The Dissertation Committee for Kimberly Elizabeth Bowen Certifies that this is the approved version of the following Dissertation:

DESIGNING PROTEASOME ADAPTORS TO DEPLETE SPECIFIC PROTEINS FROM CELLS

Committee:

Andreas Matouschek, Supervisor

Jon Huibregtse

Kevin Dalby

George Georgiou

Blerta Xhemalce

DESIGNING PROTEASOME ADAPTORS TO DEPLETE SPECIFIC PROTEINS FROM CELLS

by

Kimberly Elizabeth Bowen

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2019

Dedication

This dissertation is dedicated to my mom Carol and my sister Stephanie, whose love and support helped me through my PhD.

Acknowledgements

First and foremost, I would like to thank my supervisor, Andreas Matouschek, for his continuous support throughout my time here. I have learned so much under his tutelage and I am grateful that he gave me the opportunity to work in his lab. I would also like to thank my committee members, Dr. Kevin Dalby, Dr. George Georgiou, Dr. Jon Huibretsge, and Dr. Blerta Xhemalce, for their insightful comments and helpful tips at every meeting.

Thank you to my labmates, both past and present (in no particular order): Kirby, Houqing, Amit (and his wife Poulami and beautiful daughter Pihu), Caroline (who helped bring the lab closer together as the social chair), John, Logan, Takuya (especially for his help with my experiments), Kazu, and Frank. A special thanks to Grace Kago, who started my project and kept the lab running smoothly for my first few years. In particular, I want to thank Shameika Wilmington, whose unending patience and willingness to share her vast expertise helped me get through my PhD. I am also very appreciative of my undergrad, Xena, whom I mentored for several years. Her constant optimism, enthusiasm, and great talent were refreshing and kept me excited about my project.

Thank you also to all my friends here in Austin and back at home in Boston. There are too many to list, but you know who you are and I love you all. Special thanks to Pratibha Saxena (and her husband) for her kind words, understanding while I was TAing, and her wisdom. I will never forget how I met her, when I was in tears waiting for the decision on whether I passed my part II quals, and she came up and gave me a hug and told me everything would be alright, having never even met me. She has made an enormous impression on me and I am very grateful.

Particular thanks to Po-Tsan Ku, the graduate career counselor. His absolute faith in me and perpetual optimism were instrumental in pushing me forward. Meetings with him were akin to therapy sessions, and I never failed to feel better and lighter after our sessions. He is one of the best resources at the university, and I feel lucky to have worked with him.

As for my family, my mother and sister have boosted me up when I've felt down, shared in my joys and woes and have been my pillars of support for all my life—I am forever grateful to them. I also appreciate Vic and Mark, who have made them so happy and been very encouraging in their own right. Further thanks to my grandmother Leila (who bought her graduation tickets well ahead of time and thus kept me to my timeline), Uncle Brian, Aunt Cheryl and Uncle Richard, Joseph and Amanda, Auntie Christine and Uncle Bill, Andrew and the girls, and all of the family in Australia.

None of this would have possible without the love and support from above from my father, Les. His passing of cancer was the impetus for me to apply and attend graduate school so that I could continue the fight on his behalf. Not a day goes by that I don't think of him and channel him in my research. I wish he could see me now and share in the joy, but I know he is proud looking down on me.

Another special thanks to Matija, whom I met late in my grad school career but who has helped keep me grounded and focused on the end goal. His love, support, and ability to make me laugh have been instrumental in keeping me sane and helping me stay the course.

Abstract

Designing proteasome adaptors to deplete specific proteins from cells

Kimberly Elizabeth Bowen, Ph.D.

The University of Texas at Austin, 2019

Supervisor: Andreas Matouschek

Cellular protein levels are governed by their rates of synthesis and degradation, and both processes are intricately regulated. One way to study the role of a protein in the cell is to artificially deplete it and observe the effects. The most common methodology for depleting proteins inhibits expression of the target through RNA interference. However, this technique acts at the protein synthesis level and cannot be used to study long-lived proteins or post-translational modifications. Thus, a complementary approach that acts on the proteins themselves would be useful. One method is to target a protein to the cell's degradation machinery, the Ubiquitin Proteasome System (UPS). Proteins are targeted to the proteasome by the covalent attachment of ubiquitin molecules, which are recognized by the proteasome. The substrate is then translocated into the proteasome's proteolytic core and degraded into short peptides, while the ubiquitin molecules are cleaved off and recycled. Recently, methods have been developed to deplete proteins by inducing their ubiquitination, which accelerates their degradation by the proteasome. Ubiquitin is a signaling molecule for numerous cellular pathways other than proteolysis, however, and inducing ubiquitination does not always lead to degradation. Therefore, I have developed a system to degrade specific proteins in cells using chimeric adaptors that

shuttle proteins directly to the proteasome without the need for ubiquitination. I have shown that this system can be successfully applied to several proteins *in vitro* and that the adaptors can induce degradation of model and endogenous proteins in cells.

Table of Contents

List of Figures

Chapter 1: Introduction

THE UBIQUITIN PROTEASOME SYSTEM

Cellular protein concentrations are determined by their rates of synthesis and degradation. Most of the protein degradation in eukaryotic cells is controlled by the ubiquitin proteasome system (UPS). The UPS plays a central role in various cellular pathways, including cell cycle control, stress response, and signal transduction. At the center of the UPS is the proteasome complex, which is responsible for degrading hundreds of proteins in the cell, including misfolded and damaged proteins as well as regulatory proteins. Proteins are targeted to the proteasome by the covalent attachment of ubiquitin molecules, which are recognized by the proteasome. The substrate is then translocated into the proteasome's proteolytic core and degraded into short peptides, while the ubiquitin molecules are cleaved off and recycled[1]. In this section I will describe the proteasome's structure and targeting of proteins to the proteasome for degradation. For simplicity, I will use subunit names corresponding to the *Saccharomyces cerevisiae* proteasome.

Structure of the 26S proteasome

The proteasome is a ~2.5 megadalton macromolecular machine consisting of over 30 subunits. It exists predominantly in the 26S form which contains a 20S core particle, where the proteolytic activity is located, and either one or two 19S regulatory particles, where target proteins bind and are translocated into the core (Figure 1)[2,3]. The 19S particle can be further divided structurally into two subcomplexes called the base and the lid (Figure 1). The base is located on top of the core particle and contains ubiquitin-binding subunits as well as translocation machinery. The lid provides structural support

by wrapping around one side of the base and also contains the deubiquitinating enzyme (DUB) Rpn11[2]. Many other proteins make transient interactions with the proteasome as well, including assembly chaperones[4], non-stoichiometric DUBs[5,6], and ubiquitin-like domain (UBL) proteins such as Rad23[7,8].

When a protein is targeted for degradation, it is conjugated with a series of ubiquitin molecules which serve as a proteasome binding tag. The proteasome contains at least three receptors for ubiquitin: Rpn1, Rpn10, and Rpn13[2,9-11] (Figure 1). These receptors are located some distance apart from each other, potentially allowing for recognition of substrates with varying ubiquitin chain lengths and conformations[2]. Once a ubiquitinated protein binds to the proteasome, the protein is unfolded and translocated to the core through a narrow channel in the base, which contains a ring of six ATPases associated with various cellular activities (AAA+). In order to initiate translocation, the substrate protein must contain an unstructured region (or initiation region) where the AAA+ motor can engage the polypeptide with enough grip and force to unfold the substrate[12,13]. Interestingly, the sequence composition and amino acid properties of the initiation region affect whether a protein will be degraded. For example, biased amino acid sequences confer stability to a substrate[14], while higher hydrophobicity and stiffness correlate with increased degradation efficiency[12]. Many regulatory proteins contain disordered regions, and mutating or deleting these regions greatly influences the proteins' half-life[15].

Upon engagement, the base undergoes a dramatic conformational shift, which aligns the translocation channel above the core particle and positions the DUB Rpn11 directly over the channel where it cleaves off the ubiquitin chain from the substrate[2,16,17]. As the substrate is unfolded, coordinated ATP hydrolysis drives conformational changes of the AAA+ motors, which contain pore loops that interact with the substrate and

mechanically "pull" the target through the channel[3,18]. The direct interactions between motors and the substrate could explain the proteasome's sequence preferences, as the affinity of the base to the pore loops would affect their ability to translocate the substrate[14].

After translocation through the base, the substrate enters the 20S core particle through a narrow channel opening. Access to this opening, or gate, is tightly controlled such that it only opens upon RP binding and initiation of translocation[19,20]. This regulation helps prevent cleavage of functional proteins in the cell. The core consists of four heteroheptameric rings which form a barrel shape with a large channel in the center. Two of the rings are composed of homologous α subunits and they flank the central β subunit rings where the protease activity occurs (Figure 1). Three of the β subunits contain proteolytic active sites with trypsin-, chymotrypsin-, or caspase-like activities, which cleave the substrate protein into small peptides of approximately three to eight amino acids each[21]. The active sites in the core are positioned facing inwards so that only the unfolded substrate is degraded, further protecting folded cellular proteins[22].

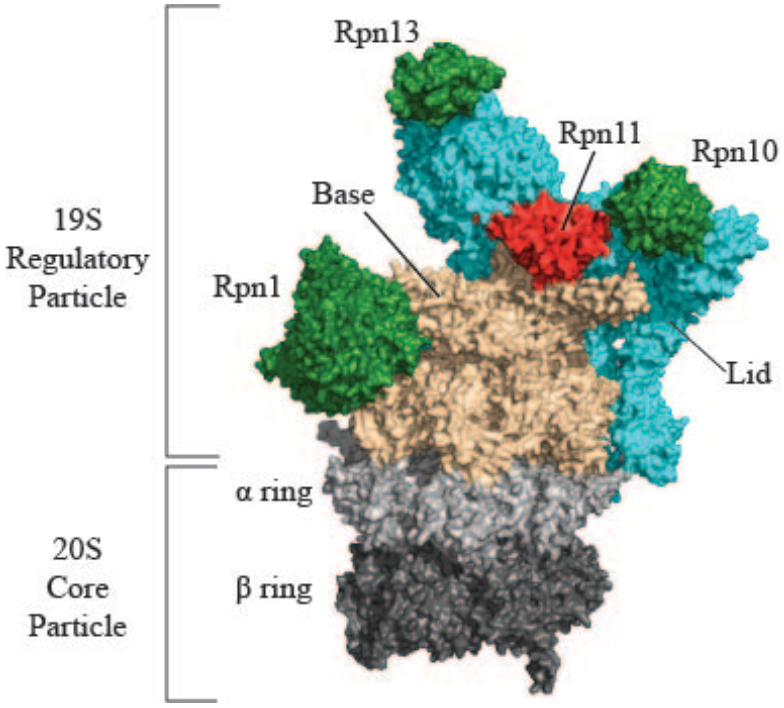


Figure 1: Structure of the 26S proteasome.

Half of the *S. cerevisiae* 26S proteasome, with one 19S regulatory particle atop the two rings in the 20S core particle. In green are the three known stoichiometric ubiquitin receptors; in beige is the base, containing the AAA+ ATPases which translocate the substrate; cyan denotes the lid, with the DUB Rpn11 in red; in light and dark grey are the α and β rings of the core particle, respectively. PDB: 4CR2.

Targeting to the proteasome

As mentioned earlier, proteins to be degraded are usually covalently tagged with one or more ubiquitin chains that bind to the proteasome. Ubiquitin is a small, ~8kDa protein that is highly evolutionarily conserved amongst eukaryotes, with only three amino acid differences between yeast and humans. Ubiquitin contains seven lysine residues where other ubiquitin molecules can be attached. Typically, a ubiquitin molecule is attached through its C-terminus to a lysine residue in the target protein, although in rare cases a cysteine residue or the N-terminus can serve as the starting point for the chain[23-26]. The presence of multiple locations for attachment within ubiquitin enables a variety of mono- or polyubiquitin patterns and geometries to be formed, allowing for a diverse "code" that is interpreted by many different players in the cell. Indeed, ubiquitination signals are used in myriad cellular pathways, such as membrane trafficking, DNA damage repair, and transcriptional regulation in addition to proteasomal targeting[27-31].

How each linkage type affects the fate of a target remains elusive. All linkage types have been observed in cells, but their detailed function is still unclear[32]. The canonical proteasome degradation signal is a polyubiquitin chain attached through K48[33,34]. The minimum requirement for degradation is four K48-linked ubiquitins[35]. K11-linked ubiquitin chains are associated with endoplasmic reticulum-associated degradation (ERAD), but usually exist in heterotypic chains, with other linkage types present, and do not lead to degradation *in vitro* as homotypic chains[32,36]. K63 chains are known to bind to the proteasome, but whether they lead to degradation is debated[4,37]. Met1-linked chains have been implicated in immune receptor signaling[38]. The roles of K6, K27, K29, and K33 chains are not fully understood[39].

Proteins are ubiquitinated through a cascade of enzymes that activate and covalently attach ubiquitin molecules onto the substrate. In most cases, three types of

enzymes are used: Enzyme 1 (E1), the ubiquitin-activating enzyme; E2, ubiquitin-conjugating enzyme; and E3, the ubiquitin ligase enzyme. First an E1 enzyme activates ubiquitin by forming a thioester bond with the C-terminus of ubiquitin, which induces a conformational change in the E1 that allows for binding to an E2 conjugating enzyme. The ubiquitin is transferred to the E2 enzyme and then to the E3 ubiquitin ligase enzyme, which attaches the C-terminus of ubiquitin to a lysine residue on the substrate protein[40-42]. In general, the E2 enzyme defines which of the seven lysine residues to attach the next ubiquitin molecule to (linkage type), while the E3 confers specificity to the target substrate[41,42]. Occasionally, an additional enzyme E4 plays a role in ubiquitination at multiple sites on the target (multiubiquitination)[43].

In eukaryotes, there are only one or two E1 enzymes and tens of E2s, but there are hundreds of E3s that recognize a diverse array of substrate proteins[44,45]. E3 enzymes can be classified into two structurally distinct categories which contain either Really Interesting New Gene (RING) or homologous to the E6-AP carboxy terminus (HECT) domains. RING domain E3s bind both the substrate and the E2 enzyme, recruiting the E2 into proximity with the substrate and allowing the transfer of ubiquitin directly from the E2 to the substrate[44]. HECT domain E3s are active players in the cascade, where the ubiquitin is transferred to the E3 active site and from there attached to the substrate[46].

Ubiquitin-independent and adaptor-mediated targeting to the proteasome

While ubiquitin chains are the primary proteasome binding tags in the cell, ubiquitination of a protein is not required or sufficient in itself for degradation. Several studies have shown, that localization to the proteasome alone is sufficient for degradation both *in vitro* and *in vivo*[13,47,48], as in the case of antizyme-mediated degradation of ornithine decarboxylase (ODC). When ODC binds to antizyme, its C-terminus is exposed

and is recognized directly by the 26S proteasome for degradation[47,49]. Antizyme also mediates the degradation of several other proteins, including nucleolar c-Myc, cyclin D1, and Aurora-A in a ubiquitin-independent manner[50-52]. In these cases, the proteasome recognizes the target directly, without the need for any proteasome-binding tag.

In addition to direct proteasome interaction with the substrate, there are proteins that act as adaptors to localize substrates to the proteasome. One class of adaptor proteins called UBL-UBA proteins contains a ubiquitin-like (UBL) domain attached to one or more ubiquitin-associated (UBA) domains. UBL domains adopt a similar structural conformation to ubiquitin and are recognized by all three ubiquitin receptors in the proteasome[53,54]. In fact, a UBL-specific recognition site distinct from its ubiquitin binding site in Rpn1 has been mapped[11,53,54].

One UBL-UBA protein is yeast Rad23, which is involved in DNA damage repair and shuttling substrates to the proteasome[55]. Rad23 contains one UBL and two UBA domains that bind multiubiquitinated proteins[56,57]. In humans, there are two functionally somewhat distinct homologs of Rad23, called hHR23a and hHR23b (Rad23b). While the homologs are redundant in the DNA repair pathway[58], only the Rad23b UBL binds efficiently to the proteasome[59]. Because Rad23 binds to the proteasome and simultaneously to ubiquitinated proteins, it is believed to act as a shuttle factor that can localize specific targets to the proteasome for degradation[7]. Rad23 itself escapes degradation and is thought to recycle and act catalytically as a proteasome adaptor[60-62].

CONTROLLING PROTEIN CONCENTRATION THROUGH TARGETED DEGRADATION

The UPS controls the concentration of hundreds of cellular proteins, and it is possible to harness the proteasome's power to artificially tune the amount of specific proteins. Being able to control a protein's concentration in cells is useful for several

reasons. First, the most common methodology for studying protein function is by depleting the protein from cells to study the effects on relevant downstream pathways[63,64]. In addition, many diseases are caused or exacerbated by the overexpression of particular proteins[65-68] or the expression of mutated proteins[69-71]. This is the basis of molecularly-targeted therapies, where the inhibition or depletion of a particular disease-causing protein can treat or even cure patients. Therefore, development of novel technologies which alter protein concentration are always in demand.

RNA interference (RNAi) is widely used to control protein expression levels by preventing the protein's synthesis[72]. In this method, a short piece of RNA complementary (either completely for silencing RNA (siRNA) or with a few mismatches for microRNA (miRNA)) to the gene of interest is introduced into the cell, where it is processed to a single strand that binds the endogenous mRNA. This prevents translation by triggering degradation or sequestration of the mRNA or direct inhibition of the translational machinery[73-75]. In this way, synthesis of the protein is greatly reduced, so basic function of the protein can be inferred by changes in downstream pathways and new phenotypes[76]. RNAi has been used to screen for protein function as well as disease targets[72,77-79]. Advances in RNA delivery technology have also made RNAi an effective therapeutic agent, with one FDA-approved RNAi therapy and clinical trials underway for an array of diseases[80,81].

While RNA interference has been used successfully for studying protein function, it has several drawbacks as a research tool and therapeutic. For example, sequence similarities among different genes can lead to off-target base pairing, causing a miRNA-like silencing effect[82,83]. This can lead to side effects, toxicity, and false positives. Decreasing protein levels by RNAi also depends on the target's intrinsic degradation rate, making naturally long-lived proteins difficult to deplete with RNAi[84]. In addition, many

proteins are regulated through post-translational modifications (PTMs), which tune their activity, stability, and interaction network[85]. These effects cannot be studied via RNAi alone because it functions at the pre-translational level[85]. Therefore, complementary methods that act on the proteins themselves have been developed, such as targeting specific proteins to the UPS for degradation.

