarified lysates were determined with a Varian Cary 100 UV-Vis Spectrophotometer (Agilent). The baseline was corrected for by using the absorbance spectra of clarified lysates from E. coli. overexpressing a non-fluorescent protein and normalizing at 540nm. The concentration of eGFP-LPETG-SpyTag in the clarified lysates was then determined from the absorbance at 488nm (ε= 56,000 M-1 cm-1)[26]. Clarified lysates containing eGFP-LPETG-SpyTag of a desired concentration were then made by diluting the original eGFP-LPETG-SpyTag clarified lysate with PBS.

To visualize the entire PBSL process, 1mL 10µM eGFP-LPETG-SpyTag clarified lysates were incubated with 4 molar equivalents of SpyCatcher-SrtA or SrtA-SpyCatcher resin for 30 min at room temperature for capture. Following capture, the resin was transferred to a Poly-Prep Chromatography Column (Bio-Rad) and washed with 3 column volumes of PBS. To initiate the sortase ligation reaction and release of eGFP-LPETGGG, 1mL PBS + 50µM CaCl2 + 2mM triglycine (Fluka) was added to the column. Following 3h at 37°C, the reaction buffer was eluted from the column followed by a second elution with 1mL PBS. Samples of protein bound to the resin pre-capture, post-capture, and post-elution were obtained by adding 1mL PBS to the resin

at each step, removing 50µL of the resulting slurry, and stripping the removed resin by incubating it with 200µL stripping buffer for 30 min at room temperature. SDS-PAGE was performed on Bolt 4-12% Bis-Tris Plus Gels (Thermo Fisher Scientific) with a Mini Gel Tank (Thermo Fisher Scientific), stained with SimplyBlue SafeStain (Invitrogen), and imaged using a Gel Logic 100 system (Kodak).

eGFP-LPETG-SpyTag capture time studies were performed as follows: 1mL 10µM eGFP-LPETG-SpyTag clarified lysates were incubated with 4 molar equivalents of SpyCatcher-SrtA or SrtA-SpyCatcher resin for the indicated amounts of time at room temperature. Following capture, the resin was transferred to a Poly-Prep Chromatography Column and washed with 3 column volumes of PBS. Protein on the resin was then stripped off by incubating the resin with 1mL stripping buffer for 20 min at room temperature twice. The concentration of the captured eGFP-LPETG-SpyTag was then determined by measuring the absorbance at 488nm in the stripping buffer. Capture % was defined to be:

$$\text{Capture } \% \ = 100 \cdot \frac{[\text{eGFP-LPETG-SpyTag}]_{\text{strip, total}}}{[\text{eGFP-LPETG-SpyTag}]_{\text{clarified lysate}}}$$

eGFP-LPETG-SpyTag capture concentration studies were performed by incubating 2mL eGFP-LPETG-SpyTag clarified lysates with 4 molar equivalents of SpyCatcher-SrtA or SrtA-SpyCatcher resin relative to 10µM eGFP-LPETG-SpyTag for 30 min at room temperature. Subsequent steps are identical to that of capture time studies.

eGFP-LPETG-SpyTag release studies were performed as follows: 1mL 7.5µM eGFP-LPETG-SpyTag clarified lysates were incubated with 4 molar equivalents of SpyCatcher-SrtA or SrtA-SpyCatcher resin for 30 min at room temperature. Following capture, the resin was transferred to a Poly-Prep Chromatography Column and washed with 3 column volumes of PBS. To initiate the sortase ligation reaction and eGFP release, 1mL PBS + 50µM CaCl2 + the indicated concentration of triglycine was added to the column. Following 1h, 2h, 3h, or 24h at

37°C, the reaction buffer was eluted from the column followed by a second elution with 1mL PBS. Protein remaining on the resin was obtained by incubating with 1mL stripping buffer for 20 min at room temperature twice. The concentration of both the released and stripped eGFP-LPETG-SpyTag was then determined by measuring the absorbance at 488nm in the stripping buffer. Release % was defined to be:

$$\text{Release }\%=100\cdot\frac{[\text{eGFP}]_{\text{release, total}}}{[\text{eGFP}]_{\text{release, total}}+[\text{eGFP}]_{\text{strip, total}}}$$

SpyCatcher-SrtA and SrtA-SpyCatcher resin storage studies were performed as follows: 1mL 10µM eGFP-LPETG-SpyTag clarified lysates were incubated with 4 molar equivalents of SpyCatcher-SrtA or SrtA-SpyCatcher resin for 30 min at room temperature. Following capture, the resin was transferred to a Poly-Prep Chromatography Column and washed with 3 column volumes of PBS. To initiate the sortase ligation reaction and eGFP release, 1mL PBS + 50µM CaCl2 + 2mM triglycine was added to the column. Following 3h at 37°C, the reaction buffer was eluted from the column followed by a second elution with 1mL PBS. Protein remaining on the resin was obtained by incubating with 1mL stripping buffer for 20 min at room temperature twice. The capture and release % were then determined in the same manner as the capture and release studies above.

MALDI-TOF studies were formed as follows: The eGFP-LPET hydrolysis product was obtained by incubating eGFP-LPETG-SpyTag captured on SpyCatcher-SrtA resin with PBS + 50µM CaCl2 at 37°C for 48h. The eGFP-LPETGGG-TAMRA product was obtained by incubating eGFP-LPETG-SpyTag captured on either SpyCatcher-SrtA or SrtA-SpyCatcher resin with PBS + 50µM CaCl2 + 200µM GGGSK(TAMRA)NH2 (Neo Scientific) at 37°C for 4h. 20mg of Sinapic acid (Sigma-Aldrich, Milwaukee, WI) was dissolved in 50:50 H2O:MeCN with 0.1% TFA and used as a matrix solution. 2µL of the sample and 2µL of the above matrix solution in a PCR tube were thoroughly mixed using a pipetter. An aliquot (0.5–2.0 µL) of this mixture was then applied to the

sample target where it is allowed to dry. Bruker OminiFlex MALDI-TOF-TOF Mass Spectrometer (Germany) was used to acquire the data.

2.6.5 SpyCatcher-SrtA and SrtA-SpyCatcher Release Modeling

Total released eGFP concentrations were fit to the sum of equations 5 and 6 in the following system of differential equations:

$$(1) \quad \frac{d[\text{eGFP-SrtA}]}{dt} = -k_1 \text{eGFP-SrtA}(t)$$

$$(2) \quad \frac{d[\text{acyl-eGFP-SrtA}]}{dt} = -k_2 \text{acyl-eGFP-SrtA}(t) - k_3 \text{acyl-eGFP-SrtA}(t)\text{GGG}(t) + k_1 \text{eGFP-SrtA}(t) + k_4 \text{eGFP-LPETGGG}(t)\text{SrtA}(t)$$

$$(3) \quad \frac{d[\text{GGG}]}{dt} = -k_3 \text{acyl-eGFP-SrtA}(t)\text{GGG}(t) + k_4 \text{eGFP-LPETGGG}(t)\text{SrtA}(t)$$

$$(4) \quad \frac{d[\text{SrtA}]}{dt} = k_2 \text{acyl-eGFP-SrtA}(t) + k_3 \text{acyl-eGFP-SrtA}(t)\text{GGG}(t) - k_4 \text{eGFP-LPETGGG}(t)\text{SrtA}(t)$$

$$(5) \quad \frac{d[\text{eGFP-LPET}]}{dt} = k_2 \text{acyl-eGFP-SrtA}(t)$$

$$(6) \quad \frac{d[\text{eGFP-LPETGGG}]}{dt} = k_3 \text{acyl-eGFP-SrtA}(t)\text{GGG}(t) - k_4 \text{eGFP-LPETGGG}(t)\text{SrtA}(t)$$

The fit was done in Mathematica using method NMinimize within NonlinearModelFit. The initial [eGFP-SrtA] was determined by adding the total released [eGFP] to the total stripped [eGFP]. The initial [SrtA] was 30μM – the initial eGFP-SrtA concentration. The initial [GGG] was varied experimentally. All other initial concentrations were set to 0.

Ligation efficiency was calculated following fitting as follows:

$$\text{Efficiency } \%(t) = 100 \cdot \frac{\text{eGFP-LPETGGG}(t)}{\text{eGFP-LPETGGG}(t) + \text{eGFP-LPET}(t)}$$

The model parameters obtained for SpyCatcher-SrtA were as follows: k1 = 1.84·10-2 s-1, k2 = 5.58·10-4 s-1, k3 = 3.42·102 M-1s-1, and k3 = 3.10·102 M-1s-1. The model parameters obtained for SrtA-SpyCatcher were as follows: k1 = 2.1·10-2 s-1, k2 = 9.13·10-4 s-1, k3 = 1.25·103 M-1s-1, and k3 = 9.03·102 M-1s-1.

2.6.6 ScFv Purification and Labeling Studies

Clarified lysates for ScFv-LPETG, ScFv-LPETG-SpyTag, and ScFv-LPETG-SrtA were all prepared in the same manner as that of SpyCatcher-SrtA and SrtA-SpyCatcher resin preparation.

ScFv-LPETG-SrtA clarified lysates were incubated with cobalt resin for 30min at room temperature for binding in Poly-Prep Chromatography Column. Afterwards, the resin was washed with 3 column volumes of PBS. ScFv-LPETG clarified lysates were incubated with cobalt resin for binding and washed in the same manner as that of ScFv-LPETG-SrtA. It was then eluted with 500µl of stripping buffer and buffer exchanged into PBS with a 3kDa MWCO Amicon Ultra centrifugal filter (EMD Millipore). SrtA was purified in the same manner as ScFv-LPETG. ScFv-LPETG-SpyTag clarified lysates were incubated with SpyCatcher-SrtA resin for 30min at room temperature for capture. The resin was then washed with 3 column volumes of PBS.

For labeling of ScFv variants, they were incubated with 500ul PBS + 50µM CaCl2 + 200µM GGGSK(TAMRA)NH2 peptide (NEO Scientific), pH 7.4 for 4-6 hours at 37°C. For ScFv-LPETG labeling reactions, increasing concentrations of SrtA were added with a SrtA:ScFv molar ratio of 1:1, 1:2, 1:5, 1:10, 1:50, and 1:100.

SDS-PAGE was performed in the same manner as eGFP-LPETG-SpyTag studies. ImageJ was used for quantification.

2.7 References

1. Krall, N., Da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113

(2016).

2. Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic labeling of proteins: Techniques and approaches. *Bioconjug. Chem.* **24**, 1277–1294 (2013).
3. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).
4. Chudasama, V., Maruani, A. & Caddick, S. Recent advances in the construction of antibody-drug conjugates. *Nat. Chem.* **8**, 114–119 (2016).
5. Agarwal, P. & Bertozzi, C. R. Site-specific antibody-drug conjugates: The nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjug. Chem.* **26**, 176–192 (2015).
6. Mazmanian, S. K., Liu, G., Ton-That, H. & Schneewind, O. Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**, 760–763 (1999).
7. Kruger, R. G. *et al.* Analysis of the Substrate Specificity of the Staphylococcus aureus Sortase Transpeptidase SrtA. *Biochemistry* **43**, 1541–1551 (2004).
8. Ton-that, H., Liu, G., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif. *Proc. Natl. Acad. Sci.* **96**, 12424–12429 (1999).
9. Ton-That, H., Mazmanian, S. K., Faull, K. F. & Schneewind, O. Anchoring of surface proteins to the cell wall of Staphylococcus aureus. Sortase catalyzed in vitro transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. *J. Biol. Chem.* **275**, 9876–81 (2000).
10. Mao, H., Hart, S. A., Schink, A. & Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **126**, 2670–2671 (2004).
11. Antos, J. M., Miller, G. M., Grotenbreg, G. M. & Ploegh, H. L. Lipid Modification of Proteins through Sortase-Catalyzed Transpeptidation. *J. Am. Chem. Soc.* **130**, 16338–16343 (2008).
12. Samantaray, S., Marathe, U., Dasgupta, S., Nandicoori, V. K. & Roy, R. P. Peptide-sugar ligation catalyzed by transpeptidase sortase: A facile approach to neoglycoconjugate synthesis. *J. Am. Chem. Soc.* **130**, 2132–2133 (2008).
13. Pritz, S. *et al.* Synthesis of Biologically Active Peptide Nucleic Acid-Peptide Conjugates by Sortase-Mediated Ligation Stephan. *J. Org. Chem.* **72**, 3909–3912 (2007).
14. Parthasarathy, R., Subramanian, S. & Boder, E. T. Sortase A as a novel molecular ‘stapler’ for sequence-specific protein conjugation. *Bioconjug. Chem.* **18**, 469–476 (2007).
15. Chan, L. *et al.* Covalent attachment of proteins to solid supports and surfaces via sortase-mediated ligation. *PLoS One* **2**, e1164 (2007).
16. Bellucci, J. J., Bhattacharyya, J. & Chilkoti, A. A noncanonical function of sortase enables

site-specific conjugation of small molecules to lysine residues in proteins. *Angew. Chem. Int. Ed.* **54**, 441–445 (2015).

17. Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci.* **108**, 11399–11404 (2011).
18. Frankel, B. A., Kruger, R. G., Robinson, D. E., Kelleher, N. L. & McCafferty, D. G. Staphylococcus aureus sortase transpeptidase SrtA: Insight into the kinetic mechanism and evidence for a reverse protonation catalytic mechanism. *Biochemistry* **44**, 11188–11200 (2005).
19. Popp, M. W., Antos, J. M., Grotenbreg, G. M., Spooner, E. & Ploegh, H. L. Sortagging: A versatile method for protein labeling. *Nat. Chem. Biol.* **3**, 707–708 (2007).
20. Heck, T., Pham, P. H., Yerlikaya, A., Thöny-Meyer, L. & Richter, M. Sortase A catalyzed reaction pathways: A comparative study with six SrtA variants. *Catal. Sci. Technol.* **4**, 2946–2956 (2014).
21. Heck, T., Pham, P. H., Hammes, F., Thöny-Meyer, L. & Richter, M. Continuous monitoring of enzymatic reactions on surfaces by real-time flow cytometry: Sortase a catalyzed protein immobilization as a case study. *Bioconjug. Chem.* **25**, 1492–1500 (2014).
22. 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. *Anal. Chem.* **85**, 11090–11097 (2013).
23. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E690-7 (2012).
24. Li, L., Fierer, J. O., Rapoport, T. A. & Howarth, M. Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* **426**, 309–317 (2014).
25. Rosen, C. B., Kwant, R. L., MacDonald, J. I., Rao, M. & Francis, M. B. Capture and Recycling of Sortase A through Site-Specific Labeling with Lithocholic Acid. *Angew. Chem. Int. Ed.* **55**, 8585–8589 (2016).
26. Cranfill, P. J. *et al.* Quantitative assessment of fluorescent proteins. *Nat. Methods* **13**, 557–562 (2016).
27. Levary, D. A., Parthasarathy, R., Boder, E. T. & Ackerman, M. E. Protein-Protein Fusion Catalyzed by Sortase A. *PLoS One* **6**, 3–8 (2011).
28. Kornberger, P. & Skerra, A. Sortase-catalyzed in vitro functionalization of a HER2-specific recombinant Fab for tumor targeting of the plant cytotoxin gelonin. *MAbs* **6**, 354–366 (2014).
29. Policarpo, R. L. *et al.* Flow-based enzymatic ligation by sortase A. *Angew. Chem. Int. Ed.* **53**, 9203–9208 (2014).

30. Li, Y. M. *et al.* Irreversible site-specific hydrazinolysis of proteins by use of sortase. *Angew. Chem. Int. Ed.* **53**, 2198–2202 (2014).

Chapter 3 – Cytosolic delivery of proteins with cell-penetrating peptides

3.1 Introduction

Currently, protein therapeutics are limited to extracellular targets because they cannot cross the plasma membrane. In general, proteins can only be delivered into living cells with physical methods such as microinjection or electroporation, but these techniques have limited throughput and cause toxicity[1]. As a result, physical methods for protein delivery are not widely utilized and are only applicable for laboratory research. Because of the foreseen impact that intracellular protein delivery can have on both protein therapeutics and basic cell biology research, there continues to be significant interest towards this goal. Immunoglobulin G (IgG) antibodies are particularly attractive for cytosolic delivery because they can be developed to bind and inhibit almost any protein with high potency and specificity[2].

One popular approach for cytosolic protein delivery involves the labeling of cargo proteins with cell-penetrating peptides (CPPs)[3], via either bioconjugation or the creation of a fusion protein. CPPs are generally short poly-cationic peptides that can efficiently induce endocytic cellular uptake of not only themselves, but also the attached cargo proteins[3–5]. Unfortunately, the endosome-lysosome system is topologically equivalent to the extracellular space and CPP-labeled proteins are extremely inefficient at escaping from endosome to enter the cytosol[4–7]. It is striking that after more than 25 years of research[8,9] and the development of more than 100 purported CPP sequences[10], CPPs have largely been limited to only delivering proteins capable of amplifying their effects such as transcription factors or enzymes[6,11–13]. Tracking large cargos conjugated to CPPs predominantly show endosomal accumulation following uptake, indicating that endosomal entrapment is the bottleneck for cytosolic delivery[6,14–16].

Recently, charge density has been identified as a major contributor to endosomal escape efficiency. Studies using a supercharged GFP with a net charge of +36 have found that it is capable of delivering proteins cargos much more efficiently than prototypical CPPs[17,18]. CPPs also appear to be more effective when linked to cargo proteins that form dimers, tetramers, or multivalent complexes[19]. Finally, dimerized CPPs appear to be capable of inducing leakages within late endosomes[20] and have significantly enhanced endosomal escape[11,21,22]. In addition to density, geometry also appears to be critical – after an optimal length, simply increasing polyarginine CPPs length actually becomes deleterious for delivery[23].

In this study, we tested whether increasing CPP valency could enable the cytosolic delivery of full-length IgG antibodies by leveraging an extremely efficient light-activated site-specific conjugation method for native IgGs (LASIC) that we have previously developed[24]. LASIC utilizes a small photoreactive antibody-binding domain (pAbBD, ~6.5kDa) derived from a thermally stable variant of the β1 domain of *Streptococcal* protein G[25] engineered to be only capable of binding to the IgG Fc region[26]. A photoreactive benzoylphenylalanine (BPA) is incorporated into pAbBD through unnatural amino acid mutagenesis[27], allowing it to be photocrosslinked to the constant region of IgGs with greater than 95% efficiency. CPP or any other functional sequences can be genetically fused to pAbBD and purified recombinantly from *E. coli* before being photocrosslinked to almost any off-the-shelf cargo IgG.

Ideal reporter systems for cytosolic protein delivery must be capable of distinguishing between cargo proteins successfully delivered to the cytosol from cargo proteins still trapped in endosomes[4]. We utilized a self-reassembling splitGFP[28] reporter system in which the larger half – splitGFP(1-10) – is expressed only in the cytosol of reporter cells while the

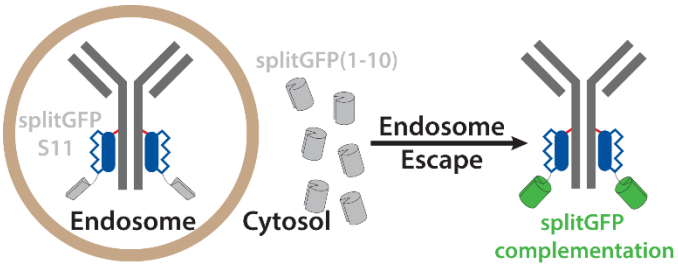

Figure 3.1 Schematic of splitGFP reporter system.

smaller half – splitGFP S11 – is fused to the pAbBD photocrosslinked to IgGs. Thus, only when cargo pAbBD or IgGs enter the cytosol will the splitGFP complement and following chromophore maturation, start fluorescing (Fig. 3.1). The splitGFP reporter system has a low false-positive rate for detecting cytosolic protein delivery[4,29] and has been well validated for screening CPPs[4,30–33].

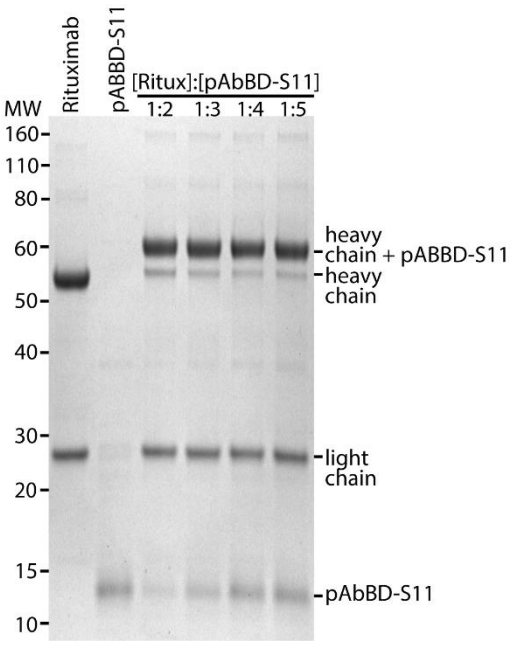

Figure 3.2 SDS-PAGE of pAbBD-S11 photocrosslinking to Rituximab.

Irradiation with 365nm UV light for 2 h at 4°C results in almost complete photocrosslinking of Rituximab heavy chains.

3.2 Results

3.2.1 Detecting cytosolic protein delivery with a splitGFP reporter system

We initially prepared pAbBD-S11 and confirmed that it could be photocrosslinked to Rituximab (Ritux) with >95% efficiency (Fig. 3.2). We use Ritux, which binds to CD20, as a model cargo IgG because it does not bind to any proteins on the surface of our reporter cells and is a clinically used IgG available in large quantities. Next, we confirmed that both pAbBD-S11 and Ritux-(pAbBD-S11)$_2$ conjugates could undergo complementation with purified splitGFP(1-10) *in vitro* (Fig. 3.3). Complete chromophore maturation required approximately 6 hours, but the degree of splitGFP complementation remained linear with respect to the amount of protein present at all times (Fig. 3.3). As expected, the splitGFP complementation fluorescence for Ritux-(pAbBD-S11)$_2$ was twice of that of pAbBD-S11 because two copies of pAbBD-S11 are photocrosslinked onto each molecule of Ritux (Fig. 3.3). We generated splitGFP(1-10) reporter cell lines and confirmed by flow cytometry as well as live cell fluorescence microscopy that both pAbBD-S11 and Ritux-(pAbBD-S11)$_2$ also underwent splitGFP complementation in cells (Fig. 3.4).

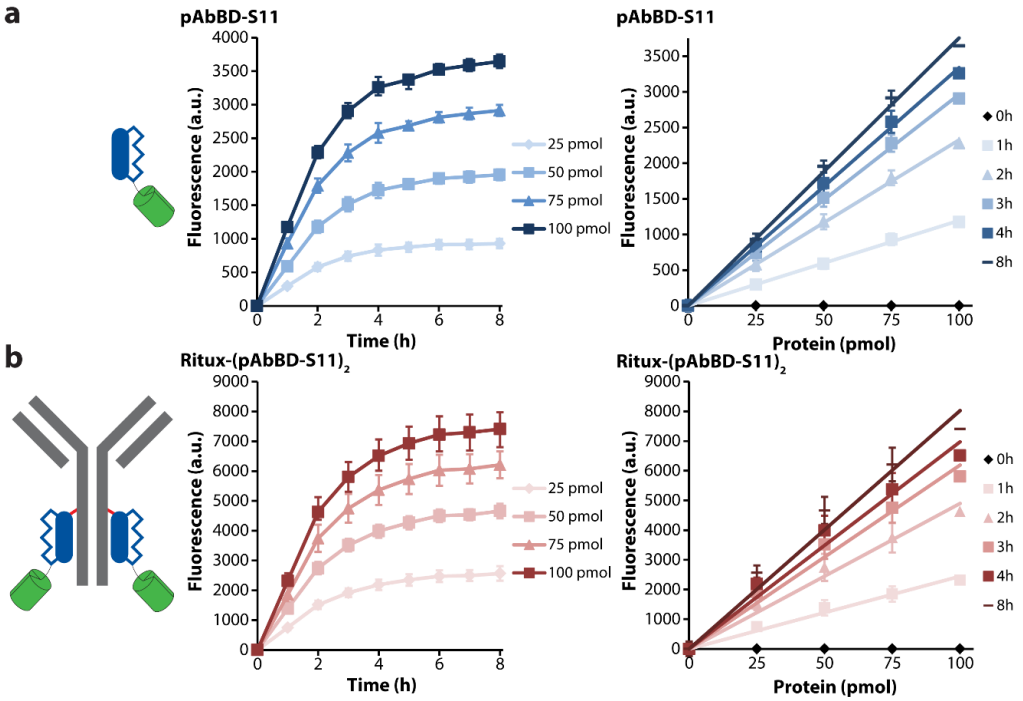

Figure 3.3 pAbBD-S11 and Ritux-(pAbBD-S11)$_2$ can complement splitGFP(1-10) in vitro.

Indicated amounts of **(a)** pAbBD-S11 and **(b)** Ritux-(pAbBD-S11)$_2$ were incubated with 400 pmol of His6-splitGFP(1-10) at 37°C. At the indicated times, splitGFP fluorescence was measured. The left panel shows the complementation kinetics. The right panel shows that the degree of complementation remains linear at all times. Fluorescence for Ritux-(pAbBD-S11)2 was approximately twice of that of pAbBD-S11 because two copies of pAbBD-S11 are photocrosslinked onto each Rituximab molecule. Data are mean + s.e.m., n=3.

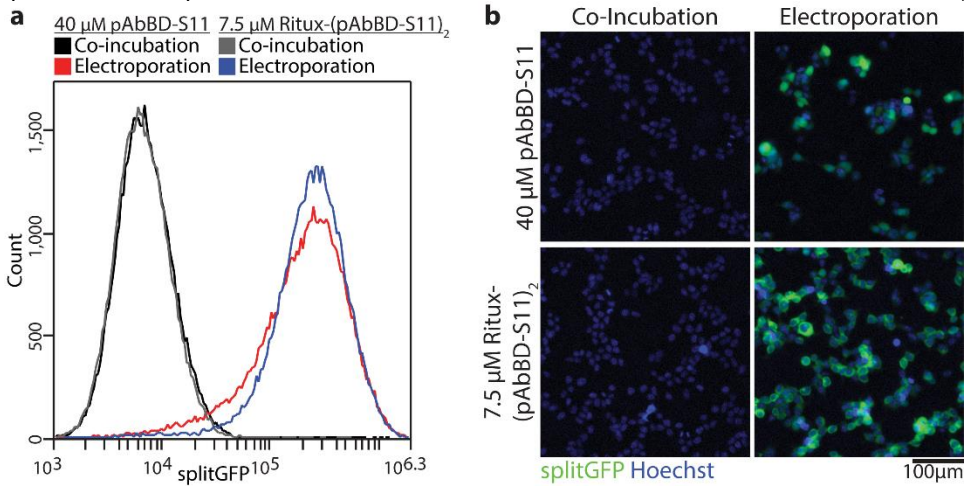

Figure 3.4 pAbBD-S11 and Ritux-(pAbBD-S11)$_2$ can complement splitGFP(1-10) in cells.

pAbBD-S11 or Ritux-(pAbBD-S11)$_2$ were electroporated into HEK293T splitGFP(1-10) cells. After 6 h to allow for splitGFP chromophore maturation, splitGFP complementation and fluorescence was assessed with either flow cytometry **(a)** or live-cell fluorescence microscopy **(b)**.

3.2.2 Cytosolic delivery of pAbBD with CPPs.

Initially, we tested whether the TAT CPP could cytosolically deliver just the pAbBD. Placing TAT at the N-terminus of pAbBD-S11 did not result in any cytosolic delivery in HEK293T splitGFP(1-10) cells, even with TAT-pAbBD-S11 concentration as high as 20µM (Fig. 3.5a). Treating cells with 200µM chloroquine, a small molecule capable of enhancing endosomal escape for cargos[6,34], resulted in a moderate splitGFP signal, indicating that N-terminal TAT was still capable of inducing endocytic uptake of TAT-pAbBD-S11 (Fig. 3.5a). Placing TAT at the C-terminus of pAbBD-S11 did not result in any cytosolic delivery (Fig. 3.5b) until pAbBD-S11-TAT concentrations reached 80µM (Supp. Fig. 3.10a). Interestingly, a greater splitGFP signal was detected for pAbBD-S11-TAT, compared with TAT-pAbBD-S11, when cells were treated with chloroquine (Fig. 3.5b). Similar trends were observed with the R8 CPP (Figure 3.6). Placing R8 at the C-terminus of pAbBD-S11 resulted in some cytosolic delivery at 80µM (Supp. Fig. 3.10b) and resulted in greater delivery with chloroquine treatment when compared to N-terminal placement

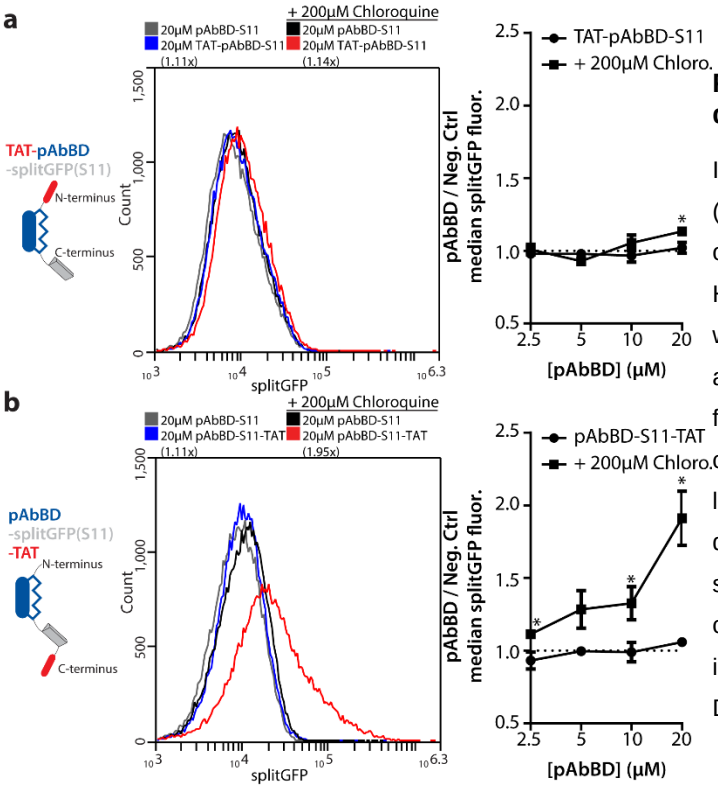

Figure 3.5 Cytosolic pAbBD delivery with the TAT CPP

Indicated concentrations of pAbBD-S11 (negative control), **(a)** TAT-pAbBD-S11, or **(b)** pAbBD-S11-TAT were added to HEK293T splitGFP(1-10) reporter cells with or without chloroquine for 6 h before assessing for splitGFP fluorescence with flow cytometry. Representative flow cytometry histograms are shown in the left panel. Flow cytometry data were quantified as the fold-increase in median splitGFP fluorescence over negative control in the right panel. The dotted line indicates no increase in fluorescence. Data are mean ± s.e.m, *p<0.05.