PROTACS: Using small molecules to induce ubiquitination

Several technologies have been introduced that decrease individual protein concentrations via the UPS. Protein-targeting chimeras (PROTACs) consist of a target-binding domain (usually a small molecule) connected through a linker to an E3-binding motif (Figure 2a). PROTACs bring a protein of interest into proximity with the E3, leading to the protein's ubiquitination and subsequent degradation[86,87]. Importantly, PROTAC-mediated degradation of the target protein is catalytic, meaning that one PROTAC molecule can lead to turnover of multiple proteins. Generally, the target-binding and E3-binding domains are small molecules that bind to an interface on the target protein or ubiquitin ligase[88]. The most common E3-binding ligand recognizes Von Hippel Lindau (VHL), which is part of an E3 ligase complex that normally targets HIF1α for degradation[89,90]. Another widely-used E3 ligase binding molecule is thalidomide (and its derivatives), which interacts with Cereblon (CRBN), an adaptor for the Cullin 4a E3 ligase complex[91,92]. A third ligand is ethyl bestatin, which binds to cellular inhibitor of apoptosis protein 1 (cIAP1), a protein that is overexpressed in certain cancers and contains an E3 ligase domain[93,94]. Finally, a PROTAC was recently developed that utilizes nutlin, a compound that binds the E3 ligase mouse double minute 2 (MDM2) and disrupts MDM2 binding to its natural substrate, p53[95]. The PROTAC created based on MDM2 actually serves double duty: it retains its ability to stabilize p53 and also induces

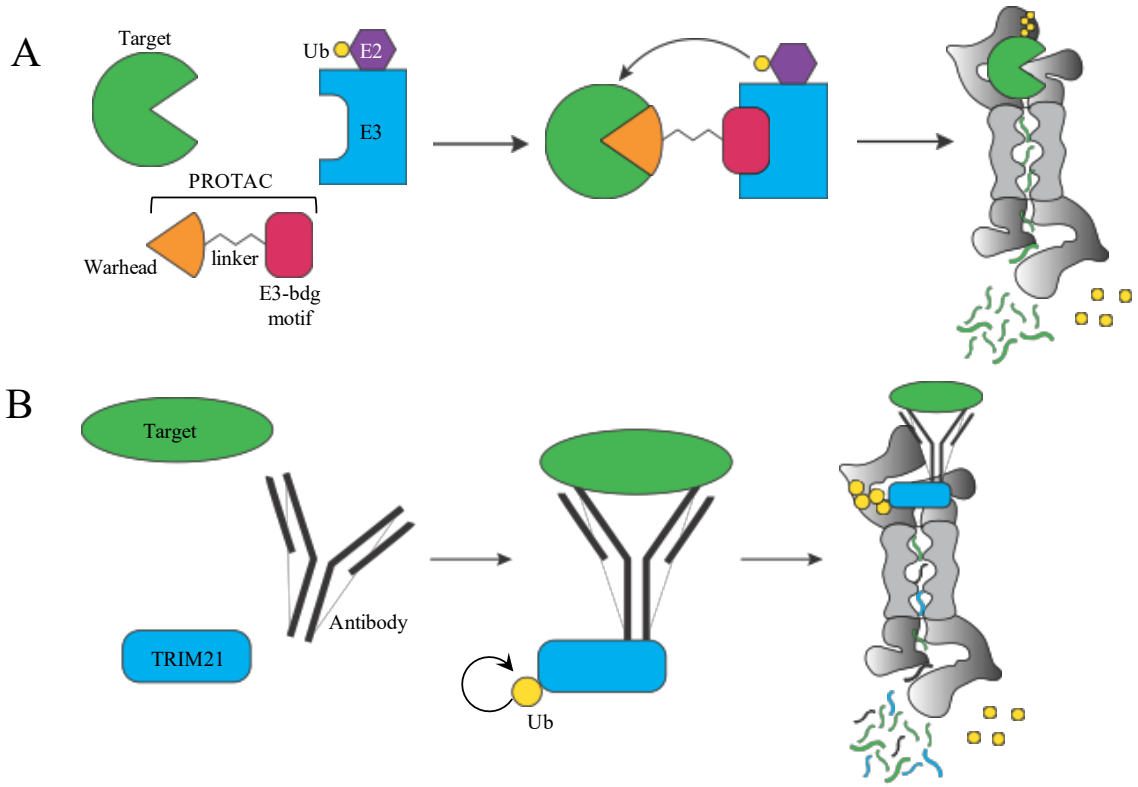


Figure 2: Induced ubiquitination and degradation of specific targets.

A) A PROTAC consists of a target binding molecule (warhead, orange) connected through a linker to an E3-binding motif (pink). The PROTAC bridges the target and E3, leading to ubiquitination of the target protein, which is then released and degraded by the proteasome. B) An antibody that binds a specific target is introduced into a cell. The E3 TRIM21 recognizes the antibody and autoubiquitinates. The whole complex is then degraded by the proteasome.

degradation of the target protein. Interestingly, changing just the E3 ligand (and therefore the recruited E3) can significantly affect degradation of a target protein. For example, a promiscuous kinase inhibitor that binds to many kinases degraded completely different sets of targets depending on whether it was attached to a VHL binder or a CRBN binder[96]. This is most likely due to the distinct structural conformations of the two E3s, which would affect the accessibility of the active site to the target[96,97].

PROTACs have been generated against a variety of target proteins classes, such as kinases, bromodomains, and hormone receptors[94,98]. Usually, the target-binding domain (also called warhead) is a previously-developed small molecule that recognizes the target protein with high affinity and specificity[99]. One PROTAC uses 4-hydroxytamoxifen, a molecule whose tight interaction with the oncoprotein estrogen receptor α (ERα) is well-established[93,100,101]. If there are no available molecules that bind to the target, high-throughput screens followed by extensive optimization are required to develop a target-binding compound[102,103]. Even with a warhead that binds tightly and selectively, the extent of PROTAC-mediated target degradation rarely correlates with warhead binding affinity profiles[96,104]. In other words, a high-affinity warhead does not necessarily lead to better degradation than a lower-affinity one. In fact, just because a molecule binds to a protein on its own does not mean it will induce degradation of the target at all when attached to the rest of the PROTAC. One study found a promiscuous kinase inhibitor that binds to hundreds of kinases on its own degraded less than half of a verified subset of targets when in PROTAC form, even though the PROTAC retained target-binding ability[96]. Surprisingly, the PROTACs that induced the highest depletion were not the tightest binders. This inconsistency makes the PROTAC development process difficult, because rational design and efficacy prediction based on established target binders is nearly impossible.

The last part of a PROTAC chimera is the linker between the warhead and E3 binding domain. The composition of the linker is extremely important, not only for degradation, but also for cell permeability and *in vivo* clearance of the PROTAC[86,105]. Initially, a peptidic linker was used, but low cell permeability and high cleavage rate led to small molecule linkers being favored[105]. Another requirement for the linker is enough length and flexibility to bridge the interaction between the target and E3 ligase and allow for sampling of the conformational landscape[104,106].

Because PROTACs are modular, all iterations of a PROTAC must be tested empirically. Each piece of the molecule (E3 binder, warhead, and linker) affects how well it binds to and ubiquitinates its targets[104,106-108]. When examining patterns to correlate binding efficiency with degradation, one study found that the major factor in ubiquitination ability was the formation of a stable ternary complex between the target, PROTAC, and E3 ligase[96,109]. The determining factor for efficient degradation was actually a stable protein-protein interaction between the target and the E3 ligase, rather than PROTAC affinity for either target or E3. Presumably, changing any of the PROTAC domains could alter the structure of the complex, which would change the availability of lysine residues in the target or active site in the E3[96,97]. Despite these challenges, PROTACs continue to be developed and have been moderately successful in preclinical studies[110]. In fact, the FDA recently approved the first clinical trial based on inducible degradation, a PROTAC that depletes androgen receptor for patients with metastatic prostate cancer[109].

"TRIM-Away": Using antibodies to bridge target and E3

Another recently-developed inducible degradation system takes advantage of an E3 ligase involved in the adaptive immune system, tri-partite motif-containing 21

(TRIM21), which recognizes the Fc region of antibodies and induces their degradation[111]. During viral infection, antibodies attach to virions and remain attached when the virions are endocytosed. In the cytosol, TRIM21 binds to the antibodies and auto-ubiquitinates. The whole complex, including the virus protein, then binds to the proteasome and is degraded[111,112].

Clift *et al* redirected this mechanism to deplete specific proteins from cells[112,113] (Figure 2b). In a system called "TRIM-Away", they overexpressed TRIM21 in cells, followed by microinjection with an antibody that binds to a protein of interest. Upon introduction of the antibody, concentration of the target protein rapidly decreased in a proteasome-dependent manner. Degradation of the target was efficient enough that phenotype changes could be readily observed. For example, in mouse oocytes overexpressing TRIM21, introduction of anti-Eg5, which targets a mitotic microtubule motor protein required for proper spindle polarity[114], led to improper spindle formation and failure to exit mitosis. Strikingly, the phenotypic effect was as pronounced as that of treatment with Eg5 inhibitor monastrol[115]. TRIM-Away was also effective at inducing the degradation of other proteins in cells, including Rec8, mTOR, and IKKα[116].

One limitation to the TRIM-Away system is that TRIM21 must be overexpressed in order to induce significant depletion[112]. The most likely explanation is that TRIM21 is degraded along with the antibody and target protein[111], so unlike PROTACs, which can turn over multiple target proteins[88], degradation in the TRIM-Away system is limited by the concentration of TRIM21 in the cell.

Other inducible degradation systems

Other methods have been developed for depleting proteins, but all of these systems require target protein modification. One technique attaches a destabilizing

domain (DD) to a protein of interest that unfolds under normal conditions, causing the protein to be ubiquitinated and degraded. Often, these domains can be stabilized with a small molecule, and thus the stability of the fusion protein can be tuned. An example of this is the attachment of a mutated form of FKBP12 to the highly stable yellow fluorescent protein (YFP). This fusion is constitutively unstable, but addition of a small molecule that binds FKBP12 stabilizes the chimera and confers resistance to degradation[117].

It is also possible to target a protein indirectly to the proteasome by taking advantage of natural protein-protein interactions. For example, the bacterial proteins barnase and barstar interact tightly in the cell. When a ubiquitin chain is attached to barnase and an initiation region to barstar, a complex is formed that is recognized by the proteasome and barstar is degraded[118]. Similarly, our lab previously constructed an inducible protein degradation system that causes the depletion of ectopically expressed substrates[119]. This system takes advantage of chemical inducers of dimerization, which bind two proteins simultaneously. Rad23b's UBL was attached to one binding partner, and as a model target GFP was attached to the other partner. These proteins were co-expressed in cells, and in the presence of the small dimerization molecule, a complex was formed, shuttled to the proteasome, and the GFP domain was degraded.

Concluding remarks

Recently, there has been much excitement in the inducible degradation field. By depleting proteins, rather than their precursors, the downstream effects of changing protein concentration can be directly measured. In addition, most cellular proteins are modified post-translationally, which affects their function[85]. Often, it is a particular subpopulation of the target that is relevant to disease or research, so the ability to study

just the modified form is useful. This can only be achieved by depleting the protein after translation, which limits the usefulness of methods like RNAi and CRISPR and makes inducible degradation an attractive option.

All of the technologies described here have been successful in depleting proteins either *in vivo* or *in vitro*. However, the methods that degrade endogenous proteins without target modification rely on ubiquitination of the target. This is an important distinction because of the potential consequences of interfering with the cell's ubiquitin system. Ubiquitin is a signaling molecule for numerous cellular pathways other than proteolysis, including DNA damage repair, transcriptional regulation, and membrane trafficking[27]. There are many ways to modify a protein with ubiquitin, and the type of modification affects the fate of the target. Therefore, inducing a protein's ubiquitination does not always lead to degradation. The ubiquitin code is still poorly understood and manipulating ubiquitination risks unintended side effects. Finally, depleting the ubiquitin pool affects cells in a pleiotropic manner and can lead to toxicity[120,121].

To address these limitations, we designed proteasome adaptors that contain a UBL domain attached through a linker to an affinity domain that recognizes a protein of interest. The UBL from Rad23 binds to the proteasome and has been used successfully in other systems to target proteins for degradation[119]. Because Rad23 is not efficiently degraded in the cell[62], it should confer resistance to degradation to the entire adaptor. The affinity domain we chose is called a monobody, a small protein based on the tenth domain of fibronectin type III that has been adapted to bind specific proteins[122]. As proof of concept, we used monobodies that bind the Src homology 2 (SH2) domains of Abelson kinase (Abl) and Shp2 phosphatase, both oncology drug targets in their own right[123-126]. In principle, these adaptors should seek out the target of interest and localize it to the proteasome for degradation in a catalytic manner, without interfering with ubiquitination.

Chapter 2: Results

DESIGN AND *IN VITRO* CHARACTERIZATION OF MODEL SUBSTRATES/ADAPTORS

Background

To investigate whether our proteasome adaptors can induce degradation of proteins, we first tested the system *in vitro* with purified model substrates, adaptors, and proteasome. To do so, we attached a proteasome-binding domain, the UBL domain from Rad23, to an affinity scaffold that recognizes a particular target. Many types of affinity scaffolds have been developed, including nanobodies[127-129], Affibodies[130], designed ankyrin repeat domains (DARPins)[131,132], and others. We chose a small, single-chain affinity reagent called a monobody, which is a stably folding protein derived from the tenth domain of fibronection type III[122,133,134]. Monobodies contain seven beta strands connected on one side through flexible loops of approximately five to eight amino acids long[133]. Similar to the CDR loops in antibodies, the sequences of these loops can be randomized to create a library and then selected to bind specific proteins of interest using phage and yeast display methods[135]. These evolved monobodies bind their targets with high affinity and sufficient selectivity to discriminate between members of homologous protein families and even PTMs[136-138]. Unlike antibodies, monobodies contain no disulfide bonds and are easily expressed in bacterial systems as stably folded proteins[122] (personal observations).

We chose to test our system on two well-studied proteins that have been prominent targets in cancer drug development efforts: Abelson kinase (Abl) and Shp2 phosphatase[139-141]. Abl is a non-receptor tyrosine kinase that plays a key role in myriad pathways, including cytoskeleton rearrangement, DNA damage repair, receptor tyrosine kinase signaling, chromatin remodeling, and transcriptional regulation, to name a few [142].

Just over 1100 amino acids long, Abl protein is approximately 122 kDa and contains three N-terminal folded domains: Src-homology (SH) 3, SH2, and a kinase (SH1) domain that make up about half of the protein[143]. At the very N-terminus is a variable region ("cap") that changes depending on exon 1 splicing[144]. One form contains a myristoylation site that, once modified, induces a conformational change in Abl that leads to its autoinhibition[143,145]. The cap in both spliced forms also interacts directly with the SH3 and kinase domains to inhibit the function of Abl[145]. This region is critically important to prevent oncogenicity and is discussed in more detail below. At the C-terminus is a partially folded actin binding domain (ABD) that interacts with the cytoskeleton and enables rearrangement and organization[146]. The ABD also contains a nuclear export signal[147]. Between the folded domains is a long unstructured stretch that contains many localization and modification regions. There are three nuclear localization signals, three PxxP domains (proline-rich regions that bind to SH3 domains), and three DNA-binding motifs that recognize A/T-rich regions, in addition to several phosphorylation sites[148].

The variety of binding motifs and localization signals indicates the complexity of Abl signaling, the full scope of which is still being elucidated[142]. It has been established that: Abl stimulates actin polymerization by activating regulators of Arp2/3[149]; Abl phosphorylates the C-terminal repeated domain of RNA polymerase II, indicating a role in transcription elongation[150]; Abl also interacts with the histone acetyltransferase Tip60 upon DNA damage by ionizing radiation, which suggests a part in DNA damage response and histone remodeling[151]; Abl also binds the receptor tyrosine kinase (RTK) adaptor protein Grb2, implicating a function in RTK signaling[124].

Interestingly, though Abl is typically thought of as a potent oncoprotein, overexpression of wild-type Abl1 does not lead to cancer[142]. However, in chronic myeloid leukemia (CML), a chromosomal translocation event causes Abl kinase to be

translated as a fusion protein C-terminally to Breakpoint Cluster Region (BCR-Abl). Although BCR has been studied mostly in the context of its fusion to Abl, BCR is a signaling protein in its own right that contains N-terminal serine/threonine kinase and dimerization domains upstream of both guanine nucleotide exchange factor (GEF) and GTPase-activating (GAP) domains[142,152,153]. The translocation occurs between chromosomes 22 and 9 and forms a neo-chromosome dubbed the "Philadelphia chromosome"[154]. The breakpoint in BCR can occur at several sites, each of which leads to a different sized fusion protein[155]. Amazingly, however, the Abl portion of the chimeric mRNA always begins at exon 2[148]. The reason for this is unknown; while the chromosomal translocation usually occurs between exon 1a and 1b of Abl, the resulting mRNA invariably excludes exon 1[148]. As mentioned earlier, Abl normally exists in an autoinhibited state with the N-terminal domain (encoded by exon 1) bound to the functional domains; when expressed as a fusion protein with BCR, the self-regulation is abrogated because the main autoinhibitory mechanism is deleted. The most significant consequence of this is a marked increase in kinase activity, including phosphorylation of other proteins as well as autophosphorylation[148]. Three main routes to transformation occur upon BCR-Abl expression: decreased adhesion to the extracellular matrix and bone marrow stroma cells[156]; increased mitogen-activated signaling through upregulated MAP kinase, Jak/STAT, and PI3 kinase pathways[157-162]; and inhibition of apoptosis[163,164]. Taken together, the deregulation of Abl signaling upon expression of BCR-Abl is the causative agent for CML and other leukemias[69,139].

The other target we chose to investigate was Shp2 phosphatase. Shp2 is an approximately 68 kDa protein consisting of three domains: two tandem SH2 domains (N-SH2 and C-SH2, respectively) connected to a phosphatase domain. Like Abl, Shp2 is normally autoinhibited in a conformation where the SH2 domains bind to the phosphatase

domain and obstruct the catalytic site[165]. Shp2 is activated by engaging phosphotyrosine motifs through its SH2 domains[166]. Upon binding, the intramolecular interaction is disrupted and the active site is exposed[166]. While phosphatases as a class of proteins are often considered tumor suppressors due to inhibition of cell growth signal pathways[167], Shp2 is somewhat unique in that it was the first reported tyrosine phosphatase oncoprotein[168,169]. Overexpression of Shp2 is associated with several types of cancers, including breast cancer[68,170,171] and glioblastoma multiforme[172], and activating mutations in Shp2 cause Noonan and LEOPARD syndromes as well as childhood cancers[173]. Shp2's oncogenic propensity is mostly due to its role in Ras activation[174]. Shp2 binds to RTKs and recruits the adaptor proteins Grb2 and Grb2 associated binder (GAB)[175,176]. Shp2 dephosphorylates Ras, which attenuates the interaction with its GTPase-activating protein (GAP)[172]. This in turn maintains Ras in an active state and promotes downstream proliferative signaling through the mitogen activated protein kinase (MAPK) pathway (@@Milburn Science 1990). Inhibition of Shp2 has been shown to reduce growth and metastasis of cancers that rely on RTK signaling[168,172,174].

To target these two proteins for degradation, we generated adaptors containing monobodies that recognize the Src-homology 2 (SH2) domains of Abl or Shp2[137,138]. In principle, the adaptors should bind to the target protein, bring it to the proteasome, then recycle back into the cell (Figure 3a, b).

Construction of adaptors and model substrates

In order for a protein to be effectively degraded, the substrate must be positioned properly on the proteasome so that the initiation region is accessible for translocation[15,118,177]. Therefore, we designed and tested several conformations of adaptor and substrate domains (Figure 3a). We made simplified model substrates containing just

the SH2 domain of Abl. As a readout for substrate concentration, we fused monomeric superfolder (ms) GFP to the C-terminus of the SH2 domain, followed by a 35-amino acid unstructured initiation region derived from cytochrome b_2 that is known to support efficient degradation (the whole substrate is referred to as SH2-GFP-tail)[13,14] (Figure 3). We used msGFP because it is a very stable protein but can be degraded when attached to an initiation region[178]. We also tested a circular permutant of GFP (Cp8) that is less stable but found no difference in degradation patterns (Appendix A)[179,180].

To optimize model substrate positioning, we created four iterations of the adaptor by changing the arrangement of the two adaptor domains and the linker between them (Figure 3a). The first configuration contained a simple Gly-Ser/Thr linker (GGSGGT) with the *S. cerevisiae* Rad23 UBL domain (aa 1-77) at the N-terminus and the Abl SH2 monobody at the C-terminus. We then interchanged the UBL and monobody domains and also replaced the short linker in both configurations with a longer flexible linker derived from the region between the UBL and first UBA domain of Rad23 (aa 77-144). We chose this linker because it is known to be resistant to proteasomal degradation and should thus confer stability to the adaptor[62]. A longer linker also gives flexibility to the adaptor, which in theory allows the adaptor-substrate complex to sample multiple positions on the proteasome. As a negative control, we designed a non-binding adaptor in which the monobody was mutated at Tyr87 to alanine. This completely abolished binding to the SH2 domain[137]. We also created and tested 4 geometries of the model substrate, the results of which are discussed in Appendix A.

I first optimized the expression and purification of the adaptors and substrates. I fused the constructs to a C-terminal chitin-binding domain (CBD), which binds to chitin resin and self-cleaves in the presence of reducing agent[181]. I also tried substituting just a His_6 tag on the N- or C-terminus, but found that the yield, purity, and stability of the CBD

fusions were much greater, despite the fact that the CBD fusions required overnight cleavage in reducing conditions. The adaptors also contained an N-terminal FLAG tag and the substrates contained an N-terminal HA tag so that they could be monitored by Western blot. Next, I purified FLAG-tagged (3xFLAG-Rpn11) proteasome from yeast using an established method (see Materials and Methods), and found that the final product was highly stable and retained activity after freeze-thaw or left overnight at 4°C. Because the proteasome is a very large complex, and most likely copurifies with proteasome-binding proteins (e.g. Ubp6), the concentration of proteasome was estimated based on a size of 2.5 MDa but was by no means an exact concentration.

After purification of all components, we first tested the UBL-GGSGGT-monobody adaptor *in vitro* with the Abl SH2 model substrate on a plate reader, with the decrease in GFP fluorescence as a readout for substrate depletion. We incubated adaptor and substrate together before adding proteasome and immediately began the assay. We found that the model substrate was degraded only in the presence of adaptor and proteasome (Figure 4a). This suggested that degradation of substrate was dependent on both adaptor and proteasome. In addition, degradation of substrate increased at higher concentrations of adaptor, although there was a point at which more adaptor led to decreased substrate depletion. One explanation is that at a high enough concentration of adaptor, the proteasome is saturated and model substrate degradation rate is determined only by substrate turnover. Another explanation could be related to the "hook effect" which occurs when one protein acts as a linker between two others. At high enough concentrations of the linker protein, formation of the ternary complex actually decreases due to competition for binding between intermediate species[182,183]. This results in a falsely low final readout. In this case, the adaptor acts as the linker protein between the substrate and proteasome. At high adaptor concentrations, there is less

substrate•adaptor•proteasome complex formed, which decreases both the total amount of degradation and the initial rate. This was most noticeable when the nanomoles of substrate degraded were plotted against the concentrations of adaptor (Figure 4b). At low concentrations of adaptor (below the concentration of substrate), there is a superstoichiometric amount of substrate degraded. Above the concentration of substrate, degradation levels off and decreases slightly at the highest concentrations of adaptor.

The model substrate was degraded completely, as no intermediate fragments were detected in a Western blot against the N-terminal hemagglutinin (HA) tag on the substrate (Figure 5, above). Over the course of 75 minutes, nearly all of the substrate was depleted, as compared to proteasome loading control (Figure 5, below). As expected, the FLAG-tagged adaptor was not degraded over the course of the assay (Figure 5, above).

We next tested the other adaptor configurations (Figure 6). Surprisingly, all of the adaptor configurations were able to deplete the substrate to a certain extent, suggesting that the adaptors are flexible enough to position the substrate properly even with a short linker, or else the adaptor can bind to multiple locations on the proteasome to optimize positioning. We found that the substrate was degraded most and had the highest initial rate in the presence of the original adaptor design, UBL-GGSGGT-monobody (Figure 6). We therefore used the original geometry for the rest of the studies. Having a longer linker was worse than a short one for both domain configurations, suggesting that there is an optimum spacing between the UBL and monobody. This was somewhat surprising, because we used the native linker between Rad23's UBL and the first UBA domain. The linker seems to be important for the endogenous protein, so we expected it to help in our adaptors[62]. This discrepancy could be explained by the fact that native Rad23 binds to ubiquitinated proteins with varying chain length and must be flexible enough position the substrates at the translocation entrance[184].

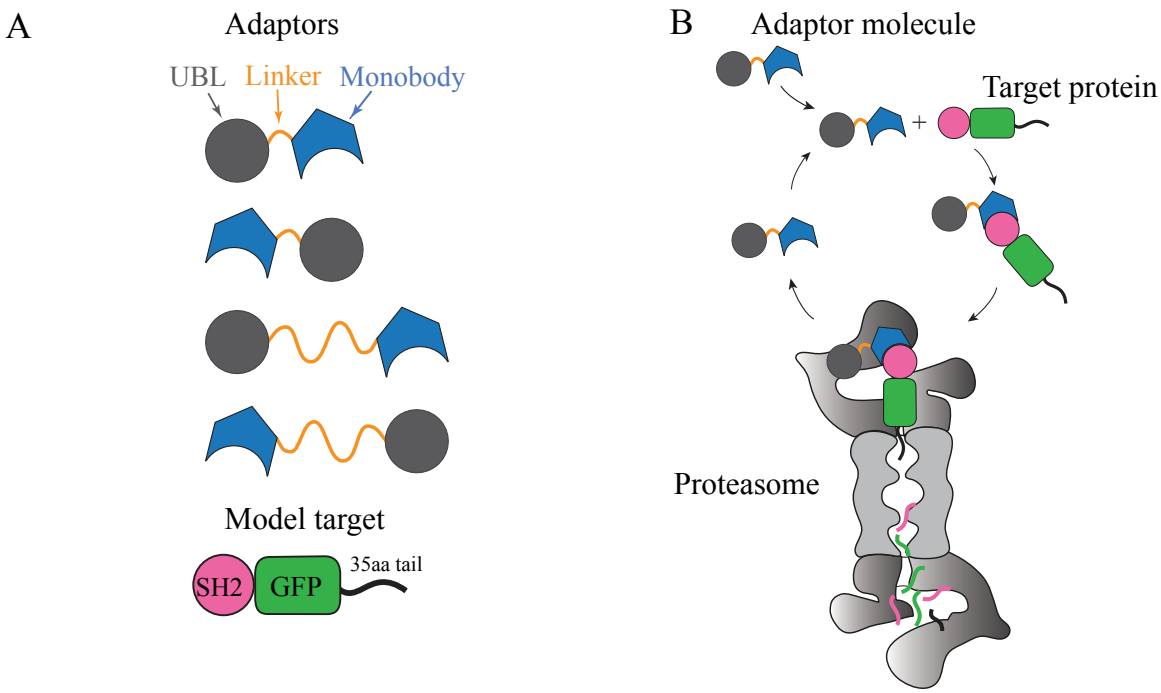


Figure 3: Model adaptor configurations and schematic

A) Adaptor and model substrate configurations. B) Schematic of model substrate and adaptor; adaptor binds to target through affinity domain (monobody) and then binds to proteasome via Ubiquitin-Like (UBL) domain. Substrate is degraded, while adaptor is not.

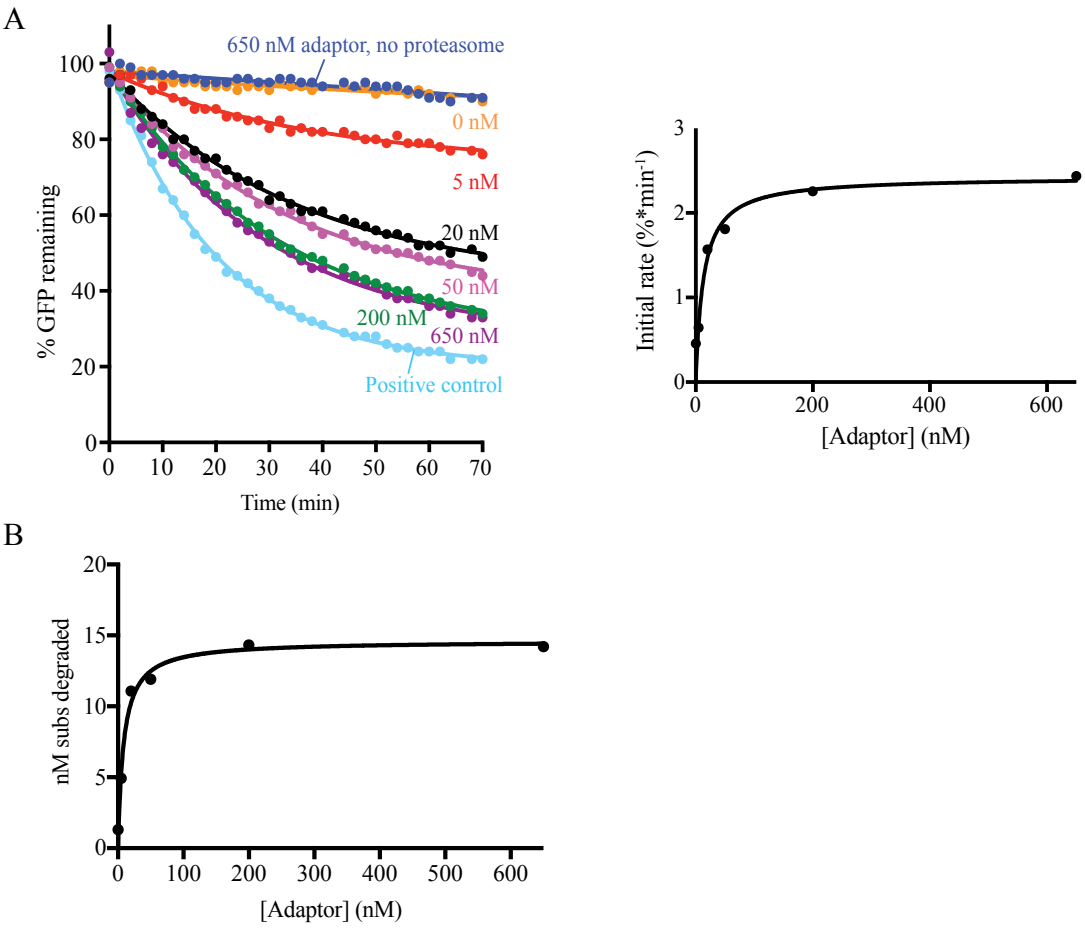


Figure 4: Model substrate is degraded *in vitro* in presence of adaptor.

A) Left: Fluorescence assay monitoring substrate through GFP fluorescence over time. 20 nM $SH2^{AbL}$-GFP-tail was incubated with indicated concentrations of adaptor in the presence or absence of 40 nM proteasome. Positive control: UBL-GFP-90aa tail. Right: Initial rate of degradation increases with concentration of adaptor until saturation of binding occurs. B) Nanomoles of substrate degraded plotted versus adaptor concentration.

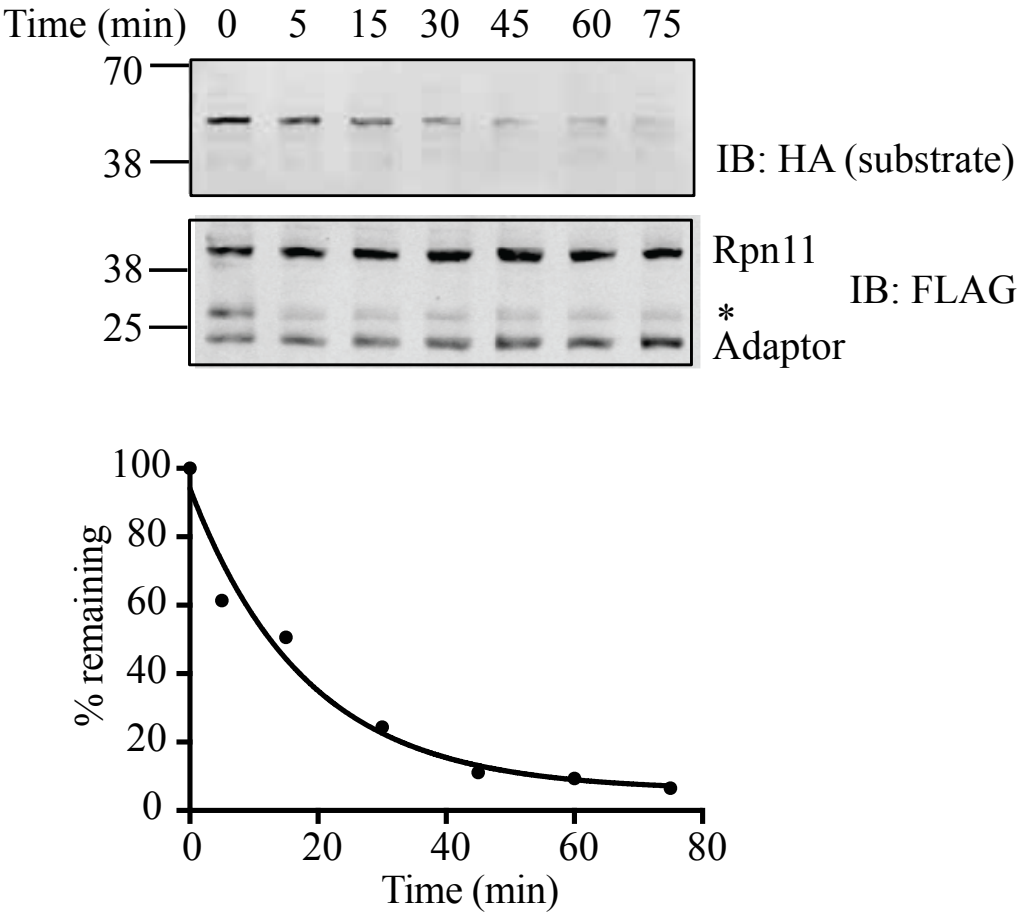


Figure 5: Western blot confirmation of model substrate depletion

Above: Western blot of HA-tagged model substrate and FLAG-tagged adaptor and proteasome (Rpn11). 20nM substrate were incubated with 160nM adaptor and 25nM proteasome and samples were taken at indicated time points. Asterisk indicates non-specific band. Below: Quantification of model substrate bands normalized to the first lane as a percentage of Rpn11 band.

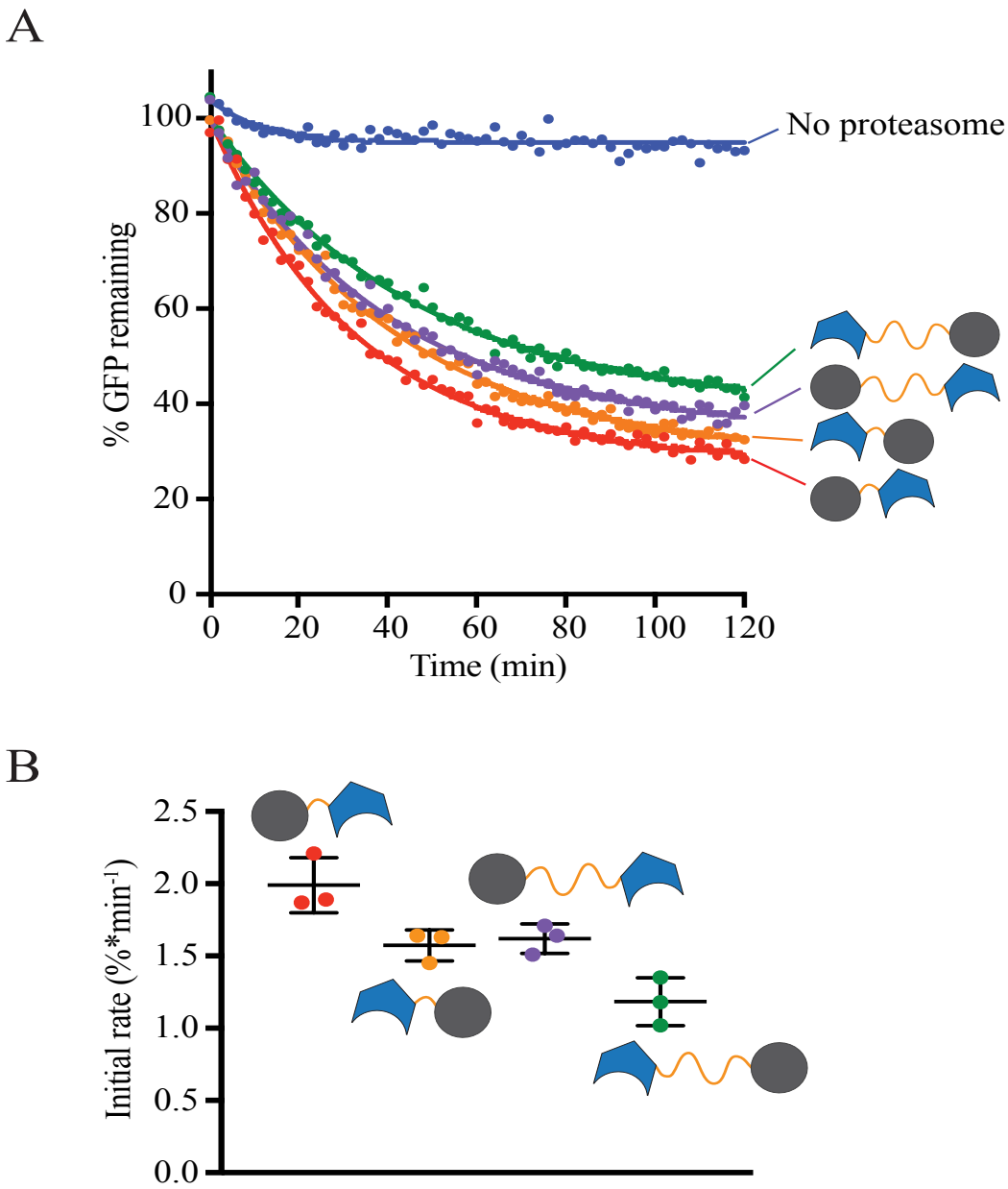


Figure 6: Testing adaptor configuration *in vitro*.

A) Fluorescence plate reader assay with 20nM $SH2^{AbL}$-GFP-tail, 160nM adaptor, with and without 25nM proteasome. B) Initial rates of different adaptor geometries confirm that UBL-mB is the fastest adaptor. Bars represent mean initial rate of three separate reactions +/- SD.

TESTING ADAPTOR-MEDIATED DEGRADATION OF MODEL PROTEINS IN CELLS

Adaptor transfection leads to depletion of model substrate in HEK293 cells

After successfully degrading the SH2-GFP-tail model protein *in vitro*, we next tested the system in HEK293 cells. We integrated the Abl model substrate (SH2-GFP-tail) under a cytomegalovirus (CMV) promoter into HEK293 cells using the Flp-In system[119] (Figure 7a) and confirmed its stable expression via Western blot (Figure 7b). The cells showed green fluorescence (Figure 8a); however, as the cells were passed, the fluorescence decreased and split into high, medium, and low fluorescence populations, which became more distinct as the passage number increased (not shown). This could be due to CMV promoter silencing, a common phenomenon observed in human cells when the CMV promoter is methylated, and it increases over time and passage number[185]. Therefore, we only used cells at low passage.

We then modified the *in vitro* adaptor construct for the cellular assay in two ways. First, we replaced the yeast Rad23 UBL with the human homolog Rad23b (hHR23B) UBL domain, which binds to the human proteasome[59]. We also added an mCherry fluorescent readout after an internal ribosome entry site (IRES)[119] so we could analyze cells that expressed the adaptor by flow cytometry (Figure 7a).

Substrate-binding or non-binding (mutated monobody) adaptor constructs were transfected into integrated or host cells and 72 hours later the cellular fluorescence was measured (Figure 8a). Only cells that were mCherry-positive (expressing adaptor) were analyzed. We found a four-fold decrease in median GFP fluorescence (substrate levels) in the cells with binding adaptor as compared to non-binding adaptor (Figure 8b), indicating that the model protein was depleted by approximately 70% only in the presence of binding adaptor. We then sorted and collected the adaptor-transfected (mCherry-positive)

cells and analyzed substrate amount by Western blot (Figure 8c). In agreement with the flow cytometry data, we saw approximately 60% less substrate in the presence of binding adaptor as compared to non-binding adaptor, though this experiment was performed once. Again, the model protein was degraded completely as no intermediate fragments were detected (Figure 8c). As expected, the FLAG-tagged adaptor was not degraded and remained stable in the cells (Figure 8c).

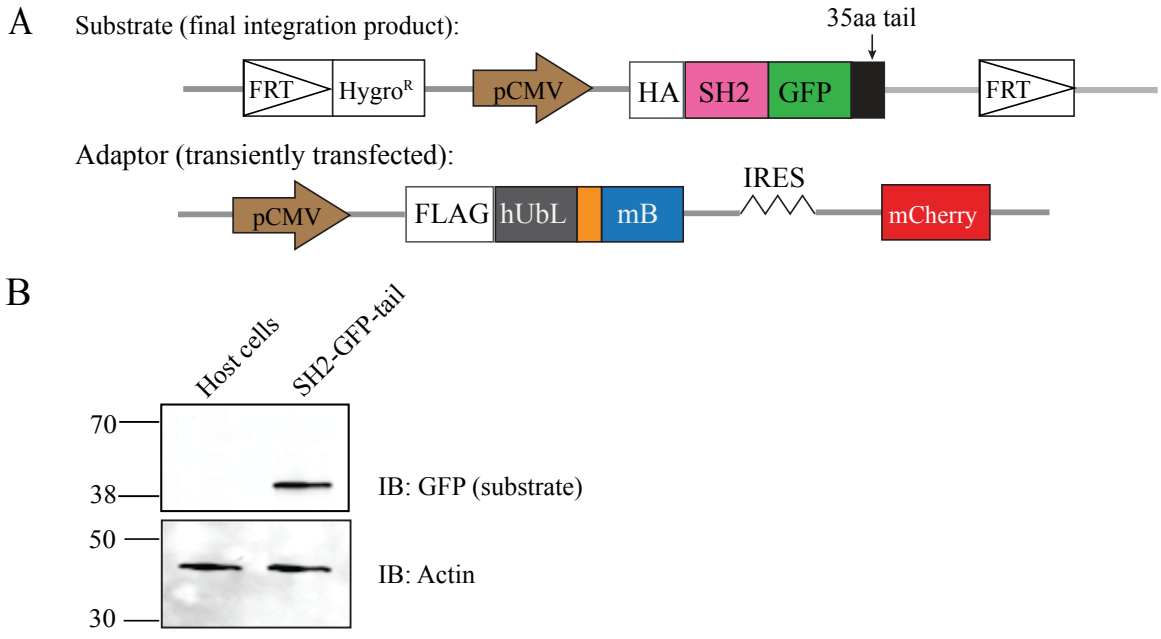


Figure 7: Design of model substrate and adaptor constructs and expression in cells

A) Schematic of integrated substrate construct and transiently transfected adaptor plasmid. Both adaptor and substrate were expressed under a CMV promoter. GFP-positive cells were selected for using hygromycin. B) Western blot confirming that substrate was integrated.