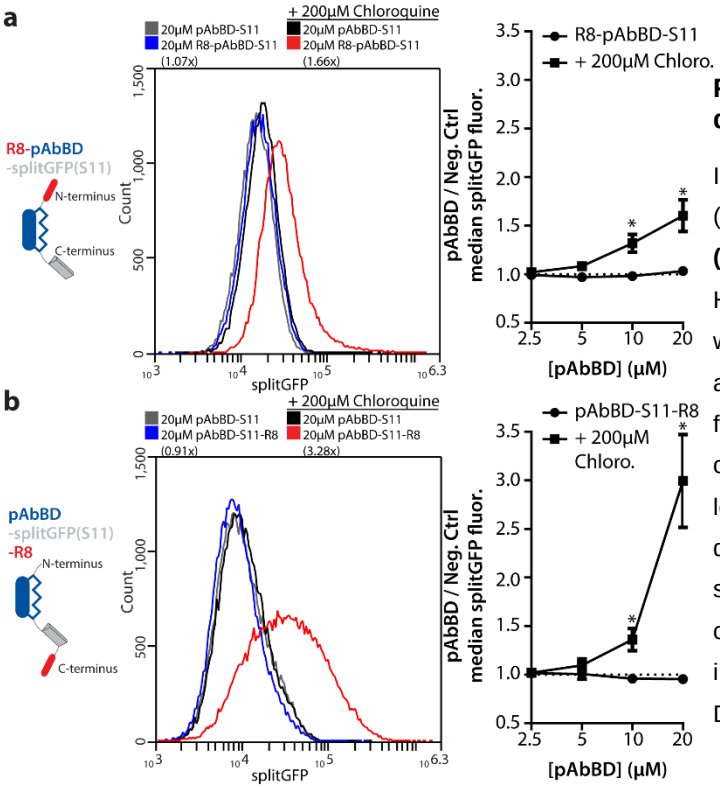

Figure 3.6 Cytosolic pAbBD delivery with the R8 CPP

Indicated concentrations of pAbBD-S11 (negative control), **(a)** R8-pAbBD-S11, or **(b)** pAbBD-S11-R8 were added to HEK293T splitGFP(1-10) reporter cells with or without chloroquine for 6 h before assessing for splitGFP fluorescence with flow cytometry. Representative flow cytometry histograms are shown in the left panel. Flow cytometry data were quantified as the fold-increase in median splitGFP fluorescence over negative control in the right panel. The dotted line indicates no increase in fluorescence. Data are mean ± s.e.m, *p<0.05.

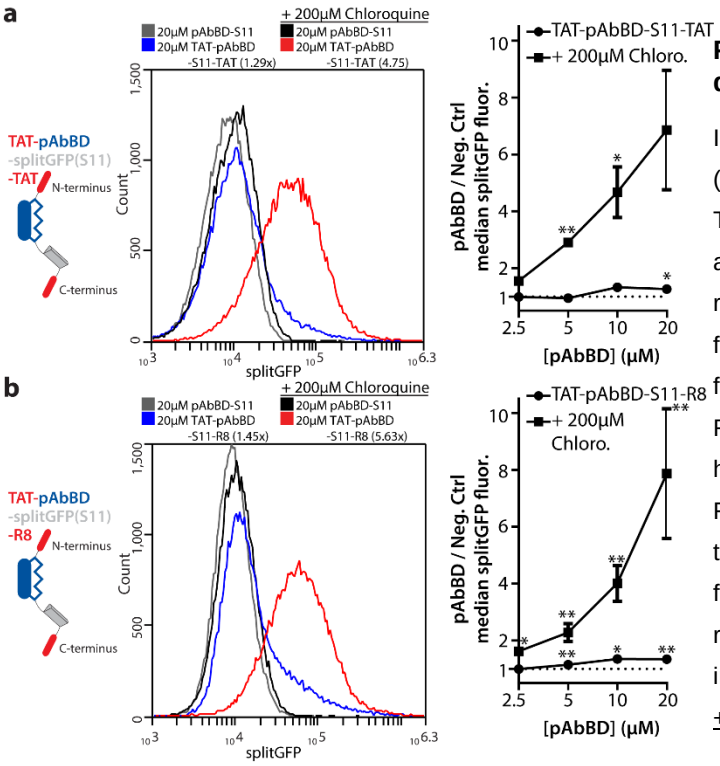

Figure 3.7 Cytosolic pAbBD delivery with dual CPPs.

Indicated concentrations of pAbBD-S11 (negative control), **(a)** TAT-pAbBD-S11-TAT, or **(b)** R8-pAbBD-S11-TAT were added to HEK293T splitGFP(1-10) reporter cells with or without chloroquine for 6 h before assessing for splitGFP fluorescence with flow cytometry. Representative flow cytometry histograms are shown in the left panel. Flow cytometry data were quantified as the fold-increase in median splitGFP fluorescence over negative control in the right panel. The dotted line indicates no increase in fluorescence. Data are mean ± s.e.m, *p<0.05 **p<0.01.

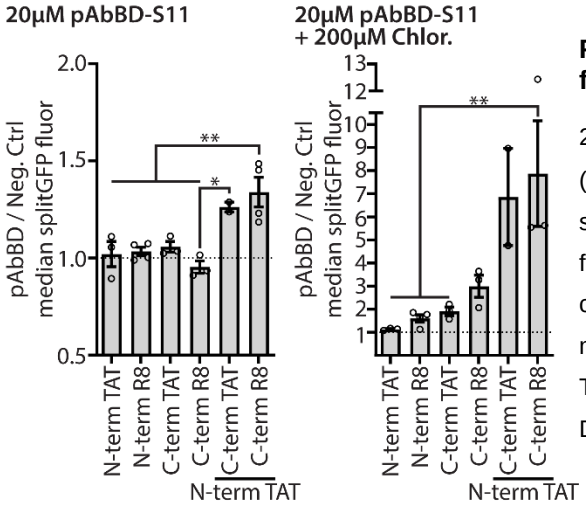

Figure 3.8 Comparisons of different CPP formats for cytosolic pAbBD delivery.

20μM of the indicated protein or pAbBD-S11 (negative control) were added to HEK293T splitGFP(1-10) reporter cells for 6 h before assessing for splitGFP fluorescence with flow cytometry. Flow cytometry data was quantified as the fold-increase in median splitGFP fluorescence over negative control. The dotted line indicates no increase in fluorescence. Data are mean ± s.e.m, *p<0.05 **p<0.01.

(Fig. 3.6). Because chloroquine treatment does not guarantee disruption of all cargo-containing endosomes, it is unclear whether improvements in cytosolic delivery when CPPs are moved from the N- to the C-terminus is due to increased cellular uptake, improved endosomal escape efficiency, or both. Next, we tested whether cytosolic delivery could be improved if a CPP was placed at both the N- and C-terminus of pAbBD. Surprisingly, this approach dramatically improved cytosolic delivery – cytosolic delivery was detectable at only 20μM for TAT-pAbBD-S11-TAT (Fig. 3.7a) and at only 5μM for TAT-pAbBD-S11-R8 (Fig. 3.7b), compared with 80μM for pAbBD-S11-TAT or pAbBD-S11-R8 (Supp. Fig. 3.10). This improvement was also evident in cells treated with chloroquine (Fig. 3.7). Comparisons between all tested CPP formats revealed that placing a CPP at both the N-terminus and C-terminus improved cytosolic delivery to a greater degree than simply adding the effects of an N-terminal CPP with a C-terminal CPP (Fig. 3.8).

3.2.3 CPPs fail to cytosolically deliver antibodies

Having identified that placing CPPs at both termini of pAbBD significantly improved cytosolic delivery, we wondered whether CPPs could also enable cytosolic delivery of antibodies. We were able to efficiently photocrosslink TAT-pAbBD-S11, pAbBD-S11-R8, and TAT-pAbBD-S11-R8 to Ritux and then tested Ritux-(TAT-pAbBD-S11)$_2$, Ritux-(pAbBD-S11-R8)$_2$, and Ritux-(TAT-pAbBD-S11-R8)$_2$ conjugates for delivery in HEK293T splitGFP(1-10) cells. Unfortunately,

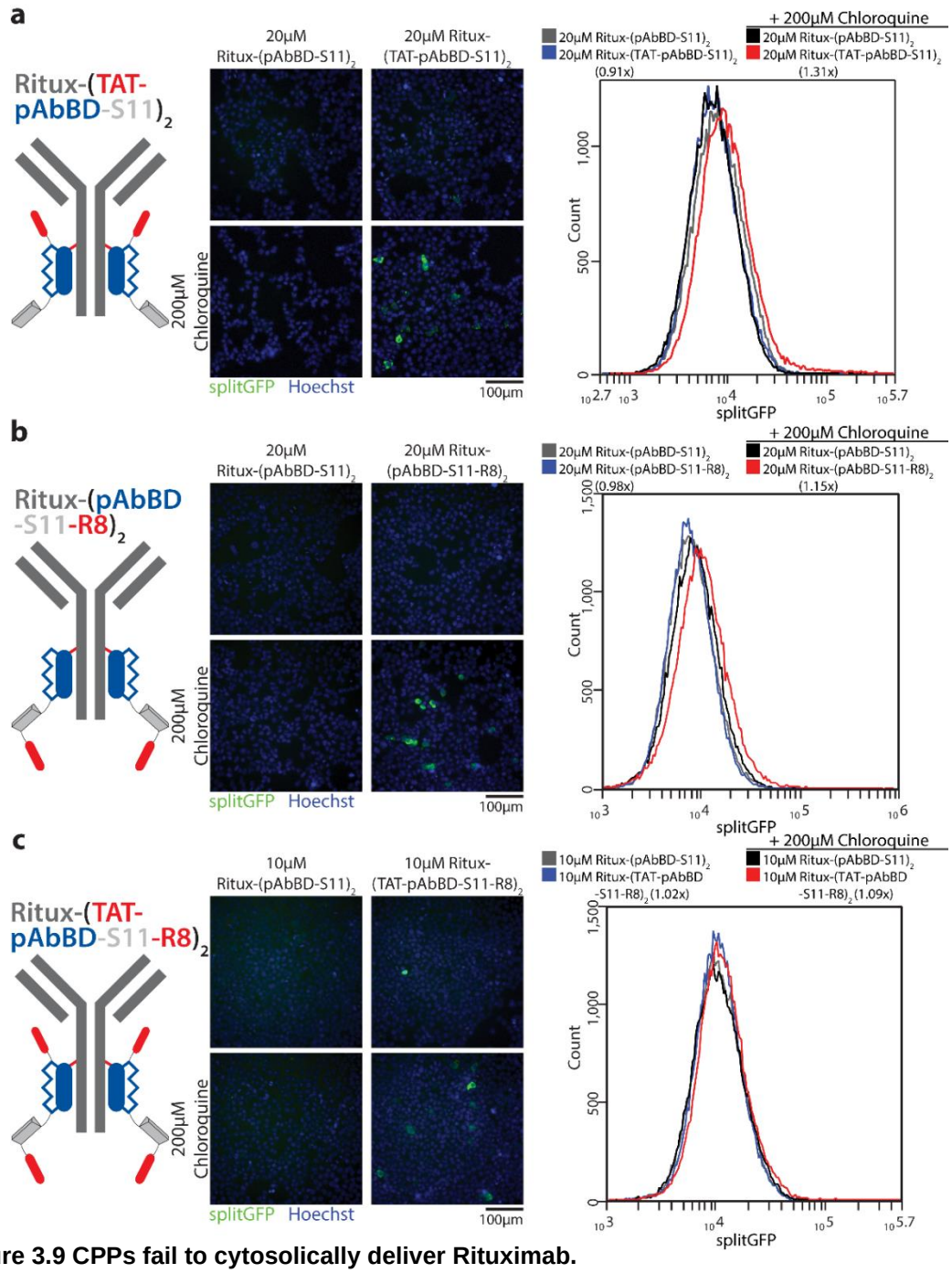

Figure 3.9 CPPs fail to cytosolically deliver Rituximab.

Indicated concentrations of Ritux-(pAbBD-S11)$_2$ (negative control), **(a)** Ritux-(TAT-pAbBD-S11)$_2$, **(b)** Ritux-(pAbBD-S11-R8)$_2$, or **(c)** Ritux-(TAT-pAbBD-S11-R8)$_2$ were added to HEK293T splitGFP(1-10) reporter cells with or without chloroquine for 6 h before assessing for splitGFP fluorescence. Representative live cell fluorescence microscopy images are shown in the left panel. Representative flow cytometry histograms of splitGFP fluorescence are shown in the right panel.

we could not detect any splitGFP fluorescence by live cell fluorescence microscopy or flow cytometry (Fig. 3.9). Following chloroquine treatment, we could detect a few rare cells with robust splitGFP fluorescence via fluorescence microscopy, but not flow cytometry (Fig. 3.9).

3.3 Discussion

CPPs have been extremely popular vehicles for delivering small peptide cargos[3,10] and are the subject of intense research for the delivery of larger macromolecules[4,5,9]. In this study, although we identified parameters that improved CPP-mediated delivery of a small protein, they were ineffective at cytosolically delivering a large IgG cargo.

We found that C-terminal CPPs performed much better than N-terminal CPPs. The mechanistic basis for this trend and whether it will hold for other cargo proteins remains unclear. If CPP placement continues to significantly perturb cytosolic delivery efficiency irrespective of the cargo protein, future studies should focus on determining whether these effects are due to changes in cellular uptake or endosomal escape. Total cellular uptake can be determined with cargos labeled with fluorophores and can then be correlated with cytosolic delivery to qualitatively estimate endosomal escape efficiency. It would also be interesting to determine whether cyclized CPPs[35] with constrained conformations are significantly perturbed by its termini placement. We also found that placing CPPs at both termini of a protein significantly improved cytosolic delivery efficiency. This is consistent with previous reports that increasing CPP valency improves endosomal escape. Again, though, future studies are needed to understand its mechanistic basis.

Similar to previous studies[3,16], we also found that increasing cargo size had a profound deleterious effect on delivery efficiency. Thus, rather than full length IgGs, CPP-mediated delivery should instead initially focus on smaller protein scaffolds such as nanobodies or antibody-like binding proteins such as DARPins, affibodies, or monobodies.

3.4 Supplementary Figure

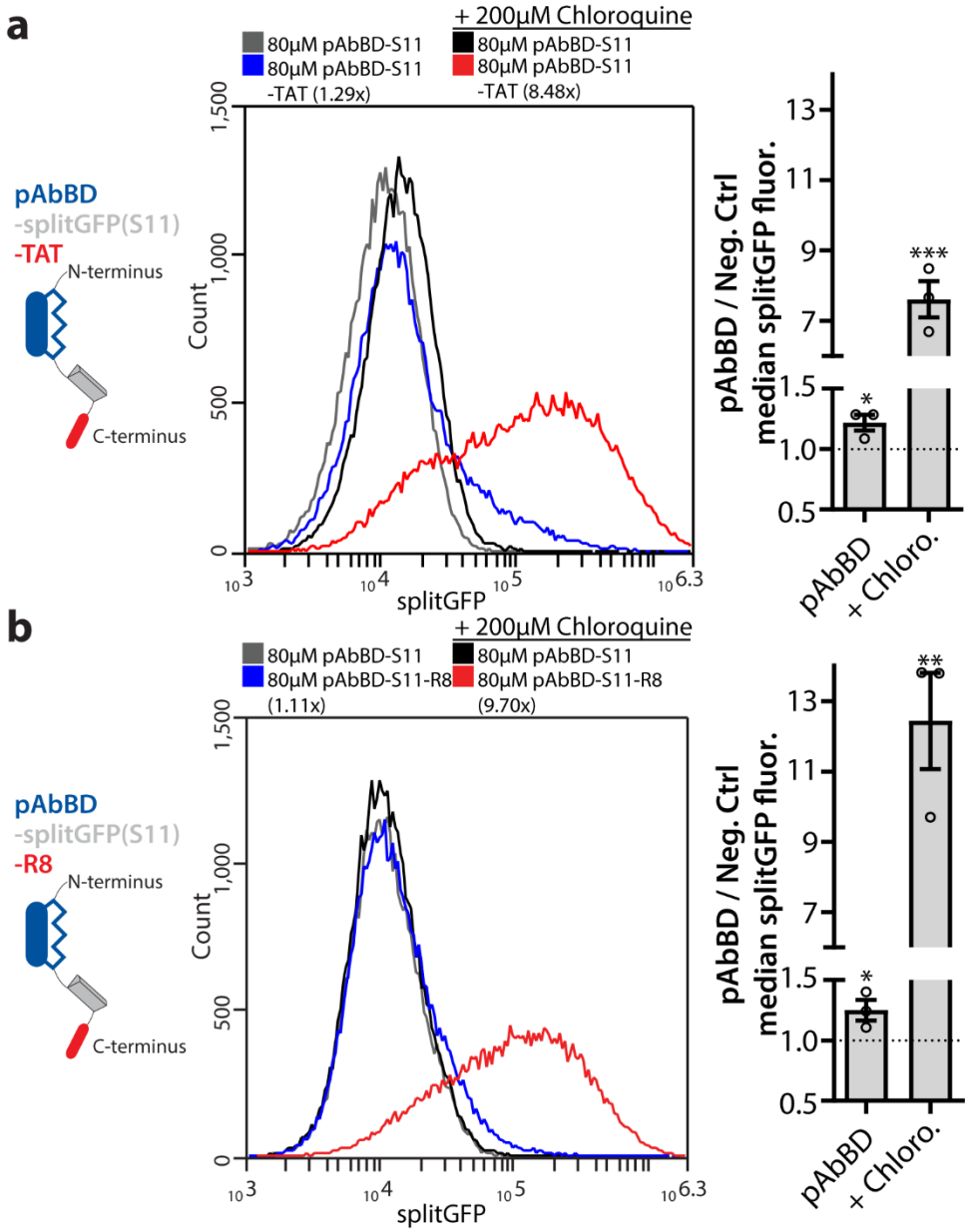

Figure 3.10 Cytosolic delivery with 80μM pAbBD-CPP

80μM pAbBD-S11 (negative control), **(a)** pAbBD-S11-TAT, or **(b)** pAbBD-S11-R8 were added to HEK293T splitGFP(1-10) reporter cells with or without chloroquine for 6 h before assessing for splitGFP fluorescence. Representative flow cytometry histograms of splitGFP fluorescence are shown. Flow cytometry data were quantified as the fold-increase in median splitGFP fluorescence over negative control in the right panel. The dotted line indicates no increase in fluorescence. Data are mean ± s.e.m, *p<0.05 **p<0.01 ***p<0.001.

3.5 Methods

3.5.1 Cloning

For pRSET His6-splitGFP(1-10), an *E. coli* optimized DNA sequence encoding for His6-splitGFP(1-10) was synthesized (Integrated DNA Technologies, IDT) and inserted between the NdeI and HindIII restriction sites of pRSET-A (Invitrogen).

For pRSET pAbBD-S11, pRSET TAT-pAbBD-S11 and pRSET R8-pAbBD-S11, *E. coli* optimized DNA sequences were synthesized and inserted between the NdeI and NcoI restriction sites of pRSET-A via In-Fusion Cloning. pEVOL-pBpF was a gift from Peter Schultz (Addgene plasmid # 31190 ; http://n2t.net/addgene:31190 ; RRID:Addgene_31190).

For pRSET SpyCatcher-SrtA-His12 (used in PBSL, see pAbBD expression and purification), pRSET SpyCatcher-SrtA-His6 was used as a starting point[36]. Overhang PCR was used to make a DNA sequence encoding for SpyCatcher-SrtA-His12, which was then inserted between the NdeI and NcoI restriction sites of pRSET-A.

All plasmids were verified by Sanger sequencing. Protein and DNA sequences for all purified proteins are listed in Table 3.1.

Table 3.1 List of purified proteins used for CPP studies

For pAbBD amino acid sequences, BPA is denoted as **B**. The CPP sequence is red. Linker sequences are colored brown. The S11 splitGFP reporter peptide is colored green. Sequences used for PBSL that are cleaved off during pAbBD purification are colored purple.

Plasmid Name	Protein Sequence	DNA Sequence (Open Reading Frame)
pRSET SpyCatcher-SrtA-His12	MSGSGDSATHIKFSK RDEDGKELAGATMEL RDSSGKTISTWISDG QVKDFYLYPGKYTFV ETAAPDGYEVATAITF TVNEQGQVTVNGGG SGGSSGGGSGGSGG SGGSGGSQAKPQIPK	ATGTCTGGTTCTGGAGACAGCGCCACTCACATTAAATTCAGTAAAC GCGATGAAGACGGCAAGGAACTCGCGGGTGCAACTATGGAACTG CGGGACTCTTCTGGTAAAACGATTAGTACCTGGATTTCCGATGGAC AGGTTAAAGATTTTTATTTGTACCCAGGTAAATATACGTTTGTAGAA ACCGCCGCGCCAGATGGCTACGAAGTGGCTACGGCCATTACCTTC ACGGTTAATGAACAAGGCCAAGTCACGGTCAATGGCGGTGGTAGT GGTGGGAGCAGTGGCGGTGGTAGTGGTGGCTCTGGCGGTTCTGG TGGCAGTGGCGGTAGCCAAGCTAAACCTCAAATTCCGAAAGATAA

	DKSKVAGYIEIPDADIK EPVYPGPATPEQLNR GVSFAEENESLDDQN ISIAGHTFIDRPNYQFT NLKAAKKGSMVYFKV GNETRKYKMTSIRDV KPTDVEVLDEQKGKD KQLTLITCDDYNEKTG VWEKRKIFVATEVKH HHHHHHHHHHHH	ATCAAAAGTGGCAGGCTATATTGAAATTCCAGATGCTGATATTAAA GAACCAGTATATCCAGGACCAGCAACACCTGAACAATTAAATAGAG GTGTAAGCTTTGCAGAAGAAAATGAATCACTAGATGATCAAAATAT TTCAATTGCAGGACACACTTTCATTGACCGTCCGAACTATCAATTTA CAAATCTTAAAGCAGCCAAAAAAGGTAGTATGGTGTACTTTAAAGT TGGTAATGAAACACGTAAGTATAAAATGACAAGTATAAGAGATGTT AAGCCAACAGATGTAGAAGTTCTAGATGAACAAAAAGGTAAAGATA AACAATTAACATTAATTACTTGTGATGATTACAATGAAAAGACAGGC GTTTGGGAAAAACGTAAAATCTTTGTAGCTACAGAAGTCAAACATC ACCACCATCATCACCACCATCATCATCACCACTGA
pRSET His6-splitGFP(1-10)	MHHHHHHSKGEELFT GVVPILVELDGDVNG HKFSVRGEGEGDATI GKLTLKFICTTGKLPV PWPTLVTTLTYGVQC FSRYPDHMKRHDFFK SAMPEGYVQERTISF KDDGKYKTRAVVKFE GDTLVNRIELKGTDFK EDGNILGHKLEYNFN SHNVYITADKQKNGIK ANFTVRHNVEDGSVQ LADHYQQNTPIGDGP VLLPDNHYLSTQTVLS KDPNEK	ATGCACCATCATCACCATCACTCTAAAGGTGAAGAACTGTTCACCG GTGTTGTTCCGATCCTGGTTGAACTGGACGGTGACGTTAACGGTC ACAAATTCTCTGTTCGTGGTGAAGGTGAAGGTGACGCGACCATCG GTAAACTGACCCTGAAATTCATCTGCACCACCGGTAAACTGCCGG TTCCGTGGCCGACCCTGGTTACCACCCTGACCTACGGTGTTCAGT GCTTCTCTCGTTACCCGGACCACATGAAACGTCACGACTTCTTCAA ATCTGCTATGCCGGAAGGTTACGTTCAGGAACGTACCATCTCTTTC AAAGACGACGGTAAATACAAAACCCGTGCTGTTGTTAAATTCGAAG GTGACACCCTGGTTAACCGTATCGAACTGAAAGGTACCGACTTCA AAGAAGACGGTAACATCCTGGGTCACAAACTGGAATACAACTTCAA CTCTCACAACGTTTACATCACCGCTGACAAACAGAAAAACGGTATC AAAGCTAACTTCACCGTTCGTCACAACGTTGAAGACGGTTCTGTTC AGCTGGCTGACCACTACCAGCAGAACACCCCGATCGGTGACGGT CCGGTTCTGCTGCCGGACAACCACTACCTGTCTACCCAGACCGTT CTGTCTAAAGACCCGAACGAAAAATAA
pRSET pAbBD-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSG GSSGGGGSGGRDHM VLHEYVNAAGITSGS GGSLPETGGGSGGS GGSGGSGGSAHIVM VDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACT GATCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCT GTTGACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACG ACTACGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAA CCTTCACCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGC GGTGGCTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACAT GGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTC GGGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTG GCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTC GATGCTTACAAACCGACGAAATAA
pRSET TAT-pAbBD-S11	MAGRKKRRQRRRGG SSGGSGGTFKLIINGK TLKGEITIEAVDA**B**EA EKIFKQYANDYGIDGE WTYDDATKTFTVTES SGGGGSGGRDHMVL HEYVNAAGITSGSGG SLPETGGGSGGSGG SGGSGGSAHIVMVDA YKPTK	ATGGCAGGGCGAAAAAAACGACGACAACGACGACGAGGGGGGAG CTCTGGCGGTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAA AACCCTGAAAGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGA AGCTGAAAAAATCTTCAAACAGTACGCTAACGACTACGGTATCGAC GGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACC GAATCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGT GTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGG TGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCG GTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATG CTTACAAACCGACGAAATAA
pRSET R8-pAbBD-S11	MRRRRRRRRGGSSG GSGGTFKLIINGKTLK GEITIEAVDA**B**EAEKIF KQYANDYGIDGEWTY DDATKTFTVTESSGG GGSGGRDHMVLHEY VNAAGITSGSGGSLP ETGGGSGGSGGSGG SGGSAHIVMVDAYKP TK	ATGCGCCGCCGACGGCGTCGCCGACGGGGTGGGAGCTCTGGCG GTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAAAACCCTGAA AGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGAAGCTGAAAA AATCTTCAAACAGTACGCTAACGACTACGGTATCGACGGTGAATG GACCTACGACGACGCTACCAAAACCTTCACCGTTACCGAATCCTC AGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACG AATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTC TGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGT GGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAA CCGACGAAATAA

3.5.2 pAbBD expression and purification

All pAbBD variants were purified via PBSL[36]. Using PBSL to purify proteins consists of 2 steps – 1) PBSL resin preparation and 2) target protein purification. PBSL resin was prepared as previously described with 2 minor modifications due to the use of SpyCatcher-SrtA-His12[36]. Pellets were lysed by resuspending in lysis buffer (PBS + 1% w/v N-Octyl-β-D-1-thioglucopyranoside (OTG) + 200 µg/mL lysozyme + 4 µg/mL DNAseI + EDTA-free protease inhibitor cocktail (Roche)) and rotating for 30 min at room temperature (RT). Following binding, the resin was washed with PBS + 20 mM imidazole 3 times followed by PBS once.

Plasmid encoding for various pAbBD variants were transformed in conjunction with pEVOL-pBpF into T7 express competent E. coli cells (New England Biolabs). Starter cultures were grown in LB + 100 µg/mL ampicillin (amp) + 25 µg/mL chloramphenicol (cam) at 37°C with shaking until OD600≈0.6. The starter culture was added at a 1:1000 dilution to autoinduction media (Formedium AIMLB0210 autoinduction media LB Broth Base including trace elements supplemented with 0.6% v/v glycerol and 100 µg/mL amp) further supplemented with 25 µg/mL cam + 0.1% w/v arabinose + 3.33 µM 4-Benzoyl-L-phenylalanine (BPA, Bachem). All pAbBD variants were grown at 37°C with shaking for 24 h. Expression culture were then pelleted and stored at -20°C.

Frozen pellets were lysed by resuspending in lysis buffer for 30 min at RT. Afterwards, lysates were frozen at -80°C and then thawed in a 37°C water bath. The lysates were clarified by centrifuging for 15 min at >14,000xg and discarding the pellet. Clarified lysates were incubated with the SpyCatcher-SrtA-His12 resin prepared above while rotating for 25 min at RT. Following binding, the resin was transferred to a Poly-Prep chromatography column (Biorad) and washed with 1 column volume (CV) of PBS, 1 CV of PBS + 20 mM imidazole, and 1 CV of PBS + 1 M NaCl + 20 mM imidazole. pAbBD variants were then eluted from the resin by adding PBS + 250 µM CaCl2 + 2 mM Gly-Gly-Gly (triglycine) and incubating at 25°C for 3 h. To ligate CPPs onto the

C-terminus of pAbBD variants, PBS + 250 µM CaCl2 + 1 M NaCl + 200 µM GGG-TAT (GGGGRKKRRQRRR) or GGG-R8 (GGGRRRRRRRR) was added to the resin and incubated at 37°C for 4 h.

Following elution, pAbBD variants were buffer exchanged into PBS and concentrated to >0.5 mg/mL via a 10k MWCO Amicon Ultra centrifugal filter (MilliporeSigma). The final protein was analyzed by SDS-PAGE for purity, tested for splitGFP complementation, stored at -80°C, and tolerates freeze-thaw cycles well.

3.5.3 His6-splitGFP(1-10) expression, purification, and complementation

pRSET His6-splitGFP(1-10) was transformed into T7 express competent *E. coli* cells. Starter cultures were grown in LB + 100 µg/mL amp at 37°C with shaking until OD_{600}≈0.6. The starter culture was added at a 1:1000 dilution to autoinduction media and grown at 25°C in an orbital shaking incubator for 40 h. Expression cultures were then pelleted and stored at -20°C.

Frozen pellets were lysed with TNG buffer (100 mM Tris-HCl, 150 mM NaCl, 10% v/v glycerol, pH 7.4) + 1% w/v OTG + 200 µg/mL lysozyme + 4 µg/mL DNAseI + EDTA-free protease inhibitor cocktail and then clarified in the same manner as pAbBD purification. Clarified lysates were incubated with cobalt resin (Talon, Clontech) while rotating for 30 min at RT. Following binding, the resin was transferred to a Poly-Prep chromatography column and washed with 3 CV of TNG buffer + 10 mM imidazole and then eluted with TNG buffer + 200 mM imidazole. Elutions were buffer exchanged into TNG buffer and concentrated to ≥1 mg/mL via a 10k MWCO Amicon Ultra centrifugal filter. The final protein was analyzed by SDS-PAGE for purity, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

Complementation reactions were done in 96-well black flat bottom plates. pAbBD-S11 or Ritux-$(pAbBD\text{-}S11)_2$ were mixed with 400 pmol His6-splitGFP(1-10) in TNG buffer (100 µl final volume) and incubated at 37°C. splitGFP fluorescence was measured using a Tecan Infinite

m200 microplate reader. A reaction containing only 400 pmol His6-splitGFP(1-10) was used to subtract out the fluorescence background. For protein quality control, 100 pmol or 50 pmol of freshly purified pAbBD-S11 or IgG-(pAbBD-S11)$_2$, respectively, was mixed with 400pmol of His6-splitGFP(1-10) and its splitGFP fluorescence was compared against that of a pAbBD-S11 standard.

3.5.4 Photocrosslinking pAbBD variants to Rituximab

Clinical grade Rituximab was provided by Eline Luning Prak (University of Pennsylvania).

For photocrosslinking, pAbBD variants were added to IgGs at a 2:1 molar ratio in PBS. IgG concentration was kept $\leq$5 μM and the pAbBD-IgG uncrosslinked mixture was aliquoted in 2 mL clear polypropylene microcentrifuge tubes. The mixture was then placed in an ice bath and irradiated for 3 h with 365 nm UV light using a UVP CL-1000L UV crosslinker placed in a 4°C cold room. After photocrosslinking, IgG-pAbBD$_2$ conjugates were washed with PBS 3 times and then concentrated to $\geq$10 μM via a 100k MWCO Amicon Ultra Centrifugal filter to remove any uncrosslinked pAbBD. SDS-PAGE was used to confirm that >95% of IgG heavy chains were photocrosslinked and that any excess pAbBD was removed. The final protein was then tested for splitGFP complementation and stored at 4°C for short durations (<1 week) and at -80°C for long durations.

3.5.5 Cell culture

HEK293T splitGFP(1-10) cells were provided by Philip Zoltick (Children's Hospital of Philadelphia) and maintained in DMEM (Corning 10-013-CV) supplemented with 10% FBS (Corning 35-010-CV), 2 μg/mL puromycin (puro) (Takara 631305), and 100 U/mL penicillin-streptomycin (ThermoFisher 15140122). Cells were cultured at 37°C in a 5% CO_2 humidified incubator. All cell lines tested negative for mycoplasma contamination.