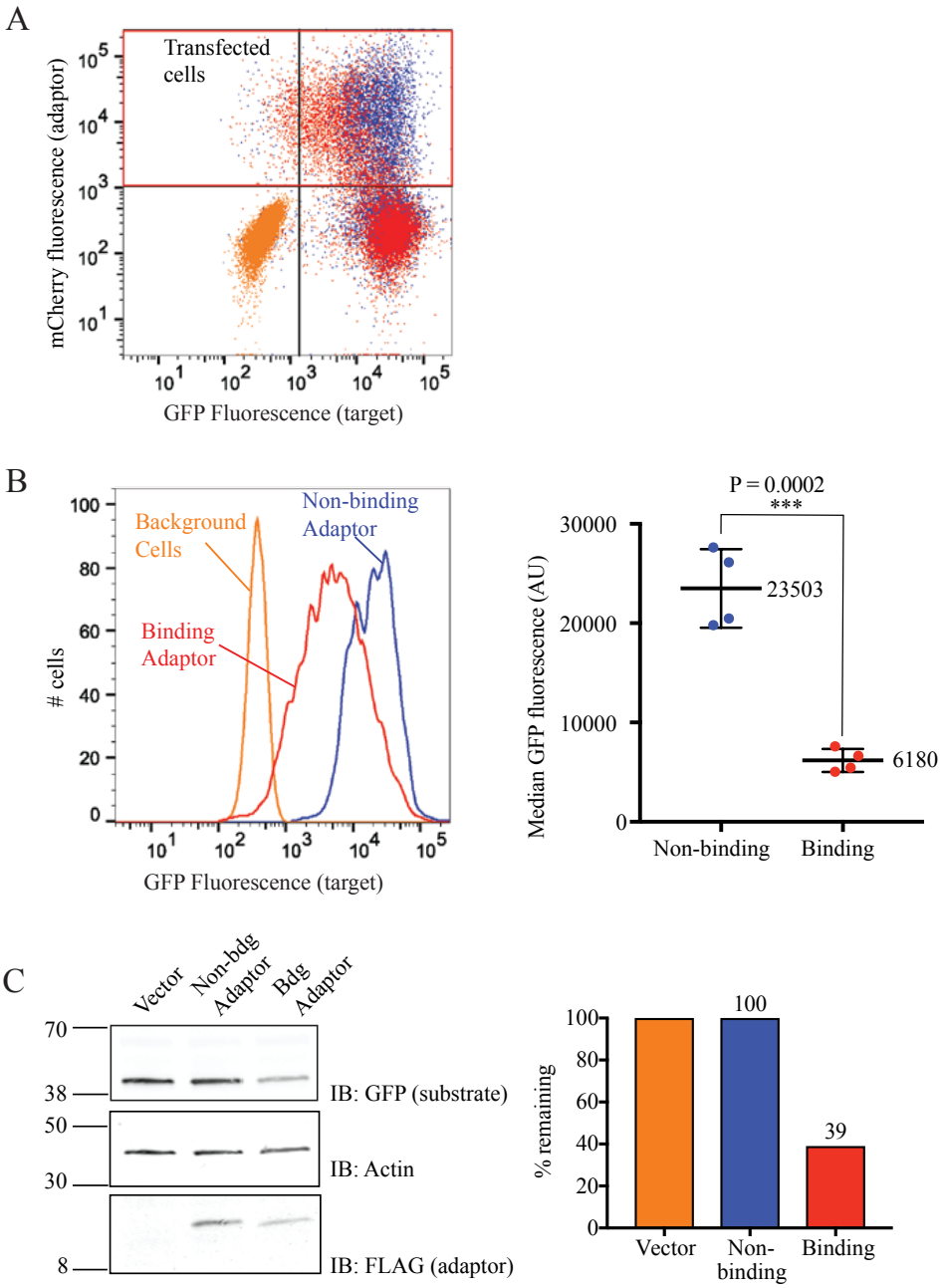


Figure 8: Degradation of Abl model substrate in cells.

A) Overlaid flow cytometry scatter plots with GFP fluorescence (substrate) on the x-axis and mCherry fluorescence (adaptor) on the y-axis. Each dot represents one cell. Orange:

Figure 8: continued page 31

Host HEK293 Flp-In™ cells transfected with empty vector indicate background fluorescence. Blue: Integrated cells transfected with non-binding (UBL-HA4^{Y87A}) adaptor. Red: Integrated cells transfected with binding (UBL-HA4) adaptor. Red box indicates transfected (mCherry positive) cells. B) Left: Histogram of GFP fluorescence of host cells and integrated cells transfected with non-binding and binding adaptors. Colors as in A. Right: Median GFP fluorescence values after binding or non-binding adaptor transfections plotted for four independent experiments. Bars represent mean +/-SD. P value was derived from unpaired two-tailed t-test. C) Left: Transfected cells (red box in B) were sorted and collected, and cell lysates were immunoblotted for GFP (substrate), FLAG (adaptor), and actin. Right: ImageJ analysis of blot on left representing % GFP signal remaining after empty vector, non-binding, and binding adaptor transfections, normalized to actin, where vector was set to 100%.

TESTING ADAPTORS ON SHP2 PHOSPHATASE

Developing and testing Shp2 model substrates *in vitro*

For the *in vitro* and initial in-cell tests of our adaptors, we used the SH2 domain from Abl kinase as our model substrate and the Abl-binding monobody in the adaptor. To test the versatility of our system, we chose two other monobodies that bind to a different target, Shp2 phosphatase. Shp2 contains two tandem SH2 domains ($SH2^N$ and $SH2^C$), each of which is recognized by a distinct monobody[138]. Therefore, we created two model substrates, Shp2 $SH2^N$- or Shp2 $SH2^C$-GFP-tail to test both *in vitro* and in cells.

Upon purification of the $SH2^N$-GFP-tail, I found that the protein was unstable, despite trying several methods of purification (His-tagged followed by size exclusion or chitin purification) and rounds of optimization. The protein was degraded by proteasome independently of the adaptor (although not without proteasome), suggesting that the substrate was at least partially unfolded (data not shown). It is possible that even in the absence of adaptor, the proteasome has some affinity to the initiation region, as it has been shown that the proteasome shows preference for certain sequences, and our initiation region was created with that in mind[14].

The $SH2^C$-GFP-tail substrate was more stable when purified, so I used that substrate to test the versatility of the system *in vitro*. I created an adaptor that binds to the C-terminal Shp2 $SH2^C$ via the monobody Cs3[138]. The substrate degraded approximately 15% in the absence of adaptor or presence of non-binding adaptor (Figure 9). However, the model substrate degraded further in the presence of binding adaptor and proteasome (Figure 9). There was dose-dependence observed, which saturated at approximately 40nM adaptor. The saturation of the Shp2 model substrate degradation occurred at a five-fold lower concentration of Shp2 adaptor than the Abl adaptor for the Abl model

substrate (Figure 4, 9a). This agrees with the five-fold lower K_d of Cs3 to Shp2 $SH2^C$ than the Abl monobody to $SH2^{Abl}$.[137,138]

To see if the Shp2 adaptor acted in a catalytic manner, I plotted the amount of substrate degraded versus the amount of adaptor present (Figure 9c). As in the Abl experiments, the amount of substrate degraded increased with increasing adaptor up to a point (~40nM adaptor). At low concentrations of adaptor (less than substrate), the adaptor degraded a superstoichiometric amount of substrate. However, once the adaptor concentration became higher than the initial substrate concentration, the amount of substrate degraded leveled off and adding more adaptor did not increase depletion. This could be due to either saturation of binding to the proteasome (which was 25nM in the reaction) or to the hook effect, where high concentrations of a linker protein can be autoinhibitory[182,186]. Thus, there exists an optimum concentration of adaptor to induce the maximum amount of substrate depletion.

Shp2 Model Substrates Are Degraded in Cells Upon Adaptor Transfection

I next tested whether the Shp2 adaptors could deplete the Shp2 model substrates in cells. I created stable HEK293 cells containing model substrates with either the N-terminal or C-terminal SH2 domain (Shp2 $SH2^N$- or Shp2 $SH2^C$-GFP-tail) with the Flp-In System as in Figure 3a. I transfected the corresponding adaptor constructs, which bound to either the $SH2^N$ or $SH2^C$ domain (but not both[138]) and contained the mCherry fluorescent readout as in Figure 3a. As a control, I used the same non-binding (mutated) adaptor as for the Abl experiments. Upon transfection of $SH2^N$-binding adaptor, there was a three-fold decrease in $SH2^N$-GFP-tail fluorescence in transfected (mCherry-positive) cells as compared to non-binding adaptor (Figure 10a). When $SH2^C$-binding adaptor was transfected into cells expressing $SH2^C$ substrate, GFP fluorescence decreased

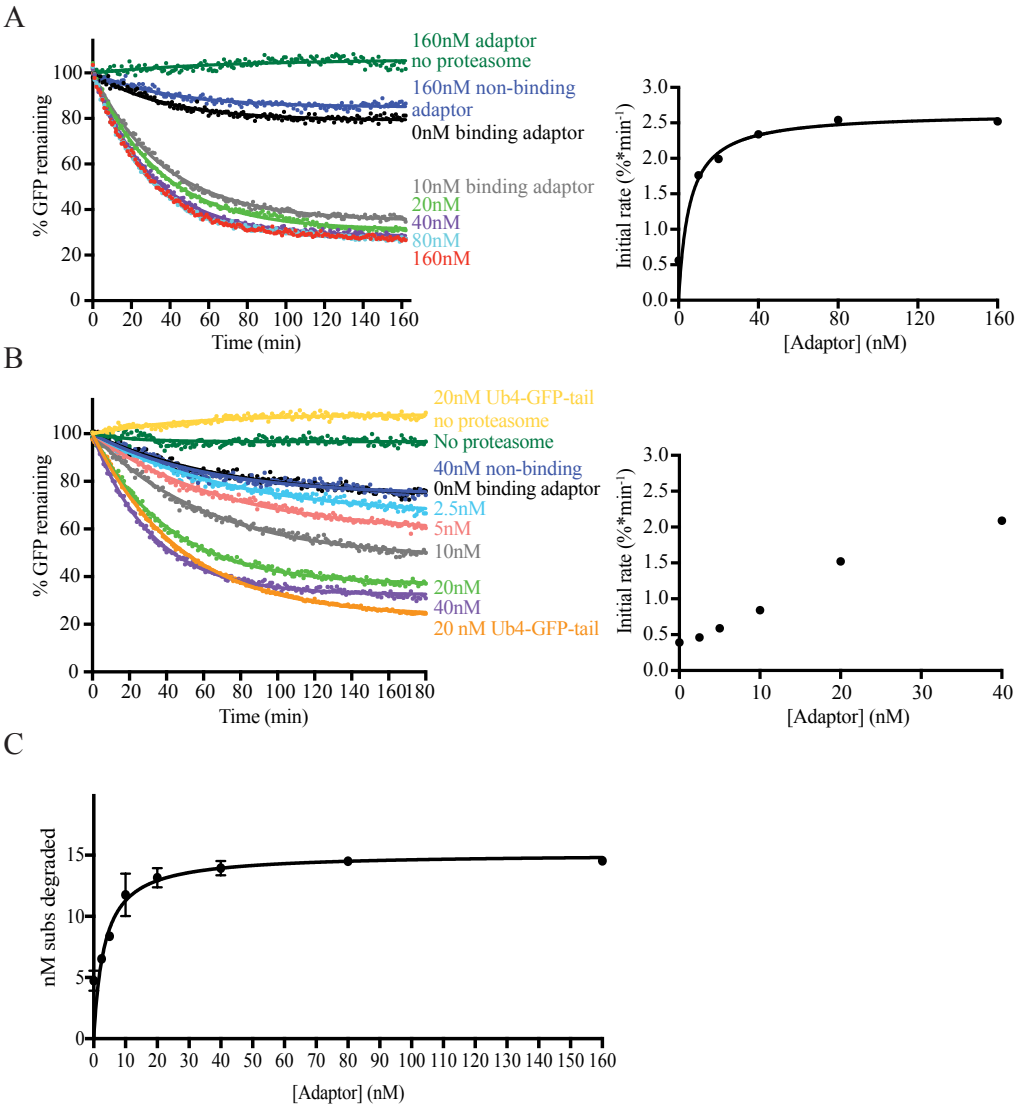


Figure 9: Degradation of Shp2 model substrates *in vitro*.

A) Left: Fluorescence assay monitoring substrate through GFP fluorescence over time. 20 nM $SH2^{AbL}$-GFP-tail was incubated with indicated concentrations of binding and non-binding adaptor in the presence or absence of 25nM proteasome. Positive control: UBL-GFP-90aa tail. Right: initial rate of degradation plotted against concentration of adaptor. B: As in A, but for lower concentrations of adaptor. C) Nanomoles of substrate degraded plotted versus adaptor concentration.

by 2.8-fold in transfected cells (Figure 10b). I then sorted and collected the transfected cells and probed substrate expression levels on a Western blot with a GFP antibody. The depletion of substrate was confirmed by Western blot (Figure 10c, n = 1).

Testing the adaptors on endogenous Shp2

The final step in this project was to verify whether the adaptors could lead to depletion of full-length, endogenous protein. I chose to use Shp2 as the target because it is ubiquitously expressed as a cytosolic protein[187,188] and is readily detectable in HEK293T cells with Western blot (Figure 11a). Using immunoprecipitation I confirmed that the $SH2^N$- and $SH2^C$-binding adaptors bound to full-length Shp2, but non-binding adaptor does not (Figure 11a). All adaptors (including non-binding) bind to the proteasome (Figure 11a). When I pulled down the adaptor via its FLAG tag, both Shp2 and proteasome pulled down with it. However, when I pulled down Shp2, no proteasome bad was detected. Presumably, this is because once the Shp2-adaptor complex binds the proteasome, Shp2 is immediately translocated and degraded.

I transfected $SH2^N$-binding or non-binding adaptor into HEK293T cells and found that the adaptor depletes endogenous Shp2 levels by about 50% in Western blot assay (Figure 11b, c). As an orthogonal readout for Shp2 levels, I also developed a protocol for immunofluorescence followed by flow cytometry; however, the antibody bound at the same region as the binding adaptor and therefore I could not use the results from this assay (described further in Appendix B). Despite this, the Western blot assay consistently showed depletion of endogenous Shp2 from cells (Figure 11b, c).

There are several possible reasons why the adaptors did not deplete endogenous proteins as well as model proteins. First, the adaptors may not bind as tightly to the full-length Shp2 protein as to the SH2 model protein. Indeed, the monobodies themselves

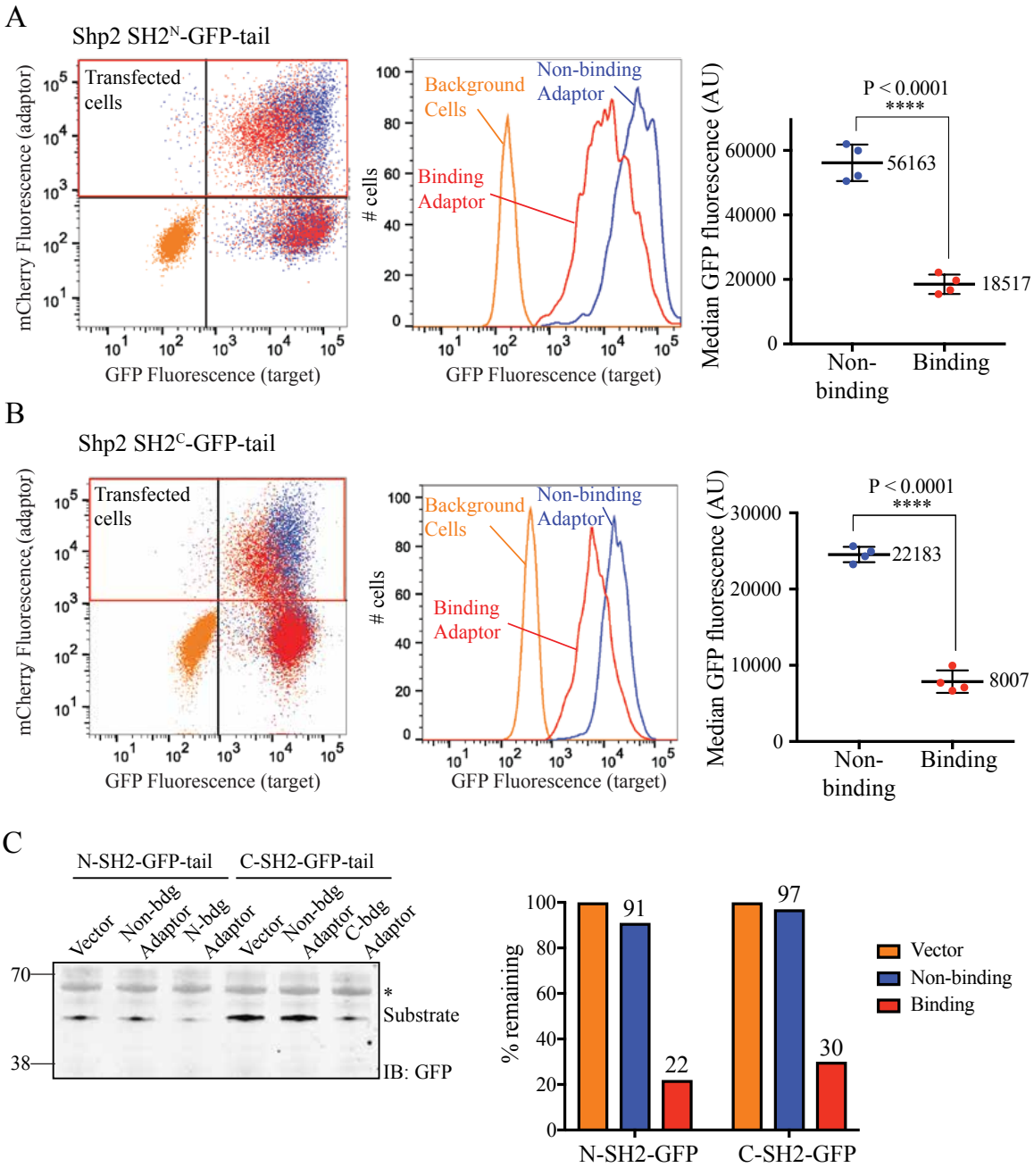
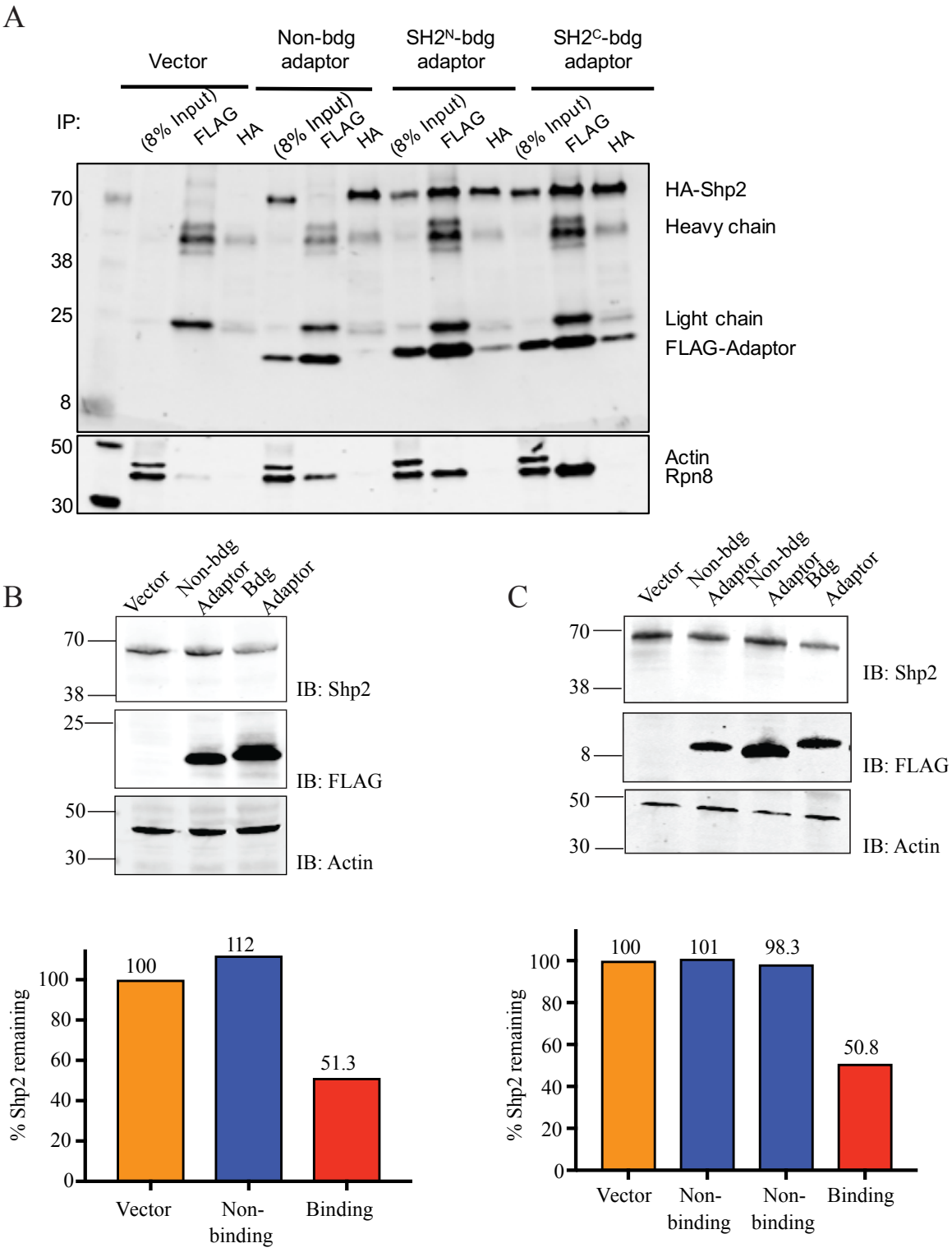


Figure 10: Degradation of Shp2 model substrates in cells.

A) Left: Overlaid flow cytometry scatter plots of Shp2 SH2^{N}-GFP-tail integrated cells with GFP fluorescence (substrate) on the x-axis and mCherry fluorescence (adaptor) on

Figure 10: continued page 37

the y-axis. Orange: Host HEK293 Flp-In™ cells transfected with empty vector indicate background fluorescence. Blue: Integrated cells transfected with non-binding (UBL-HA4^{Y87A}) adaptor. Red: Integrated cells transfected with binding (UBL-HA4) adaptor. Red box indicates transfected (mCherry positive) cells. Middle: Histogram of GFP fluorescence of host cells and transfected integrated cells with non-binding and binding adaptors. Colors as in B. Right: Median GFP fluorescence values after binding or non-binding adaptor transfections plotted for four independent experiments. Bars represent mean +/-SD. P value was derived from unpaired two-tailed t-test. B) As in A, for Shp2 SH2^{C}-GFP-tail substrate. C) Left: Transfected cells (red box in C) were sorted and collected, and cell lysates were immunoblotted for GFP (substrate) and FLAG (adaptor). Asterisk indicates BSA from leftover media in sample. Right: ImageJ analysis of blot on left representing % GFP signal remaining after empty vector, non-binding, and binding adaptor transfections, where vector was set to 100%.



(Caption next page)

Figure 11: Shp2 adaptors bind to and degrade endogenous Shp2.

A) 293T cells were transfected with vector, non-binding adaptor-IRES-HA-Shp2, Shp2 $SH2^N$-binding adaptor-IRES-HA-Shp2, or Shp2 $SH2^C$-binding adaptor-IRES-HA-Shp2. Lysates were immunoprecipitated with either HA beads (pulling down Shp2) or FLAG beads (pulling down adaptor) according to manufacturer's instructions and blotted for FLAG, HA, proteasome (Rpn8), and actin. B and C) Top: Western blots of two independent transfection experiments. 293T cells were transfected with vector, non-binding, or Shp2 $SH2^N$-binding adaptor. Lysates were blotted for Shp2, FLAG (adaptor), and actin. Bottom: ImageJ quantification of Shp2 band intensity normalized to actin and vector bands.

bind more tightly *in vitro* to a single SH2 domain than to tandem SH2 domains[138]. In addition, cellular Shp2 exists in a "closed", auto-inhibited conformation until it binds to a phosphorylated protein, upon which the SH2 domains dissociate from the phosphatase domain and reveal the active site[165]. The Shp2 monobodies preferentially bind the closed conformation[138], which may not be the optimal conformation for degradation. Further, while Shp2 contains a C-terminal unstructured region[165], this may not serve as an efficient initiation region for the proteasome.

CONCLUSIONS

I have developed a method for the depletion of proteins from cells by targeting them directly to the proteasome. This system uses protein adaptors which bind to the proteasome through a UBL domain and simultaneously bind to a protein of interest through an affinity domain. I showed that this method works well to degrade model substrates by ~70-80% *in vitro*. The most likely explanation for the 20% of substrate remaining is that some of the substrate was oligomerized or aggregated. All fluorescent proteins are prone to oligomerization and aggregation[189], which are difficult for the proteasome to degrade[190].

Interestingly, we saw that the amount of depletion and initial rate increased with adaptor concentration up to a certain point, after which substrate degradation decreased. One explanation for this could be related to the "hook effect" which occurs when one protein acts as a linker between two others. At high enough concentrations of the linker protein, formation of the ternary complex actually decreases due to competition between intermediate species[183,186], resulting in falsely low final readout. In this case, the adaptor acts as the linker protein between the substrate and proteasome. At high adaptor concentrations, there is less substrate•adaptor•proteasome complex formed, which

decreases both the total amount of degradation and the initial rate. We also found that the configuration of the adaptor domains (UBL, linker, monobody) affected how well the model substrates were degraded. This could be due to the positioning of the substrate on the proteasome; the adaptor-substrate complex must bind in such a way as to make the initiation region accessible for unfolding and translocation and only certain configurations were able to do so.

The adaptors also depleted integrated model substrates in HEK293 cells by 60-75%. While this was not quite as much as in the *in vitro* studies, this could be explained by differential expression patterns of adaptor and substrate. The substrate was diffuse throughout the cell (data not shown), but because we used an independent readout for adaptor expression (mCherry expressed from an IRES promoter), we were unable to monitor the location of the adaptor. It is possible that the adaptor remained in the cytosol, in which case any nuclear substrate would remain undegraded.

Finally, the adaptors were able to deplete endogenous Shp2 levels by ~50%. One explanation is that the adaptors may not recognize full-length Shp2 protein as easily as SH2 model protein. *In vitro*, the monobodies showed poorer binding to a model substrate containing the tandem SH2 domains than to either SH2 by itself[138]. In addition, most of the Shp2 population in cells exists in a "closed", auto-inhibited conformation until it binds to a phosphorylated protein. The whole protein then undergoes a conformational change where the SH2 domains dissociate from the phosphatase domain to reveal the active site[165]. The Shp2 monobodies preferentially bind the closed conformation[138], which may not be the optimal conformation for degradation. Further, while Shp2 contains a C-terminal unstructured region of approximately 70 amino acids[165], this may not serve as an efficient initiation region for the proteasome.

One feature of this system is that it bypasses the need for ubiquitination because our adaptors were designed to interact directly with the proteasome. It may be advantageous to have a depletion method that does not rely on ubiquitination because of the pleiotropic effects of depleting the ubiquitin pool. Our system is versatile as well, as the target-binding domain can be switched out to virtually any affinity scaffold. In this work, we showed depletion of three different model proteins with three affinity domains as well as an endogenous protein. Future work includes testing our adaptors with other affinity scaffolds, such as nanobodies. A further avenue is to explore delivery of the adaptors into cells at the protein, rather than plasmid, level.

Chapter 3: Conclusions and Future Directions

The ability to deplete a protein from cells is extremely useful for many reasons, including studying the protein's function and treating diseases caused by malfunctioning or overexpressed proteins. The most common methodology uses RNAi to inhibit synthesis of specific proteins. However, inhibiting the expression of a protein does not allow for analysis at the post-translational level. Recently, the idea to deplete the protein via the ubiquitin proteasome system (UPS) has gained traction and several systems have been developed to do so. Here, I have presented a novel technology for degradation of specific proteins, both *in vitro* and in cells. The mechanism uses proteasome adaptors that recognize particular targets and shuttle them to the proteasome for degradation.

Our system centers around the 26S proteasome, a macromolecular, ATP-dependent protease which is responsible for degrading hundreds of cellular proteins in a tightly-controlled manner. The proteasome is integral to nearly every process in the cell, including cell cycle, transcription and translation, and signaling networks. As a main regulator of protein concentration, the proteasome must be versatile enough to degrade all types of domains, yet specific enough to degrade only proteins that are unnecessary or malfunctioning. Proper timing and specificity of proteasomal degradation is extremely important, and dysregulation of the UPS has been implicated in many diseases. For example, neurodegenerative diseases can be caused by accumulation of aggregated proteins. Although there is evidence that the protein aggregates are tagged with ubiquitin, the proteasome is unable to degrade them[191,192]. It is possible that the ubiquitinated aggregates never reach the proteasome; therefore, shuttling them directly to the proteasome could help to clear the aggregates.

The proteasome has also been targeted in drug development for many years. Proteasome inhibitors have been successful in treating various diseases such as cancer [193].

For example, Velcade (Bortezomib) has seen clinical success in the treatment of multiple myeloma[194]. The rationale for inhibiting the proteasome is that damaged proteins, which are more prevalent in cancer cells than healthy ones, will build up in the cancerous cells and trigger the apoptotic pathways[195].

Recently, the notion of harnessing the proteasome's power to degrade specific proteins has been examined. Several technologies have been developed that induce degradation of certain proteins. An advantage to depleting the proteins, rather than inhibiting them, is that they are completely destroyed, and therefore have no potential to cause further harm in the cell. The best-characterized technique involves PROTACs, small molecules which link a target with an E3 ligase[88]. The E3 ubiquitinates the target which leads to the target's degradation. Various PROTACs have been created using different E3 ligases, and one was recently approved to begin clinical trials[86,109]. An advantage to PROTACs is that as small molecules, PROTACs have increased cell permeability over protein therapeutics. Also, the PROTAC is not degraded with the target and acts catalytically to induce ubiquitination of multiple proteins[88]. PROTACs have been limited in their versatility, however. They contain 3 modules, a warhead (target-binding domain), an E3-binding domain, and a linker in between. Each of the domains must be optimized in the context of the entire molecule[106]. The E3, and any associated E2s, must have access to a lysine residue on the target. Therefore, the optimal ternary conformation is extremely important but is difficult to predict. In addition, PROTACs are limited by the availability of both E3-binding and target-binding ligands. Developing new binding molecules is time intensive and cost prohibitive[102], and the mere fact that a compound binds does not mean that it will make an effective PROTAC module[104]. Finally, PROTACs rely on ubiquitination of the target. Ubiquitin is a signaling protein involved in myriad pathways, and a clear understanding of the ubiquitin code remains

elusive. Manipulating the ubiquitin pool may cause pleiotropic effects that cannot be predicted[121]. Nonetheless, PROTACs have overcome many obstacles and remain the most promising therapeutic for inducible degradation.

Another recently developed methodology uses an E3, TRIM21, that naturally targets antigen-bound antibodies to the proteasome for destruction[111]. In this system, an antibody that recognizes a particular protein is introduced into the cell and is in turn recognized by TRIM21. The E3 then autoubiquitinates and the whole complex binds to the proteasome and is degraded. Initial experiments show promise for inducing the degradation of a wide variety of proteins. An advantage is that antibodies are relatively easy to generate and commercially available and validated for thousands of proteins. Again, however, the system relies on ubiquitination of the target and has the same potential risks as PROTACs. Another disadvantage is that the E3 ligase is degraded along with the antibody and target and therefore TRIM21 must be overexpressed to be effective[112].

Here I have introduced a new method for inducing degradation of specific proteins based on proteasome adaptors. I have constructed chimeric shuttle proteins (adaptors) that consist of a recognition domain to select the target protein and a proteasome-binding element that brings the target to the cell's degradation machinery for hydrolysis. I analyzed the system both *in vitro* and in cells and showed depletion for three model proteins and an endogenous protein.

This strategy diverges from traditional protein depletion methods in several ways. For one, the adaptors are engineered to function as catalytic molecules that link a target directly to the proteasome, thus circumventing the ubiquitination pathway. I based the adaptor design on a natural proteasome adaptor, human Rad23b (hHR23B), which brings ubiquitinated proteins to the proteasome but escapes degradation itself[62]. Conferring

stability on the adaptors allows them to turn over multiple substrates, as shown in Figures 4 and 9. The catalytic nature means that a lower concentration of adaptor is required to degrade the substrate as compared to traditional, stoichiometric inhibitors. Using a UBL domain instead of ubiquitination could help prevent any potential negative effects of manipulating the ubiquitin pool. In addition, because many disease-related cellular aggregates colocalize with ubiquitin but fail to degrade, having a different method to target them to the proteasome may help to clear them.

The adaptors can also be engineered to recognize nearly any protein of interest. For target binding, I used previously-designed affinity domains, monobodies, which are derived from the tenth domain of fibronectin type III and can be evolved *in vitro* to bind almost any protein and even to recognize specific PTMs[133,136,196,197]. Further, adaptors act directly on the target protein, so depletion is not limited by the target's intrinsic turnover rate. In principle, the adaptors should work in a complementary way to RNAi, as they act through different cellular mechanisms. Using both technologies in combination promises synergistic effects[198].

This system is versatile as well, as the target-binding domain can be switched out to virtually any affinity scaffold. In this work, I showed depletion of three different model proteins with three affinity domains. However, all of the domains were based on the monobody scaffold. There are many other protein interaction domains that have been developed against myriad proteins[199,200]. Many of these interaction domains were designed to inhibit the function of the target protein, but the scaffolds for these adaptors need not be limited to inhibitors. Future work includes testing our adaptors with other affinity scaffolds, such as nanobodies. A further avenue is to explore delivery of the adaptors into cells at the protein, rather than plasmid, level. This would be useful because delivery of the protein allows for control of adaptor concentration in the cell. The

adaptors were transiently transfected into cells, and thus the amount of adaptor could not be tuned. Another option would be to express the adaptors under control of an inducible promoter such as Tet On and titrate their levels with different amounts of inducer.

In conclusion, I have developed a method to artificially deplete specific proteins from cells using the intrinsic cellular degradation machinery, the 26S proteasome. The technology is based on proteasome adaptors, consisting of a UBL attached to an affinity domain, which bind to a protein of interest and shuttle it to the proteasome for destruction. This system is ubiquitin-independent, versatile, and robust. The adaptors developed in this study induced depletion of three different proteins both *in vitro* and in cells. Beyond just showing degradation of specific targets, my findings also underscore the importance of proper substrate positioning on the proteasome. This system could be used in many ways, including studying protein function and stopping the action of disease-causing proteins.

Chapter 4: Materials and Methods

In vitro *substrate and adaptor plasmid design:* *In vitro* constructs were cloned into the second multiple cloning site in a pETDUET vector with a C-terminal Chitin Intein binding domain tag followed by a 6x His tag and expressed from the T7 promoter. Constructs were cloned via assembly PCR. Substrates contained an N-terminal HA tag followed by the SH2 domain (Abl1: aa 127-217; Shp2 $SH2^N$: aa 1-103; Shp2 $SH2^C$: aa 97-217) and monomeric superfolder GFP (msGFP) with a C-terminal 35 amino acid tail derived from cytochrome b_2 that served as an initiation region[13,177]. Adaptors had an N-terminal FLAG tag followed by the UBL from *S. cerevisiae* Rad23 (aa 1-77) fused to a monobody through a GGSGGT linker. The binding monobodies were HA4 (Abl1), Nsa1 (Shp2 $SH2^N$), and Cs3 (Shp2 $SH2^C$) and the non-binding monobody was HA4(Y87A) which does not bind to Abl SH2 domain[137,138].

Antibodies: Primary antibodies used were Abl (mouse, monoclonal, clone 8E9, Santa Cruz), HA (mouse, monoclonal, Covance), FLAG (mouse, monoclonal, Sigma), Shp2 (mouse, monoclonal, BD), GFP (mouse, monoclonal, Clontech), Actin (rabbit, monoclonal, Sigma). Secondary antibodies were goat anti-mouse-IR 800 dye conjugated (LI-COR), goat anti-rabbit-IR 700 dye conjugated (LI-COR).

Protein Purification: Yeast proteasome was purified from S. cerevisiae strain YYS40 (*MATa rpn11::RPN11 3 × FLAG-HIS3 leu2 his3 trp1 ade2 can1 ssd1*) by immunoaffinity chromatography using FLAG-conjugated M2 agarose affinity beads (Sigma)[180]. Proteasome preparations were analyzed by SDS-PAGE and compared to published compositions. Each proteasome preparation was checked for activity by testing degradation of the proteasome substrate UBL-GFP-95aa tail.

The substrates and adaptors used for the *in vitro* inhibition assays were overexpressed in and purified from E. coli strains BL21(DE3)pLysS or Rosetta(DE3)pLysS (Novagen). Bacterial strains were grown in 1 L of 2x YT Media at 37 °C to an optical density of 0.6-1.0. Protein expression was induced with 0.3 mM isopropyl β-D-1- thiogalactopyranoside (IPTG), and incubation continued overnight at 16 °C. Cells were harvested by centrifugation, re-suspended in cold 20 mM TRIS pH 7.4, 500 mM NaCl, 10% glycerol and homogenized using the Avestin Emulsiflex C3 (2 passes) at 15,000 psi. The lysate was cleared by centrifugation at 38,000 x g for 45 minutes. The cell lysate was incubated with Ni-NTA metal affinity beads (GE Healthcare) and nutated for 1 hour at 4 °C. This mixture was applied to a gravity column and washed with NPI buffer (50 mM sodium phosphate pH 7.4, 300 mM NaCl) plus 10 mM imidazole, then NPI plus 20 mM imidazole. The protein was eluted from the resin with NPI plus 250 mM imidazole. The elution was desalted by PD-10 desalting column (GE Healthcare) and changed into Chitin-column binding buffer (CBB) containing 20 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol, and 0.5 mM EDTA. This was bound to Chitin beads (NEB) then washed with 50 bed volumes of CBB. The protein was eluted after an overnight incubation in CBB + 0.1M DTT, and buffer exchanged into a storage buffer containing 20 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol. Protein concentration was measured by using either absorbance at 280 nm and the extinction coefficient from the proteins' sequence (ExPASy's ProtParam) or by the Pierce 660 nm protein kit (Thermo Scientific). The identity and purity of purified proteins was confirmed by SDS-PAGE.

Substrate Degradation Assay: The degradation of fluorescent substrate *in vitro* was monitored by fluorescence intensity and performed in 384-well plate on plate reader

(Infinite M1000 PRO, Tecan) as previously published[180]. Briefly, substrate and adaptor were diluted separately in assay buffer (10 mM Tris HCl, 50mM MgAc, 5% glycerol), then 10 μL of adaptor and substrate were mixed and incubated at 30 °C for 5 minutes. Proteasome was diluted in assay buffer plus ATP regeneration system[180]. 20 μL of proteasome was added to the adaptor/substrate mixture and the plate was immediately put into the plate reader. Fluorescence intensity was read out every minute at the excitation wavelength of 388 nm and the emission wavelength of 420 nm for 1-3 hours. Each assay was repeated at least two times. Initial degradation rates are given by the slope of the decay curves at time zero and are calculated as the product of the amplitude and the rate constant of the decay curve determined by nonlinear fitting to a single exponential decay in the software package Prism (Graphpad, version 8.0.0).

For *in vitro* time course assay, 160 nM adaptor was incubated with 20 nM substrate and 25 nM proteasome at 30 °C. 15 μL were taken out at indicated times and immediately mixed with sample buffer and boiled to quench the reaction. Samples were loaded onto 10% Tris Tricine gels and blotted with indicated antibodies.

Mammalian Construct Design: SH2-GFP-tail substrates were cloned into pcDNA5/FRT/TO plasmid (Life Technologies) under a human cytomegalovirus (CMV) promoter for integration into Flp-In cells. For adaptors, the human UBL domain from hHR23b (aa 1-79) was substituted for the yeast UBL in the *in vitro* assays. Adaptor constructs were cloned into pcDNA3 (Life Technologies) under a CMV promoter and contained an N-terminal FLAG tag and C-terminal Internal Ribosome Entry Site (IRES) followed by mCherry.