3.5.6 Protein delivery

80,000 HEK293T splitGFP(1-10) cells were seeded in each well of a 48-well flat-bottom tissue culture plate ~16 h prior to delivery. For delivery, the media was replaced with 200 µl fresh media containing the indicated concentration of protein to be delivered with or without 200 µM chloroquine. Each delivery condition was paired with the same concentration of pAbBD-S11 or Ritux-(pAbBD-S11)$_2$ as a negative control.

3.5.7 Fluorescence microscopy

Protein delivery was performed as described above. 30 min prior to microscopy, 50 µg/mL Hoechst 33258 (ThermoFisher) was added to the cells to stain the nuclei. The media was then replaced with live cell imaging solution (ThermoFisher, A14291D). Microscopy images were acquired on an Olympus IX 81 inverted fluorescence microscope with a motorized stage, a 20x 0.4 N.A. dry objective (Olympus), a SOLA light engine excitation source (Lumencor), and an Andor iXon EM+ back-illuminated EMCCD camera using Metamorph software (Molecular Devices). After microscopy, ImageJ was used to background subtract images, equalize channels, and merge channels.

3.5.8 Flow cytometry

Protein delivery was performed as described above. Following delivery, cells were detached by incubating with 0.05% trypsin for 3 min at 37°C, resuspended in flow cytometry buffer (DPBS + 1 mM EDTA + 1% w/v BSA) and placed on ice until data collection. Flow cytometry was performed on a BD Accuri C6 Plus flow cytometer (BD Biosciences). Analysis was done using the included BD Accuri C6 Plus software.

3.6 References

1. Stewart, M. P. *et al.* In vitro and ex vivo strategies for intracellular delivery. *Nature* **538**, 183–192 (2016).
2. Carter, P. J. & Lazar, G. A. Next generation antibody drugs: pursuit of the ‘high-hanging

fruit'. *Nat. Rev. Drug Discov.* **17**, 197–223 (2018).

3. Peraro, L. & Kritzer, J. A. Emerging Methods and Design Principles for Cell-Penetrant Peptides. *Angew. Chem. Int. Ed.* **57**, 11868–11881 (2018).
4. Deprey, K., Becker, L., Kritzer, J. & Plückthun, A. Trapped! A Critical Evaluation of Methods for Measuring Total Cellular Uptake versus Cytosolic Localization. *Bioconjug. Chem.* **30**, 1006–1027 (2019).
5. Lönn, P. & Dowdy, S. F. Cationic PTD/CPP-mediated macromolecular delivery: charging into the cell. *Expert Opin. Drug Deliv.* **12**, 1627–1636 (2015).
6. Wadia, J. S., Stan, R. V. & Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **10**, 310–315 (2004).
7. Verdurmen, W. P. R., Mazlami, M. & Plückthun, A. A quantitative comparison of cytosolic delivery via different protein uptake systems. *Sci. Rep.* **7**, 1–13 (2017).
8. Derossi, D., Joliot, A. H., Chassaing, G. & Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* **269**, 10444–50 (1994).
9. Koren, E. & Torchilin, V. P. Cell-penetrating peptides: Breaking through to the other side. *Trends Mol. Med.* **18**, 385–393 (2012).
10. Milletti, F. Cell-penetrating peptides: Classes, origin, and current landscape. *Drug Discov. Today* **17**, 850–860 (2012).
11. Erazo-Oliveras, A. *et al.* Protein delivery into live cells by incubation with an endosomolytic agent. *Nat. Methods* **11**, 861–867 (2014).
12. Li, M. *et al.* Discovery and Characterization of a Peptide That Enhances Endosomal Escape of Delivered Proteins in Vitro and in Vivo. *J. Am. Chem. Soc.* **137**, 14084–14093 (2015).
13. Akishiba, M. *et al.* Cytosolic antibody delivery by lipid-sensitive endosomolytic peptide. *Nat. Chem.* **9**, 751–761 (2017).
14. Erazo-Oliveras, A., Muthukrishnan, N., Baker, R., Wang, T. Y. & Pellois, J. P. Improving the endosomal escape of cell-penetrating peptides and their cargos: Strategies and challenges. *Pharmaceuticals* **5**, 1177–1209 (2012).
15. Gillmeister, M. P., Betenbaugh, M. J. & Fishman, P. S. Cellular trafficking and photochemical internalization of cell penetrating peptide linked cargo proteins: A dual fluorescent labeling study. *Bioconjug. Chem.* **22**, 556–566 (2011).
16. Tünnemann, G. *et al.* Cargo-dependent mode of uptake and bioavailability of TAT-containing proteins and peptides in living cells. *FASEB J.* **20**, 1775–1784 (2006).
17. Cronican, J. J. *et al.* Potent Delivery of Functional Proteins into Mammalian Cells in Vitro

and in Vivo Using a Supercharged Protein. *ACS Chem. Biol.* **5**, 747–752 (2010).

18. Cronican, J. J. *et al.* A class of human proteins that deliver functional proteins into mammalian cells in vitro and in vivo. *Chem. Biol.* **18**, 833–838 (2011).
19. Sung, M., Poon, G. M. K. & Gariépy, J. The importance of valency in enhancing the import and cell routing potential of protein transduction domain-containing molecules. *Biochim. Biophys. Acta* **1758**, 355–363 (2006).
20. Erazo-Oliveras, A. *et al.* The Late Endosome and Its Lipid BMP Act as Gateways for Efficient Cytosolic Access of the Delivery Agent dfTAT and Its Macromolecular Cargos. *Cell Chem. Biol.* **23**, 598–607 (2016).
21. Najjar, K. *et al.* Unlocking Endosomal Entrapment with Supercharged Arginine-Rich Peptides. *Bioconjug. Chem.* **28**, 2932–2941 (2017).
22. Brock, D. J. *et al.* Efficient cell delivery mediated by lipid-specific endosomal escape of supercharged branched peptides. *Traffic* **19**, 421–435 (2018).
23. Futaki, S. *et al.* Arginine-rich Peptides. *J. Biol. Chem.* **276**, 5836–5840 (2001).
24. Hui, J. Z., Tamsen, S., Song, Y. & Tsourkas, A. LASIC: Light Activated Site-Specific Conjugation of Native IgGs. *Bioconjug. Chem.* **26**, 1456–1460 (2015).
25. Malakauskas, S. M. & Mayo, S. L. Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5**, 470–5 (1998).
26. Chung, B. H. *et al.* Photoactivable antibody binding protein: Site-selective and covalent coupling of antibody. *Anal. Chem.* **81**, 936–942 (2009).
27. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proc. Natl. Acad. Sci.* **99**, 11020–11024 (2002).
28. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102–107 (2005).
29. Smith, S. A., Selby, L. I., Johnston, A. P. R. & Such, G. K. The Endosomal Escape of Nanoparticles: Toward More Efficient Cellular Delivery. *Bioconjug. Chem.* **30**, 263–272 (2019).
30. Milech, N. *et al.* GFP-complementation assay to detect functional CPP and protein delivery into living cells. *Sci. Rep.* **5**, 1–11 (2015).
31. Schmidt, S. *et al.* Detecting Cytosolic Peptide Delivery with the GFP Complementation Assay in the Low Micromolar Range. *Angew. Chem. Int. Ed.* **54**, 15105–15108 (2015).
32. Kauffman, W. B., Guha, S. & Wimley, W. C. Synthetic molecular evolution of hybrid cell penetrating peptides. *Nat. Commun.* **9**, (2018).

33. Schmidt, S. *et al.* Identification of Short Hydrophobic Cell-Penetrating Peptides for Cytosolic Peptide Delivery by Rational Design. *Bioconjug. Chem.* **28**, 382–389 (2017).
34. Caron, N. J., Quenneville, S. P. & Tremblay, J. P. Endosome disruption enhances the functional nuclear delivery of Tat-fusion proteins. *Biochem. Biophys. Res. Commun.* **319**, 12–20 (2004).
35. Herce, H. D. *et al.* Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells. *Nat. Chem.* **9**, 762–771 (2017).
36. Wang, H. H., Altun, B., Nwe, K. & Tsourkas, A. Proximity-Based Sortase-Mediated Ligation. *Angew. Chem. Int. Ed.* **56**, 5349–5352 (2017).

Chapter 4 – Cytosolic delivery of inhibitory antibodies with cationic lipids

4.1 Abstract

The inability of antibodies to enter the cytosol limits inhibitory antibodies to extracellular targets. Given that two-thirds of drug targets are intracellular, any system that can efficiently deliver antibodies into the cytosol would dramatically increase their therapeutic utility. Furthermore, cytosolic antibodies would offer unique opportunities to directly modulate and study intracellular protein function. Here we demonstrate that immunoglobulin G (IgG) antibodies that are conjugated with anionic polypeptides (ApPs) can be complexed with cationic lipids through electrostatic interactions, enabling close to 90% cytosolic delivery efficiency with only 500 nM IgG. The ApP is fused to a small photoreactive antibody-binding domain (pAbBD) that can be site-specifically photocrosslinked to nearly all off-the-shelf IgGs. We show that cytosolically delivered IgGs can inhibit the drug efflux pump multidrug resistance-associated protein 1 (MRP1) and the transcription factor NFκB. This work establishes a new approach for using existing antibody collections to modulate intracellular protein function and provides the foundations for therapeutic cytosolic antibodies.

4.2 Introduction

Antibodies have become important research tools and a powerful class of therapeutics because they can be developed to bind nearly any exposed protein epitope with high affinity and specificity through either traditional immunization or *in vitro* display approaches[1]. By binding to an appropriate epitope, antibodies can also directly inhibit their antigen's biological activity by either sterically blocking the antigen from binding to interaction partners or locking the antigen in an inactive conformation[1-2]. In contrast to small-molecule drugs, which must bind to small pockets crucial for their target protein's function[3], IgG antibodies bind to antigens via much larger and

flatter contact areas, making them attractive for inhibiting traditionally undruggable targets[2,4]. Indeed, cytosolic antibody-dependent inhibition of major oncology drug targets such as Ras has been long known to be possible with microinjected antibodies[5-6]. The lack of an adequate cytosolic delivery system, however, has limited therapeutic antibodies to only targets accessible from the extracellular space.

Unsurprisingly, many approaches have been explored for cytosolic antibody delivery[7-8]. Physical delivery techniques such as microinjection[5-6,9] or electroporation[9-10] are very effective, but are low-throughput, can result in significant toxicity, and have no therapeutic translatability. Antibody fragments can be expressed intracellularly as intrabodies, but the coding sequence for many IgG antibodies aren't known, antibody fragments often fold poorly in the cytosol, and this approach suffers from the same therapeutic drawbacks as gene therapy[7-8,11]. Antibody-like binding proteins can be fused to cell-penetrating peptides (CPPs), which are effective at inducing endocytic uptake, but are extremely inefficient at inducing endosomal escape of cargo proteins and thus, require high protein concentrations ($\geq$5 μM) to achieve detectable delivery[12-13]. Finally, antibodies with an innate ability to enter the cytosol have been recently developed, but their delivery functionality lies within the light chain variable region, which imposes significant constraints when generating new antibodies[14].

Carrier-mediated approaches for cytosolic protein delivery, in which cargo proteins are encapsulated by delivery lipids or polymers, are attractive because they can potentially capitalize on advances in non-viral nucleic acid delivery formulations that have already been proven effective in humans[15]. Although carrier-mediated cytosolic antibody delivery has been reported multiple times[7-8], they must be evaluated cautiously. A stringent assessment of commercially available carrier-mediated antibody delivery platforms revealed that none were capable of cytosolic delivery to >6% of cells[10]. However, recent progress in delivering the Cas9 protein, which are similar in size to IgG antibodies, with lipid nanoparticles for genome editing[16-17]

suggests that carrier-mediated approaches are still viable strategies for cytosolic antibody delivery.

Inspired by strategies for complexing proteins with cationic lipids[16,18], polymers[19-20], and nanoparticles[21], we hypothesized that IgGs functionalized with ApPs could mimic the polyanionic nature of nucleic acids and be complexed with cationic lipids designed for nucleic acid delivery. Rather than engineering IgGs directly, we fused ApPs to a pAbBD that could be photocrosslinked to each heavy chain of an IgG to create highly negatively charged IgG-(pAbBD-ApP)$_2$ conjugates without perturbing binding affinity (Fig. 4.1)[22]. Because functionality is built into the pAbBD rather than the IgG, cargo IgGs can be easily exchanged without genetic re-engineering, allowing most off-the-shelf IgGs to be easily functionalized.

Here we report that our cytosolic antibody delivery approach enables close to 90% delivery efficiency at a concentration of only 500 nM IgG. Our modular antibody functionalization strategy is compatible with IgGs from many different species and isotypes and our complexation approach is compatible with a diverse set of cationic lipids. Finally, we demonstrate that cytosolically delivered IgGs are functional and can inhibit not only the drug efflux pump MRP1 to sensitize cancer cells to chemotherapeutic drugs, but also the transcription factor NFκB. These findings extend the potential applications of existing IgG antibody collections and present a straightforward framework for therapeutic cytosolic protein delivery.

4.3 Results

4.3.1 Stringent detection of cytosolic protein delivery

When developing intracellular protein delivery technologies, it is imperative to use an assay that only detects cargo proteins that have been successfully delivered to the cytosol. Endosome escape is the major bottleneck for cytosolic delivery, resulting in large false-positive rates for assays that measure total cellular uptake of cargo proteins[23-24]. Accordingly, we used a

stringent self-assembling splitGFP[25] reporter system in which one half of the splitGFP, the S11 peptide, is fused to pAbBD-ApP whereas the other half, splitGFP(1-10), is expressed in reporter cells[26-29]. Only once IgG-(pAbBD-ApP-S11)$_2$ is successfully delivered into the cytosol does splitGFP complementation and turn-on fluorescence occur (Fig. 4.1). Furthermore, fluorescence intensity is directly correlated with the amount of protein cytosolically delivered.

We prepared pAbBD-ApP-S11 with 10 to 30 residues-long poly-aspartate or poly-glutamate ApPs and confirmed that they could photocrosslink to Rituximab (Ritux) (Supplementary Fig. 4.7a, b). We chose Rituximab because its antigen, CD20, is not expressed in our reporter cells, removing a potential confounding factor for delivery. To validate that our protein cargos are compatible with the splitGFP reporter system, we confirmed that splitGFP complementation occurred when either pAbBD-S11 or Ritux-(pAbBD-S11)$_2$ were incubated with purified splitGFP(1-10) or physically delivered into HEK293T splitGFP(1-10) reporter cells by

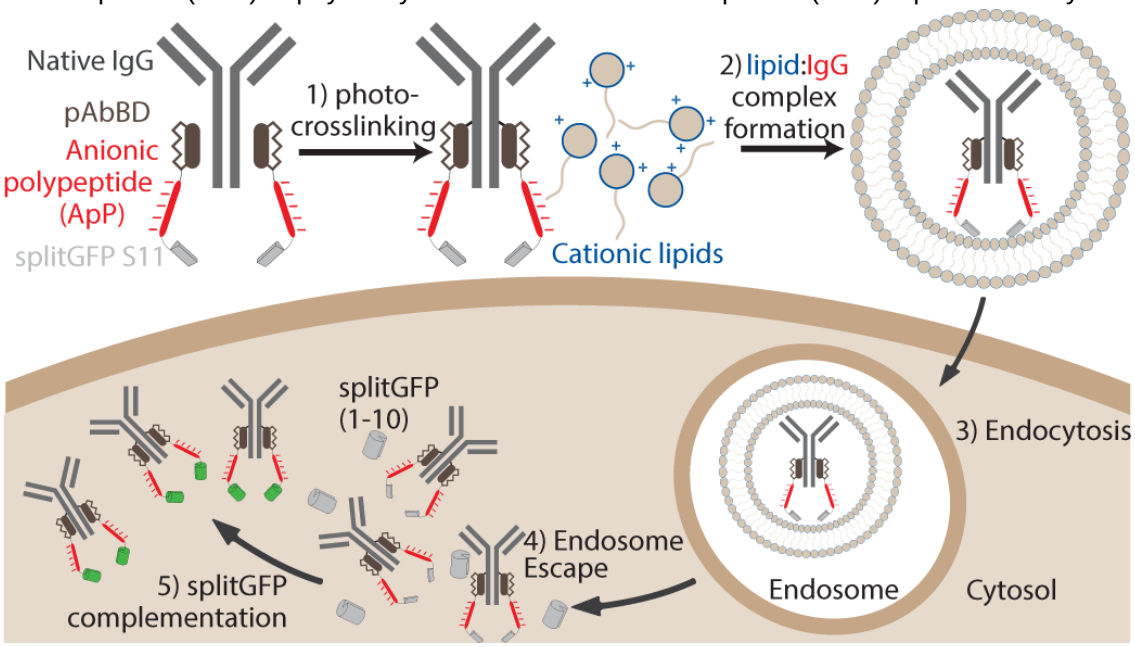

Figure 4.1 Schematic of antibody delivery approach.

pAbBD is purified as a fusion to an ApP as well as the splitGFP S11 reporter peptide and then photocrosslinked to a cargo IgG (1). IgG-(pAbBD-ApP-S11)$_2$ conjugates are complexed with cationic delivery lipids (2), which are taken up by cells via endocytosis (3). Delivery lipids promote endosome escape (4) and cytosolic IgG-(pAbBD-ApP-S11)$_2$ delivery, which can be detected by splitGFP complementation and turn-on fluorescence (5).

electroporation (Chapter 3 Fig. 3.4 and 3.5). Due to the time required for chromophore maturation (Chapter 3 Fig. 3.4), reporter cells were assessed for splitGFP fluorescence 6 hours following electroporation with either flow cytometry or live-cell fluorescence microscopy. As expected, fluorescence microscopy revealed diffuse splitGFP fluorescence with both pAbBD-S11 and Ritux-$(pAbBD\text{-}S11)_2$, but fluorescence was depleted from the nucleus only with Ritux-$(pAbBD\text{-}S11)_2$ (Chapter 3 Fig. 3.5b). Because pAbBD-S11 is only ~11kDa, it is capable of passively translocating across the nuclear pore complex, whereas Ritux-$(pAbBD\text{-}S11)_2$ is ~170kDa and is very inefficient at passive nuclear translocation[30].

4.3.2 Cytosolic pAbBD delivery

Initially, we wondered whether pAbBD-ApP-S11 alone, without IgG, could be cytosolically delivered into HEK293T splitGFP(1-10) cells once complexed with a commercially available cationic transfection lipid, Lipofectamine (Lipo) 2000. We tested simple poly-aspartate or poly-glutamate ApPs that were 10 to 30 residues-long. Live-cell fluorescence microscopy showed robust splitGFP fluorescence, indicating substantial cytosolic delivery, once poly-aspartate ApPs were at least 15 residues-long (D15), with a peak at 20 aspartate residues (D20) (Fig. 4.2a). With poly-glutamate ApPs, splitGFP fluorescence increased with length up to 30 glutamate residues (E30) (Fig. 4.2b). Even though the base pAbBD-S11 has a net charge of -7, no cytosolic delivery could be detected without ApPs (Fig. 4.2a-b).

Using flow cytometry, we quantified splitGFP fluorescence as either the percentage of splitGFP-positive cells, which reflects delivery efficiency, or the fold-increase in splitGFP fluorescence, which reflects the amount of protein delivered (Fig. 4.2c-d, Supplementary Fig. 4.8). Flow cytometry confirmed the trends seen between ApP length and delivery efficiency identified via microscopy. Increasing the ratio of cationic lipid to protein generally improved delivery efficiency, but at the cost of increased toxicity (Supplementary Fig. 4.8). We were able to identify regimes, though, where viability remained greater than 90% with excellent delivery efficiency

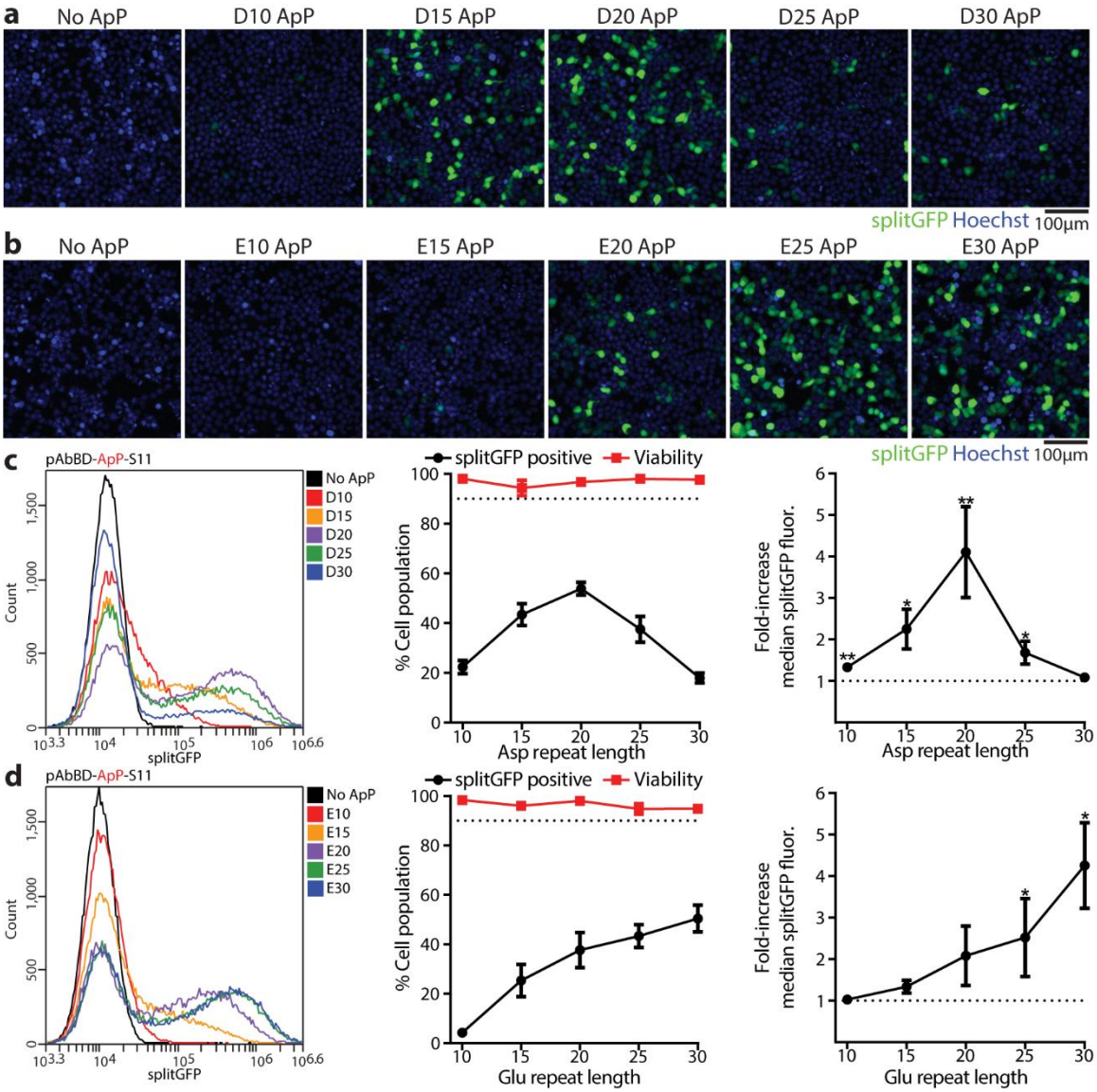

Figure 4.2 Optimizing ApPs for cytosolic pAbBD delivery.

500 nM pAbBD-S11 (negative control) or pAbBD-ApP-S11 with either poly-aspartate or poly-glutamate ApPs 10 15, 20, 25, or 30 residues-long were complexed with 2 µl Lipo 2000 and added to HEK293T splitGFP(1-10) cells for 6 h. **a-b**, Representative live-cell fluorescence microscopy images following delivery with poly-aspartate **(a)** and poly-glutamate **(b)** ApPs shows diffuse splitGFP fluorescence indicating significant cytosolic delivery. **c-d**, Flow cytometry of splitGFP fluorescence following delivery with poly-aspartate **(c)** and poly-glutamate **(d)** ApPs. Left panel: representative flow cytometry histograms. Center panel: percent of cells splitGFP-positive. Right panel: fold-increase in median splitGFP fluorescence over negative control. The dotted line indicates either 90% of the cell population (center panel) or no increase in fluorescence (right panel). Viability was determined with the LDH assay. Data are mean ± s.e.m., n=4, *p<0.05 **p<0.01 (one-sided one sample *t*-test of log-ratios).

(Fig. 4.2c-d, Supplementary Fig. 4.8). The best poly-glutamate and poly-aspartate ApPs were D20 and E30 with delivery efficiencies of 53.9 ± 2.6% and 50.5 ± 5.4%, respectively, when 500 nM pAbBD-ApP-S11 was complexed with 2 µl Lipo 2000 (Fig. 4.2c-d).

Together, these results demonstrate that a small protein, pAbBD-S11 (~11kDa), can be efficiently delivered into cytosol simply by fusing it to poly-aspartate or poly-glutamate ApPs via complexation with cationic lipids. We believe that this strategy can be easily adopted for cytosolic delivery of small antibody-like binding proteins such as affibodies (~6kDa), monobodies (~10kDa), or nanobodies (~15kDa), but ApP length may need to be re-optimized for maximal delivery.

4.3.3 Cytosolic IgG delivery

Next, we tested whether Ritux-(pAbBD-ApP-S11)$_2$ could also be cytosolically delivered when complexed with Lipo 2000. As expected, no splitGFP fluorescence could be detected by microscopy with Ritux-(pAbBD-S11)$_2$, which has a base net charge of +4 (Fig. 4.3a-b). Once both poly-aspartate and poly-glutamate ApPs reached at least 20-residues long, however, microscopy revealed diffuse splitGFP fluorescence with nuclear depletion (Fig. 4.3a-b). Because nuclear depletion indicates that the S11 reporter peptide remained linked to Ritux-(pAbBD-ApP-S11)$_2$ following endosomal escape, we are confident that the splitGFP fluorescence is reflective of significant cytosolic delivery of Ritux-(pAbBD-ApP-S11)$_2$.

Flow cytometry of splitGFP fluorescence corresponded well to microscopy, but additionally revealed that delivery peaked with 25 aspartates (D25), but plateaued with 20 glutamates (E20) (Fig. 4.3c-d, Supplementary Fig. 4.9). Lipo RNAiMax, a cationic lipid designed for siRNA delivery, was more effective than Lipo 2000 at delivering Ritux-(pAbBD-ApP-S11)$_2$ with short ApPs, but worse with longer ApPs (Supplementary Fig. 4.10). Similarly to pAbBD-ApP-S11 delivery, increasing the ratio of either lipids to Ritux-(pAbBD-ApP-S11)$_2$ generally improved delivery, but also increased toxicity (Supplementary Fig. 4.9, 4.10). After significant optimization, we achieved a maximal delivery efficiency of 65.7 ± 3.6% in HEK293T splitGFP(1-10) cells with

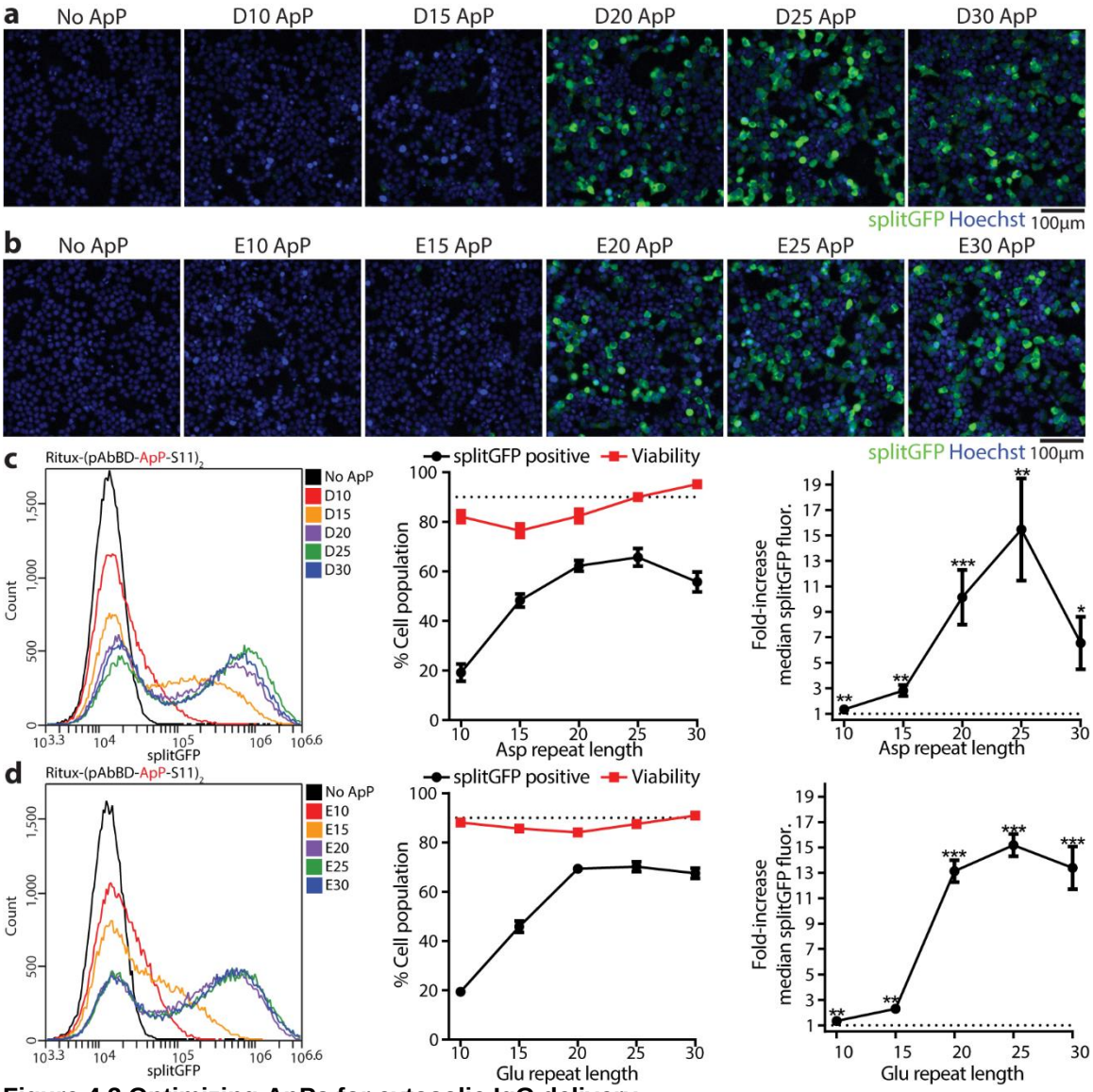

Figure 4.3 Optimizing ApPs for cytosolic IgG delivery.

500 nM Ritux-(pAbBD-S11)$_2$ (negative control) or Ritux-(pAbBD-ApP-S11)$_2$ with either poly-aspartate or poly-glutamate ApPs 10 15, 20, 25, or 30 residues-long were complexed with 2 µl Lipo 2000 and added to HEK293T splitGFP(1-10) cells for 6 h. **a-b**, Representative live-cell fluorescence microscopy images following delivery with poly-aspartate **(a)** and poly-glutamate **(b)** ApPs shows diffuse splitGFP fluorescence with nuclear depletion indicating significant cytosolic delivery. **c-d**, Flow cytometry of splitGFP fluorescence following delivery with poly-aspartate **(c)** and poly-glutamate **(d)** ApPs. Left panel: representative flow cytometry histograms. Center panel: percent of cells splitGFP-positive. Right panel: fold-increase in median splitGFP fluorescence over negative control. The dotted line indicates either 90% of the cell population (center panel) or no increase in fluorescence (right panel). Viability was determined with the LDH assay. Data are mean ± s.e.m., n=4, *p<0.05 **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

>90% viability when 500 nM Ritux-(pAbBD-D25-S11)$_2$ was complexed with 2 µl Lipo 2000 (Fig. 4.3d). Interestingly, delivery efficiencies for Ritux-(pAbBD-ApP-S11)$_2$ were higher than that of pAbBD-ApP-S11 alone, likely due to having 2 ApPs linked to each Rituximab molecule.

To demonstrate the generalizability of our IgG delivery approach, we tested several other commercially available cationic lipids and reporter cell lines. All conditions resulted in efficient cytosolic delivery of Ritux-(pAbBD-D25-S11)$_2$ and Ritux-(pAbBD-E25-S11)$_2$, albeit to varying degrees (Fig. 4.4a-c, Supplementary Fig. 4.11a-c, 4.12, 4.13). Delivery efficiency was

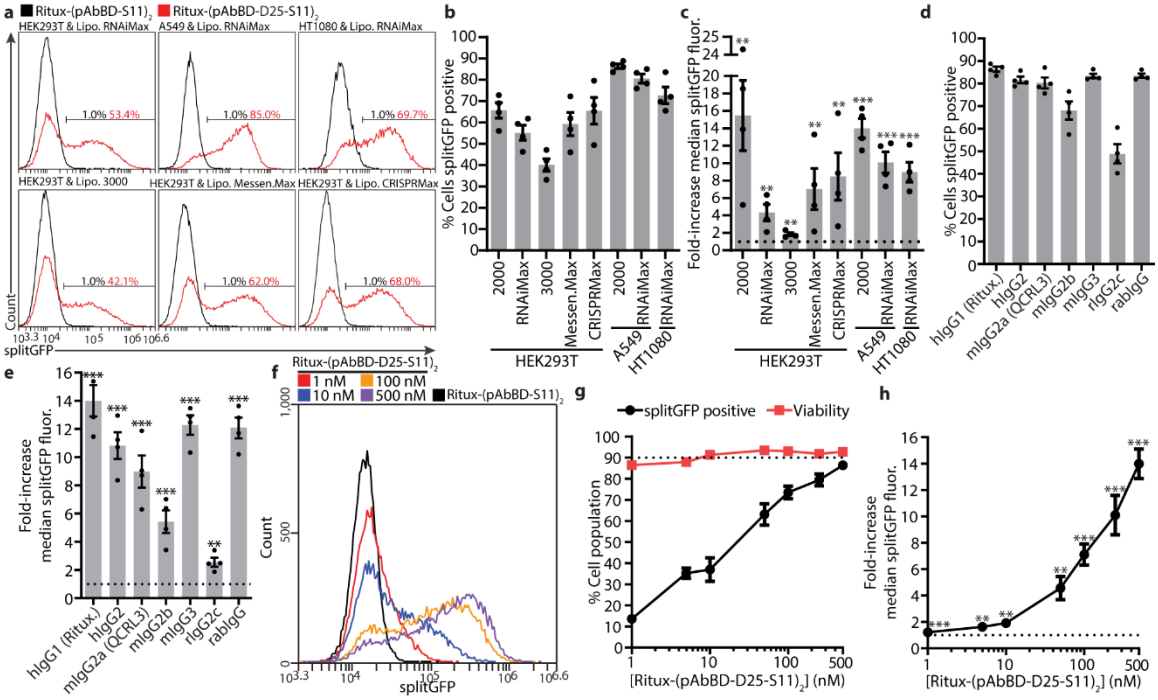

Figure 4.4 IgG delivery scope.

a-c, 500 nM Ritux-(pAbBD-S11)$_2$ (negative control) or Ritux-(pAbBD-D25-S11)$_2$ was complexed with 2 µl of the indicated cationic lipid and added to the indicated splitGFP(1-10) reporter cells for 6 h. Representative flow cytometry histograms of splitGFP fluorescence are shown in **(a)**. Flow cytometry data were quantified as the percent of cells splitGFP-positive **(b)** and the fold-increase in median splitGFP fluorescence over negative control **(c). d, e,** Same is for **b, c,** but 500 nM IgG-(pAbBD-D25-S11)$_2$ of the indicated species and isotype was complexed with 2 µl Lipo 2000 and added to A549 splitGFP(1-10) cells. **f-h,** Same as for **a-c**, but indicated concentrations of Ritux-(pAbBD-D25-S11)$_2$ were complexed with 2 µl Lipo 2000 and added to A549 splitGFP(1-10) cells. The dotted line indicates either no increase in fluorescence **(c, e, h)** or 90% of the cell population **(g)**. Viability was determined with the LDH assay. Data are mean ± s.e.m., n=4, **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

significantly higher in HT1080 and A549 splitGFP(1-10) cells, reaching up to 72.7 ± 3.9% and 86.4 ± 1.1%, respectively, with 500 nM Ritux-(pAbBD-D25-S11)$_2$ (Fig. 4.4b, Supplementary Fig. 4.13). Other than Rituximab (human IgG, hIgG1), pAbBD-D25-S11 could also photocrosslink and cytosolically deliver hIgG2, mouse IgG2a (mIgG2a), mIgG2b, mIgG3, rat IgG2c (rIgG2c), and rabbit IgG (rabIgG) into A549 splitGFP(1-10) cells (Fig. 4.4d, e). Delivery efficiency remained high with all tested IgG species and isotypes, except for rIgG2c, which had a moderate delivery efficiency (Fig. 4.4d).

Finally, we measured how delivery efficiency varied with Ritux-(pAbBD-D25-S11)$_2$ and Ritux-(pAbBD-E25-S11)$_2$ dose (Fig. 4.4f-h, Supplementary Fig. 4.11d-f). At 100nM Ritux-(pAbBD-D25-S11)$_2$, a five-fold decrease in concentration, delivery efficiency was only slightly reduced to 73.5 ± 3.0% in A549 splitGFP(1-10) cells with Lipo 2000 (Fig. 4.4g). Even at 1nM Ritux-(pAbBD-D25-S11)$_2$ or Ritux-(pAbBD-E25-S11)$_2$, cytosolic delivery was still detectable, albeit low (Fig. 4.4f-h, Supplementary Fig. 4.11d-f).

Collectively, these results demonstrate the versatility of our IgG delivery approach with regards to the following factors. (i) We can deliver off-the-shelf IgGs from a variety of species and isotypes. (ii) It is compatible with all tested cationic lipids and cell lines thus far. (iii) Delivery efficiency is maintained at low IgG-(pAbBD-ApP-S11)$_2$ concentrations. We observe that although increasing the net negative charge of cargo proteins is critical for complexation and delivery, it eventually becomes unproductive or even deleterious for delivery efficiency.

4.3.4 A cytosolically delivered IgG can inhibit MRP1, a drug efflux pump

Having shown robust delivery, we next sought to demonstrate the utility of cytosolic IgGs by inhibiting MRP1, a drug-export pump associated with chemotherapy resistance[31]. We chose MRP1 because it can be inhibited by a well-characterized monoclonal IgG, QCRL3, which inhibits via binding to one of MRP1's cytosolic nucleotide binding domains[32-34].

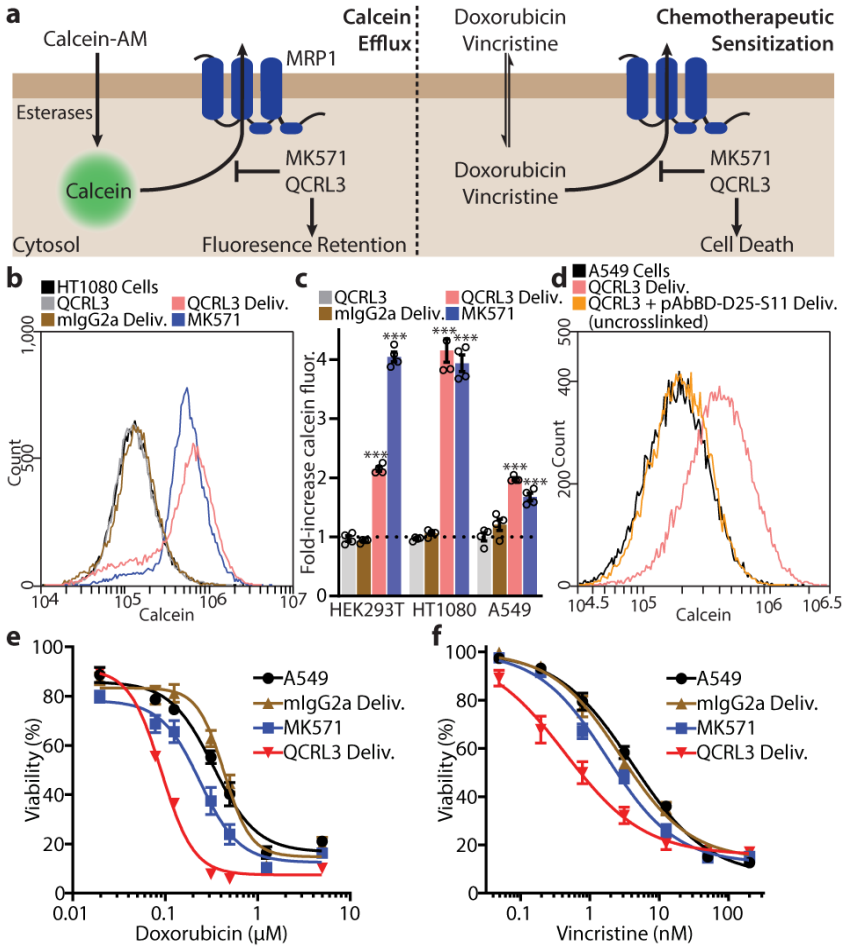

Figure 4.5 Cytosolic QCRL3 delivery inhibits MRP1.

a, Schematic of assays that assess MRP1 inhibition. In the calcein efflux assay, cells are first loaded with calcein, a fluorescent membrane-impermeable MRP1 substrate. Cells with high MRP1 activity will rapidly export calcein, whereas MRP1 inhibition results in calcein fluorescence retention. In the chemotherapeutic sensitization assay, MRP1 inhibition results in greater intracellular accumulation of doxorubicin or vincristine, resulting in sensitization to both compounds. **b,** Representative flow cytometry histograms of calcein fluorescence after 16 h of export in calcein-loaded HT1080 cells treated with 20 μM MK571, 500 nM QCRL3-(pAbBD-D25-S11)$_2$, 500 nM cytosolically delivered mIgG2a-(pAbBD-D25-S11)$_2$, or 500 nM cytosolically delivered QCRL3-(pAbBD-D25-S11)$_2$. **c,** Calcein-efflux assay quantification across HEK293T, HT1080, and A549 cell lines. Only QCRL3 delivery and MK571 treatment resulted in calcein fluorescence retention. Data are mean ± s.e.m., n=4, ***p<0.001 (one-sided one sample *t*-test of log-ratios). **d,** Same as for **b**, but in calcein-loaded A549 cells treated with cytosolic delivery of 500nM QCRL3 with or without photocrosslinking to pAbBD-D25-S11. Calcein fluorescence retention is only seen with photocrosslinked QCRL3, indicating that photocrosslinking is necessary for delivery. **e, f,** A549 cell sensitivity to doxorubicin **(e)** or vincristine **(f)** following treatment with 20 μM MK571, 500 nM cytosolically delivered mIgG2a-(pAbBD-D25-S11)$_2$, or 500 nM cytosolically delivered QCRL3-(pAbBD-D25-S11)$_2$. Data are mean ± s.e.m, n=4.

Initially, we used the calcein-efflux assay to assess MRP1 activity in which MRP1 inhibition results in calcein fluorescence retention (Fig. 4.5a). 500 nM QCRL3-(pAbBD-D25-S11)$_2$ delivery resulted in calcein retention in HEK293T, HT1080, and A549 cells (Fig. 4.5b-c), indicating inhibition of endogenous MRP1 activity. No inhibition was observed following delivery of the mIgG2a-(pAbBD-D25-S11)$_2$ isotype control or simply incubating cells with QCRL3-(pAbBD-D25-S11)$_2$. QCRL3-(pAbBD-D25-S11)$_2$ delivery performed as well as MK571, a non-selective small-molecule MRP1 inhibitor, in all except for HEK293T cells (Fig. 4.5c)[31]. We attribute this to the high expression of other efflux pumps that are also inhibited by MK571 in HEK293T cells. Finally, photocrosslinking is necessary for IgG delivery, as delivery of QCRL3 mixed with pAbBD-D25-S11 did not result in calcein retention (Fig. 4.5d). This suggests that during complexation, non-covalent interactions between pAbBD and IgGs are disrupted.

Next, in a more therapeutically relevant context, we attempted to sensitize A549 cells to doxorubicin and vincristine, which are chemotherapeutic drugs known to be MRP1 substrates[31,35]. 500 nM QCRL3-(pAbBD-D25-S11)$_2$ delivery was able to sensitize A549 cells to doxorubicin by 3.7 $\pm$ 0.45 fold and vincristine by 9.0 $\pm$ 2.0 fold (Fig. 4.5e, f). In comparison, MK571 treatment resulted in only moderate sensitization and delivery of the mIgG2a isotype control resulted in no sensitization (Fig. 4.5e, f). Thus, cytosolically delivered IgGs are functional and capable of inhibiting therapeutic targets.

4.3.5 Cytosolically delivered IgGs can inhibit the transcription factor NFκB.

Finally, we investigated whether cytosolically delivered IgGs could inhibit protein-protein interactions, which are particularly difficult to drug with small-molecules. We targeted the transcription factor NFκB, which is a heterodimer between p50 and RelA (p65) whose nuclear localization signals (NLS) are normally masked by IκBα, sequestering NFκB in the cytosol. With TNFα stimulation, IκBα is degraded, allowing NFκB to enter the nucleus to stimulate transcription[36]. We hypothesized that an anti-RelA NLS IgG could sterically block RelA from engaging with its cognate nuclear import factor (NIF) to prevent RelA nuclear translocation and

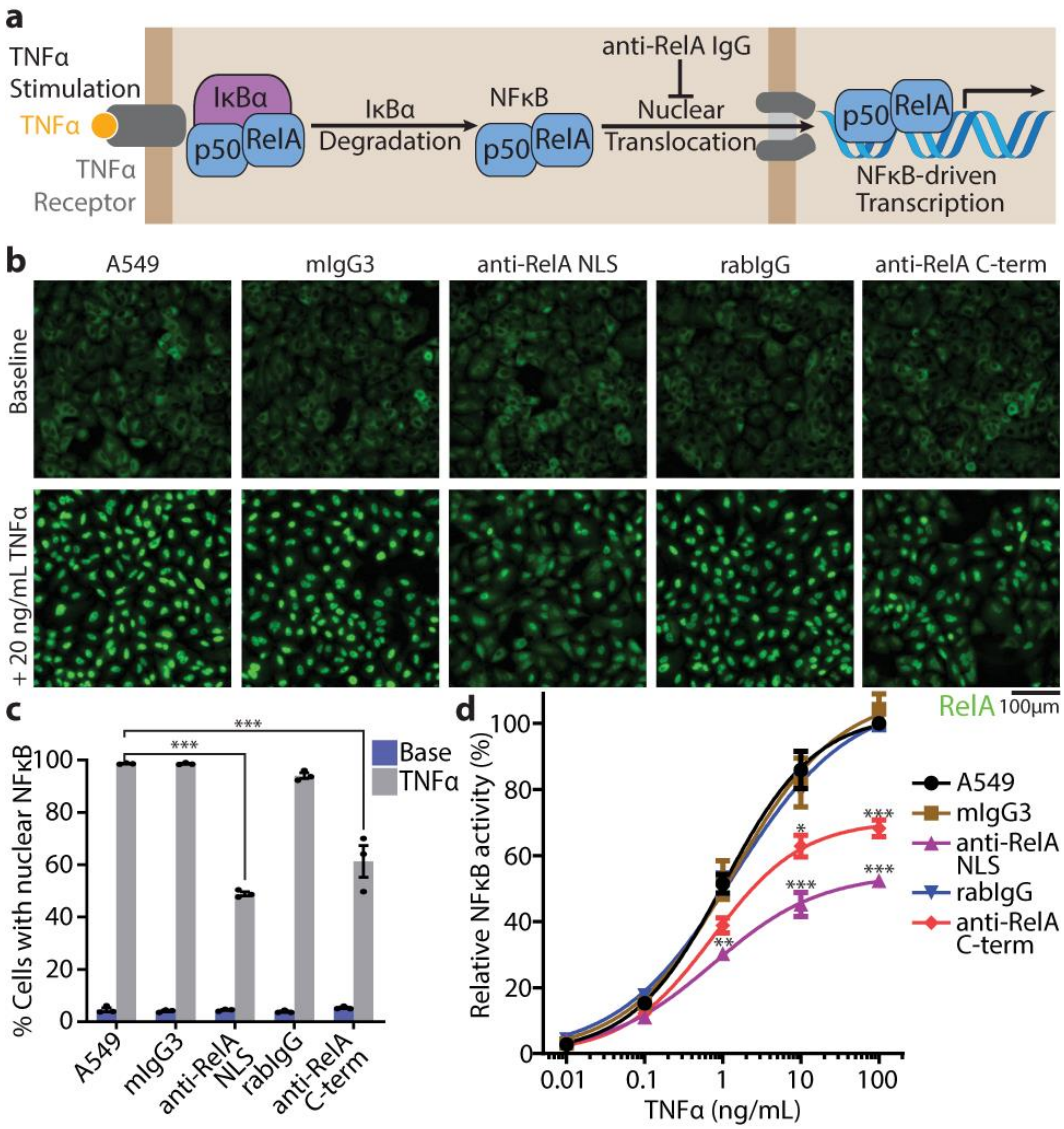

Figure 4.6 Cytosolic anti-RelA IgG delivery inhibits NFκB.

a, Schematic of NFκB inhibition. anti-RelA IgGs inhibit NFκB transcriptional activity by preventing its nuclear translocation following TNFα stimulation. **b, c,** Representative immunofluorescence images **(b)** and quantification **(c)** of RelA nuclear translocation following delivery of the indicated 150 nM IgG-(pAbBD-D25-S11)$_2$ antibody and TNFα treatment. Only delivery of anti-RelA IgGs reduced RelA nuclear translocation. Data are mean ± s.e.m, n=3, ***p<0.001 (one-way ANOVA). **d,** A549 cells were transiently transfected with a NFκB-driven firefly luciferase reporter plasmid. NFκB transcriptional activity was detected by luminescence following delivery of the indicated 150 nM IgG-(pAbBD-D25-S11)$_2$ antibody and TNFα treatment. Only delivery of anti-RelA IgGs inhibited NFκB transcriptional activity. Data are mean ± s.e.m, n=3, *p<0.05 **p<0.01 ***p<0.001 (one-way ANOVA).

NFκB-mediated transcription (Fig. 4.6a). We also tested an anti-RelA C-terminus IgG that bound to an epitope distinct from the NLS.

Initially, we delivered 150 nM anti-RelA NLS-(pAbBD-D25-S11)$_2$, anti-RelA C-term-(pAbBD-D25-S11)$_2$, or their isotype controls (mIgG3 for anti-RelA NLS, rabIgG for anti-RelA C-term) into A549 cells and assessed for RelA nuclear translocation following TNFα stimulation. Immunofluorescence revealed that both anti-RelA NLS and anti-RelA C-term delivery reduced RelA nuclear translocation to 48.0 $\pm$ 0.8% and 60.1 $\pm$ 5.9%, respectively (Fig. 4.6b-c, Supplementary Fig. 4.14). Next, by using a NFκB-driven luciferase reporter plasmid, we showed that delivery of both anti-RelA NLS and anti-RelA C-term reduced NFκB transcriptional activity to 52.4 $\pm$ 1.1% and 68.3 $\pm$ 2.6% of that of normal cells, respectively (100 ng/mL TNFα) (Fig. 4.6d). This degree of inhibition is excellent considering the delivery efficiencies with 150 nM anti-RelA NLS and anti-RelA C-term were 70.9 $\pm$ 0.8% for and 29.7 $\pm$ 4.5%, respectively (Supplementary Fig. 4.15). Because anti-RelA C-term IgG delivery was capable of inhibiting NFκB, antibody binding to non-NLS epitopes may be sufficient to sterically block nuclear translocation of many target proteins. We anticipate that cytosolic IgG-dependent cytoplasmic sequestration can be easily adopted to inhibit other transcription factors or nuclear proteins.

4.4 Discussion

Efficient cytosolic delivery of proteins has long been sought after not only for its profound therapeutic implications, but also for expanding the toolbox that one can use for perturbing biological systems. By leveraging a modular antibody functionalization technology[22], we were able to easily append ApPs onto off-the-shelf IgGs to enable their complexation with a variety of commercially available cationic lipids and efficiently deliver them into the cytosol of cells. Cytosolically delivered IgGs remain functional and are capable of inhibiting diverse proteins such as MRP1 and NFκB.

When compared to CPP-mediated delivery of small peptides or proteins, our approach enables cytosolic delivery of a much larger IgG cargo with similar or greater efficiencies at a

~100-fold lower concentration[26-29]. It is difficult to directly compare our technology to previous carrier-mediated approaches due to the use of different reporter systems. We note, however, that previous studies have relied on either a reporter that only detects total cellular uptake[19] or by delivering proteins capable of greatly amplifying their signal, such as Cre-recombinase or enzymes[16,18,20-21]. In contrast, our functional assays – MRP1 and NFκB inhibition – are more stringent and require delivery of close to stoichiometric amounts of IgG relative to their target, which is more representative of most protein inhibition assays.

Intrabodies have long been used to modulate cellular biology, ranging from simply inhibiting target proteins[7-8,10,37-38] to marking target proteins for degradation[39-40]. However, those approaches require either expertise in *in vitro* display technologies or significant genetic modifications of target proteins and cell lines, which can be major barriers for easy adaptation by researchers. Because IgG antibodies have been invaluable research tools for decades, there exists large and well-validated antibody collections[41-42] that, using our modular IgG functionalization and cytosolic delivery approach, can now be repurposed for perturbing the activity of intracellular proteins. Although IgG binding is not guaranteed to inhibit all target proteins, we believe that the large size of IgGs is sufficient to sterically block many biological interactions. Finally, since IgG antibody generation is often one of the initial steps when studying poorly characterized proteins, our technology may be invaluable for validating biological hypotheses in cases where no tools currently exist[38].

Our approach does have two main limitations. First, cytosolic IgGs have been reported to be capable of engaging the TRIM21 E3 ubiquitin ligase to degrade their target proteins[9], but pAbBD photocrosslinking sterically blocks the TRIM21 binding[22,43-44] site and prevents our cytosolically delivered IgGs from harnessing any endogenous protein degradation machinery. This could be remedied by using alternative antibody-binding domains that can be photocrosslinked to IgGs outside of the Fc region[45]. Second, in this study we complexed IgGs with commercially available cationic lipids. Although the resulting complexes perform excellently

against cultured cells, their poor pharmacokinetics render them unsuitable for *in vivo* studies. Our electrostatics-based complexation strategy, though, should be compatible with most *in vivo* delivery formulations, such as those containing ionizable and poly(ethylene glycol)-conjugated lipids[46].

In summary, we have designed and rigorously validated a modular cytosolic IgG delivery platform that allows previously developed IgG collections to be now used as intracellular tools. Future studies may also focus on using our complexation strategy to identify optimal *in vivo* IgG or protein delivery formulations, which could ultimately lead to cytosolic protein delivery as a new therapeutic modality.

4.5 Supplementary Information

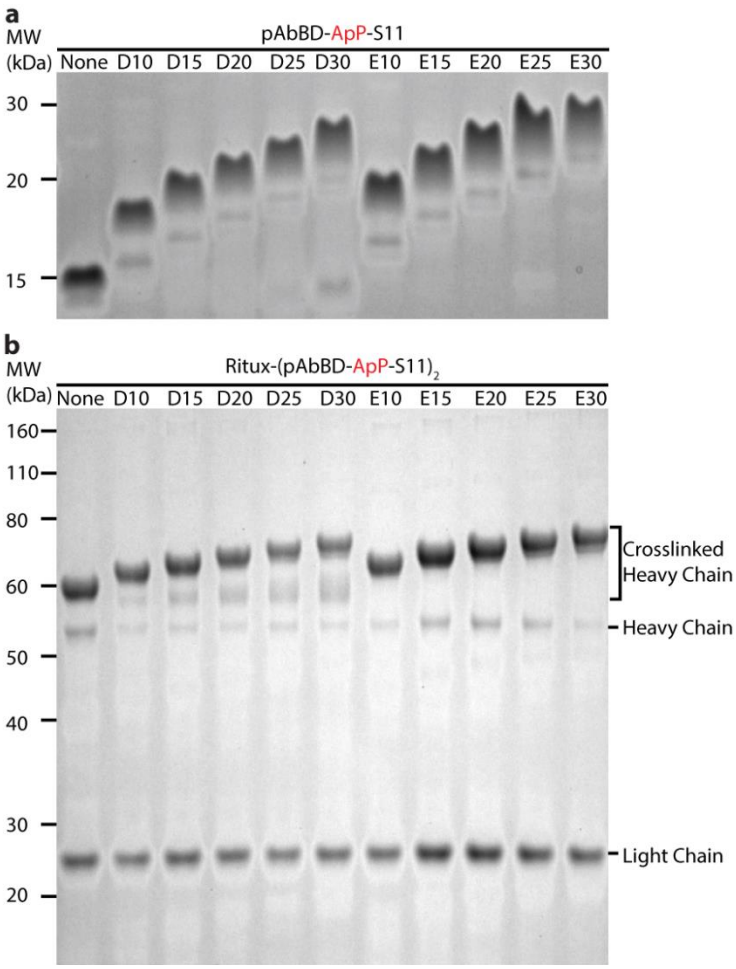

Figure 4.7 SDS-PAGE of pAbBD-ApP-S11 and Ritux-(pAbBD-ApP-S11)$_2$.

a, b) SDS-PAGE of pAbBD-ApP-S11 **(a)** or Ritux-(pAbBD-ApP-S11)$_2$ **(b)** with no ApP or 10, 15, 20, 25, or 30 residues-long poly-aspartate or poly-glutamate ApPs.

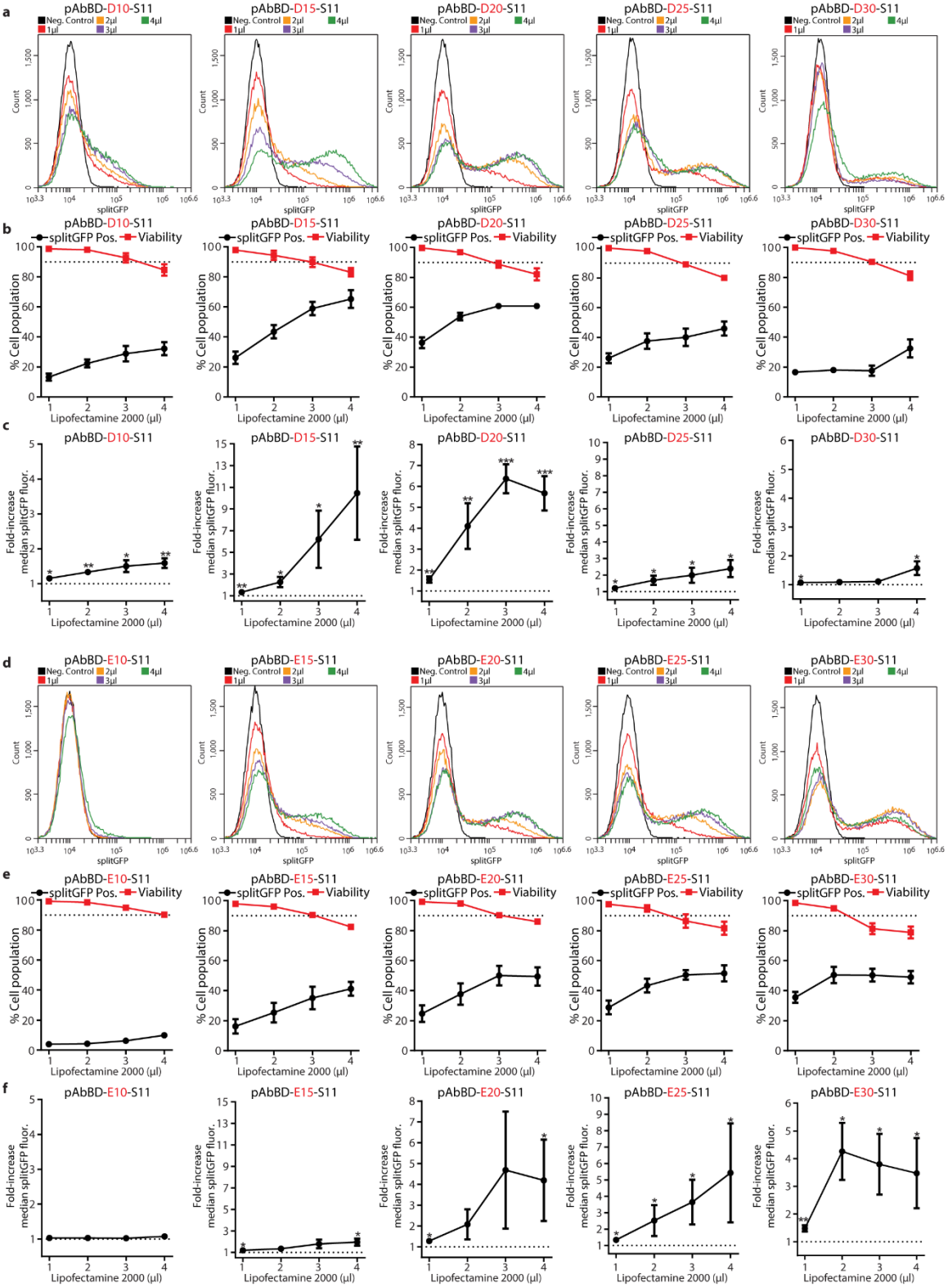
a
pAbBD-D10-S11
pAbBD-D15-S11
pAbBD-D20-S11
pAbBD-D25-S11
pAbBD-D30-S11
Neg. Control
1µl
2µl
3µl
4µl
Count
splitGFP
b
splitGFP Pos.
Viability
% Cell population
Lipofectamine 2000 (µl)
c
Fold-increase median splitGFP fluor.
d
pAbBD-E10-S11
pAbBD-E15-S11
pAbBD-E20-S11
pAbBD-E25-S11
pAbBD-E30-S11
e
f

Figure 4.8 Detailed flow cytometry data for optimizing ApPs for pAbBD delivery.