Stable Cell Generation: Stable HEK293 Flp-In cells expressing the substrate (SH2-GFP-tail) were generated using the Flp-In system (Life Technologies) according to manufacturer's instructions. Briefly, Flp-In 293 host cells were transfected with 1 μg substrate DNA in pcDNA5/FRT/TO plasmid and 9 μg pOG44 plasmid encoding Flp recombinase (Life Technologies) with Lipofectamine 2000 in a 6-well plate. 24 hours later, media was replaced with Dulbecco's Modified Eagle's Medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco), 1% penicillin/streptomycin (pen/strep, Gibco) and 1% glutamine (Q, Gibco) and cells were allowed to recover overnight. Cells were trypsinized the next day and plated on 10 cm^2 plates in fresh medium. Medium was replaced 24 hours later with DMEM supplemented with 10% FBS, 1% pen/strep/Q, and 200 μg/mL hygromycin B (Gibco). Medium was replaced every two to three days with fresh selective medium until colonies formed. All colonies were considered to be isogenic and the whole plate was trypsinized and plated into fresh selective medium in a 75 cm^2 dish, then subcultured according to manufacturer's instructions. Aliquots of cells were frozen in FBS supplemented with 10% DMSO by cooling at -1 °C/min at -80 °C and stored in a -150 °C freezer until further use.

Transfections: HEK293 Flp-In Host and SH2-GFP-tail stable cells were transfected with either binding (FLAG-hUbL-monobody-IRES-mCherry) or non-binding adaptor (FLAG-hUbL-HA4(Y87A)-IRES-mCherry) cloned into pcDNA3 (Life Technologies) or empty pcDNA3 with Lipofectamine 2000 reagent according to manufacturer's instructions. 48-72 hours after transfection, cells were washed with ice-cold D-PBS (Gibco) and harvested by trypsinization and centrifugation.

Immunoblot analysis: Cells were lysed in Lysis Buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.5% NP-40) supplemented with 1 mM DTT and Protease Inhibitor Cocktail Set V (Sigma-Aldrich) and cleared by centrifugation. For immunoblot analysis, 4 μg cleared lysate (as measured by Pierce 660nm protein assay (Thermo Scientific)) were loaded on a 4-12% Tris-Glycine gel (Bio-Rad) then transferred onto nitrocellulose membranes using the Trans-Blot Turbo system according to manufacturers instructions. Membranes were blocked with Odyssey blocking buffer (LI-COR) and stained with indicated antibodies. Blots were scanned using the Odyssey imaging system (LI-COR) and quantified with ImageJ software. Substrate bands were normalized to Actin and/or vector control as indicated. For SH2-GFP-tail integrated transfections, cells were first sorted on a BD FACSARIA Fusion flow cytometer. 100,000 mCherry-positive (as indicated in figures) cells were collected and immunoblotted as above.

Flow cytometry: Transfected cells were washed in ice-cold D-PBS (Gibco) and trypsinized then neutralized with D-MEM with no phenol red (Gibco) supplemented with 10% FBS (Gibco). Cells were analyzed on a BD Fortessa with 10,000 cells typically represented. Data analysis was performed with FlowJo v10.

Appendix A: Testing model substrate configuration

When we initially set up the *in vitro* experiments, we wanted to test several domain configurations of the model substrate. The Abl SH2 domain is stable when purified on its own[137], but is located between two folded domains in the full-length endogenous protein. Adding a GFP domain could change the stability of the SH2 depending on which terminus it was put. Further, in order to ensure that degradation could occur, we needed to add an initiation region for the proteasome to engage[177]. The location of the 35 amino acid tail could drastically affect the degradation of the model protein. Therefore, we created and assayed four configurations of the model substrate (Figure A1a): SH2-GFP-tail (SG35), tail-GFP-SH2 (35GS), tail-SH2-GFP (35SG), and GFP-SH2-tail (GS35). In all cases, there was a GGSGGT linker between the SH2 and GFP domains and an HA tag on the opposite terminus from the tail.

Surprisingly, we found that only two configurations degraded at all: SG35 and 35GS (Figure A1b, c). 35GS degraded about half as well as SG35. Clearly, only when the initiation region was attached to the GFP domain was any degradation seen. The most degradation occurred when the initiation region was at the C-terminus. This is in agreement with findings from others in our lab, who showed that this particular sequence serves as a better initiation region when at the C-terminus of the folded domain (unpublished observations).

We next investigated why only two of the conformations worked in our assay and came up with several hypotheses. First, the positioning of the substrate on the proteasome may not be optimal in the case of the non-degraded substrates. We showed previously that the geometry of the ternary complex is important, as changing the configuration of

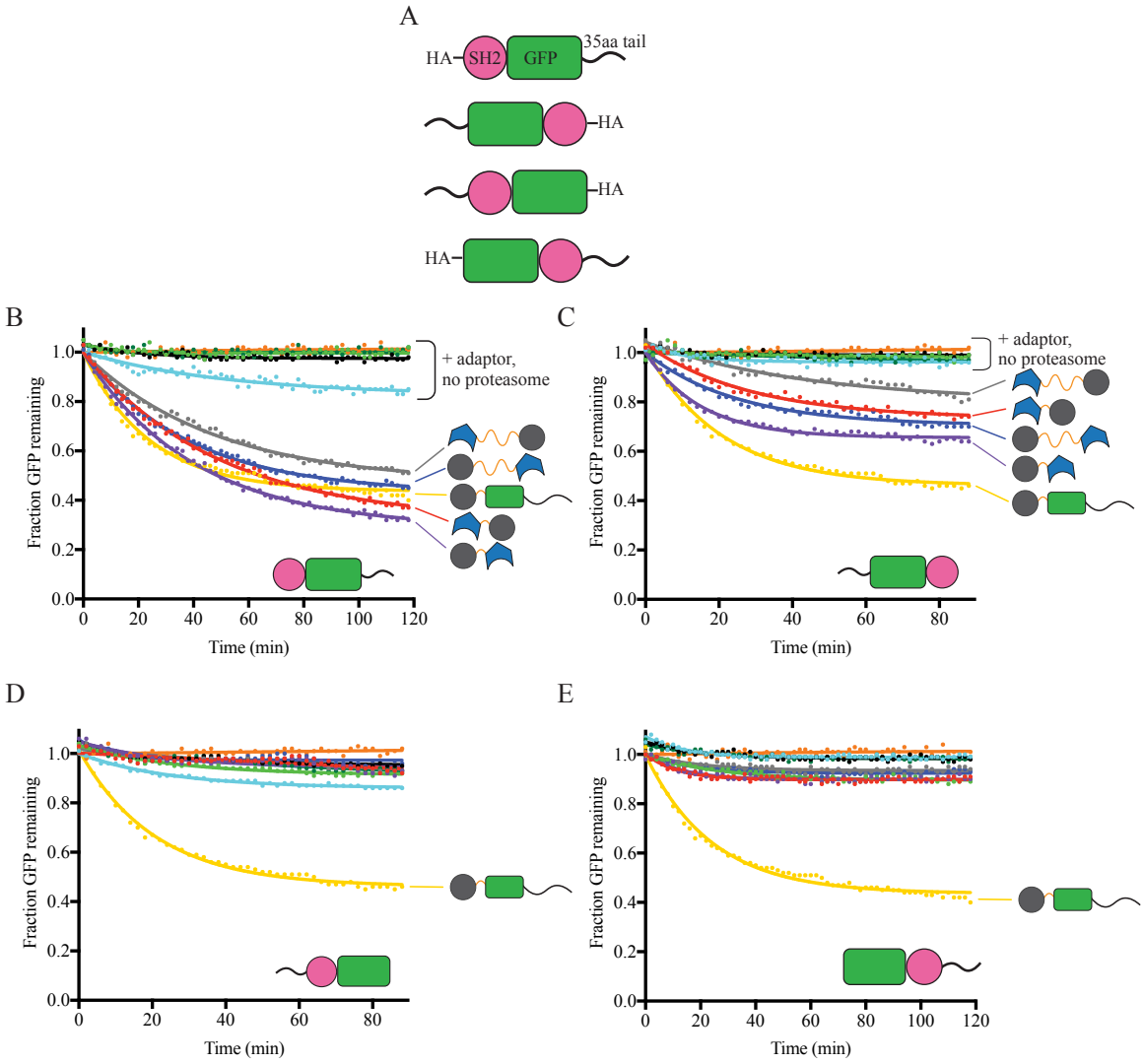


Figure A1: Degradation of different model substrate geometries.

A) Schematic of model substrate configurations. Pink circle represents SH2 domain, green square represents GFP, and black line represents 35 amino acid initiation region. B-E) Fluorescence assay monitoring substrate through GFP fluorescence over time. 20 nM indicated substrate was incubated with 160nM of different adaptor configurations in the presence or absence of 25nM proteasome. Positive control: UBL-GFP-90aa tail.

the adaptor affected degradation ability (Figure 6). Changing the substrate configuration would also change the position of the substrate on the proteasome. In the case of the two configurations that remained stable in our assay, the adaptor, bound to the SH2 domain, would be very close to the initiation region and may sterically hinder the translocation[201]. We reasoned that perhaps a longer linker between the adaptor domains would help position the substrate correctly. However, none of the adaptors were able to induce degradation of those two substrates (Figure A1d, e).

It is also possible that the substrate was getting only partially degraded. In our plate reader assays, we are not directly measuring the concentration of the SH2 domain, but rather the fluorescence of the GFP domain. If the proteasome got "stuck" after the SH2 domain, the GFP signal would remain stable. While the proteasome generally degrades processively, there are examples of partial degradation in cells when the proteasome encounters a stable folded domain[201]. For example, the structure of the enzyme dihydrofolate reductase (DHFR) is greatly stabilized when the ligand methotrexate is bound. In the presence of methotrexate, the proteasome will pause degradation when it encounters a DHFR domain, no matter where in the substrate the domain is put[201]. To test whether this was the case, we took the plate reader samples at the end of the experiment and ran them on an SDS-PAGE gel then analyzed the fluorescence using a Typhoon imager (Figure A2). There was no fragment with our best model substrate, SG35 (Figure A2a). We saw a lower molecular weight species appear at the end of the assay for two of the substrates, 35GS and GS35 Figure A2b and d, respectively). The full-length substrates were 48 kDa (GFP = 27kDa, SH2 = 17kDa, 35 aa tail = 4.4kDa) and the fragments were ~30kDa. We were surprised to see the fragment in 35GS, because it degraded about 40% in our assay. If the proteasome degraded the GFP portion of that substrate, it would not have appeared on the Typhoon assay. It's

possible that there was non-specific cleavage between the domains. The appearance of the ~30kDa fragment in the GS35 sample suggested that the proteasome did degrade the SH2 domain but could not degrade the GFP domain. In the case of 35SG, there was a fragment even without the proteasome (Figure A3c). This suggested that the substrate may be prone to proteolytic cleavage between the domains. Indeed, 35SG was less stable during purification, with some precipitation observed, further suggesting it is sensitive to proteases.

Finally, we hypothesized that the 35aa tail was not long enough to reach the translocation channel. As mentioned before, the positioning of the substrate on the proteasome is very important, and while lengthening the adaptor did not induce degradation, we reasoned that perhaps lengthening the initiation region would help[177]. We therefore replaced the 35aa IR with a 90aa tail (derived from the N-terminus of cytochrome b_2)[177]. This helped slightly, as about 20% of the substrate was degraded (Figure A3).

While we were not able to pinpoint the exact cause of substrate stability when the initiation region was attached to the SH2 domain, it is reasonable to assume that a combination of the three hypotheses is responsible. When the SH2 domain is next to the initiation region, the adaptor may not be able to position the substrate so that the tail is accessible to the proteasome. Also, the adaptor binds to the SH2 domain with low nanomolar affinity[137] and may stabilize the domain structure to the extent that the proteasome cannot degrade it. Increasing the length of the initiation region did help somewhat, suggesting that an extended tail allows the proteasome to engage the substrate. The appearance of a low molecular weight band at the end of the assay also suggests some intrinsic substrate instability or susceptibility to protease cleavage.

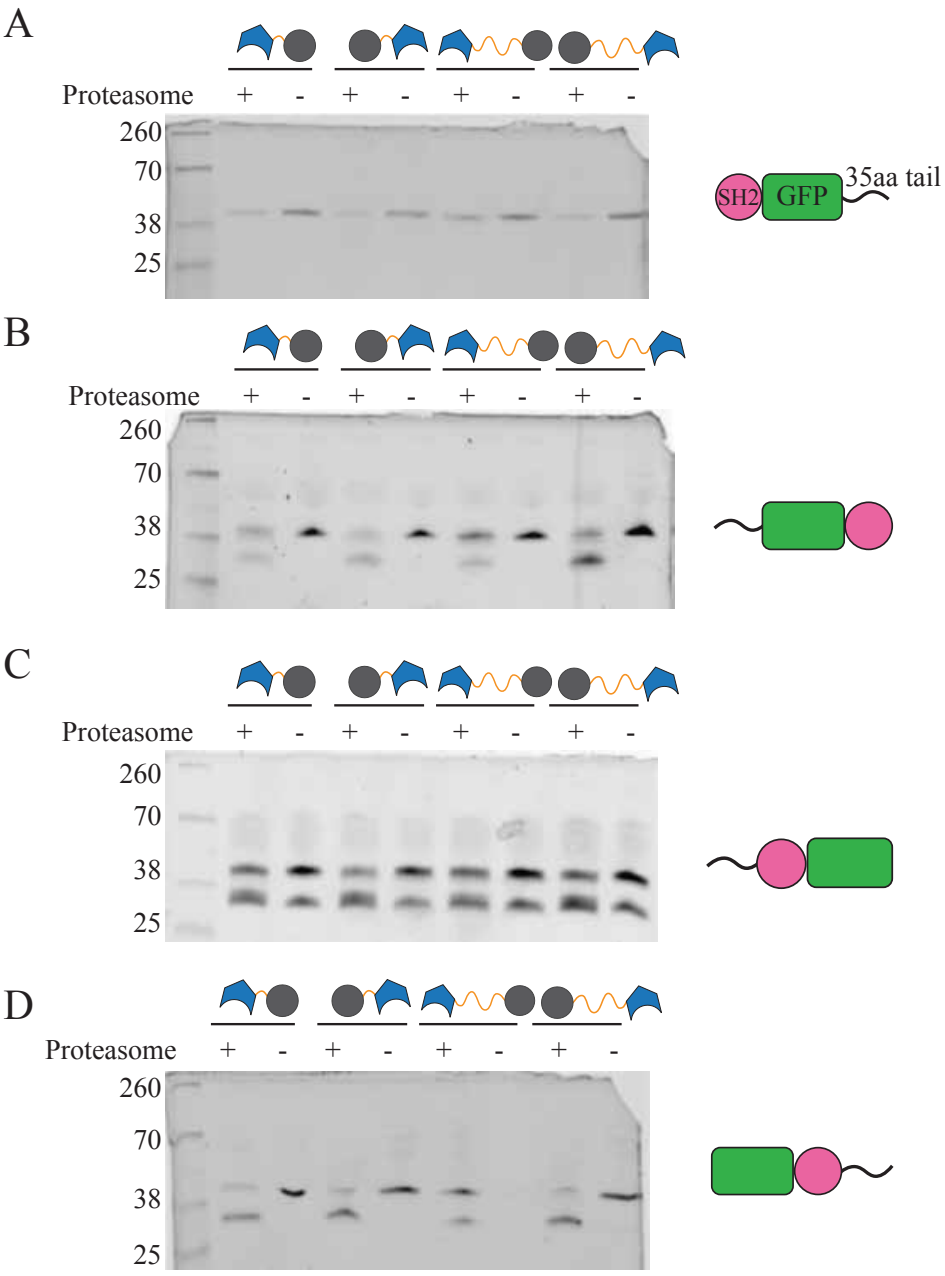


Figure A2: Fluorescence imaging of assay endpoints.

A-D) Fluorescence images from Typhoon taken with a 488nm excitation filter. Indicated substrates (20nM) were incubated with different configurations of adaptor (160nM) with or without the proteasome (25nM).

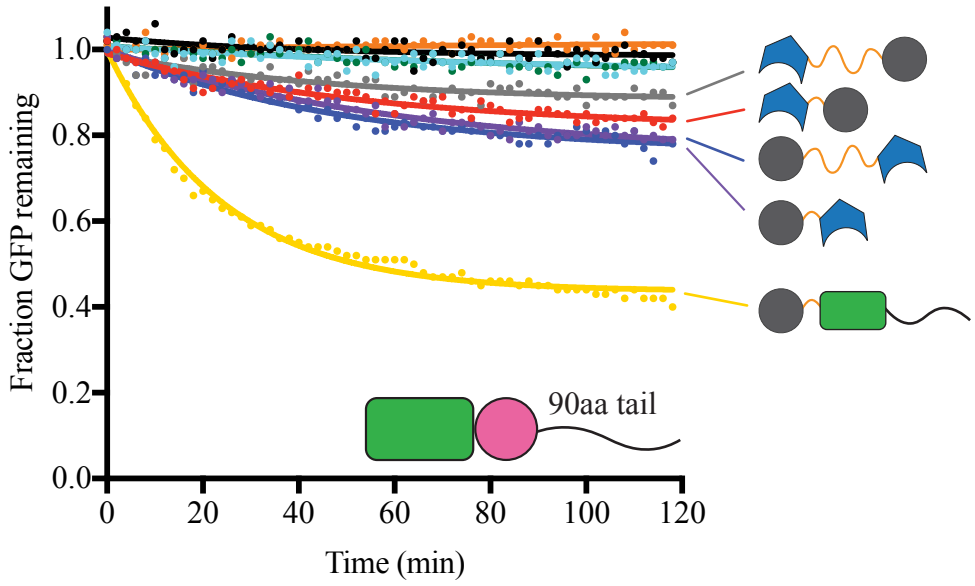


Figure A3: Lengthening the initiation region slightly increases degradation.

Fluorescence assay measuring the fraction of GFP signal remaining over time. 20nM substrate with 90aa tail was incubated with 160nM of different adaptor configurations and 25nM proteasome. Positive control: UBL-GFP-90aa tail.

Appendix B: Flow cytometry on endogenous Shp2

After confirming that our technology is capable of degrading different model substrates both *in vitro* and in cells, the next step was to test the system on endogenous proteins. I chose Shp2 because it is ubiquitously expressed in the cell and readily detectable with an antibody (see Figure 11). To analyze cellular protein concentration, Western blots are widely used because they are simple to perform and show a visual of the protein separated on a gel. Results can be normalized and quantified by including a reference protein, such as actin, whose concentration is fairly consistent between cells that have undergone the same treatment. Western blots require two antibodies, a primary that recognizes the protein of interest and a secondary that recognizes the primary species and is conjugated to a dye or fluorophore that can be used to visualize the protein. With the recent advent of fluorescence or infrared labeling, Western blot quantitation has become more reliable than the enzymatic chemiluminescent reaction[202]. However, quantification of protein bands still relies on pixel counting in an image, which can become saturated and skew the results. Also, using a housekeeping gene as a loading reference is not reliable between cell lines and differing treatment regimens[203]. Therefore, I wanted to use an orthogonal technique to the Western blotting to quantify the depletion of endogenous Shp2 in cells.

Immunofluorescence staining followed by flow cytometry is a method for detecting either surface or intracellular proteins[204]. After cells are collected, they are fixed and permeabilized, then stained with primary antibody followed by fluorophore-conjugated secondary. The stained cells are then analyzed by flow cytometry (see Chapter 4, Methods). Using this protocol measures the secondary antibody directly within the cell and can be normalized to background fluorescence. Fixing the cells

prevents protease action that can occur in lysate during sample preparation for Western blots.

I first validated the Shp2 antibody (mouse anti-Shp2, BD Biosciences) by transfecting the cells with siRNA (Dharmacon) and staining with a secondary conjugated to AlexaFluor-488 (goat anti-mouse, GE Life Sciences). Untreated cells showed an increase in 488 fluorescence as compared to cells without primary, and siRNA-treated cells had a decreased fluorescence (Figure B1a). This was confirmed by Western blot as well (Figure B1b).