Related to Figure 4.2. **a-f)** 500 nM pAbBD-ApP-S11 with no ApP (negative control), poly-aspartate **(a-c)**, or poly-glutamate **(d-f)** ApPs 10, 15, 20, 25, or 30 residues-long were complexed with varying amounts of Lipo 2000 (1 to 4 µl) and added to HEK293T splitGFP(1-10) cells for 6 h. Representative flow cytometry histograms of splitGFP fluorescence are shown in **(a)** and **(d)**. Flow cytometry data were quantified as the percent of cells splitGFP-positive **(b, e)** and the fold-increase in median splitGFP fluorescence over negative control **(c, f)**. The dotted line indicates either 90% of the cell population **(b, e)** or no increase in fluorescence **(c, f)**. Viability was determined with the LDH assay. Data are mean $\pm$ s.e.m., n=4, *p<0.05 **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

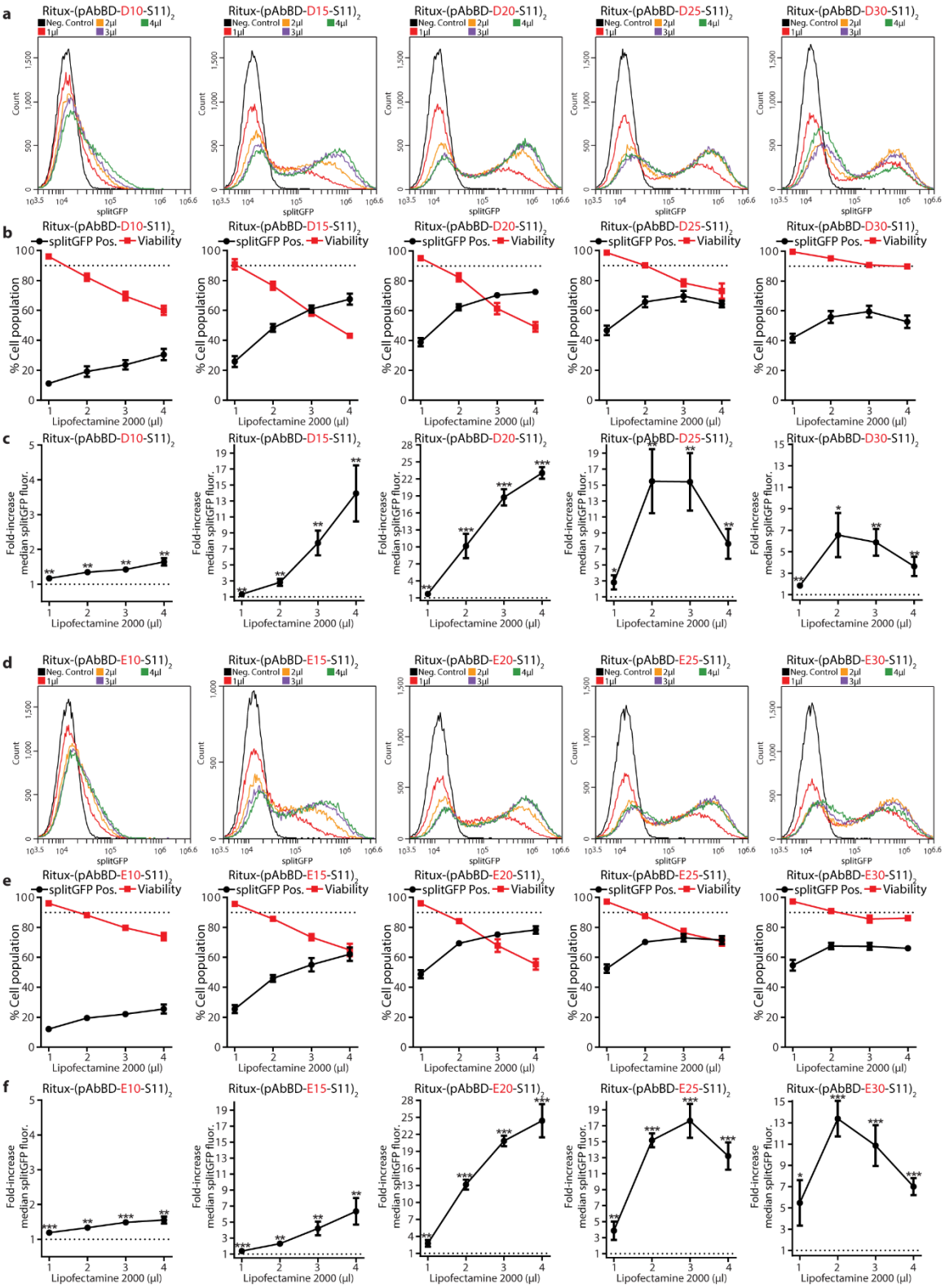
a
Ritux-(pAbBD-D10-S11)2
Ritux-(pAbBD-D15-S11)2
Ritux-(pAbBD-D20-S11)2
Ritux-(pAbBD-D25-S11)2
Ritux-(pAbBD-D30-S11)2
Neg. Control
1µl
2µl
3µl
4µl
Count
splitGFP
b
splitGFP Pos.
Viability
% Cell population
Lipofectamine 2000 (µl)
c
Fold-increase median splitGFP fluor.
d
Ritux-(pAbBD-E10-S11)2
Ritux-(pAbBD-E15-S11)2
Ritux-(pAbBD-E20-S11)2
Ritux-(pAbBD-E25-S11)2
Ritux-(pAbBD-E30-S11)2
e
f

Figure 4.9 Detailed flow cytometry data for optimizing ApPs for Ritux-(pAbBD)$_2$ delivery with Lipo 2000.

Related to Figure 4.3. **a-f)** 500 nM Ritux-(pAbBD-ApP-S11)$_2$ with no ApP (negative control), poly-aspartate **(a-c)**, or poly-glutamate **(d-f)** ApPs 10, 15, 20, 25, or 30 residues-long were complexed with varying amounts of Lipo 2000 (1 to 4 µl) and added to HEK293T splitGFP(1-10) cells for 6 h. Representative flow cytometry histograms of splitGFP fluorescence are shown in **(a)** and **(d)**. Flow cytometry data were quantified as the percent of cells splitGFP-positive **(b, e)** and the fold-increase in median splitGFP fluorescence over negative control **(c, f)**. The dotted line indicates either 90% of the cell population **(b, e)** or no increase in fluorescence **(c, f)**. Viability was determined with the LDH assay. Data are mean $\pm$ s.e.m., n=4, *p<0.05 **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

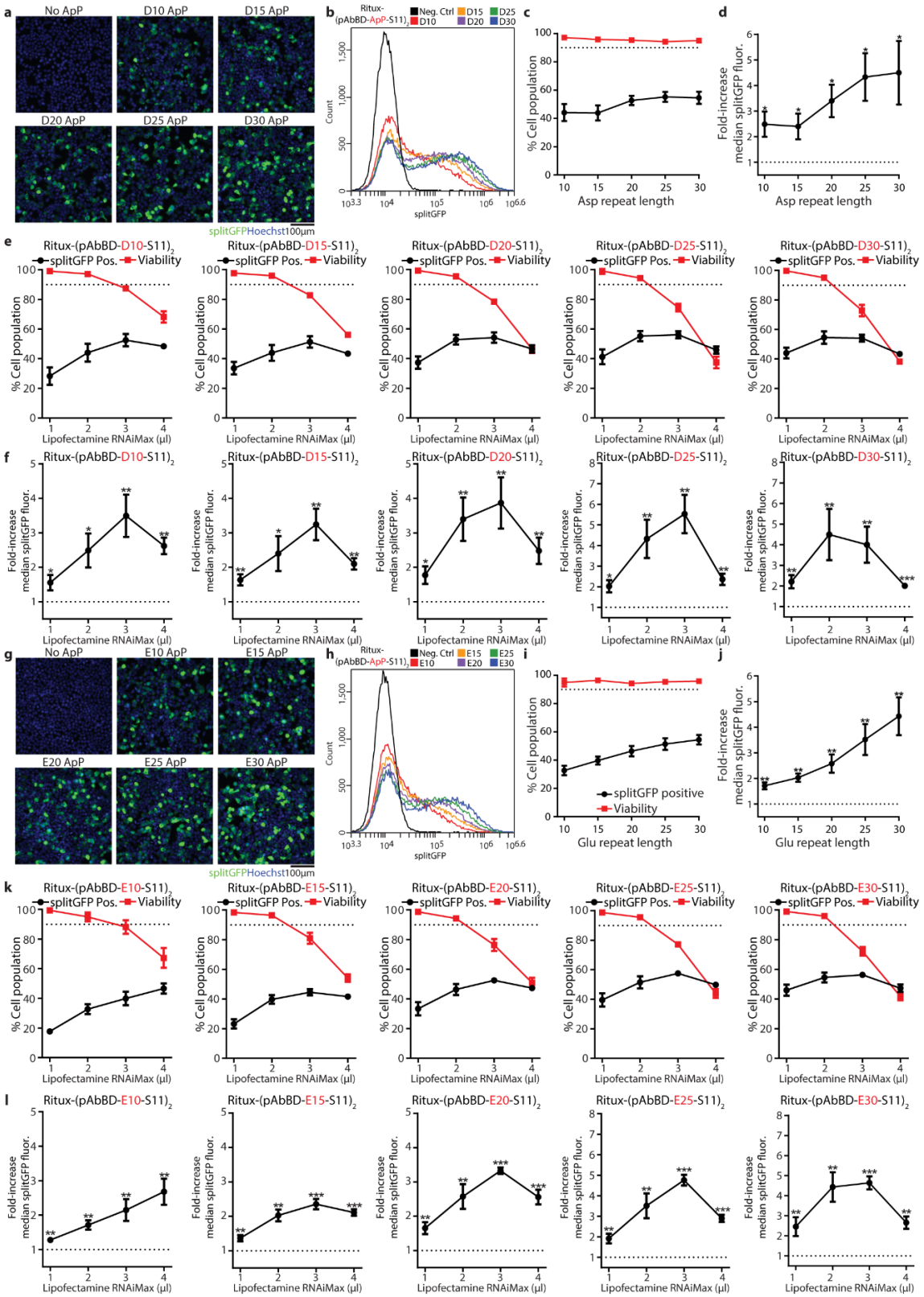
a
No ApP
D10 ApP
D15 ApP
D20 ApP
D25 ApP
D30 ApP
splitGFPHoechst100µm
b
Ritux-(pAbBD-ApP-S11)2
Neg. Ctrl
D10
D15
D20
D25
D30
Count
splitGFP
c
% Cell population
Asp repeat length
d
Fold-increase median splitGFP fluor.
Asp repeat length
e
Ritux-(pAbBD-D10-S11)2
Ritux-(pAbBD-D15-S11)2
Ritux-(pAbBD-D20-S11)2
Ritux-(pAbBD-D25-S11)2
Ritux-(pAbBD-D30-S11)2
splitGFP Pos.
Viability
% Cell population
Lipofectamine RNAiMax (µl)
f
Fold-increase median splitGFP fluor.
Lipofectamine RNAiMax (µl)
g
No ApP
E10 ApP
E15 ApP
E20 ApP
E25 ApP
E30 ApP
splitGFPHoechst100µm
h
Ritux-(pAbBD-ApP-S11)2
Neg. Ctrl
E10
E15
E20
E25
E30
Count
splitGFP
i
% Cell population
splitGFP positive
Viability
Glu repeat length
j
Fold-increase median splitGFP fluor.
Glu repeat length
k
Ritux-(pAbBD-E10-S11)2
Ritux-(pAbBD-E15-S11)2
Ritux-(pAbBD-E20-S11)2
Ritux-(pAbBD-E25-S11)2
Ritux-(pAbBD-E30-S11)2
splitGFP Pos.
Viability
% Cell population
Lipofectamine RNAiMax (µl)
l
Fold-increase median splitGFP fluor.
Lipofectamine RNAiMax (µl)

Figure 4.10 Optimizing ApPs for cytosolic IgG delivery with Lipo RNAiMax.

a-l) 500 nM Ritux-(pAbBD-ApP-S11)$_2$ with no ApP (negative control), poly-aspartate **(a-f)**, and poly-glutamate **(g-l)** ApPs 10, 15, 20, 25, or 30 residues-long were complexed with Lipo RNAiMax and added to HEK293T splitGFP(1-10) cells for 6 h. Representative live-cell fluorescence microscopy **(a, g)** and flow cytometry histograms of splitGFP fluorescence **(b, h)** are shown for when Ritux-(pAbBD-ApP-S11)$_2$ was complexed with 2 µl Lipo RNAiMax. Flow cytometry data were quantified at either a fixed 2 µl Lipo RNAiMax **(c, d, i, j)** or when Lipo RNAiMax varied from 1 to 4 µl **(e, f, k, l)**. Flow cytometry data were quantified as the percent of cells splitGFP-positive **(c, e, i, k)** and the fold-increase in median splitGFP fluorescence over negative control **(d, f, j, l)**. The dotted line indicates either 90% of the cell population **(c, e, i, k)** or no increase in fluorescence **(d, f, j, l)**. Viability was determined with the LDH assay. Data are mean $\pm$ s.e.m., n=4, *p<0.05 **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

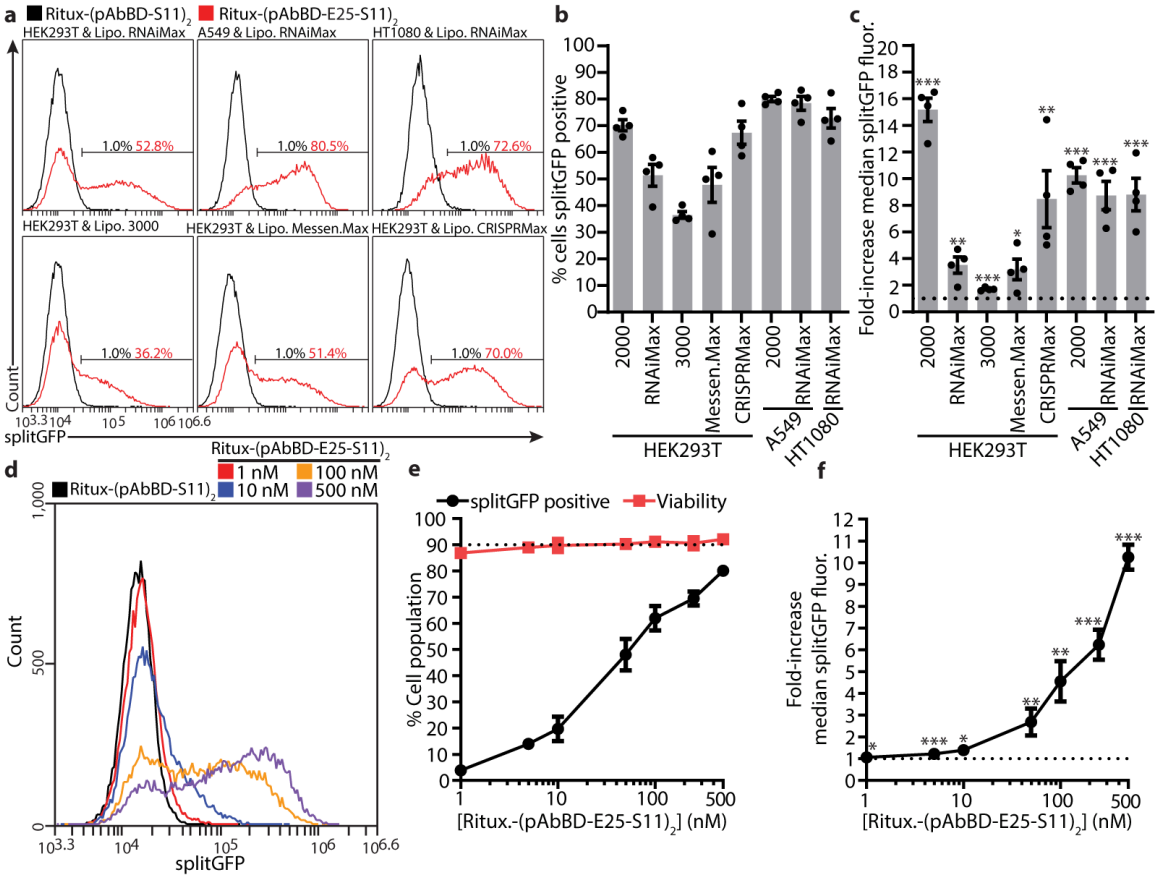

Figure 4.11 IgG delivery scope with a 25 residues-long poly-glutamate ApP.

Related to Fig 4.4. **a-c)** 500 nM Ritux-(pAbBD-S11)$_2$ (negative control) or Ritux-(pAbBD-E25-S11)$_2$ was complexed with 2 µl of the indicated cationic lipid and added to the indicated splitGFP(1-10) reporter cells for 6 h. Representative flow cytometry histograms of splitGFP fluorescence are shown in **(a)**. Flow cytometry data were quantified as the percent of cells splitGFP-positive **(b)** and the fold-increase in median splitGFP fluorescence over negative control **(c). d-f)** Same as for **a-c**, but indicated concentrations of Ritux-(pAbBD-E25-S11)$_2$ was complexed with 2 µl Lipo 2000 and added to A549 splitGFP(1-10) cells. The dotted line indicates either no increase in fluorescence **(c, f)** or 90% of the cell population **(e)**. Viability was determined with the LDH assay. Data are means ± s.e.m., n=4, *p<0.05 **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

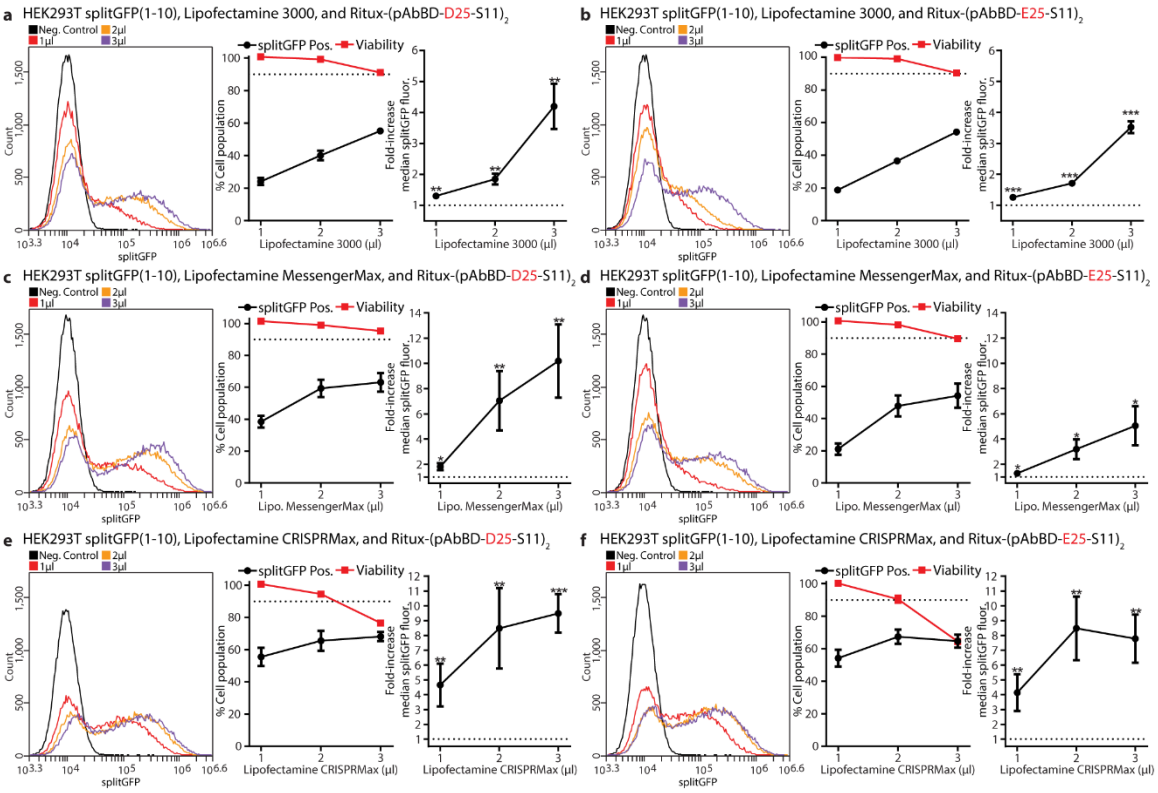

Figure 4.12 Detailed flow cytometry data for cytosolic IgG delivery with additional cationic lipids in HEK293T splitGFP(1-10) cells.

Related to Fig 4.4 and Supplementary Fig. 4.11. **a-f)** 500 nM Ritux-(pAbBD-D25-S11)$_2$ **(a, c, e)**, Ritux-(pAbBD-E25-S11)$_2$ **(b, d, f)**, or Ritux-(pAbBD-S11)$_2$ (negative control) was complexed with 1 to 3 µl of Lipo 3000 **(a, b)**, Lipo MessengerMax **(c, d)**, or Lipo CRISPRMax **(e, f)** and added to HEK293T splitGFP(1-10) cells for 6 h. splitGFP fluorescence was determined by flow cytometry. A representative histogram of splitGFP fluorescence is shown in the left panel. Flow cytometry data were quantified as the percent of cells splitGFP-positive (middle panel) and the fold-increase in median splitGFP fluorescence over negative control (right panel). The dotted line indicates either 90% of the cell population (middle panel) or no increase in fluorescence (right panel). Viability was determined with the LDH assay. Data are means ± s.e.m., n=4, *p<0.05 **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

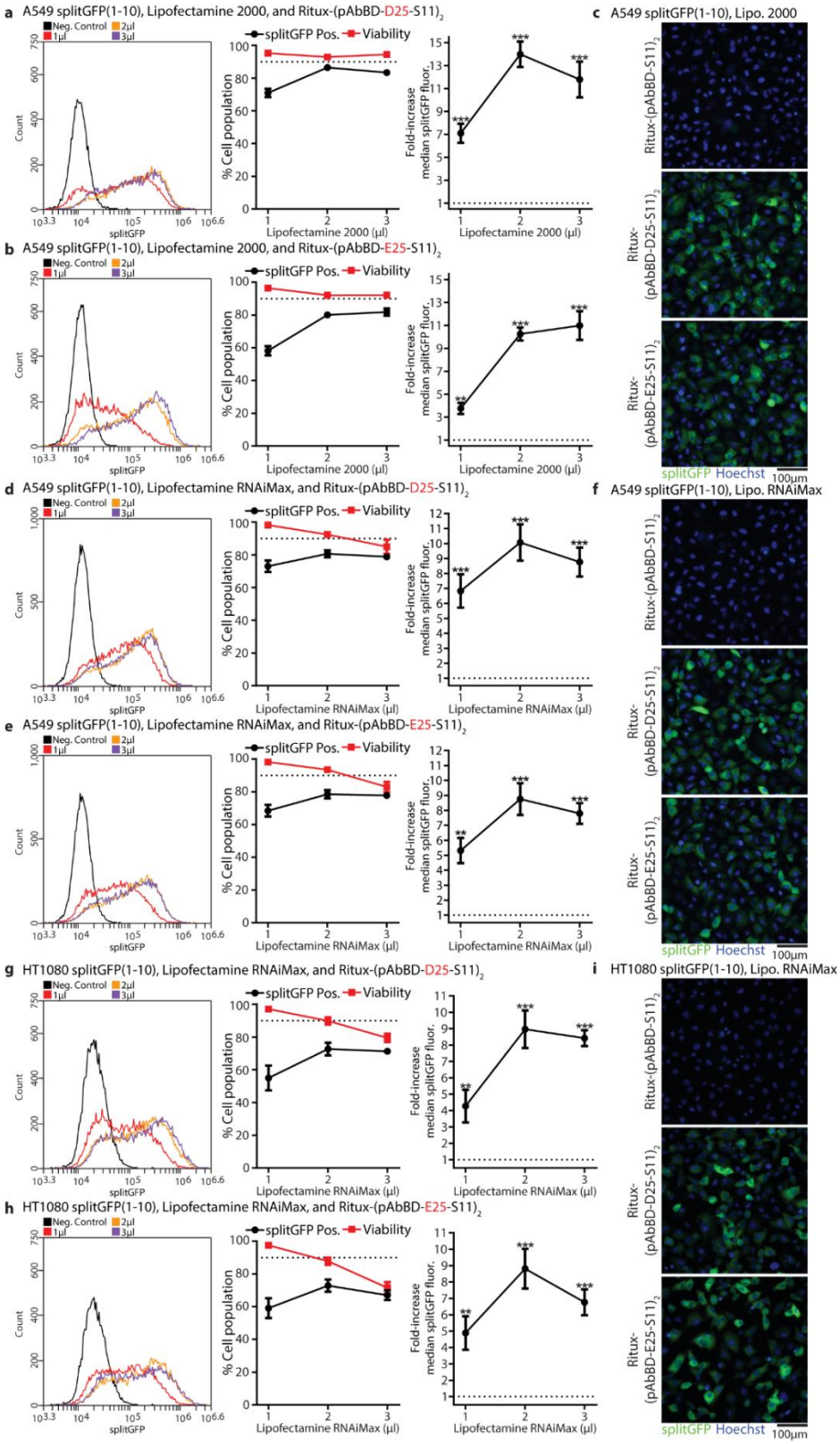
a A549 splitGFP(1-10), Lipofectamine 2000, and Ritux-(pAbBD-D25-S11)2
b A549 splitGFP(1-10), Lipofectamine 2000, and Ritux-(pAbBD-E25-S11)2
c A549 splitGFP(1-10), Lipo. 2000
d A549 splitGFP(1-10), Lipofectamine RNAiMax, and Ritux-(pAbBD-D25-S11)2
e A549 splitGFP(1-10), Lipofectamine RNAiMax, and Ritux-(pAbBD-E25-S11)2
f A549 splitGFP(1-10), Lipo. RNAiMax
g HT1080 splitGFP(1-10), Lipofectamine RNAiMax, and Ritux-(pAbBD-D25-S11)2
h HT1080 splitGFP(1-10), Lipofectamine RNAiMax, and Ritux-(pAbBD-E25-S11)2
i HT1080 splitGFP(1-10), Lipo. RNAiMax
Neg. Control
1µl
2µl
3µl
splitGFP Pos.
Viability
Count
splitGFP
% Cell population
Fold-increase median splitGFP fluor.
Lipofectamine 2000 (µl)
Lipofectamine RNAiMax (µl)
Ritux-(pAbBD-S11)2
Ritux-(pAbBD-D25-S11)2
Ritux-(pAbBD-E25-S11)2
splitGFP Hoechst
100µm

Figure 4.13 Detailed flow cytometry data for cytosolic IgG delivery with additional cationic lipids in A549 and HT1080 splitGFP(1-10) cells.

Related to Fig 4.4 and Supplementary Fig. 4.11. **a-c)** 500 nM Ritux-(pAbBD-D25-S11)$_2$ **(a, c)**, Ritux-(pAbBD-E25-S11)$_2$ **(b, c)**, or Ritux-(pAbBD-S11)$_2$ (negative control) was complexed with 1 to 3 µl of Lipo 2000 and added to A549 splitGFP(1-10) cells for 6 h. Afterwards, splitGFP fluorescence was determined by flow cytometry **(a, b)** or live-cell fluorescence microscopy **(c)**. For flow cytometry, the left panel shows a representative histogram of splitGFP fluorescence. Flow cytometry data were quantified as the percent of cells splitGFP-positive (middle panel) and the fold-increase in median splitGFP fluorescence over negative control (right panel). The dotted line indicates either 90% of the cell population (middle panel) or no increase in fluorescence (right panel). Viability was determined with the LDH assay. **d-f)** Same as for **a-c**, but with Lipo RNAiMax. **g-i)** Same as for **a-c**, but with Lipo RNAiMax and HT1080 splitGFP(1-10) cells. Data are means $\pm$ s.e.m., n=4, **p<0.01 ***p<0.001 (one-sided one sample *t*-test of log-ratios).

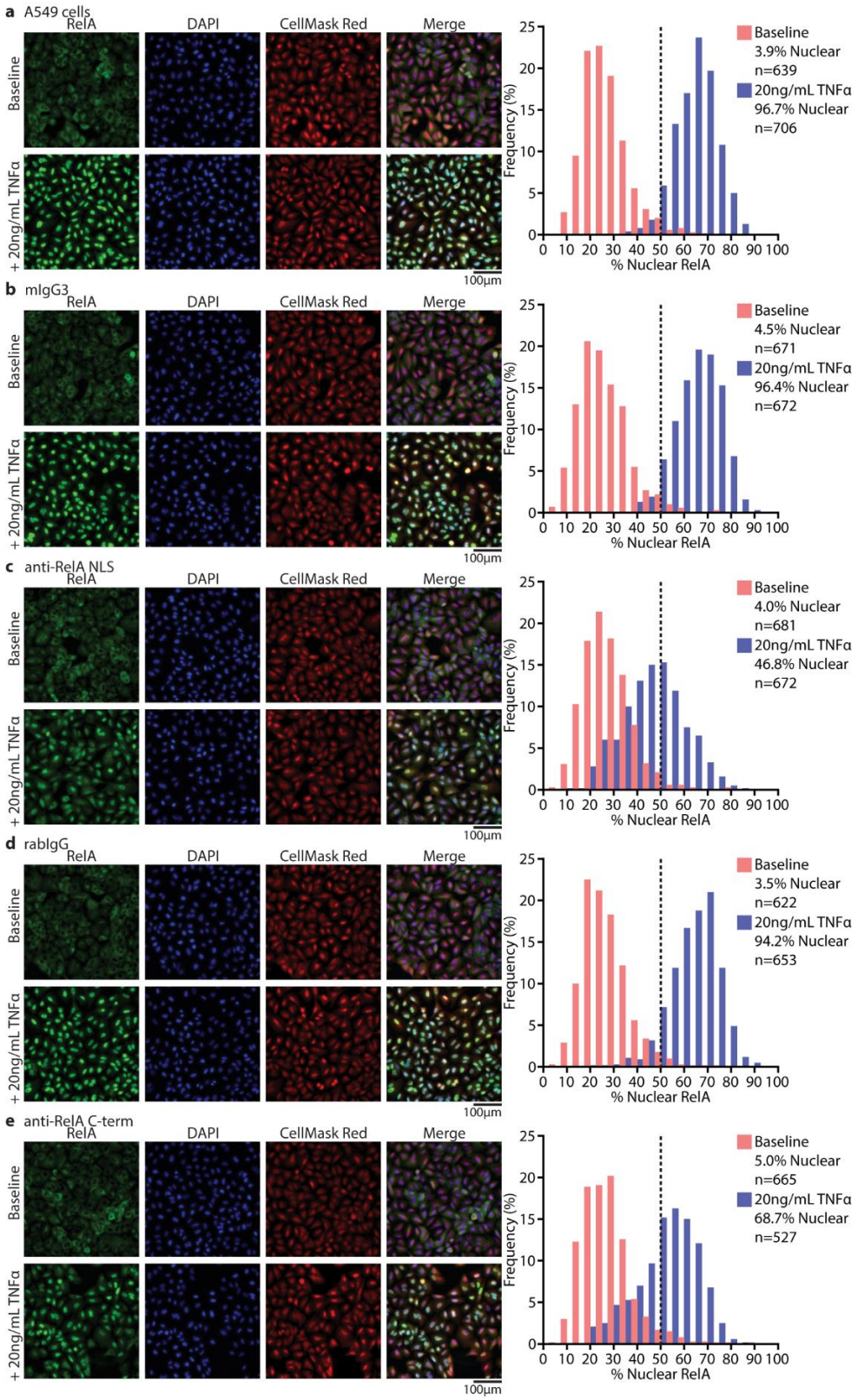
a A549 cells
RelA
DAPI
CellMask Red
Merge
Baseline
+ 20ng/mL TNFα
100µm
Frequency (%)
% Nuclear RelA
Baseline
3.9% Nuclear
n=639
20ng/mL TNFα
96.7% Nuclear
n=706
b mIgG3
Baseline
4.5% Nuclear
n=671
20ng/mL TNFα
96.4% Nuclear
n=672
c anti-RelA NLS
Baseline
4.0% Nuclear
n=681
20ng/mL TNFα
46.8% Nuclear
n=672
d rabIgG
Baseline
3.5% Nuclear
n=622
20ng/mL TNFα
94.2% Nuclear
n=653
e anti-RelA C-term
Baseline
5.0% Nuclear
n=665
20ng/mL TNFα
68.7% Nuclear
n=527

Figure 4.14 RelA immunofluorescence quantification.

Related to Figure 4.5. **a-e)** Representative immunofluorescence images of A549 cells with or without TNFα stimulation are shown without protein delivery **(a)** or with 150 nM mIgG3-(pAbBD-D25-S11)$_2$ (anti-RelA NLS isotype control) **(b)**, anti-RelA NLS IgG-(pAbBD-D25-S11)$_2$ **(c)**, rabIgG-(pAbBD-D25-S11)$_2$ (anti-RelA C-term isotype control) **(d)**, or anti-RelA C-term IgG-(pAbBD-D25-S11)$_2$ **(e)** delivered with 2 µl Lipo RNAiMax. CellProfiler was used for automated image analysis. The DAPI channel was used for nuclear segmentation whereas the CellMask Red channel was used for cellular segmentation. For each cell, the nuclear RelA fluorescence intensity was normalized to the cellular RelA fluorescence intensity. At least 5 image sets were taken for each biological replicate. Histograms of normalized nuclear RelA fluorescence are shown for one biological replicate for each delivery condition. 50% nuclear RelA was used as a cutoff for denoting a cell as having nuclear RelA.