After antibody validation, I transfected cells with either $SH2^N$-binding or non-binding adaptor and performed the assay. There was a decrease in cell fluorescence in transfected (mCherry positive) cells containing the binding adaptor as compared to non-binding (Figure B2a). The primary antibody was raised against the N-terminus of Shp2 (where the SH2 domains are), and it was possible that the antibody bound Shp2 in the same region as the adaptor. Because the cells were fixed (and crosslinked), any interactions between the adaptor and Shp2 would remain intact during the assay. This is as opposed to Western blotting, where all interactions are destroyed upon denaturation. It was therefore plausible that the decrease in fluorescence was due to the antibody's inability to bind to Shp2, rather than adaptor-mediated Shp2 depletion. As a control, I transfected the cells with monobody only (no UBL) so the Shp2 would not be targeted to the proteasome. Indeed, the cells showed nearly the same decrease in fluorescence with monobody only as compared to adaptor (Figure B2b). I concluded that the antibody and monobody did bind to the same place and the effect was actually from competition of binding rather than degradation. I saw the same effect with the $SH2^C$-binding adaptor (data not shown). I also tested an antibody that bound to the Shp2 phosphatase domain but was unable to validate the antibody specificity with siRNA (data not shown). In the

end, I used Western blot only to show depletion of endogenous Shp2 and saw ~50% degradation (Figure 11).

To ensure that the adaptor-induced degradation of Shp2 was proteasome-dependent, I performed the flow cytometry assay in the presence of proteasome inhibitor (100 nM PS341 (bortezomib)). 24h after transfecting the cells with adaptor, I added the proteasome inhibitor and harvested the cells 24h later; however, I did not see any increase in Shp2 fluorescence (which would indicate stabilized Shp2) via flow cytometry (data not shown). I believe that this is due to the fact that my assay was invalid and was not measuring degradation, but rather antibody binding. I will perform the proteasome inhibition experiment on the model substrates.

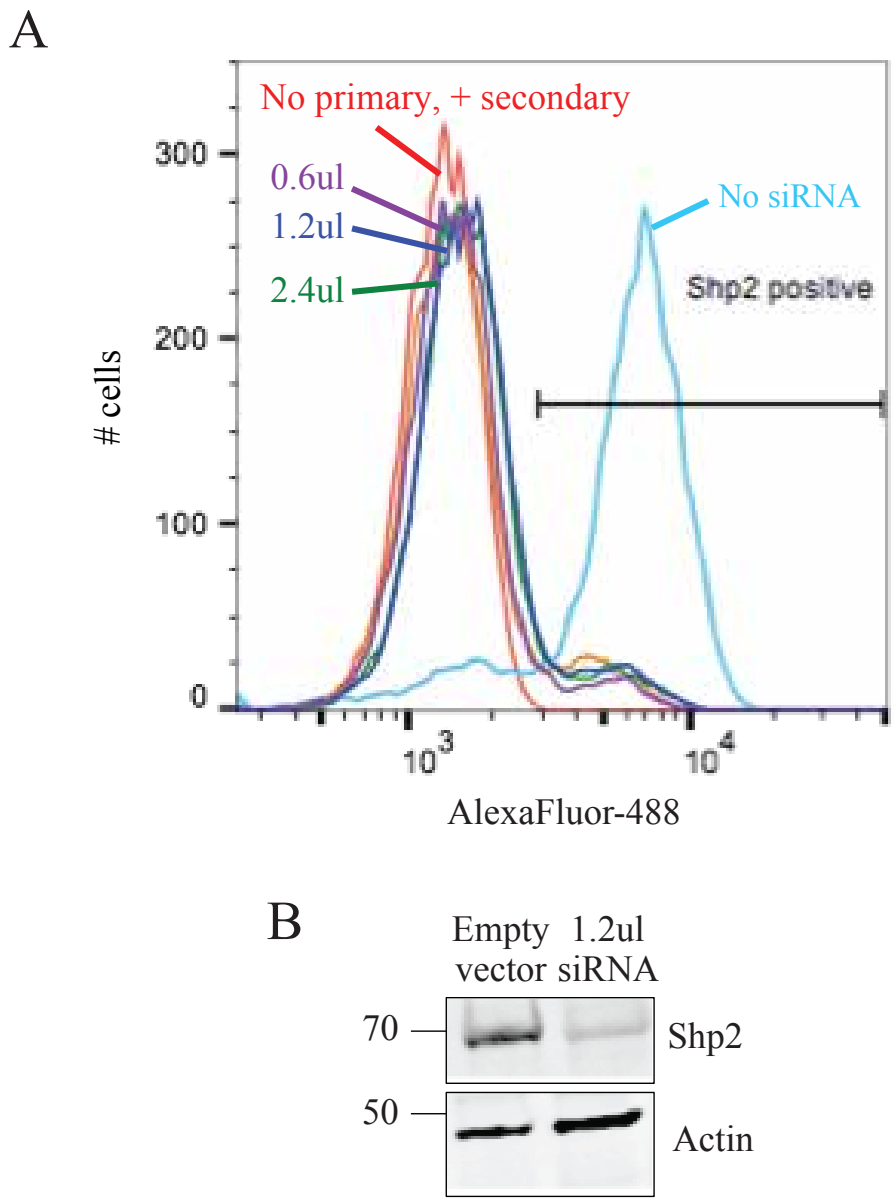
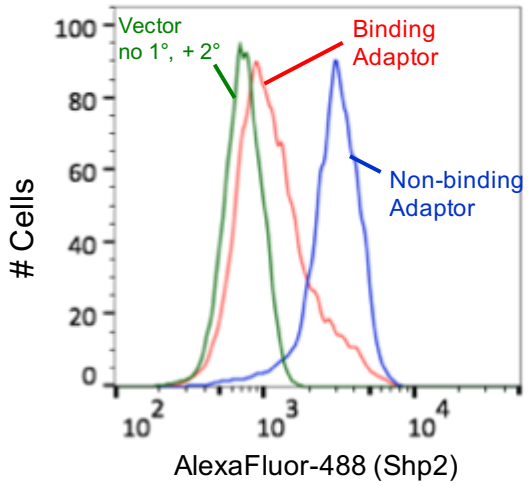


Figure B1: Validation of anti-Shp2 antibody

A) Histogram of cells treated with either 0.6, 1.2, or 2.4ul siRNA (20uM). Light blue line: cells transfected with empty pcDNA3 vector. Red line: cells stained with secondary antibody only (for background fluorescence). B) Western blot of cells transfected with empty vector or 1.2ul siRNA.

A



B

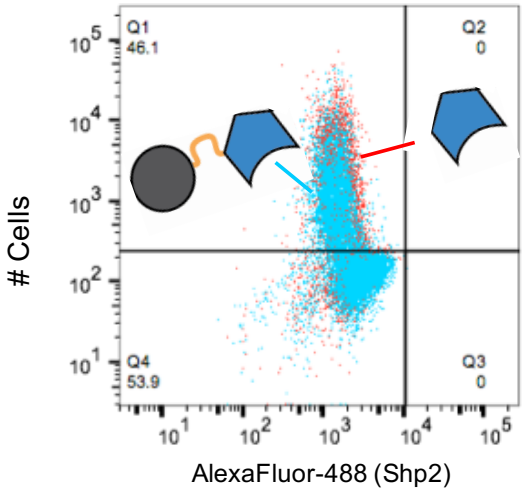


Figure B2: Shp2 antibody and adaptor compete for binding

A) Histogram of transfected cells with either non-binding (blue), $SH2^N$-binding (red) adaptor, or empty vector. B) Immunofluorescence of HEK293 cells transfected with either $SH2^N$-binding adaptor (blue) or monobody only (red).

Glossary

Abl: Abelson kinase

ABD: Actin-binding domain

cIAP: Cellular inhibitor of apoptosis

Cp8: GFP circular permutant 8

DD: Destabilizing domain

DHFR: Dihydrofolate reductase

DUB: Deubiquitinating enzyme

ERAD: Endoplasmic reticulum associated degradation

ERα: Estrogen receptor α

HEK293: Human Embryonic Kidney 293 cells

hHR23b: human homolog Rad23b

IRES: Internal ribosome entry site

MDM2: Mouse double minute 2

miRNA: micro RNA

PROTAC: Proteolysis targeting chimera

RNAi: RNA interference

SH2: Src homology domain 2

Shp2: SH2 domain containing phosphatase 2

siRNA: silencing RNA

TRIM21

UBA: Ubiquitin-associated domain

UBL: Ubiquitin-like domain

UPS: Ubiquitin proteasome system

VHL: Von Hippel Lindau

YFP/GFP: Yellow/Green fluorescent protein

References

1. Tomko, R.J. & Hochstrasser, M. Molecular architecture and assembly of the eukaryotic proteasome. *Annu Rev Biochem* **82**, 415-445 (2013).

2. Bard, J.A.M. *et al.* Structure and Function of the 26S Proteasome. *Annu Rev Biochem* **87**, 697-724 (2018).

3. de la Peña, A.H., Goodall, E.A., Gates, S.N., Lander, G.C. & Martin, A. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis–driven translocation. *Science* **301**, eeav0725 (2018).

4. Saeki, Y. *et al.* Lysine 63-linked polyubiquitin chain may serve as a targeting signal for the 26S proteasome. *EMBO J* **28**, 359-71 (2009).

5. Koulich, E., Li, X. & DeMartino, G.N. Relative structural and functional roles of multiple deubiquitylating proteins associated with mammalian 26S proteasome. *Mol Biol Cell* **19**, 1072-82 (2008).

6. Lee, B.H. *et al.* Enhancement of proteasome activity by a small-molecule inhibitor of USP14. *Nature* **467**, 179-84 (2010).

7. Schauber, C. *et al.* Rad23 links DNA repair to the ubiquitin/proteasome pathway. *Nature* **391**, 715-718 (1998).

8. Hartmann-Petersen, R. & Gordon, C. Proteins interacting with the 26S proteasome. *Cell Mol Life Sci* **61**, 1589-95 (2004).

9. Elsasser, S. *et al.* Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat Cell Biol* **4**, 725-730 (2002).

10. Hartmann-Petersen, R. & Gordon, C. Protein degradation: recognition of ubiquitinylated substrates. *Curr Biol* **14**, R754-6 (2004).

11. Shi, Y. *et al.* Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351**(2016).

12. Yu, H. *et al.* Conserved Sequence Preferences Contribute to Substrate Recognition by the Proteasome. *J Biol Chem* **291**, 14526-39 (2016).

13. Yu, H., Kago, G., Yellman, C.M. & Matouschek, A. Ubiquitin-like domains can target to the proteasome but proteolysis requires a disordered region. *EMBO J* **35**, 1522-1536 (2016).

14. Fishbain, S. *et al.* Sequence composition of disordered regions fine-tunes protein half-life. *Nat Struct Mol Biol* **22**, 214-221 (2015).

15. van der Lee, R. *et al.* Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep* **8**, 1832-1844 (2014).

16. Worden, E.J., Padovani, C. & Martin, A. Structure of the Rpn11-Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nat Struct Mol Biol* **21**, 220-7 (2014).

17. Worden, E.J., Dong, K.C. & Martin, A. An AAA Motor-Driven Mechanical Switch in Rpn11 Controls Deubiquitination at the 26S Proteasome. *Mol cell* **67**, 799-811 (2017).

18. Dong, Y. *et al.* Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **565**, 49-55 (2019).

19. Chen, S. *et al.* Structural basis for dynamic regulation of the human 26S proteasome. *Proc Natl Acad Sci U S A* **113**, 12991-12996 (2016).

20. Eisele, M.R. *et al.* Expanded Coverage of the 26S Proteasome Conformational Landscape Reveals Mechanisms of Peptidase Gating. *Cell Rep* **24**, 1301-1315 (2018).

21. Jung, T. & Grune, T. Structure of the proteasome. *Prog Mol Biol Transl Sci* **109**, 1-39 (2012).

22. Unno, M. *et al.* The structure of the mammalian 20S proteasome at 2.75 A resolution. *Structure* **10**, 609-18 (2002).

23. Cadwell, K. & Coscoy, L. Ubiquitination on nonlysine residues by a viral E3 ubiquitin ligase. *Science* **309**, 127-30 (2005).

24. Ciechanover, A. & Schwartz, A.L. The ubiquitin system: pathogenesis of human diseases and drug targeting. *Biochim Biophys Acta* **1695**, 3-17 (2004).

25. Ciechanover, A. & Ben-Saadon, R. N-terminal ubiquitination: more protein substrates join in. *Trends Cell Biol* **14**, 103-6 (2004).

26. Pickart, C.M. & Fushman, D. Polyubiquitin chains: polymeric protein signals. *Curr Opin Chem Biol* **8**, 610-6 (2004).

27. Komander, D. & Rape, M. The ubiquitin code. *Annu Rev Biochem* **81**, 203-229 (2012).

28. Soetens, O., De Craene, J.O. & Andre, B. Ubiquitin is required for sorting to the vacuole of the yeast general amino acid permease, Gap1. *J Biol Chem* **276**, 43949-57 (2001).

29. Galan, J.M. & Haguenauer-Tsapis, R. Ubiquitin lys63 is involved in ubiquitination of a yeast plasma membrane protein. *EMBO J* **16**, 5847-54 (1997).

30. Spence, J., Sadis, S., Haas, A.L. & Finley, D. A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol Cell Biol* **15**, 1265-73 (1995).

31. Mukhopadhyay, D. & Riezman, H. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science* **315**, 201-5 (2007).

32. Xu, P. *et al.* Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation. *Cell* **137**, 133-45 (2009).

33. Pickart, C.M. Ubiquitin in chains. *Trends Biochem Sci* **25**, 544-8 (2000).

34. Chau, V. *et al.* A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **243**, 1576-83 (1989).

35. Thrower, J.S., Hoffman, L., Rechsteiner, M. & Pickart, C.M. Recognition of the polyubiquitin proteolytic signal. *EMBO J* **19**, 94-102 (2000).

36. Matsumoto, M.L. *et al.* K11-linked polyubiquitination in cell cycle control revealed by a K11 linkage-specific antibody. *Mol Cell* **39**, 477-84 (2010).

37. Nathan, J.A., Kim, H.T., Ting, L., Gygi, S.P. & Goldberg, A.L. Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? *EMBO J* **32**, 552-65 (2013).

38. Hrdinka, M. & Gyrd-Hansen, M. The Met1-Linked Ubiquitin Machinery: Emerging Themes of (De)regulation. *Mol Cell* **68**, 265-280 (2017).

39. Kulathu, Y. & Komander, D. Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat Rev Mol Cell Biol* **13**, 508-23 (2012).

40. Weissman, A.M. Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* **2**, 169-78 (2001).

41. Liu, F. & Walters, K.J. Multitasking with ubiquitin through multivalent interactions. *Trends Biochem Sci* **35**, 352-60 (2010).

42. Pickart, C.M. Mechanisms underlying ubiquitination. *Annu Rev Biochem* **70**, 503-33 (2001).

43. Hoppe, T. Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. *Trends Biochem Sci* **30**, 183-7 (2005).

44. Deshaies, R.J. & Joazeiro, C.A. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**, 399-434 (2009).

45. Li, W. *et al.* Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling. *PLoS One* **3**, e1487 (2008).

46. Rotin, D. & Kumar, S. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* **10**, 398-409 (2009).

47. Murakami, Y., Matsufuji, S., Hayashi, S., Tanahashi, N. & Tanaka, K. Degradation of ornithine decarboxylase by the 26S proteasome. *Biochem Biophys Res Commun* **267**, 1-6 (2000).

48. Janse, D.M., Crosas, B., Finley, D. & Church, G.M. Localization to the proteasome is sufficient for degradation. *J Biol Chem* **279**, 21415-20 (2004).

49. Murakami, Y. *et al.* Ornithine decarboxylase is degraded by the 26S proteasome without ubiquitination. *Nature* **360**, 597-9 (1992).

50. Lim, S.K. & Gopalan, G. Antizyme1 mediates AURKAIP1-dependent degradation of Aurora-A. *Oncogene* **26**, 6593-603 (2007).

51. Newman, R.M. *et al.* Antizyme targets cyclin D1 for degradation. A novel mechanism for cell growth repression. *J Biol Chem* **279**, 41504-11 (2004).

52. Murai, N., Murakami, Y., Tajima, A. & Matsufuji, S. Novel ubiquitin-independent nucleolar c-Myc degradation pathway mediated by antizyme 2. *Sci Rep* **8**, 3005 (2018).

53. Saeki, Y., Saitoh, A., Toh-e, A. & Yokosawa, H. Ubiquitin-like proteins and Rpn10 play cooperative roles in ubiquitin-dependent proteolysis. *Biochem Biophys Res Commun* **293**, 986-92 (2002).

54. Husnjak, K. *et al.* Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**, 481-8 (2008).

55. Batty, D.P. & Wood, R.D. Damage recognition in nucleotide excision repair of DNA. *Gene* **241**, 193-204 (2000).

56. Chen, L. & Madura, K. Rad23 promotes the targeting of proteolytic substrates to the proteasome. *Mol Cell Biol* **22**, 4902-13 (2002).

57. Ortolan, T.G. *et al.* The DNA repair protein rad23 is a negative regulator of multi-ubiquitin chain assembly. *Nat Cell Biol* **2**, 601-8 (2000).

58. Sugasawa, K. *et al.* Two human homologs of Rad23 are functionally interchangeable in complex formation and stimulation of XPC repair activity. *Mol Cell Biol* **17**, 6924-31 (1997).

59. Chen, L. & Madura, K. Evidence for distinct functions for human DNA repair factors hHR23A and hHR23B. *FEBS Lett* **580**, 3401-8 (2006).

60. Elsasser, S. & Finley, D. Delivery of ubiquitinated substrates to protein-unfolding machines. *Nat Cell Biol* **7**, 742-749 (2005).

61. Watkins, J.F., Sung, P., Prakash, L. & Prakash, S. The Saccharomyces cerevisiae DNA repair gene RAD23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Mol Cell Biol* **13**, 7757-7765 (1993).

62. Fishbain, S., Prakash, S., Herrig, A., Elsasser, S. & Matouschek, A. Rad23 escapes degradation because it lacks a proteasome initiation region. *Nat Commun* **2**, 192 (2011).

63. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431-7 (2004).

64. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).

65. Bianchini, G., Balko, J.M., Mayer, I.A., Sanders, M.E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol* **13**, 674-690 (2016).

66. Albert, F.W. & Kruglyak, L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet* **16**, 197-212 (2015).

67. Bresnick, A.R., Weber, D.J. & Zimmer, D.B. S100 proteins in cancer. *Nat Rev Cancer* **15**, 96-109 (2015).

68. Aceto, N. *et al.* Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop. *Nat Med* **18**, 529-37 (2012).

69. Ren, R. Mechanisms of BCR-ABL in the pathogenesis of chronic myelogenous leukaemia. *Nat Rev Cancer* **5**, 172-183 (2005).

70. Ward, P.S. *et al.* The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* **17**, 225-34 (2010).

71. Chan, G., Kalaitzidis, D. & Neel, B.G. The tyrosine phosphatase Shp2 (PTPN11) in cancer. *Cancer Metastasis Rev* **27**, 179-92 (2008).

72. Hannon, G.J. & Rossi, J.J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371-8 (2004).

73. Mello, C.C. & Conte, D., Jr. Revealing the world of RNA interference. *Nature* **431**, 338-42 (2004).

74. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-11 (1998).

75. Hannon, G.J. RNA interference. *Nature* **418**, 244-51 (2002).

76. Gonsalves, F.C. *et al.* An RNAi-based chemical genetic screen identifies three small-molecule inhibitors of the Wnt/wingless signaling pathway. *Proc Natl Acad Sci U S A* **108**, 5954-63 (2011).

77. Kamath, R.S. *et al.* Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421**, 231-7 (2003).

78. Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524-8 (2011).

79. Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-48 (2009).

80. Morrison, C. Alnylam prepares to land first RNAi drug approval. *Nat Rev Drug Discov* **17**, 156-157 (2018).

81. Chen, X. *et al.* RNA interference-based therapy and its delivery systems. *Cancer Metastasis Rev* **37**, 107-124 (2018).

82. Scacheri, P.C. *et al.* Short interfering RNAs can induce unexpected and divergent changes in the levels of untargeted proteins in mammalian cells. *Proc Natl Acad Sci U S A* **101**, 1892-7 (2004).

83. Jackson, A.L. *et al.* Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA* **12**, 1179-87 (2006).

84. Zhang, J., Zheng, N. & Zhou, P. Exploring the functional complexity of cellular proteins by protein knockout. *Proc Natl Acad Sci U S A* **100**, 14127-14132 (2011).

85. Mann, M. & Jensen, O.N. Proteomic analysis of post-translational modifications. *Nat Biotech* **21**, 255-261 (2003).

86. Sakamoto, K.M. *et al.* Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci U S A* **98**, 8554-9 (2001).

87. Bhattacharyya, S., Yu, H., Mim, C. & Matouschek, A. Regulated protein turnover: snapshots of the proteasome in action. *Nat Rev Mol Cell Biol* **15**, 122-133 (2014).