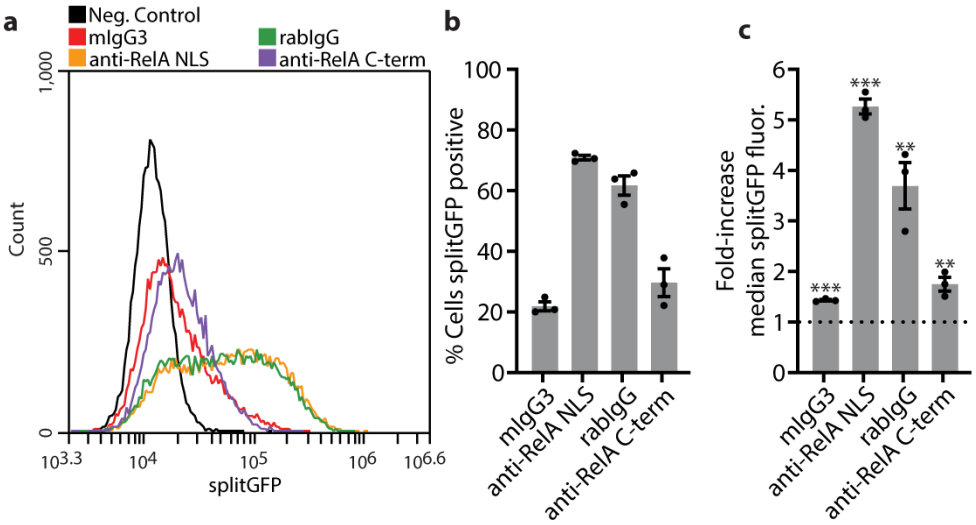

Figure 4.15 Cytosolic delivery efficiencies for NFκB experiment IgGs.

Related to Fig 4.6. **a-c)** 150 nM mIgG3-(pAbBD-D25-S11)$_2$, anti-RelA NLS IgG-(pAbBD-D25-S11)$_2$, rabIgG-(pAbBD-D25-S11)$_2$, anti-RelA C-term IgG-(pAbBD-D25-S11)$_2$, or Ritux-(pAbBD-S11)$_2$ (negative control) was complexed with 2 µl Lipo RNAiMax and added to A549 splitGFP(1-10) cells for 6 h. A representative flow cytometry histogram of splitGFP fluorescence is shown in **a**. Flow cytometry was quantified as the percent of splitGFP-positive cells **(b)** and the fold-increase in median splitGFP fluorescence over negative control **(c)**. Delivery efficiency is greater with anti-RelA NLS than its isotype control (mIgG3), but lower with anti-RelA C-term than its isotype control (rabIgG). The dotted line indicates no increase in fluorescence. Data are means ± SEM, n=3, **p<0.01, ***p<0.001 (one-sided one sample *t*-test of log-ratios).

Table 4.1 List of purified proteins used for cytosolic delivery of IgGs.

For pAbBD amino acid sequences, the unnatural amino acid benzoylphenylalanine is denoted as **B**. ApP sequences are colored red. Linker sequences are colored brown. The S11 splitGFP reporter peptide is colored green. Sequences used for PBSL that are cleaved off during pAbBD purification are colored purple.

Plasmid Name	Protein Sequence	DNA Sequence (Open Reading Frame)
pRSET SpyCatcher-SrtA-His12	MSGSGDSATHIKFSK RDEDGKELAGATMEL RDSSGKTISTWISDG QVKDFYLYPGKYTFV ETAAPDGYEVATAITF TVNEQGQVTVNGGG SGGSSGGGSGGSGG SGGSGGSQAKPQIPK DKSKVAGYIEIPDADIK EPVYPGPATPEQLNR GVSFAEENESLDDQN ISIAGHTFIDRPNYQFT NLKAAKKGSMVYFKV GNETRKYKMTSIRDV KPTDVEVLDEQKGKD KQLTLITCDDYNEKTG VWEKRKIFVATEVKH HHHHHHHHHHH	ATGTCTGGTTCTGGAGACAGCGCCACTCACATTAAATTCAGTAAACG CGATGAAGACGGCAAGGAACTCGCGGGTGCAACTATGGAACTGCGG GACTCTTCTGGTAAAACGATTAGTACCTGGATTTCCGATGGACAGGT TAAAGATTTTTATTTGTACCCAGGTAAATATACGTTTGTAGAAACCGC CGCGCCAGATGGCTACGAAGTGGCTACGGCCATTACCTTCACGGTT AATGAACAAGGCCAAGTCACGGTCAATGGCGGTGGTAGTGGTGGGA GCAGTGGCGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTG GCGGTAGCCAAGCTAAACCTCAAATTCCGAAAGATAAATCAAAAGTG GCAGGCTATATTGAAATTCCAGATGCTGATATTAAAGAACCAGTATAT CCAGGACCAGCAACACCTGAACAATTAAATAGAGGTGTAAGCTTTGC AGAAGAAAATGAATCACTAGATGATCAAAATATTTCAATTGCAGGACA CACTTTCATTGACCGTCCGAACTATCAATTTACAAATCTTAAAGCAGC CAAAAAAGGTAGTATGGTGTACTTTAAAGTTGGTAATGAAACACGTAA GTATAAAATGACAAGTATAAGAGATGTTAAGCCAACAGATGTAGAAGT TCTAGATGAACAAAAAGGTAAAGATAAACAATTAACATTAATTACTTGT GATGATTACAATGAAAAGACAGGCGTTTGGGAAAAACGTAAAATCTTT GTAGCTACAGAAGTCAAACATCACCACCATCATCACCACCATCATCAT CACCACTGA
pRSET His6-splitGFP(1-10)	MHHHHHHSKGEELFT GVVPILVELDGDVNG HKFSVRGEGEGDATI GKLTLKFICTTGKLPV PWPTLVTTLTYGVQC FSRYPDHMKRHDFFK SAMPEGYVQERTISF KDDGKYKTRAVVKFE GDTLVNRIELKGTDFK EDGNILGHKLEYNFN SHNVYITADKQKNGIK ANFTVRHNVEDGSVQ LADHYQQNTPIGDGP VLLPDNHYLSTQTVLS KDPNEK	ATGCACCATCATCACCATCACTCTAAAGGTGAAGAACTGTTCACCGG TGTTGTTCCGATCCTGGTTGAACTGGACGGTGACGTTAACGGTCACA AATTCTCTGTTCGTGGTGAAGGTGAAGGTGACGCGACCATCGGTAAA CTGACCCTGAAATTCATCTGCACCACCGGTAAACTGCCGGTTCCGTG GCCGACCCTGGTTACCACCCTGACCTACGGTGTTCAGTGCTTCTCTC GTTACCCGGACCACATGAAACGTCACGACTTCTTCAAATCTGCTATG CCGGAAGGTTACGTTCAGGAACGTACCATCTCTTTCAAAGACGACGG TAAATACAAAACCCGTGCTGTTGTTAAATTCGAAGGTGACACCCTGGT TAACCGTATCGAACTGAAAGGTACCGACTTCAAAGAAGACGGTAACA TCCTGGGTCACAAACTGGAATACAACTTCAACTCTCACAACGTTTACA TCACCGCTGACAAACAGAAAAACGGTATCAAAGCTAACTTCACCGTT CGTCACAACGTTGAAGACGGTTCTGTTCAGCTGGCTGACCACTACCA GCAGAACACCCCGATCGGTGACGGTCCGGTTCTGCTGCCGGACAAC CACTACCTGTCTACCCAGACCGTTCTGTCTAAAGACCCGAACGAAAA ATAA
pRSET pAbBD-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSG GSSGGGGSGGRDHM VLHEYVNAAGITSGS GGSLPETGGGSGGS GGSGGSGGSAHIVM VDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGA TCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTT GACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTA CGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCA CCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGGTGGCT CCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACA CGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTC TGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGG CAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGA CGAAATAA

pRSET pAbBD-D10-S11	MAGGSSGGSGGTFKLIINGKTLKGEITIEAVDA**B**EAEKIFKQYANDYGIDGEWTYDDATKTFTVTEGGSGGGGSSDDDDDDDDDSSGGGGSGGRDHMVLHEYVNAAGITSGSGGSLPETGGGSGGSGGSGGGGGSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTACGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGATGACGATGATGACGACGATGACGACGATTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAATAA
pRSET pAbBD-D15-S11	MAGGSSGGSGGTFKLIINGKTLKGEITIEAVDA**B**EAEKIFKQYANDYGIDGEWTYDDATKTFTVTEGGSGGGGSSDDDDDDDDDDDDDDDSSGGGGSGGRDHMVLHEYVNAAGITSGSGGSLPETGGGSGGSGGSGGGSGGSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTACGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGATGATGATGACGACGATGACGATGATGACGACGATGACGACGATTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAATAA
pRSET pAbBD-D20-S11	MAGGSSGGSGGTFKLIINGKTLKGEITIEAVDA**B**EAEKIFKQYANDYGIDGEWTYDDATKTFTVTEGGSGGGGSSDDDDDDDDDDDDDDDDDDDDSSGGGGSGGRDHMVLHEYVNAAGITSGSGGSLPETGGGSGGSGGSGGGSGGSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTACGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGACGACGACGACGACGATGATGATGACGACGATGACGATGATGACGACGATGACGACGATTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAATAA
pRSET pAbBD-D25-S11	MAGGSSGGSGGTFKLIINGKTLKGEITIEAVDA**B**EAEKIFKQYANDYGIDGEWTYDDATKTFTVTEGGSGGGGSSDDDDDDDDDDDDDDDDDDDDDDDDDSSGGGGSGGRDHMVLHEYVNAAGITSGSGGSLPETGGGSGGSGGSGGGSGGSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTACGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGATGATGATGATGATGACGACGACGACGACGATGATGATGACGACGATGACGATGACGACGATTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAATAA
pRSET pAbBD-D30-S11	MAGGSSGGSGGTFKLIINGKTLKGEITIEAVDA**B**EAEKIFKQYANDYGIDGEWTYDDATKTFTVTEGGSGGGGSSDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDSSGGGGSGGRDHMVLHEYVNAAGITSGSGGSLPETGGGSGGSGGSGGGSGGSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGATCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTTGACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTACGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCACCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGATGATGATGATGACGATGATGATGATGACGACGACGACGACGATGATGATGACGACGATGACGATGATGACGACGATGACGACGATTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAATAA

pRSET pAbBD-E10-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSE EEEEEEEEESSGGGG SGGRDHMVLHEYVN AAGITSGSGGSLPET GGGSGGSGGSGGSG GSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGA TCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTT GACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTA CGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCA CCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGAAGAGG AGGAGGAAGAAGAAGAGGAGGAGTCCTCAGGGGGTGGCGGGAGCG GCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGG TATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTA GTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATAT CGTTATGGTCGATGCTTACAAACCGACGAAATAA
pRSET pAbBD-E15-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSE EEEEEEEEEEEEEES SGGGGSGGRDHMVL HEYVNAAGITSGSGG SLPETGGGSGGSGG SGGSGGSAHIVMVDA YKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGA TCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTT GACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTA CGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCA CCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGAGGAGG AGGAGGAGGAAGAGGAGGAGGAAGAAGAAGAGGAGGAGTCCTCAG GGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATA CGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCG GAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTG GCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAA TAA
pRSET pAbBD-E20-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSE EEEEEEEEEEEEEEE EEEESSGGGGSGGR DHMVLHEYVNAAGIT SGSGGSLPETGGGS GGSGGSGGSGGSAH IVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGA TCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTT GACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTA CGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCA CCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGAAGAAG AAGAAGAAGAGGAGGAGGAGGAGGAAGAGGAGGAGGAAGAAGAAG AGGAGGAGTCCTCAGGGGGTGGCGGGAGCGGCGGTCGTGATCACA TGGTGTTACACGAATACGTCAACGCGGCAGGTATTACTAGTGGTTCG GGTGGTTCTCTGCCGGAAACCGGTGGTGGTAGTGGTGGCTCTGGCG GTTCTGGTGGCAGTGGCGGTAGCGCGCATATCGTTATGGTCGATGC TTACAAACCGACGAAATAA
pRSET pAbBD-E25-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSE EEEEEEEEEEEEEEE EEEEEEEEESSGGGG SGGRDHMVLHEYVN AAGITSGSGGSLPET GGGSGGSGGSGGSG GSAHIVMVDAYKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGA TCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTT GACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTA CGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCA CCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGAAGAAG AGGAAGAGGAAGAAGAAGAAGAAGAGGAGGAGGAGGAGGAAGAGG AGGAGGAAGAAGAAGAGGAGGAGTCCTCAGGGGGTGGCGGGAGCG GCGGTCGTGATCACATGGTGTTACACGAATACGTCAACGCGGCAGG TATTACTAGTGGTTCGGGTGGTTCTCTGCCGGAAACCGGTGGTGGTA GTGGTGGCTCTGGCGGTTCTGGTGGCAGTGGCGGTAGCGCGCATAT CGTTATGGTCGATGCTTACAAACCGACGAAATAA
pRSET pAbBD-E30-S11	MAGGSSGGSGGTFK LIINGKTLKGEITIEAVD A**B**EAEKIFKQYANDY GIDGEWTYDDATKTF TVTEGGSGGGGSSE EEEEEEEEEEEEEEE EEEEEEEEEEEEEES SGGGGSGGRDHMVL HEYVNAAGITSGSGG SLPETGGGSGGSGG SGGSGGSAHIVMVDA YKPTK	ATGGCGGGTGGGAGCTCTGGCGGTTCCGGCGGTACCTTCAAACTGA TCATCAACGGTAAAACCCTGAAAGGTGAAATCACCATCGAAGCTGTT GACGCTTAGGAAGCTGAAAAAATCTTCAAACAGTACGCTAACGACTA CGGTATCGACGGTGAATGGACCTACGACGACGCTACCAAAACCTTCA CCGTTACCGAAGGTGGCTCGGGCGGGGGTGGCTCGAGCGAAGAAG AAGAGGAGGAAGAAGAGGAAGAGGAAGAAGAAGAAGAAGAGGAGG AGGAGGAGGAAGAGGAGGAGGAAGAAGAAGAGGAGGAGTCCTCAG GGGGTGGCGGGAGCGGCGGTCGTGATCACATGGTGTTACACGAATA CGTCAACGCGGCAGGTATTACTAGTGGTTCGGGTGGTTCTCTGCCG GAAACCGGTGGTGGTAGTGGTGGCTCTGGCGGTTCTGGTGGCAGTG GCGGTAGCGCGCATATCGTTATGGTCGATGCTTACAAACCGACGAAA TAA

4.6 Methods

4.6.1 Plasmids and Cloning

For pRSET pAbBD-S11, an *E. coli* optimized DNA sequence encoding for pAbBD-S11 was synthesized (Integrated DNA Technologies, IDT) and inserted between the NdeI and NcoI restriction sites of pRSET-A (Invitrogen) via In-Fusion Cloning (Clontech). Short glycine- and serine-rich linkers were placed N-terminal to pAbBD, between pAbBD and the S11 splitGFP peptide, and between S11 and sequences used for proximity-based sortase-mediated ligation (PBSL, see pAbBD expression and purification). A SacI restriction site was placed in the DNA sequence encoding for the linker N-terminal to pAbBD and both the XhoI as well as Bsu36I restriction sites were placed in the DNA sequence encoding for the linker between pAbBD and S11 to assist with cloning of pAbBD-S11 variants.

Plasmids for pAbBD-ApP-S11 variants were made by inserting DNA sequences encoding for D10, D15, D20, D25, D30, E10, E15, E20, E25, and E30 ApPs between the XhoI and Bsu36I restriction sites of pRSET pAbBD-S11.

For pRSET His6-splitGFP(1-10), an *E. coli* optimized DNA sequence encoding for His6-splitGFP(1-10) was synthesized and inserted between the NdeI and HindIII restriction sites of pRSET-A.

For pRSET SpyCatcher-SrtA-His12 (used in PBSL, see pAbBD expression and purification), pRSET SpyCatcher-SrtA-His6 was used as a starting point[47]. Overhang PCR was used to make a DNA sequence encoding for SpyCatcher-SrtA-His12, which was then inserted between the NdeI and NcoI restriction sites of pRSET-A.

pEVOL-pBpF was a gift from Peter Schultz (Addgene plasmid # 31190 ; http://n2t.net/addgene:31190 ; RRID:Addgene_31190).

All plasmids were verified by Sanger sequencing. Protein and DNA sequences for all purified proteins are listed in Supplementary Table 4.1.

4.6.2 pAbBD expression and purification

pAbBD expression and purification was performed as described in subchapter 3.4.2 with the following modifications. All pAbBD variants were grown at 37°C with shaking for 24 h, except for pAbBD-D30-S11 and pAbBD-E30-S11, which were grown at 25°C with shaking for 48 h.

4.6.3 His6-splitGFP(1-10) expression, purification, and complementation

His6-splitGFP(1-10) expression, purification, and complementation was performed as described in subchapter 3.4.3.

4.6.4 Photocrosslinking pAbBD variants to IgGs

Clinical grade Rituximab was provided by Eline Luning Prak (University of Pennsylvania). The hybridoma for QCRL3 was obtained from ATCC (HB-11766) and the antibody was purified by the Cell Center at the University of Pennsylvania. Human IgG2 (hIgG2) isotype control (BE0301), mouse IgG2a (mIgG2a) isotype control (BE0085), mouse IgG2b (mIgG2b) isotype control (BE0086), and rabbit IgG (rabIgG) isotype control (BE0095) was purchased from BioXCell. Mouse IgG3 (mIgG3) isotype control (401302) used in delivery studies and rat IgG2c (rIgG2c) isotype control (400702) was purchased from BioLegend. mIgG3 isotype control used in NFκB studies was purchased from BD Biosciences (553486). Anti-RelA NLS monoclonal mouse IgG3 was purchased from MilliporeSigma (MAB3026). Anti-RelA C-term polyclonal rabbit IgG was purchased from ThermoFisher (51-0500).

Photocrosslinking was performed as described in subchapter 3.4.4.

4.6.5 Cell culture

HEK293T (ATCC CRL-3216), HT1080 (ATCC CCL-121), and A549 (ATCC CCL-185) cell lines were maintained in DMEM (Corning 10-013-CV) supplemented with 10% FBS (Corning 35-010-CV) and 100 U/mL penicillin-streptomycin (ThermoFisher 15140122). HEK293T splitGFP(1-10), HT1080 splitGFP(1-10), and A549 splitGFP(1-10) cell lines were maintained in the same

media as above + 2 μg/mL puromycin (puro) (Takara 631305). During all delivery experiments, cells were maintained in media without antibiotics or selection media. Cells were cultured at 37°C in a 5% CO_2 humidified incubator. All cell lines tested negative for mycoplasma contamination.

4.6.6 splitGFP(1-10) reporter cell line generation

Lentivirus expressing splitGFP(1-10)-IRES-[Puromycin resistance] as well as freshly transduced HEK293T splitGFP(1-10) cells were provided by Philip Zoltick (Children's Hospital of Philadelphia). HT1080 and A549 cell lines were transduced with the splitGFP(1-10) lentivirus at a multiplicity of infection of 10 in the presence of 8 μg/mL polybrene. 3 days following transduction, all reporter cell lines were selected for puro resistance by culturing in media with 2 μg/mL puro for 1 week. Afterwards, 20 μM pAbBD-S11 was electroporated into each cell line using the Neon transfection system (ThermoFisher) following the manufacturer's instructions. Electroporation settings were as follows: 150,000 HEK293T splitGFP(1-10) cells in buffer R with 10 μl tips and the 1150 V 20 ms 2 pulses setting, 100,000 H1080 splitGFP(1-10) cells in buffer R with 10 μl tips and the 1400 V 30ms 1 pulse setting, and 500,000 A549 splitGFP(1-10) cells in buffer T with 100 μl tips and the 1150 V 30 ms 2 pulses setting. At least 10^6 cells were electroporated and then incubated at 37°C for 6 h before using fluorescence activated cell sorting (FACS) to collect the brightest splitGFP-positive cells (top 20%). FACS was performed with a BD FACSJazz (BD Biosciences) with assistance from the Flow Cytometry and Cell Sorting Facility at the University of Pennsylvania.

To confirm that pAbBD-S11 and Ritux-$(pAbBD\text{-}S11)_2$ were compatible with the splitGFP reporter system, HEK293T splitGFP(1-10) cells were electroporated with 40 μM pAbBD-S11 or 7.5 μM Ritux-$(pAbBD\text{-}S11)_2$ using the Neon transfection system and the same settings as above. The cells were then incubated at 37°C for 6 h before being analyzed for splitGFP fluorescence by live-cell fluorescence microscopy or flow cytometry as described below.

4.6.7 Protein delivery

Cells were seeded to 70-80% confluency in a 48-well flat-bottom tissue culture plate. Just prior to delivery, the media was replaced with 180 µl fresh media. The protein to be delivered was diluted in OptiMEM I (ThermoFisher 31985-070) to 20x final protein concentration in 10 µl total volume. The indicated transfection reagent was diluted in OptiMEM to 10 µl total volume and incubated at RT for 10 min. The diluted protein and diluted cationic lipid was then mixed and incubated at RT for 10 min. Finally, the mixture was immediately added to reporter cell lines to a final volume of 200 µl (180µl media + 20 µl protein/cationic lipid mixture). Reporter cell lines were incubated at 37°C for 6 h before splitGFP complementation was determined by live-cell fluorescence microscopy or flow cytometry. Each delivery condition was paired with the same concentration of pAbBD-S11 or Ritux-(pAbBD-S11)$_2$ with the same amount of the indicated cationic lipid as a negative control. Viability was determined using the LDH assay (Dojindo Molecular Technologies) per the manufacturer's recommended protocol.

4.6.8 Fluorescence microscopy

Fluorescence microscopy was performed as described in subchapter 3.4.7.

4.6.9 Flow cytometry

Flow cytometry was performed as described in subchapter 3.4.8. The borders of a splitGFP positive cell population gate were delineated such that only the top 1% of negative control cells would fall within it.

4.6.10 Calcein efflux assay

Cells were seeded to 70-80% confluency in a 48-well flat-bottom tissue culture plate. Just prior to delivery, the media was replaced with 180 µl fresh media. 500 nM (final concentration in 200 µl media) mIgG2a-(pAbBD-D25-S11)$_2$ or QCRL3-(pAbBD-D25-S11)$_2$ was complexed with 2 µl Lipofectamine 2000 (HEK293T cells), 1.5 µl Lipofectamine RNAiMax (HT1080 cells), or 2 µl

Lipofectamine 2000 (A549 cells) and then added to cells as described above (protein delivery). For controls, 20 µl OptiMEM (cells only), 20µl 5µM QCRL3-(pAbBD-D25-S11)$_2$ in OptiMEM (QCRL3 co-incubation), or 20 µl 200 µM MK571 (Cayman Chemical) in OptiMEM were added to cells. After incubating at 37°C for 6 h, the media was replaced with 250 µl media containing 0.1 µM, 0.05 µM, or 0.25 µM Calcein-AM (ThermoFisher, C3100M) for HEK293T, HT1080, and A549 cells, respectively. After loading cells with calcein for 30 min at 37°C, the calcein-loading media was replaced with 250 µl fresh media and then the cells were incubated at 37°C to allow for export. For cells treated with MK571, 20 µM MK571 was kept in the media during calcein loading and export. After 16 h of calcein export, cells were analyzed by flow cytometry as described above (flow cytometry). Two technical replicates were used for each biological replicate.

4.6.11 Doxorubicin and vincristine sensitization

For A549 cells only and 20 µM MK571 treatment conditions, 5,000 (doxorubicin) or 2,500 (vincristine) A549 cells were seeded into each well of a 96-well flat bottom tissue culture plate containing serial dilutions of either doxorubicin (LC Laboratories) or vincristine (Cayman Chemical) in 100 µl media with or without 20 µM MK571. For IgG delivery conditions, A549 cells were seeded to 70-80% confluency in a 48-well flat-bottom tissue culture plate. Just prior to delivery, the media was replaced 180µl fresh media. 500 nM (final concentration in 200 µl media) mIgG2a-(pAbBD-D25-S11)$_2$ or QCRL3-(pAbBD-D25-S11)$_2$ was complexed with 2 µl Lipofectamine 2000 and added to cells as described above (protein delivery). After incubating cells with the complexed IgGs at 37°C for 4 h, the A549 cells were then reseeded into wells of a 96-well plate as described above. Following 48 h for doxorubicin or 72 h for vincristine, cell proliferation was determined using the MTT assay (Roche) per the manufacturer's protocol. Wells containing media only were used for background subtraction. Cell proliferation was then normalized to each condition in which cells were seeded without any chemotherapeutics. Three technical replicates were used for each biological replicate.

4.6.12 RelA (NFκB) immunofluorescence

A549 cells were seeded to 70-80% confluency in a 96-well flat clear bottom black tissue culture plate. Just prior to delivery, the media in each well was replaced with 90µl fresh media. 150 nM (final concentration in 100 µl media) of each tested IgG-(pAbBD-D25-S11)$_2$ was complexed with 1 µl Lipofectamine RNAiMax and added to cells as described above (protein delivery) scaled down to 100 µl final volume. For the A549 cells only control, 10 µl OptiMEM was added to the cells. After 6 h incubation at 37°C, the media was changed to 100 µl media with or without 20 ng/mL human TNFα (Cell Signaling Technology (CST), 8902SC) and incubated for 30 min at 37°C.

Afterwards, the cells were fixed with 4% formaldehyde at RT for 12 min and then washed with DPBS 3 times. The cells were then permeabilized and blocked by adding DPBS + 5% goat serum (CST 5425S) + 0.3% Triton X-100 and incubating for 1 h at RT. Next, the cells were stained with an anti-p65 Alexa 488 conjugated monoclonal antibody (CST 49445S) diluted at 1:100 in DPBS + 1% BSA + 0.3% Triton X-100 at 4°C overnight. Following the overnight incubation, the cells were washed with DPBS and then stained with 0.05 µg/mL HCS CellMask Red (ThermoFisher H32712) for 30 min at RT. After another DPBS wash, the cells were stained with 2 µg/mL DAPI (ThermoFisher) for 7 min at RT. After a final DPBS wash, the cells were imaged in DPBS. Microscopy images were acquired using the same setup as described above (fluorescence microscopy). At least 5 representative images were obtained per condition.

ImageJ was used to background subtract all microscopy images. CellProfiler 3.1.8[48] was used quantify the fraction of RelA that was in the nucleus of each cell. The CellProfiler pipeline used is available at https://upenn.app.box.com/s/kk9fu5yf0smax0ogj3ejwjlxvl3reihc.

4.6.13 NFκB transcriptional activity

A549 cells were seeded to 70-80% confluency in a 6-well tissue-culture plate and then 2.5 µg plasmid at a 9:1 pGL4.32[luc2P/NF-κB-RE/Hygro] (Promega E849A) reporter plasmid to

pNL1.1.TK[Nluc/TK] (Promega N150A) control plasmid ratio was transiently transfected into the cells using Lipofectamine 3000 (ThermoFisher) per the manufacturer's protocol (7.5 µl Lipofectamine 3000 reagent per well). 8 h following transfection, cells were reseeded at 17,500 cells per well in a 96-well flat bottom white tissue-culture plate and allowed to adhere overnight at 37°C. The A549 cells only control and IgG delivery was performed as described above (RelA immunofluorescence). After 4 h incubation at 37°C, the media was changed to 100 µl phenol-red free DMEM (ThermoFisher 31053028) supplemented with 10% FBS as well as indicated amounts of TNFα and incubated at 37°C for 5 h. Luminescence from both the firefly reporter luciferase as well as the NanoLuc control luciferase was determined by using the Nano-Glo Dual-Luciferase reporter assay system (Promega) per the manufacturer's instructions. The NanoLuc luciferase signal was used to correct the firefly luciferase signal for technical variability. Cells incubated with no TNFα was used to subtract out background bioluminescence. Firefly luciferase signals were then normalized to that of A549 cells only with 100 ng/mL TNFα stimulation. Two technical replicates were used for each biological replicate.

4.6.14 Statistical Analysis

Prism 7 (GraphPad Software) was used for statistical analysis. Data are presented as mean ± s.e.m. Significance testing for ratios were performed using a one-tailed *t*-test on the log-ratio. Comparisons between groups were made using a one-way ANOVA with Dunnet's correction.

4.7 References

1. Bradbury, A.R.M., Sidhu, S., Dübel, S. & McCafferty, J. Beyond natural antibodies: the power of in vitro display technologies. *Nat. Biotechnol.* **29**, 245-254 (2011).
2. Carter, P.J. & Lazar, G.A. Next generation antibody drugs: pursuit of the 'high-hanging fruit.' *Nat. Rev. Drug Discov.* **17**, 197-223 (2018).
3. Cheng, A.C., et al. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **25**, 71-75 (2007).

4. Ramaraj, T., Angel, T., Dratz, E.A., Jesaitis, A.J. & Mumey, B. Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures. *Biochim. Biophys. Acta* **1824**, 520-532 (2012).
5. Mulcahy, L.S., Smith, M.R. & Stacey, D.W. Requirements for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* **313**, 241-243 (1985).
6. Davies, H., et al. Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).
7. Marschall, A.L.J., Frenzel, A., Schirrmann, T., Schüngel, M. & Dübel, S. Targeting antibodies to the cytoplasm. *mAbs* **3**, 3-16 (2011).
8. Slastnikova, T.A., Ulasov, A.V., Rosenkranz, A.A. & Sobolev, A.S. Targeted intracellular delivery of antibodies: the state of the art. *Front. Pharmacol.* **9**, 1208 (2018).
9. Clift, D., et al. A method for the acute and rapid degradation of endogenous proteins. *Cell* **172**, 1692-1706 (2017).
10. Marschall, A.L.J., et al. Delivery of antibodies to the cytosol. *mAbs* **6**, 943-956 (2014).
11. Wörn, A. & Plückthun, A. Stability engineering of antibody single-chain Fv fragments. *J. Mol. Biol.* **305**, 989-1010 (2001).
12. Verdurmen, W.P.R., Mazlami, M. & Plückthun, A. A quantitative comparison of cytosolic delivery via different protein uptake systems. *Sci. Rep.* **7**, 13194 (2017).
13. Herce, H.D., et al. Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells. *Nat. Chem.* **9**, 762-771 (2017).
14. Shin S., et al. Antibody targeting intracellular oncogenic Ras mutants exerts anti-tumor effects after systemic administration. *Nat. Commun*. **8**, 15090 (2017).
15. Adams, D., et al. Patisiran, an RNAi therapeutic, for hereditary transthyretin amyloidosis. *N. Engl. J. Med.* **379**, 11-21 (2018).
16. Zuris, J. A., et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **33**, 73-80 (2015).
17. Gao, X., et al. Treatment of autosomal dominant hearing loss by in vivo delivery of genome editing agents. *Nature* **553**, 217-221 (2018).
18. Eltoukhy, A.A., et al. Nucleic acid-mediated intracellular protein delivery by lipid-like nanoparticles. *Biomaterials* **35**, 6454-6461 (2014).
19. Lee, Y., et al. Efficient delivery of bioactive antibodies into the cytoplasm of living cells by charge-conversional polyion complex micelles. *Angew. Chem. Int. Ed.* **49**, 2552-2555 (2010).

20. Postupalenko, V., et al. Protein delivery system containing a nickel-immobilized polymer for multimerization of affinity-purified his-tagged proteins enhances cytosolic transfer. *Angew. Chem. Int. Ed.* **54**, 10583-10586 (2015).
21. Mout, R., et al. General strategy for direct cytosolic protein delivery via protein-nanoparticle co-engineering. *ACS Nano* **11**, 6416-6421 (2017).
22. Hui, J.Z., Tamsen, S., Song, Y. & Tsourkas, A. LASIC: light activated site-specific conjugation of native IgGs. *Bioconjugate Chem.* **26**, 1456-1460 (2015).
23. Deprey, K., Becker, L., Kritzer, J. & Plückthun, A. Trapped! A critical evaluation of methods for measuring total cellular uptake versus cytosolic localization. *Bioconjugate Chem.* **30**, 1006-1027 (2019).
24. Smith, S.A., Selby, L.I., Johnston, A.P.R. & Such, G.K. The endosomal escape of nanoparticles: toward more efficient cellular delivery. *Bioconjugate Chem.* **30**, 263-272 (2019).
25. Cabantous, S., Terwilliger, T.C. & Waldo, G.S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23**, 102-107 (2005).
26. Milech, N., et al. GFP-complementation assay to detect functional CPP and protein delivery into living cells. *Sci. Rep.* **5**, 18329 (2015).
27. Schmidt, S., et al. Detecting cytosolic peptide delivery with the GFP complementation assay in the low micromolar range. *Angew. Chem. Int. Ed.* **54**, 15105-15108 (2015).
28. P. Lönn, et al., Enhancing endosomal escape for intracellular delivery of macromolecular biologic therapeutics. *Sci. Rep.* **6**, 32301 (2016).
29. Kauffman, W.B., Guha, S. & Wimley, W.C. Synthetic molecular evolution of hybrid cell penetrating peptides. *Nat. Commun.* **9**, 2568 (2018).
30. Timney, B.L., Simple rules for passive diffusion through the nuclear pore complex. *J. Cell Biol.* **215**, 57-76 (2016).
31. Cole, S.P.C. Targeting multidrug resistance protein 1 (MRP1, ABCC1): past, present, and future. *Annu. Rev. Pharmacol. Toxicol.* **54**, 95-117 (2014).
32. Loe, D.W., Almquist, K.C., Deeley, R.G. & Cole, S.P.C. Multidrug resistance protein (MRP)-mediated transport of leukotriene C_4 and chemotherapeutic agents in membrane vesicles. *J. Biol. Chem.* **271**, 9675-9682 (1996).
33. Loe, D.W., Deeley, R.G. & Cole, S.P.C. Characterization of vincristine transport by the Mr 190,000 multidrug resistance protein (MRP): evidence for cotransport with reduced glutathione. *Cancer Res.* **58**, 5130-5136 (1998).

34. Hipfner, D.R., et al. Monoclonal antibodies that inhibit the transport function of the 190-kDa multidrug resistance protein, MRP. *J. Biol. Chem.* **274**, 15420-15426 (1999).
35. Cole, S.P.C., et al. Pharmacological characterization of multidrug resistance MRP-transfected human tumor cells. *Cancer Res.* **54**, 5902-5910 (1994).
36. Wu, H. & Napetschnig, J. Molecular basis of NF-κB signaling. *Annu. Rev. Biophys.* **42**, 443-468 (2013).
37. Guillard, S., et al. Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange. *Nat. Commun*. **8**, 16111 (2017).
38. Gupta, A., et al. Facile target validation in an animal model with intracellularly expressed monobodies. *Nat. Chem. Biol.* **14**, 895-900 (2018).
39. Caussinus, E., Oguz, K. & Affolter, M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat. Struct. Mol. Biol.* **19**, 117-121 (2012).
40. Kanner, S.A., Morgenstern, T. & Colecraft, H.M. Sculpting ion channel functional expression with engineered ubiquitin ligases. *eLife* **6**, e29744 (2017).
41. Marcon, E., et al. Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat. Methods* **12**, 725-731 (2015).
42. Venkataraman, A., et al. A toolbox of immunoprecipitation-grade monoclonal antibodies to human transcription factors. *Nat. Methods* **15**, 330-338 (2018).
43. James, L.C., Keeble, A.H., Khan, Z., Rhoades, D.A. & Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci. USA* **104**, 6200-6205 (2007).
44. Sauer-Eriksson, A.E., Kleywegt, G.J., Uhlén, M. & Jones, T.A. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **15**, 265-278 (1995).
45. Choe, W., Durgannavar, T.A. & Chung, S.J. Fc-binding ligands of immunoglobulin G: an overview of high affinity proteins and peptides. *Materials (Basel)* **9**, 994 (2016).
46. Kanasty, R., Dorkin, J.R., Vegas, A. & Anderson, D. Delivery materials for siRNA therapeutics. *Nat. Mater.* **12**, 967-977 (2013).
47. Wang, H.H., Altun, B., Nwe, K. & Tsourkas, A. Proximity-based sortase-mediated ligation. *Angew. Chem. Int. Ed.* **56**, 5349-5352 (2017).
48. Carpenter, A.E., et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7,** R100 (2006).

Chapter 5 – Summary and Future Directions

5.1 Proximity-based sortase-mediated ligation

5.1.1 Summary

All approaches for site-specific protein bioconjugation have tradeoffs between their perturbation to the target protein, flexibility in labeling location, labeling efficiency, and ease of use. Canonical sortase-mediated ligation (SML) minimally perturbs protein function with the introduction of a small recognition motif, but is limited to labeling proteins at either the N- or C-terminus and suffers from poor labeling efficiency[1]. Even sortase A (SrtA) variants engineered to have improved catalytic activity require precise optimization of labeling times to minimize formation of side-products[2,3]. The Tsourkas lab has previously developed sortase-tag expressed protein ligation (STEPL)[4], which dramatically improved labeling efficiency and ease by tethering target proteins to SrtA to increase the effective concentration of the target protein. Unfortunately, STEPL required target proteins to be expressed as a fusion to SrtA, which, for certain target proteins, can deleteriously affect yield.

We have developed proximity-based sortase-mediated ligation (PBSL),[5] which utilizes the SpyCatcher-SpyTag domain-peptide pair[6,7] to minimize its target protein expression footprint while retaining the increases in labeling efficiency obtained with STEPL. Although PBSL requires the additional step of preparing SpyCatcher-SrtA pre-immobilized on Co^{2+} resin, SpyCatcher-SrtA expresses at extremely high yields in *E. coli* and the resin can be stored at 4°C for months without any reduction in activity. PBSL is still less labor intensive than SML because similarly to STEPL, PBSL combines target protein purification, the SrtA ligation reaction, and separation of the labeled protein from SrtA into a single process. We demonstrated the utility of PBSL by fluorescently labeling an anti-CD3 scFv[5] and functionalizing a photoreactive antibody binding domain (pAbBD) with a C-terminal cell-penetrating peptide (CPP, chapter 3), two proteins that express very poorly in the STEPL system.

5.1.2 High purity protein purification with PBSL

Recombinant protein purification frequently requires compromise between yield, purity, and speed. The most widespread technique is immobilized metal affinity chromatography (IMAC), which uses Ni^{2+} or Co^{2+} resin to purify His-tagged proteins. However, IMAC often results in only moderate purity because many endogenous *E. coli* proteins can non-specifically bind to IMAC resins[8]. Washing IMAC resins with low concentrations of imidazole can modestly improve purity, but reduces yield. Fast protein liquid chromatography (FPLC) post-IMAC can dramatically improve purity, but increases complexity and requires time-intensive method development. High-affinity interactions or covalent linkages such as that of between biotin and streptavidin or HaloTag and HaloLink[9], respectively, can be employed for protein purifications to enable stringent washes, greatly improving purity. However, these approaches require either harsh conditions[10] or additional protease cleavage steps[9] to elute proteins.

We originally developed PBSL to enable site-specific labeling of proteins at their C-terminus, but PBSL can also be adapted for protein purification by using triglycine (Gly-Gly-Gly, GGG) in the ligation reaction rather than functionalized peptides. With some minor modifications, PBSL can achieve similar yields compared to that of IMAC, but with higher purity. Although endogenous proteins can still bind to SpyCatcher-SrtA-His6 resin, the ligation reaction buffer contain no imidazole, which minimizes the amount of contaminants released. Furthermore, a long His-tag (His12) can be attached to SpyCatcher-SrtA, which dramatically increases its affinity to the Co^{2+} resin. Contaminant proteins that non-specifically bind to and then are released from the resin during the ligation reaction can then be removed by washing with moderate concentrations of imidazole without reducing protein yield.

We were able to purify SpyCatcher-SrtA-His12 with high yield (>40 mg L^{-1} *E. coli* culture), immobilize it on Co^{2+} resin, and confirm that it had a maximal capture efficiency (~80%) and release efficiency similar to that of SpyCathcer-SrtA-His6. We began by purifying eGFP to

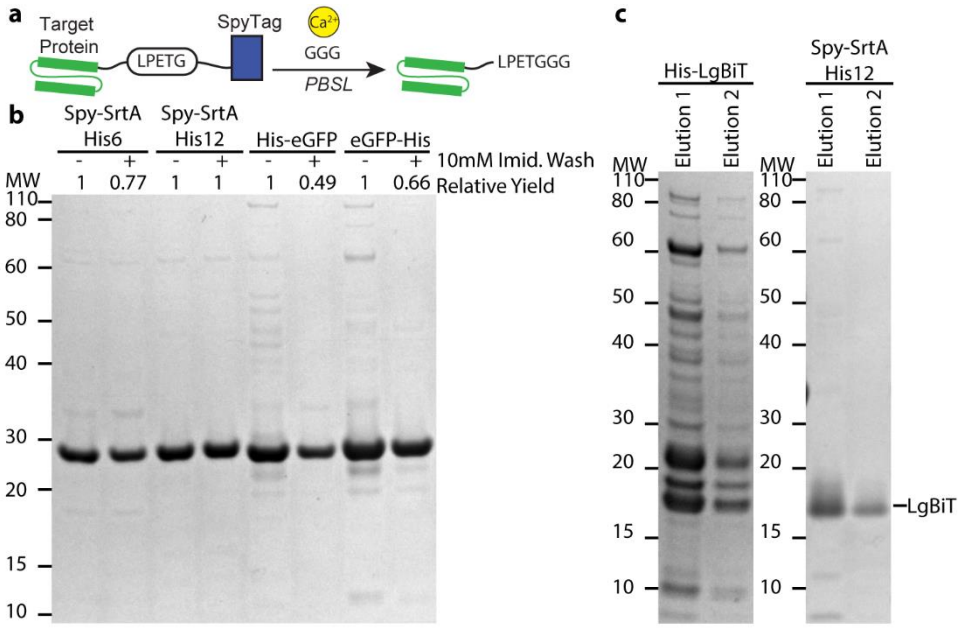

Figure 5.1 Comparing C-terminal PBSL purification purity and yield against IMAC.

a) Schematic of C-terminal PBSL protein purification. **b)** SDS-PAGE comparing the yield and purity of eGFP obtained via C-terminal PBSL purification or IMAC. Clarified lysates containing equivalent amounts of eGFP-LPETG-SpyTag, His6-eGFP, or eGFP-His6 were used for purification. **c)** SDS-PAGE comparing the purity of LgBiT obtained via C-terminal PBSL purification with SpyCatcher-SrtA-His12 resin against IMAC.

benchmark C-terminal PBSL purification (Fig. 5.1a) against IMAC and found that PBSL resulted in greater purity when the resin was washed without any imidazole (Fig. 5.1b). When washing with 10mM imidazole, there was no decrease in yield with SpyCatcher-SrtA-His12 resin, but we observed significant reductions in yield with the SpyCathcher-SrtA-His6 resin and IMAC (Fig. 5.1b). Even after a 10 mM imidazole wash, C-terminal PBSL purification with SpyCatcher-SrtA-His12 still resulted in slightly improved purity when compared with IMAC (Fig. 5.1b). Interestingly, when purifying LgBiT, which is the large half of a split luciferase[11], C-terminal PBSL purification resulted in dramatically improved purity when compared to IMAC (Fig. 5.1c). This suggests that initial studies with eGFP may underestimate the improvements in purity that can be obtained with C-terminal PBSL purification.

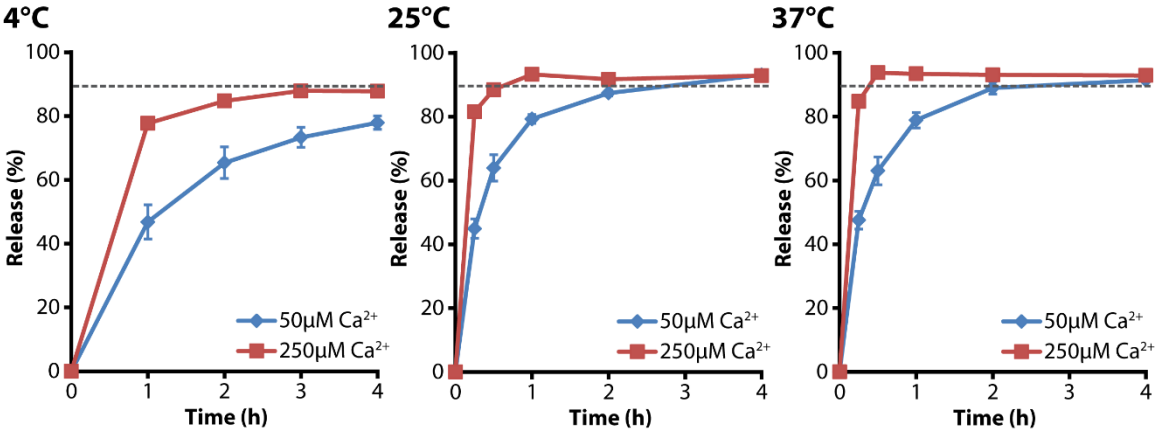

Figure 5.2 C-terminal PBSL purification elution kinetics.

Elution kinetics for eGFP-LPETG-SpyTag captured on SpyCatcher-SrtA-His12 resin using a PBS + 2mM GGG + 50μM or 250μM elution buffer at the indicated temperatures. The dashed line indicates 90% release. Data are mean ± s.e.m.

For IMAC, proteins are rapidly eluted with high concentrations of imidazole, even at low temperatures (4°C). Because C-terminal PBSL purification elutes proteins via an enzymatic reaction, it is more strongly affected by changes in temperature. We originally optimized PBSL for elutions at 37°C[5], which is not practical for many aggregation-prone proteins, so we sought to identify fast and efficient elution conditions that could be performed at 25°C or 4°C. We prepared SpyCatcher-SrtA-His12 resin, fully loaded it with eGFP, and then determined the kinetics using a PBS + 50μM $CaCl_2$ + 2mM GGG elution buffer (Fig. 5.2). Under these conditions, at 25°C and 37°C, 90% elution could be achieved in 2h whereas at 4°C, only 80% elution could be achieved in 4h (Fig. 5.2). When we increased the calcium concentration to 250μM, elution kinetics improved, resulting in 90% elution within 30min at 25°C and 37°C as well as 90% elution within 3h at 4°C (Fig. 5.2). Although still not as rapid as IMAC elutions, optimized C-terminal PBSL elution conditions should be sufficient for even aggregation-prone proteins.

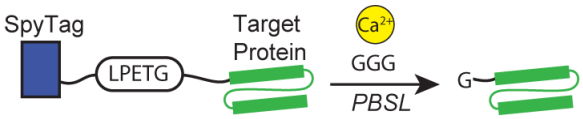

Figure 5.3 Schematic of N-terminal PBSL purification.

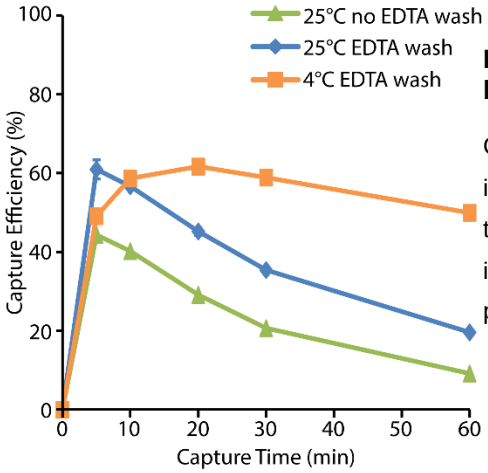

Figure 5.4 C-terminal PBSL purification capture kinetics.

Clarified lysates containing SpyTag-LPETG-eGFP were incubated with SpyCatcher-SrtA-His12 resin at the indicated temperature for the indicated amount of time. EDTA wash indicates lysates that were prepared after washing bacteria pellet with PBS + 100mM EDTA.

Next, we wondered whether PBSL could be used to purify proteins containing an N-terminal SpyTag and SrtA recognition motif. N-terminal PBSL purification would be almost traceless since proteins would only contain an N-terminal Gly following elution (Fig. 5.3). This is comparable to purifying proteins using an N-terminal affinity tag that is subsequently cleaved off with TEV protease. Although eGFP could be purified using N-terminal PBSL purification and the SpyCatcher-SrtA-His12 resin, maximal capture efficiency (~45%) was significantly lower for N-terminal PBSL purification than for C-terminal PBSL purification (~80%) (Fig. 5.4, 25°C no EDTA wash). Furthermore, capture efficiency peaked after 5 min before rapidly decreasing (Fig. 5.4, 25°C no EDTA wash), whereas capture efficiency was stable for C-terminal PBSL purification (Chapter 2 Fig. 2.3a). We attributed this decrease in capture efficiency to increased background SrtA cleavage. Since SrtA activity is Ca^{2+} dependent, we attempted to remove any extracellular Ca^{2+} carried over from protein expression by washing the bacteria pellet with EDTA prior to lysis. This moderately improved maximal capture efficiency to ~60%, but capture efficiency was still heavily sensitive to capture time (Fig. 5.4, 25°C EDTA wash). Next, we tested whether capturing at 4°C instead of 25°C would improve capture efficiency and found that although maximal capture efficiency still peaked at ~60%, capture efficiency was less sensitive to time (Fig. 5.4, 4°C EDTA wash). We believe that this improvement results from SpyTag-SpyCatcher isopeptide bond formation being less temperature sensitive[6] than background SrtA activity. Although additional

parameters could be adjusted to increase N-terminal PBSL purification capture efficiency, it is unclear how much additional improvement can be achieved.

Future studies should determine whether N-terminal PBSL purification offer similar improvements in purity and yield as that of C-terminal PBSL purification. Given that N-terminal PBSL suffers from background SrtA activity, it may have much faster elution kinetics than that of C-terminal PBSL purification, which would be advantageous. Finally, both N- and C-terminal PBSL purification should be benchmarked against IMAC with a panel of proteins within *E. coli*, *S. cerevisiae*, and mammalian lysate backgrounds.

5.2 Cytosolic delivery of proteins with cell-penetrating peptides

5.2.1 Summary

Cell-penetrating peptides (CPPs) have been extensively studied for their potential to deliver small peptides or protein cargos into the cytosol of cells[12,13], but they suffer from poor efficacy due to their inability to efficiently induce endosomal escape[14,15]. Inspired by reports that dimerizing or multimerizing CPPs can significantly improve their endosomal escape efficiency[16–19], we placed either the TAT or R8 CPP at both termini of a small photoreactive antibody-binding domain (pAbBD) and found it dramatically improved cytosolic delivery. Unfortunately, these improvements did not translate to a much larger IgG cargo. Regardless, we have identified a simple design principle for improving CPP efficacy that may be applicable to other small protein cargos.

5.2.2 Screening for improved cell-penetrating peptide and endosomolytic peptide pairs

Endosomolytic peptides (ELPs) show promise for improving endosomal escape of cargo proteins when used in conjugation with CPPs. ELPs are short, negatively charged peptides that at low pH, such as within the lumen of maturing endosomes, are capable of binding to and disrupting the endosome membrane to induce permeabilization[20–23]. In a seminal paper, Dowdy and coworkers demonstrated the feasibility of pairing CPPs with ELPs by enhancing the delivery

of a CPP-labeled Cre-recombinase by co-incubating it with a CPP-ELP peptide[14]. The ELP can also be directly fused to a protein cargo – this approach was used to deliver p53 in cancer cells[24] and Cre-recombinase in mouse neonatal cochleas[25]. Although these examples show the potential of using CPPs in conjunction with ELPs, they were still delivering cargos capable of amplifying their effects. Since inhibitory proteins must achieve cytosolic concentrations that are stoichiometric relative to their target to be efficacious, new CPP-ELP pairs with significantly improved cytosolic delivery must be developed for CPP-mediated cytosolic protein delivery to be widely adopted.

Currently, no screen of CPP-ELP pairs has been conducted due to three major challenges. First, both peptides are usually 10 to 15 amino acids long, rendering synthesized peptide libraries of just CPPs fused to ELPs costly. Addition of reporter moieties, critical for any screen, only increases synthesis complexity and cost. Second, the mechanism for CPP-mediated cytosolic delivery is different between small peptides and larger protein cargos[26]. Thus, the peptides will have to be conjugated to the cargo of interest, which may be difficult to achieve in a high-throughput and efficient manner. Third, directly fusing CPPs to ELPs may introduce unwanted electrostatic interactions between the two components. CPPs depend upon having their positively charged side groups accessible, so placing a negatively charged ELP nearby may shield the CPP similar to that of masking sequences used in activatable cell-penetrating peptides (aCPPs)[27–29]. Thus, CPPs and ELPs have to be conjugated to cargos in a manner that separates them physically.

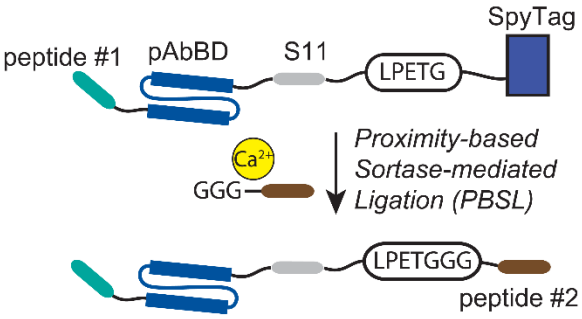

Figure 5.5 CPP and ELP library generation.

We previously attempted to screen CPP-ELP pairs for delivery of full-length IgG antibodies by generating a library of pAbBD variants containing different CPP-ELP pairs. We

envisioned first purifying a panel of pAbBD variants containing different CPPs or ELPs at the N-terminus. Then, the partner peptide can be ligated to the C-terminus of pAbBD (e.g. an ELP with an N-terminal CPP) using PBSL (Fig. 5.5). By using a two-step process, library generation is greatly simplified. Instead of cloning and optimizing the purification strategy for 100 pAbBD variants to test 10 N-terminal peptides paired with 10 C-terminal peptides, our strategy only requires purifying 10 pAbBD variants with different N-terminal peptides and synthesizing 10 different C-terminal peptides. After producing a library of pAbBD variants, we could then photocrosslink them onto IgGs to screen for CPP-ELP pairs that could efficiently cytosolically deliver IgGs into the cytosol.

We planned to construct an initial library consisting of the TAT[30], polyarginine (R8)[31], penetratin[32], transportan[33], MAP[34], and M-lycotoxin L17E[22] CPPs paired with the HA2[14], aurein 1.2[25], or INF7[35] ELPs. Unfortunately, we were only able to purify pAbBD with an N-terminal TAT or R8. Even after significant optimization, all other CPP or ELP sequences resulted in either extremely poor purification yields or truncation products, respectively. Next, we synthesized the TAT, R8, penetratin, and transportan CPPs as peptides compatible with PBSL. Both TAT and R8 were successfully ligated onto the C-terminus of pAbBD via PBSL, but penetratin and transportan caused pAbBD to aggregate following ligation. As a result, we were unable to test most of the library for cytosolic delivery.

Although our initial attempt at screening CPP-ELP pairs with pAbBD failed due to technical reasons, we believe that conducting the screen could be successful with a designed ankyrin repeat protein (DARPin) scaffold. Similarly to antibodies, DARPins can be developed to bind and inhibit target proteins through *in vitro* display[36], making them excellent therapeutic cytosolic protein scaffolds. Furthermore, DARPin scaffolds have been designed for increased stability[36,37] and may be more tolerant of having CPP fused to their N-terminus. We believe that the truncation products associated with ELP fusions were due to cryptic protease sites found in

pAbBD, which would likely not be present in DARPins. Finally, in chapter 3, we found that CPP placement at both termini could deliver pAbBD, but failed to do so for IgGs-$(pAbBD)_2$ due to the large size of IgGs. Thus, it is much more likely for the screen to identify CPP-ELP pairs capable of efficiently delivering DARPins into the cytosol because they are much more compact.

5.3 Cytosolic delivery of inhibitory antibodies with cationic lipids

5.3.1 Summary

We have made a significant advance towards realizing therapeutic cytosolic antibodies by demonstrating that IgG antibodies conjugated with anionic polypeptides (ApPs) can be complexed with a variety of cationic delivery lipids to enable efficient cytosolic delivery. We used a stringent splitGFP reporter system[38] to quantify delivery efficiency and found that when compared with CPP-mediated delivery of peptides[39–43], our approach can achieve similar deliver efficiencies with a larger antibody cargo at ~100-fold lower concentrations. In chapter 3, we found that CPP-mediated delivery efficiency plummets as cargo size increases, suggesting that the ~100-fold improvement figure is a significant underestimation. Finally, because ApPs were conjugated to IgGs by fusing them to a photoreactive antibody binding domain (pAbBD)[44], we were able to easily demonstrate that cytosolically delivered IgGs were functional by inhibiting the drug export pump multidrug resistance-associated protein 1 (MRP1) as well as the transcription factor NFκB.

5.3.2 Additional delivery characterization

We used a splitGFP reporter system that normally only reports on cytosolic protein delivery, but it could also be adapted to determine both the cellular uptake efficiency (fraction of protein added to cells that is taken up via endocytosis) as well as the endosomal escape efficiency (fraction of proteins taken up via endocytosis that reaches the cytosol). In the future, we can use these additional parameters to better understand how different delivery lipid formulations, protein cargos, ApPs, or cell lines affect cytosolic delivery.

To obtain cellular uptake efficiency, note that we know how much protein is added and can correlate that to a splitGFP fluorescence signal (signal #1). We can correlate total cellular protein uptake to a splitGFP fluorescence signal by lysing cells following delivery, adding excess splitGFP(1-10) to the lysates, and then measuring fluorescence once sufficient time has passed to ensure chromophore maturation (signal #2). The lysis step releases any proteins trapped in endosomes, allowing them to subsequently undergo splitGFP complementation. Cellular uptake efficiency is then just signal #2 divided by signal #1. To obtain endosomal escape efficiency, we can correlate the amount of cytosolically delivered proteins to a splitGFP fluorescence signal by lysing cells following delivery and then immediately measuring fluorescence (signal #3). Due to the kinetics of splitGFP chromophore maturation, immediately measuring fluorescence will ensure that endosomal proteins do not contribute to the signal. Endosomal escape efficiency is then just signal #3 divided by signal #2.

Previous studies[45–47] have indicated that IgGs typically have an approximately 24h half-life in the cytosol of cells, but it would be useful to independently determine the half-life of IgGs that we cytosolically deliver. For example, it is possible that pAbBD photocrosslinking perturbs IgG stability in the cytosol. We would like to avoid determining half-life by tracking splitGFP fluorescence due to the delays introduced by chromophore maturation and the potential for splitGFP complementation to perturb half-life. Thus, the most useful approach for determining IgG half-life would be to track how long cytosolically delivered IgGs can inhibit a given protein. One can rapidly determine the extent of QCRL3-mediated inhibition of MRP1 by tracking doxorubicin accumulation within cells with flow cytometry.

5.3.3 Improving cytosolic delivery

Although we were capable of cytosolically delivering IgGs into cells with up to 86% delivery efficiency, a subpopulation consistently remains recalcitrant to delivery. We hypothesize that this could be partially due to cell-cycle effects, which could be tested for by serial treatment

with protein-lipid complexes or synchronizing cells and then testing for delivery efficiency at various cell cycle stages. There are also countless other parameters that we could adjust to affect protein-lipid complexation. We could test additional commercially available cationic transfection reagents, change the complexation pH, or use either N-terminal or different ApPs. For the latter, we should test whether ApPs consisting of a mixture of poly-aspartates/glutamates (e.g. $(DE)_{15}$) or ApPs containing non-anionic residues (e.g. $(DDX)_{15}$ or $(EEG)_{15}$) would affect delivery. Liu and coworkers recently screened a panel of anionic human proteins and found that genome editing proteins could be fused to prothymosin alpha (ProTα) and then complexed with cationic lipids for cytosolic delivery[48]. Interestingly, they also found that ProTα performed better than mixed poly-aspartate/glutamate sequences[48]. Thus, future studies should determine how ProTα compares against our poly-aspartate or poly-glutamate ApPs in cytosolic IgG delivery.

5.3.4 Cytosolic delivery of nanobodies and antibody-like binding proteins

In chapter 4, we demonstrated that in addition to IgG antibodies, our approach for cytosolic protein delivery is also applicable to small proteins such as pAbBD by itself. Thus, we wondered whether we could also efficiently deliver nanobodies and antibody-like binding proteins into the cytosol as well. Small protein binders have short serum half-lives, limiting their therapeutic uses, but that is no longer a limitation when they are encapsulated in nanocarriers because pharmacokinetics will then be predominantly determined by properties associated with the nanocarrier[49]. By removing the bulkiness associated the IgG Fc region, small protein binders can be easily recombinantly produced in *E. coli* and are more amenable to protein engineering. Ultimately, if therapeutic cytosolic proteins become a reality, they are much more likely to be based on smaller protein scaffolds.

We chose an anti-GFP nanobody (αGFPnb)[50], an anti-Taq polymerase affibody (αTaqAffi)[51], and an anti-KRas DARPin (DARPinK27)[52] as prototypical small protein binders and purified them as fusions with the S11 splitGFP reporter peptide as well as a 30 residues-long

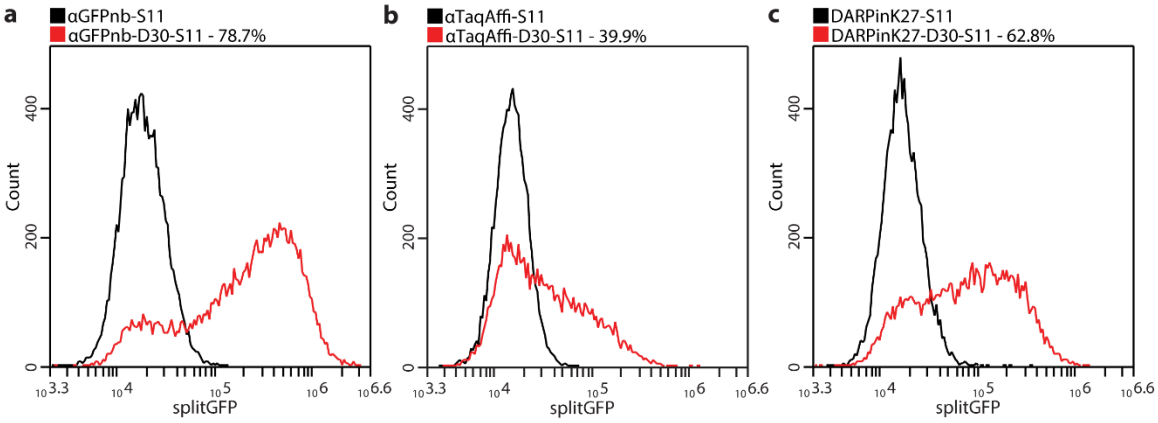

Figure 5.6 Cytosolic delivery of αGFPnb, αTaqAffi, and DARPinK27.

500nM **(a)** αGFPnb-ApP-S11, **(b)** αTaqAffi-ApP-S11, or **(c)** DARPinK27-ApP-S11 with either no ApP (negative control) or a D30 ApP were complexed with 2µl Lipo 2000 and added to A549 splitGFP(1-10) cells for 6h. Representative flow cytometry histograms of splitGFP fluorescence are shown.

poly-aspartate ApP (D30). We also purified versions without any ApP to serve as negative controls. After complexing with 2µl Lipofectamine (Lipo) 2000, 500nM αGFPnb-D30-S11, αTaqAffi-D30-S11, or DARPinK27-D30-S11 were cytosolically delivered into A549 splitGFP(1-10) cells with efficiencies of 78.7%, 39.9%, and 62.8%, respectively (Fig. 5.6). Next, we wanted to demonstrate that our cytosolically delivered small protein binders are functional. Although the αGFPnb and the αTaqAffi were chosen as model proteins, DARPinK27 can inhibit Ras[52].

The Ras family consists of a family of 3 oncogenes – KRas, NRas, and HRas – that is found to be mutated in almost a third of all cancers[53]. The Ras proteins are GTPases that normally shuttles between a GDP-bound inactive form and a GTP-bound active form, but oncogenic mutations often preferentially lock Ras in its active form[53–55]. Activated Ras induces signaling through a variety of effector channels such as the RAF/MEK/ERK and the PI3K/Akt pathway to promote cell proliferation and inhibit apoptosis[56]. Because Ras is mutated so frequently in cancer and is a regulator of so many pathways that contribute to cancer cell survival, Ras has been the focus of intense drug discovery efforts. Since its discovery in the 1980s, Ras has been considered to be nearly undruggable. Recently, KRas-G12C inhibitors have been

developed[57] and are being tested in clinical trials, but the KRas-G12C mutation represents a very small fraction of all oncogenic Ras mutations[53]. Thus, there is still a huge unmet need for pan-Ras inhibitors.

Minter and coworkers identified DARPinK27 through phage display and found that it preferentially binds to GDP-bound KRas to prevent KRas nucleotide exchange[52]. So far, DARPinK27 has been found to inhibit wild type KRas, KRas-G12C, KRas-G12D, KRas-G12V, wild type NRas, NRas-G12D, NRas-Q61K, and NRas-Q61R[52]. DARPinK27 overexpression results in the blockade of KRas-induced activation of both the RAF/MEK/ERK and the PI3K/Akt pathways[52,58].

To assess KRas signaling activity, we stimulated cells with human epidermal growth factor (hEGF) and determined whether the RAF/MEK/ERK pathway has been activated by western blotting for phosphorylated ERK[54]. As expected, treating cells with Trametinib, a MEK inhibitor[59], abolished ERK phosphorylation, but cytosolically delivering DARPinK27n3, which is incapable of binding to KRas, or incubating cells with either the control or active DARPinK27 did not affect pERK levels (Fig. 5.7). Excitingly, cytosolic delivery of 500nM DARPinK27-D30-S11 reduced pERK levels by ~65% (Fig. 5.7), indicating successful inhibition of KRas-G12S signaling.

Future studies should further validate that we are inhibiting KRas signaling with cytosolically delivered DARPinK27 by checking whether we are inhibiting the activation of other

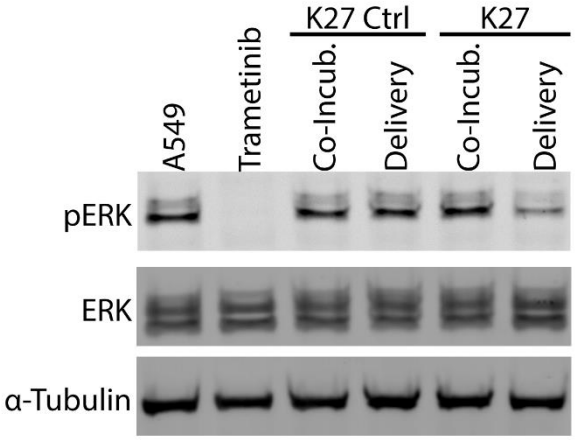

Figure 5.7 DARPinK27 inhibits KRas-G12S signaling in A549 cells.

500nM DARPinK27-D30-S11 and a negative control DARPinK27n3-D30-S11 were either co-incubated or cytosolically delivered into A549 cells with Lipo 2000. 4 hours following delivery, KRAs signaling activity was determined by stimulating with 100ng/mL hEGF for 30min and then western blotting for phosphorylated ERK. 100nM Trametinib treatment was used as a positive control.

Ras effector pathways such as the PI3K/Akt pathway. Afterwards, we need to show that cytosolic DARPinK27 delivery can inhibit cancer cell growth in both cell culture and murine models. It should be noted that many protein binders have been found to be capable of inhibiting Ras via a variety of mechanisms. Minter and coworkers have also developed a series of DARPins that block Ras-Raf interactions[52]. Rabbitts and coworkers developed iDab, which prevents GTP-bound Ras from interacting with effector proteins, as well as DARPins that inhibit Ras by binding to an allosteric site[58,60]. Finally, O'Bryan and coworkers have developed a monobody that disrupts Ras dimerization[61]. Our cytosolic protein delivery system may be ideal for benchmarking these proteins to identify the most effective Ras inhibitor.

Myc is another potent oncogene considered to be largely 'undruggable[62,63]' that we could potentially inhibit using our cytosolic protein delivery technology. Myc is a transcription factor considered to be a master regulator of cell growth and found to be deregulated in up to 70% of cancers[64]. Because Myc must heterodimerize with Max to active transcription, the Myc-Max dimerization interface has been the target of numerous small molecule drug screens[62–64]. Unfortunately, small molecule inhibitors are rarely capable of inhibiting Myc transcription in cells at submicromolar concentration[62–64].

A small protein domain called Omomyc has been shown to be a potent Myc inhibitor by acting as a dominant negative – it heterodimerizes with Myc to prevent Myc-Max heterodimer formation[65]. Omomyc overexpression in mice is relatively well tolerated and can induce rapid regression of lung adenocarcinomas[66]. Although it has been reported that Omomyc has an intrinsic ability to enter the cell, systemic Omomyc treatment resulted in relatively modest inhibition of tumor growth in mice[67]. This suggests that there is still considerable room for improving cytosolic Omomyc delivery, which our delivery approach would be well suited for.

5.3.5 Antibody-induced protein degradation

Although the current process underpinning therapeutic antibody development can efficiently generate large collections of IgGs that bind to a given target, identifying an IgG that both tightly binds and inhibits its target is not guaranteed. For example, structural features critical for a target protein's function may be too buried for antibodies to access. This can be addressed if antibodies delivered to the cytosol could be engineered to induce degradation of target proteins – that way, any IgG that binds to a target will then be capable of inhibiting it by depletion. Furthermore, degradative antibodies could catalyze multiple rounds of target protein degradation, potentially allowing for complete target inhibition with sub-stoichiometric amounts of cytosolically delivered antibodies.

Therapeutically inhibiting proteins via small-molecule induced degradation has already been shown to be a robust strategy with proteolysis-targeting chimaera (PROTAC) and hydrophobic tagging technologies[68]. Both technologies utilize bifunctional molecules that bind and degrade target proteins by recruiting endogenous protein quality control machinery to them[68]. Unfortunately, developing small molecules that tightly bind to targets is challenging – even if any binding site is acceptable – and currently, most PROTACs still target protein classes that contain easily druggable binding pockets[68]. Furthermore, because PROTACs are bifunctional, they are usually rather large for small-molecule drugs, which can significantly decrease cell permeability[68,69].

Schuh and coworkers[45] have demonstrated that cytosolic IgGs can naturally engage the TRIM21 E3 ubiquitin ligase via its Fc region to degrade their protein binding targets. Unfortunately, this functionality is not available to our cytosolically delivered IgGs because pAbBD photocrosslinking precludes TRIM21 binding[70,71]. We could employ two different strategies to restore our cytosolically delivered IgGs' ability to engage TRIM21. First, we could develop

alternative pAbBDs that photocrosslink to IgGs outside of the Fc region. Alternatively, we could recombinantly produce IgGs with ApPs fused directly to either the heavy or the light chain.

Rather than relying on TRIM21, we could also adapt the strategy behind the deGradFP method[72] and its variants[50,73,74] for antibody-induced protein degradation. deGradFP uses an anti-GFP nanobody fused to a domain that recruits protein degradation machinery to deplete GFP-tagged proteins[72]. In our cytosolic antibody delivery platform, pAbBD can be easily engineered to be fused to both an ApP as well as a protein degradation domain. Thus, rather than exogenously expressing proteins to use deGradFP, we can simply deliver our modified IgGs via an optimized cationic delivery lipid formulation to induce target protein degradation. We can also fuse protein degradation domains to nanobodies or antibody-like binding proteins to turn them into protein degraders as well.

Initial candidate protein degradation domains to test should include the TRIM21[45], nanoCHIP[74], VHL[75], and IpaH9.8[76] E3 ligases, all of which have been shown to be capable of degrading target proteins bound by IgGs, nanobodies, or monobodies. To ensure successful target protein degradation, two factors must be considered. First, a productive ternary complex must be formed between the target protein, the degradative antibody, and the endogenous protein degradation machinery prior to target protein degradation[68,69,77]. This may require significant optimization of linker lengths between the degradation domain and the pAbBD[78]. Second, it is important to ensure that the protein degradation domain does not also rapidly degrade the IgG that it is linked to. Previous studies suggest that IgGs are at least resistant to TRIM21-induced degradation[45], but it is unclear whether other E3 ligases might significantly reduce cytosolic IgG half-life. If either factors preclude IgG-dependent protein degradation, a nanobody or an antibody-like binding protein should be tested. With a smaller protein scaffold, ternary complex formation may be simpler to engineer. Also, with small proteins, it may be

possible to mutate any solvent-accessible lysines to arginines in order to increase their resistance to ubiquitination and degradation[50].

Finally, cytosolic antibodies could also be utilized to degrade membrane proteins. Although antibodies can naturally inhibit membrane proteins by binding to an extracellular domain, membrane protein may also be present in high concentrations within intracellular compartments. Degradation allows for depletion of both the surface as well as the intracellular pool of membrane proteins. Using protein degradation domains outlined above to ubiquitinate the cytosolic domains of membrane proteins can mark them for lysosomal sorting and degradation[79]. Alternatively, cytosolically delivered antibodies could also be labeled with lysosomal sorting motifs (LSM) to induce ubiquitin-independent lysosomal sorting for membrane proteins[79–81]. Candidate LSMs include HSC70-binding motifs[82,83] and the HIP1R sorting signal[81].

5.3.6 *In vivo* delivery

Currently, we are using commercially available cationic transfection lipids to cytosolically deliver antibodies and antibody-like binding proteins. Although they are very effective in cell culture and could be extremely useful for interrogating cell biology, ultimately, cytosolic protein delivery needs to be achieved *in vivo* to unlock its therapeutic potential. We cannot use the commercially available formulations that we have tested thus far for *in vivo* studies because they form complexes with a high positive charge (zeta potential) and are not labeled with polyethylene glycol (PEG), which results in rapid *in vivo* clearance[84]. Thus, we must identify formulations that are compatible with our protein complexation strategy and capable of efficiently delivering proteins into the cytosol of cells within desired tissues.

The splitGFP reporter system has been extremely useful for monitoring cytosolic protein delivery in cell culture, but is not ideal for monitoring cytosolic protein delivery *in vivo* because splitGFP fluorescence penetrates poorly through tissue. Instead, a split NanoLuc luciferase (NanoBiT)[11] should be used to enable detection of cytosolic protein delivery by luminescence *in*

vivo. Similarly to splitGFP, the small half of the NanoBiT – HiBiT – can be fused to the protein cargo while the large half of the NanoBiT – LgBiT – can be expressed in only the cytosol of reporter cells[11]. HiBiT binds to LgBiT with a K_D of 0.7nM, enabling rapid reconstitution of luciferase activity whenever HiBiT-labeled protein cargo enters the cytosol of reporter cells[11]. The NanoBiT reporter system can be easily implemented for detecting cytosolic protein delivery to tumors in xenograft models by directly engineering cancer cell lines. Alternatively, adeno-associated viruses can be used to express LgBiT in the cytosol of cells for non-tumor delivery.

Modern lipid-based non-viral nucleic acid delivery systems typically consists of ionizable lipids, cholesterol, helper lipids, and PEGylated lipids[85,86]. Cationic lipids enable easy complexation with nucleic acids and have enhanced cellular uptake, but they result in the formation of positively charged nanocarriers, which is associated with increased toxicity and reduced efficacy *in vivo*[85,86]. Ionizable lipids are designed to be positively charged at low pH, but neutral at physiological pH. Thus, nucleic acids can be complexed with ionizable lipids under low pH conditions to form nanocarriers that have a relatively neutral charge at physiological pH[85,86]. Both cholesterol and helper lipids promote nanocarrier stability and increases delivery efficiency[87]. Finally, PEGylated lipids promote storage stability and can increase circulation time by reducing unwanted cellular uptake[85,86].

Currently, the field of non-viral nucleic acid delivery is filled with promising lipid-based delivery formulations found to be highly efficacious in preclinical models[85,88–92]. It remains to be seen how translatable previously optimized siRNA or mRNA delivery formulations are to proteins. Given that siRNA delivery formulations have to be re-optimized for mRNA delivery[93], significant re-optimization or screening of lipid libraries may be required to identify efficient formulations for cytosolic protein delivery. It may be possible to conduct large *in vivo* screens for cytosolic protein delivery by ligating nucleic acid barcodes or polymer-based elemental tags[94] onto cargo proteins via PBSL. Barcoded cargo proteins could then be complexed with different delivery formulations

to create a library of formulations complexed with barcoded cargos. Finally, the entire library could be administered to a single mouse and then the best formulations for delivery into specific tissues could be identified via deep sequencing[95,96] for nucleic acid barcodes or inductively coupled plasma mass cytometry (ICP-MS)[97,98] for polymer-based elemental tags.

5.4 Concluding remarks

Therapeutic cytosolic protein delivery has the potential to dramatically expand the druggable proteome. We have shown that by reducing their net charge, IgG antibodies, nanobodies, and other antibody-like binding proteins can be complexed with cationic delivery lipids via electrostatic interactions. This simple complexation strategy should enable the screening and identification of advanced delivery formulations for cytosolic protein delivery *in vivo*. If *in vivo* delivery can be achieved, not only can classically undruggable proteins be targeted with cytosolic proteins, but selective inhibition of proteins with specific post-translational modifications could also be feasible. Finally, by integrating concepts from synthetic biology[99,100], it may even be possible to design therapeutic cytosolic proteins that are activated in response to specific cellular states.

5.5 References

1. Antos, J. M., Truttmann, M. C. & Ploegh, H. L. Recent advances in sortase-catalyzed ligation methodology. *Curr. Opin. Struct. Biol.* **38**, 111–118 (2016).
2. Heck, T., Pham, P. H., Hammes, F., Thöny-Meyer, L. & Richter, M. Continuous monitoring of enzymatic reactions on surfaces by real-time flow cytometry: Sortase a catalyzed protein immobilization as a case study. *Bioconjug. Chem.* **25**, 1492–1500 (2014).
3. Heck, T., Pham, P. H., Yerlikaya, A., Thöny-Meyer, L. & Richter, M. Sortase A catalyzed reaction pathways: A comparative study with six SrtA variants. *Catal. Sci. Technol.* **4**, 2946–2956 (2014).
4. 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. *Anal. Chem.* **85**, 11090–11097 (2013).
5. Wang, H. H., Altun, B., Nwe, K. & Tsourkas, A. Proximity-Based Sortase-Mediated Ligation.

Angew. Chem. Int. Ed. **56**, 5349–5352 (2017).

6. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E690-7 (2012).
7. Li, L., Fierer, J. O., Rapoport, T. A. & Howarth, M. Structural analysis and optimization of the covalent association between SpyCatcher and a peptide tag. *J. Mol. Biol.* **426**, 309–317 (2014).
8. Structural Genomics Consortium *et al.* Protein production and purification. *Nat. Methods* **5**, 135–46 (2008).
9. Ohana, R. F. *et al.* HaloTag-based purification of functional human kinases from mammalian cells. *Protein Expr. Purif.* **76**, 154–164 (2011).
10. Rybak, J. N., Scheurer, S. B., Neri, D. & Elia, G. Purification of biotinylated proteins on streptavidin resin: A protocol for quantitative elution. *Proteomics* **4**, 2296–2299 (2004).
11. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem. Biol.* **11**, 400–408 (2016).
12. Peraro, L. & Kritzer, J. A. Emerging Methods and Design Principles for Cell-Penetrant Peptides. *Angew. Chem. Int. Ed.* **57**, 11868–11881 (2018).
13. Copolovici, D. M., Langel, K., Eriste, E. & Langel, Ü. Cell-Penetrating Peptides: Design, Synthesis, and Applications. *ACS Nano* **8**, 1972–1994 (2014).
14. Wadia, J. S., Stan, R. V. & Dowdy, S. F. Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* **10**, 310–315 (2004).
15. Lönn, P. & Dowdy, S. F. Cationic PTD/CPP-mediated macromolecular delivery: charging into the cell. *Expert Opin. Drug Deliv.* **12**, 1627–1636 (2015).
16. Erazo-Oliveras, A. *et al.* Protein delivery into live cells by incubation with an endosomolytic agent. *Nat. Methods* **11**, 861–867 (2014).
17. Erazo-Oliveras, A. *et al.* The Late Endosome and Its Lipid BMP Act as Gateways for Efficient Cytosolic Access of the Delivery Agent dfTAT and Its Macromolecular Cargos. *Cell Chem. Biol.* **23**, 598–607 (2016).
18. Najjar, K. *et al.* Unlocking Endosomal Entrapment with Supercharged Arginine-Rich Peptides. *Bioconjug. Chem.* **28**, 2932–2941 (2017).
19. Brock, D. J. *et al.* Efficient cell delivery mediated by lipid-specific endosomal escape of supercharged branched peptides. *Traffic* **19**, 421–435 (2018).
20. Subbarao, N. K., Parente, R. A., Szoka, F. C., Nadasdi, L. & Poneracz, K. pH-Dependent Bilayer Destabilization by an Amphipathic Peptide. *Biochemistry* **26**, 2964–2972 (1987).
21. Murata, M., Takahashi, S., Kagiwada, S., Suzuki, A. & Ohnishi, S. pH-Dependent membrane

fusion and vesiculation of phospholipid large unilamellar vesicles induced by amphiphilic anionic and cationic peptides. *Biochemistry* **31**, 1986–1992 (1992).

22. Akishiba, M. *et al.* Cytosolic antibody delivery by lipid-sensitive endosomolytic peptide. *Nat. Chem.* **9**, 751–761 (2017).
23. Lee, Y. J., Johnson, G. & Pellois, J. P. Modeling of the endosomolytic activity of HA2-TAT peptides with red blood cells and ghosts. *Biochemistry* **49**, 7854–7866 (2010).
24. Michiue, H. *et al.* The NH 2 Terminus of Influenza Virus Hemagglutinin-2 Subunit Peptides Enhances the Antitumor Potency of Polyarginine-mediated p53 Protein Transduction. *J. Biol. Chem.* **280**, 8285–8289 (2005).
25. Li, M. *et al.* Discovery and Characterization of a Peptide That Enhances Endosomal Escape of Delivered Proteins in Vitro and in Vivo. *J. Am. Chem. Soc.* **137**, 14084–14093 (2015).
26. Koren, E. & Torchilin, V. P. Cell-penetrating peptides: Breaking through to the other side. *Trends Mol. Med.* **18**, 385–393 (2012).
27. Wender, P. a *et al.* The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters. *Proc. Natl. Acad. Sci.* **97**, 13003–13008 (2000).
28. Rothbard, J. B., Jessop, T. C., Lewis, R. S., Murray, B. A. & Wender, P. A. Role of membrane potential and hydrogen bonding in the mechanism of translocation of guanidinium-rich peptides into cells. *J. Am. Chem. Soc.* **126**, 9506–9507 (2004).
29. Jiang, T. *et al.* Tumor imaging by means of proteolytic activation of cell-penetrating peptides. *Proc. Natl. Acad. Sci.* **101**, 17867–17872 (2004).
30. Vivès, E., Brodin, P. & Lebleu, B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J. Biol. Chem.* **272**, 16010–16017 (1997).
31. Futaki, S. *et al.* Arginine-rich Peptides. *J. Biol. Chem.* **276**, 5836–5840 (2001).
32. Derossi, D., Joliot, A. H., Chassaing, G. & Prochiantz, A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* **269**, 10444–50 (1994).
33. Soomets, U. *et al.* Deletion analogues of transportan. *Biochim. Biophys. Acta* **1467**, 165–176 (2000).
34. Klauschenz, E. *et al.* Cellular uptake of an α-helical amphipathic model peptide with the potential to deliver polar compounds into the cell interior non-endocytically. *Biochim. Biophys. Acta* **1414**, 127–139 (2002).
35. Plank, C., Oberhauser, B., Mechtler, K., Koch, C. & Wagner, E. The influence of endosome-disruptive peptides on gene transfer using synthetic virus-like gene transfer systems. *J. Biol.*

Chem. **269**, 12918–12924 (1994).

36. Plückthun, A. Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).
37. Yu, X., Yang, Y.-P., Dikici, E., Deo, S. K. & Daunert, S. Beyond Antibodies as Binding Partners: The Role of Antibody Mimetics in Bioanalysis. *Annu. Rev. Anal. Chem.* **10**, 293–320 (2017).
38. Deprey, K., Becker, L., Kritzer, J. & Plückthun, A. Trapped! A Critical Evaluation of Methods for Measuring Total Cellular Uptake versus Cytosolic Localization. *Bioconjug. Chem.* **30**, 1006–1027 (2019).
39. Kauffman, W. B., Guha, S. & Wimley, W. C. Synthetic molecular evolution of hybrid cell penetrating peptides. *Nat. Commun.* **9**, (2018).
40. Schmidt, S. *et al.* Identification of Short Hydrophobic Cell-Penetrating Peptides for Cytosolic Peptide Delivery by Rational Design. *Bioconjug. Chem.* **28**, 382–389 (2017).
41. Lönn, P. *et al.* Enhancing Endosomal Escape for Intracellular Delivery of Macromolecular Biologic Therapeutics. *Sci. Rep.* **6**, 1–9 (2016).
42. Schmidt, S. *et al.* Detecting Cytosolic Peptide Delivery with the GFP Complementation Assay in the Low Micromolar Range. *Angew. Chem. Int. Ed.* **54**, 15105–15108 (2015).
43. Milech, N. *et al.* GFP-complementation assay to detect functional CPP and protein delivery into living cells. *Sci. Rep.* **5**, 1–11 (2015).
44. Hui, J. Z., Tamsen, S., Song, Y. & Tsourkas, A. LASIC: Light Activated Site-Specific Conjugation of Native IgGs. *Bioconjug. Chem.* **26**, 1456–1460 (2015).
45. Clift, D. *et al.* A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell* **171**, 1692-1706.e18 (2017).
46. Yamaizumi, M., Uchida, T., Mekada, E. & Okada, Y. Antibodies introduced into living cells by red cell ghosts are functionally stable in the cytoplasm of the cells. *Cell* **18**, 1009–1014 (1979).
47. Zavortink, M., Thacher, T. & Rechsteiner, M. Degradation of proteins microinjected into cultured mammalian cells. *J. Cell. Physiol.* **100**, 175–185 (1979).
48. Kim, Y. B., Zhao, K. T., Thompson, D. B. & Liu, D. R. An anionic human protein mediates cationic liposome delivery of genome editing proteins into mammalian cells. *Nat. Commun.* **10**, 2905 (2019).
49. Moghimi, S. M., Hunter, A. C. & Andresen, T. L. Factors Controlling Nanoparticle Pharmacokinetics: An Integrated Analysis and Perspective. *Annu. Rev. Pharmacol. Toxicol.* **52**, 481–503 (2012).
50. Daniel, K. *et al.* Conditional control of fluorescent protein degradation by an auxin-

dependent nanobody. *Nat. Commun.* **9**, 1–11 (2018).

51. Lendel, C., Dogan, J. & Härd, T. Structural Basis for Molecular Recognition in an Affibody:Affibody Complex. *J. Mol. Biol.* **359**, 1293–1304 (2006).
52. Guillard, S. *et al.* Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange. *Nat. Commun.* **8**, 16111 (2017).
53. Li, S., Balmain, A. & Counter, C. M. A model for RAS mutation patterns in cancers: finding the sweet spot. *Nat. Rev. Cancer* **18**, 767–777 (2018).
54. Simanshu, D. K., Nissley, D. V. & McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17–33 (2017).
55. Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: Weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774 (2011).
56. Rajalingam, K., Schreck, R., Rapp, U. R. & Albert, Š. Ras oncogenes and their downstream targets. *Biochim. Biophys. Acta - Mol. Cell Res.* **1773**, 1177–1195 (2007).
57. Goody, R. S., Müller, M. P. & Rauh, D. Mutant-Specific Targeting of Ras G12C Activity by Covalently Reacting Small Molecules. *Cell Chem. Biol.* (2019). doi:10.1016/j.chembiol.2019.07.005
58. Bery, N. *et al.* KRAS-specific inhibition using a DARPin binding to a site in the allosteric lobe. *Nat. Commun.* **10**, 2607 (2019).
59. Robert, C. *et al.* Improved overall survival in melanoma with combined dabrafenib and trametinib. *N. Engl. J. Med.* **372**, 30–39 (2015).
60. Tanaka, T., Williams, R. L. & Rabbitts, T. H. Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. *EMBO J.* **26**, 3250–3259 (2007).
61. Spencer-Smith, R. *et al.* Inhibition of RAS function through targeting an allosteric regulatory site. *Nat. Chem. Biol.* **13**, 62–68 (2017).
62. Whitfield, J. R., Beaulieu, M.-E. & Soucek, L. Strategies to Inhibit Myc and Their Clinical Applicability. *Front. Cell Dev. Biol.* **5**, 10 (2017).
63. McKeown, M. R. & Bradner, J. E. Therapeutic strategies to inhibit MYC. *Cold Spring Harb. Perspect. Med.* **4**, (2014).
64. Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22–35 (2012).
65. Soucek, L. *et al.* Design and properties of a Myc derivative that efficiently homodimerizes. *Oncogene* **17**, 2463–2472 (1998).
66. Soucek, L. *et al.* Modelling Myc inhibition as a cancer therapy. *Nature* **455**, 679–683 (2008).
67. Beaulieu, M. E. *et al.* Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy. *Sci. Transl. Med.* **11**, 1–14 (2019).

68. Lai, A. C. & Crews, C. M. Induced protein degradation: An emerging drug discovery paradigm. *Nat. Rev. Drug Discov.* **16**, 101–114 (2017).
69. Churcher, I. Protac-Induced Protein Degradation in Drug Discovery: Breaking the Rules or Just Making New Ones? *J. Med. Chem.* **61**, 444–452 (2018).
70. James, L. C., Keeble, A. H., Khan, Z., Rhodes, D. A. & Trowsdale, J. Structural basis for PRYSPRY-mediated tripartite motif (TRIM) protein function. *Proc. Natl. Acad. Sci.* **104**, 6200–6205 (2007).
71. Sauer-Eriksson, A. E., Kleywegt, G. J., Uhlén, M. & Jones, T. A. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3**, 265–78 (1995).
72. Caussinus, E., Kanca, O. & Affolter, M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat. Struct. Mol. Biol.* **19**, 117–122 (2012).
73. Schmit, N. E., Neopane, K. & Hantschel, O. Targeted Protein Degradation through Cytosolic Delivery of Monobody Binders Using Bacterial Toxins. *ACS Chem. Biol.* **14**, 916–924 (2019).
74. Kanner, S. A., Morgenstern, T. & Colecraft, H. M. Sculpting ion channel functional expression with engineered ubiquitin ligases. *Elife* **6**, 1–26 (2017).
75. Fulcher, L. J. *et al.* An affinity-directed protein missile system for targeted proteolysis. *Open Biol.* **6**, (2016).
76. Ludwicki, M. B. *et al.* Broad-Spectrum Proteome Editing with an Engineered Bacterial Ubiquitin Ligase Mimic. *ACS Cent. Sci.* **5**, acscentsci.9b00127 (2019).
77. Burslem, G. M. & Crews, C. M. Small-Molecule Modulation of Protein Homeostasis. *Chem. Rev.* **117**, 11269–11301 (2017).
78. Smith, B. E. *et al.* Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase. *Nat. Commun.* **10**, 1–13 (2019).
79. Bonifacino, J. S. & Traub, L. M. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72**, 395–447 (2003).
80. Braulke, T. & Bonifacino, J. S. Sorting of lysosomal proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **1793**, 605–614 (2009).
81. Wang, H. *et al.* HIP1R targets PD-L1 to lysosomal degradation to alter T cell–mediated cytotoxicity. *Nat. Chem. Biol.* **15**, 42–50 (2019).
82. Bauer, P. O. *et al.* Harnessing chaperone-mediated autophagy for the selective degradation of mutant huntingtin protein. *Nat. Biotechnol.* **28**, 256–263 (2010).
83. Fan, X., Jin, W. Y., Lu, J., Wang, J. & Wang, Y. T. Rapid and reversible knockdown of endogenous proteins by peptide-directed lysosomal degradation. *Nat. Neurosci.* **17**, 471–480 (2014).

84. Ernsting, M. J., Murakami, M., Roy, A. & Li, S. D. Factors controlling the pharmacokinetics, biodistribution and intratumoral penetration of nanoparticles. *J. Control. Release* **172**, 782–794 (2013).
85. Hajj, K. A. & Whitehead, K. A. Tools for translation: Non-viral materials for therapeutic mRNA delivery. *Nat. Rev. Mater.* **2**, 1–17 (2017).
86. Kanasty, R., Dorkin, J. R., Vegas, A. & Anderson, D. Delivery materials for siRNA therapeutics. *Nat. Mater.* **12**, 967–977 (2013).
87. Cheng, X. & Lee, R. J. The role of helper lipids in lipid nanoparticles (LNPs) designed for oligonucleotide delivery. *Adv. Drug Deliv. Rev.* **99**, 129–137 (2016).
88. Akinc, A. *et al.* A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. *Nat. Biotechnol.* **26**, 561–569 (2008).
89. Love, K. T. *et al.* Lipid-like materials for low-dose, in vivo gene silencing. *Proc. Natl. Acad. Sci.* **107**, 1864–1869 (2010).
90. Semple, S. C. *et al.* Rational design of cationic lipids for siRNA delivery. *Nat. Biotechnol.* **28**, 172–176 (2010).
91. Whitehead, K. A. *et al.* Degradable lipid nanoparticles with predictable in vivo siRNA delivery activity. *Nat. Commun.* **5**, 1–10 (2014).
92. Kranz, L. M. *et al.* Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* **534**, 396–401 (2016).
93. Kauffman, K. J. *et al.* Optimization of Lipid Nanoparticle Formulations for mRNA Delivery in Vivo with Fractional Factorial and Definitive Screening Designs. *Nano Lett.* **15**, 7300–7306 (2015).
94. Lou, X. *et al.* Polymer-based elemental tags for sensitive bioassays. *Angew. Chem. Int. Ed.* **46**, 6111–6114 (2007).
95. Dahlman, J. E. *et al.* Barcoded nanoparticles for high throughput in vivo discovery of targeted therapeutics. *Proc. Natl. Acad. Sci.* **114**, 2060–2065 (2017).
96. Sago, C. D. *et al.* High-throughput in vivo screen of functional mRNA delivery identifies nanoparticles for endothelial cell gene editing. *Proc. Natl. Acad. Sci.* **115**, E9944–E9952 (2018).
97. Bendall, S. C. *et al.* Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science* **332**, 687–696 (2011).
98. Spitzer, M. H. & Nolan, G. P. Mass Cytometry: Single Cells, Many Features. *Cell* **165**, 780–791 (2016).
99. Langan, R. A. *et al.* De novo design of bioactive protein switches. *Nature* (2019). doi:10.1038/s41586-019-1432-8

100. Ng, A. H. *et al.* Modular and tunable biological feedback control using a de novo protein switch. *Nature* **9**, (2019).