88. Bondeson, D.P. *et al.* Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat Chem Biol* **11**, 611-7 (2015).

89. Buckley, D.L. *et al.* Targeting the von Hippel-Lindau E3 ubiquitin ligase using small molecules to disrupt the VHL/HIF-1alpha interaction. *J Am Chem Soc* **134**, 4465-8 (2012).

90. Kamura, T. *et al.* Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science* **284**, 657-61 (1999).

91. Chamberlain, P.P. *et al.* Structure of the human Cereblon-DDB1-lenalidomide complex reveals basis for responsiveness to thalidomide analogs. *Nat Struct Mol Biol* **21**, 803-9 (2014).

92. Kronke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301-5 (2014).

93. Ohoka, N. *et al.* In Vivo Knockdown of Pathogenic Proteins via Specific and Nongenetic Inhibitor of Apoptosis Protein (IAP)-dependent Protein Erasers (SNIPERs). *J Biol Chem* **292**, 4556-4570 (2017).

94. Itoh, Y., Ishikawa, M., Naito, M. & Hashimoto, Y. Protein knockdown using methyl bestatin-ligand hybrid molecules: design and synthesis of inducers of ubiquitination-mediated degradation of cellular retinoic acid-binding proteins. *J Am Chem Soc* **132**, 5820-6 (2010).

95. Hines, J., Lartigue, S., Dong, H., Qian, Y. & Crews, C.M. MDM2-Recruiting PROTAC Offers Superior, Synergistic Antiproliferative Activity via Simultaneous Degradation of BRD4 and Stabilization of p53. *Cancer Res* **79**, 251-262 (2019).

96. Bondeson, D.P. *et al.* Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead. *Cell Chem Biol* **25**, 78-87 (2018).

97. Smith, B.E. *et al.* Differential PROTAC substrate specificity dictated by orientation of recruited E3 ligase. *Nat Commun* **10**, 131 (2019).

98. Lu, J. *et al.* Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem Biol* **22**, 755-63 (2015).

99. Corson, T.W., Aberle, N. & Crews, C.M. Design and Applications of Bifunctional Small Molecules: Why Two Heads Are Better Than One. *ACS Chem Biol* **3**, 677-692 (2008).

100. Okuhira, K. *et al.* Development of hybrid small molecules that induce degradation of estrogen receptor-alpha and necrotic cell death in breast cancer cells. *Cancer Sci* **104**, 1492-8 (2013).

101. Brzozowski, A.M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753-8 (1997).

102. Wells, J.A. & McClendon, C.L. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* **450**, 1001-9 (2007).

103. Keseru, G.M. & Makara, G.M. The influence of lead discovery strategies on the properties of drug candidates. *Nat Rev Drug Discov* **8**, 203-12 (2009).

104. Chan, K.H., Zengerle, M., Testa, A. & Ciulli, A. Impact of Target Warhead and Linkage Vector on Inducing Protein Degradation: Comparison of Bromodomain and Extra-Terminal (BET) Degraders Derived from Triazolodiazepine (JQ1) and Tetrahydroquinoline (I-BET726) BET Inhibitor Scaffolds. *J Med Chem* **61**, 504-513 (2018).

105. Cyrus, K. *et al.* Impact of linker length on the activity of PROTACs. *Mol Biosyst* **7**, 359-64 (2011).

106. Nowak, R.P. *et al.* Plasticity in binding confers selectivity in ligand-induced protein degradation. *Nat Chem Biol* **14**, 706-714 (2018).

107. Zoppi, V. *et al.* Iterative design and optimization of initially inactive Proteolysis Targeting Chimeras (PROTACs) identify VZ185 as a potent, fast and selective

von Hippel-Lindau (VHL)-based dual degrader probe of BRD9 and BRD7. *J Med Chem* (2018).

108. Lai, A.C. *et al.* Modular PROTAC Design for the Degradation of Oncogenic BCR-ABL. *Angew Chem Int Ed Engl* **55**, 807-10 (2016).

109. Pettersson, M. & Crews, C. PROteolysis TArgeting Chimeras (PROTACs) — Past, present and future. *Drug discovery today* (2019).

110. Guo, J., Liu, J. & Wei, W. Degrading proteins in animals: "PROTAC"tion goes in vivo. *Cell Res* **29**, 179-180 (2019).

111. Mallery, D.L. *et al.* Antibodies mediate intracellular immunity through tripartite motif-containing 21 (TRIM21). *Proc Natl Acad Sci U S A* **107**, 19985-90 (2010).

112. Clift, D. *et al.* A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell* **171**, 1692-1706 (2017).

113. Clift, D., So, C., McEwan, W.A., James, L.C. & Schuh, M. Acute and rapid degradation of endogenous proteins by Trim-Away. *Nat Protoc* **13**, 2149-2175 (2018).

114. Kapitein, L.C. *et al.* The bipolar mitotic kinesin Eg5 moves on both microtubules that it crosslinks. *Nature* **435**, 114-8 (2005).

115. Mayer, T.U. *et al.* Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971-4 (1999).

116. Clift, D. *et al.* A Method for the Acute and Rapid Degradation of Endogenous Proteins. *Cell* **171**, 1692-1706.e18-1706.e18 (2017).

117. Banaszynski, L.A., Chen, L.C., Maynard-Smith, L.A., Ooi, A.G. & Wandless, T.J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995-1004 (2006).

118. Prakash, S., Inobe, T., Hatch, A.J. & Matouschek, A. Substrate selection by the proteasome during degradation of protein complexes. *Nat Chem Biol* **5**, 29-36 (2009).

119. Wilmington, S.R. & Matouschek, A. An Inducible System for Rapid Degradation of Specific Cellular Proteins Using Proteasome Adaptors. *PloS one* **11**, e0152679 (2016).

120. Hanna, J., Meides, A., Zhang, D.P. & Finley, D. A ubiquitin stress response induces altered proteasome composition. *Cell* **129**, 747-759 (2007).

121. Hanna, J., Leggett, D.S. & Finley, D. Ubiquitin Depletion as a Key Mediator of Toxicity by Translational Inhibitors. *Mol Cell Biol* **23**, 9251-9261 (2003).

122. Koide, A., Bailey, C.W., Huang, X. & Koide, S. The fibronectin type III domain as a scaffold for novel binding proteins. *J Mol Biol* **284**, 1141-1151 (1998).

123. Daley, G.Q., Van Etten, R.A. & Baltimore, D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* **247**, 824-30 (1990).

124. Brehme, M. *et al.* Charting the molecular network of the drug target Bcr-Abl. *Proc Natl Acad Sci U S A* **106**, 7414-9 (2009).

125. Scherr, M. *et al.* Enhanced sensitivity to inhibition of SHP2, STAT5, and Gab2 expression in chronic myeloid leukemia (CML). *Blood* **107**, 3279-87 (2006).

126. Sattler, M. *et al.* Critical role for Gab2 in transformation by BCR/ABL. *Cancer Cell* **1**, 479-92 (2002).

127. McMahon, C. *et al.* Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat Struct Mol Biol* **25**, 289-296 (2018).

128. Zuo, J. *et al.* Institute collection and analysis of Nanobodies (iCAN): a comprehensive database and analysis platform for nanobodies. *BMC Genomics* **18**, 797 (2017).

129. Liu, W. *et al.* Recent advances in the selection and identification of antigen-specific nanobodies. *Mol Immunol* **96**, 37-47 (2018).

130. Löfblom, J. *et al.* Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications. *FEBS Lett* **584**, 2670-2680 (2010).

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

132. Stumpp, M.T. & Amstutz, P. DARPins: a true alternative to antibodies. *Curr Opin Drug Discov Devel* **10**, 153-9 (2007).

133. Koide, A., Wojcik, J., Gilbreth, R.N., Hoey, R.J. & Koide, S. Teaching an old scaffold new tricks: monobodies constructed using alternative surfaces of the FN3 scaffold. *J Mol Biol* **415**, 393-405 (2012).

134. Koide, A., Gilbreth, R.N., Esaki, K., Tereshko, V. & Koide, S. High-affinity single-domain binding proteins with a binary-code interface. *Proc Natl Acad Sci U S A* **104**, 6632-6637 (2007).

135. Koide, S., Koide, A. & Lipovsek, D. Target-binding proteins based on the 10th human fibronectin type III domain ((1)(0)Fn3). *Methods Enzymol* **503**, 135-56 (2012).

136. Olson, C.A., Liao, H.I., Sun, R. & Roberts, R.W. mRNA display selection of a high-affinity, modification-specific phospho-IkappaBalpha-binding fibronectin. *ACS Chem Biol* **3**, 480-5 (2008).

137. Wojcik, J. *et al.* A potent and highly specific FN3 monobody inhibitor of the Abl SH2 domain. *Nat Struct Mol Biol* **17**, 519-527 (2010).

138. Sha, F. *et al.* Dissection of the BCR-ABL signaling network using highly specific monobody inhibitors to the SHP2 SH2 domains. *Proc Natl Acad Sci U S A* **110**, 14924-14929 (2013).

139. Grebien, F. *et al.* Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. *Cell* **147**, 306-319 (2011).

140. Garcia Fortanet, J. *et al.* Allosteric Inhibition of SHP2: Identification of a Potent, Selective, and Orally Efficacious Phosphatase Inhibitor. *Journal of medicinal chemistry* **59**, 7773-7782 (2016).

141. Chen, Y.-N.P. *et al.* Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148-152 (2016).

142. Wang, J.Y. The capable ABL: what is its biological function? *Mol Cell Biol* **34**, 1188-97 (2014).

143. Nagar, B. *et al.* Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* **112**, 859-71 (2003).

144. Ben-Neriah, Y., Bernards, A., Paskind, M., Daley, G.Q. & Baltimore, D. Alternative 5' exons in c-abl mRNA. *Cell* **44**, 577-86 (1986).

145. Pluk, H., Dorey, K. & Superti-Furga, G. Autoinhibition of c-Abl. *Cell* **108**, 247-59 (2002).

146. Hantschel, O. *et al.* Structural basis for the cytoskeletal association of Bcr-Abl/c-Abl. *Mol Cell* **19**, 461-73 (2005).

147. Taagepera, S. *et al.* Nuclear-cytoplasmic shuttling of C-ABL tyrosine kinase. *Proc Natl Acad Sci U S A* **95**, 7457-62 (1998).

148. Deininger, M.W.N., Goldman, J.M. & Melo, J.V. The molecular biology of chronic myeloid leukemia. *Blood* **96**, 3343-3356 (2000).

149. Woodring, P.J., Hunter, T. & Wang, J.Y. Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. *J Cell Sci* **116**, 2613-26 (2003).

150. Baskaran, R., Dahmus, M.E. & Wang, J.Y. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **90**, 11167-71 (1993).

151. Kaidi, A. & Jackson, S.P. KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling. *Nature* **498**, 70-4 (2013).

152. McWhirter, J.R., Galasso, D.L. & Wang, J.Y. A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. *Mol Cell Biol* **13**, 7587-95 (1993).

153. Reuther, G.W., Fu, H., Cripe, L.D., Collier, R.J. & Pendergast, A.M. Association of the protein kinases c-Bcr and Bcr-Abl with proteins of the 14-3-3 family. *Science* **266**, 129-33 (1994).

154. Kurzrock, R., Gutterman, J.U. & Talpaz, M. The molecular genetics of Philadelphia chromosome-positive leukemias. *N Engl J Med* **319**, 990-8 (1988).

155. Melo, J.V. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. *Blood* **88**, 2375-84 (1996).

156. Gordon, M.Y., Dowding, C.R., Riley, G.P., Goldman, J.M. & Greaves, M.F. Altered adhesive interactions with marrow stroma of haematopoietic progenitor cells in chronic myeloid leukaemia. *Nature* **328**, 342-4 (1987).

157. Cortez, D., Reuther, G. & Pendergast, A.M. The Bcr-Abl tyrosine kinase activates mitogenic signaling pathways and stimulates G1-to-S phase transition in hematopoietic cells. *Oncogene* **15**, 2333-42 (1997).

158. Puil, L. *et al.* Bcr-Abl oncoproteins bind directly to activators of the Ras signalling pathway. *EMBO J* **13**, 764-73 (1994).

159. Chai, S.K., Nichols, G.L. & Rothman, P. Constitutive activation of JAKs and STATs in BCR-Abl-expressing cell lines and peripheral blood cells derived from leukemic patients. *J Immunol* **159**, 4720-8 (1997).

160. Sillaber, C., Gesbert, F., Frank, D.A., Sattler, M. & Griffin, J.D. STAT5 activation contributes to growth and viability in Bcr/Abl-transformed cells. *Blood* **95**, 2118-25 (2000).

161. Skorski, T. *et al.* Phosphatidylinositol-3 kinase activity is regulated by BCR/ABL and is required for the growth of Philadelphia chromosome-positive cells. *Blood* **86**, 726-36 (1995).

162. Skorski, T. *et al.* Transformation of hematopoietic cells by BCR/ABL requires activation of a PI-3k/Akt-dependent pathway. *EMBO J* **16**, 6151-61 (1997).

163. Cortez, D., Kadlec, L. & Pendergast, A.M. Structural and signaling requirements for BCR-ABL-mediated transformation and inhibition of apoptosis. *Mol Cell Biol* **15**, 5531-41 (1995).

164. Bedi, A. *et al.* BCR-ABL-mediated inhibition of apoptosis with delay of G2/M transition after DNA damage: a mechanism of resistance to multiple anticancer agents. *Blood* **86**, 1148-58 (1995).

165. Hof, P., Pluskey, S., Dhe-Paganon, S., Eck, M.J. & Shoelson, S.E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **92**, 441-50 (1998).

166. Vogel, W., Lammers, R., Huang, J. & Ullrich, A. Activation of a phosphotyrosine phosphatase by tyrosine phosphorylation. *Science* **259**, 1611-4 (1993).

167. Laczmanska, I. & Sasiadek, M.M. Tyrosine phosphatases as a superfamily of tumor suppressors in colorectal cancer. *Acta Biochim Pol* **58**, 467-70 (2011).

168. Aceto, N. *et al.* Tyrosine phosphatase SHP2 promotes breast cancer progression and maintains tumor-initiating cells via activation of key transcription factors and a positive feedback signaling loop. *Nature medicine* **18**, 529-537 (2012).

169. Chan, R.J. & Feng, G.S. PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood* **109**, 862-7 (2007).

170. Zhang, J., Zhang, F. & Niu, R. Functions of Shp2 in cancer. *Journal of cellular and molecular medicine* **19**, 2075-2083 (2015).

171. Hu, Z. *et al.* Overexpression of SHP2 tyrosine phosphatase promotes the tumorigenesis of breast carcinoma. *Oncology reports* **32**, 205-212 (2014).

172. Bunda, S. *et al.* Inhibition of SHP2-mediated dephosphorylation of Ras suppresses oncogenesis. *Nat Commun* **6**, 8859 (2015).

173. Bentires-Alj, M. Activating Mutations of the Noonan Syndrome-Associated SHP2/PTPN11 Gene in Human Solid Tumors and Adult Acute Myelogenous Leukemia. *Cancer research* **64**, 8816-8820 (2004).

174. Chen, Y.-N.P. *et al.* Allosteric inhibition of SHP2 phosphatase inhibits cancers driven by receptor tyrosine kinases. *Nature* **535**, 148--152 (2016).

175. Dance, M., Montagner, A., Salles, J.P., Yart, A. & Raynal, P. The molecular functions of Shp2 in the Ras/Mitogen-activated protein kinase (ERK1/2) pathway. *Cell Signal* **20**, 453-9 (2008).

176. Ahmed, T.A. *et al.* SHP2 Drives Adaptive Resistance to ERK Signaling Inhibition in Molecularly Defined Subsets of ERK-Dependent Tumors. *Cell Rep* **26**, 65-78 e5 (2019).

177. Inobe, T., Fishbain, S., Prakash, S. & Matouschek, A. Defining the geometry of the two-component proteasome degron. *Nature chemical biology* **7**, 161-167 (2011).

178. Caussinus, E., Kanca, O. & Affolter, M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. *Nat Struct Mol Biol* **19**, 117-21 (2011).

179. Wohlever, M.L., Nager, A.R., Baker, T.A. & Sauer, R.T. Engineering fluorescent protein substrates for the AAA+ Lon protease. *Protein Eng Des Sel* **26**, 299-305 (2013).

180. Gautam, A.K.S., Martinez-Fonts, K. & Matouschek, A. Scalable In Vitro Proteasome Activity Assay. *Methods Mol Biol* **1844**, 321-341 (2018).

181. Mitchell, S.F. & Lorsch, J.R. Protein Affinity Purification using Intein/Chitin Binding Protein Tags. *Methods Enzymol* **559**, 111-25 (2015).

182. Roy, R.D., Rosenmund, C. & Stefan, M.I. Cooperative binding mitigates the high-dose hook effect. *BMC Syst Biol* **11**, 74 (2017).

183. Hoofnagle, A.N. & Wener, M.H. The fundamental flaws of immunoassays and potential solutions using tandem mass spectrometry. *J Immunol Methods* **347**, 3-11 (2009).

184. Elsasser, S., Chandler-Militello, D., Muller, B., Hanna, J. & Finley, D. Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome. *J Biol Chem* **279**, 26817-22 (2004).

185. Moritz, B., Becker, P.B. & Gopfert, U. CMV promoter mutants with a reduced propensity to productivity loss in CHO cells. *Sci Rep* **5**, 16952 (2015).

186. Douglass, E.F., Jr., Miller, C.J., Sparer, G., Shapiro, H. & Spiegel, D.A. A comprehensive mathematical model for three-body binding equilibria. *J Am Chem Soc* **135**, 6092-9 (2013).

187. Uhlen, M. *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* **4**, 1920-32 (2005).

188. Ahmad, S., Banville, D., Zhao, Z., Fischer, E.H. & Shen, S.H. A widely expressed human protein-tyrosine phosphatase containing src homology 2 domains. *Proc Natl Acad Sci U S A* **90**, 2197-201 (1993).

189. Katayama, H., Yamamoto, A., Mizushima, N., Yoshimori, T. & Miyawaki, A. GFP-like proteins stably accumulate in lysosomes. *Cell Struct Funct* **33**, 1-12 (2008).

190. Bence, N.F., Sampat, R.M. & Kopito, R.R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552-5 (2001).

191. Kalchman, M.A. *et al.* Huntingtin is ubiquitinated and interacts with a specific ubiquitin-conjugating enzyme. *J Biol Chem* **271**, 19385-94 (1996).

192. Morishima-Kawashima, M. *et al.* Ubiquitin is conjugated with amino-terminally processed tau in paired helical filaments. *Neuron* **10**, 1151-60 (1993).

193. Goldberg, A.L. Development of proteasome inhibitors as research tools and cancer drugs. *J Cell Biol* **199**, 583-8 (2012).

194. Kouroukis, T.C. *et al.* Bortezomib in multiple myeloma: systematic review and clinical considerations. *Curr Oncol* **21**, 573-603 (2014).

195. Kubiczkova, L., Pour, L., Sedlarikova, L., Hajek, R. & Sevcikova, S. Proteasome inhibitors - molecular basis and current perspectives in multiple myeloma. *J Cell Mol Med* **18**, 947-61 (2014).

196. Kummer, L. *et al.* Structural and functional analysis of phosphorylation-specific binders of the kinase ERK from designed ankyrin repeat protein libraries. *Proc Natl Acad Sci U S A* **109**, 2248-57 (2012).

197. Fellouse, F.A. *et al.* High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. *J Mol Biol* **373**, 924-940 (2007).

198. Hannah, J. & Zhou, P. Maximizing target protein ablation by integration of RNAi and protein knockout. *Cell Res* **21**, 1152-4 (2011).

199. Skrlec, K., Strukelj, B. & Berlec, A. Non-immunoglobulin scaffolds: a focus on their targets. *Trends Biotechnol* **33**, 408-18 (2015).

200. Wuo, M.G. & Arora, P.S. Engineered protein scaffolds as leads for synthetic inhibitors of protein-protein interactions. *Curr Opin Chem Biol* **44**, 16-22 (2018).

201. Lee, C., Schwartz, M.P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol Cell* **7**, 627-37 (2001).

202. Gingrich, J.C., Davis, D.R. & Nguyen, Q. Multiplex detection and quantitation of proteins on western blots using fluorescent probes. *Biotechniques* **29**, 636-42 (2000).

203. Dittmer, A. & Dittmer, J. Beta-actin is not a reliable loading control in Western blot analysis. *Electrophoresis* **27**, 2844-5 (2006).

204. Jung, T., Schauer, U., Heusser, C., Neumann, C. & Rieger, C. Detection of intracellular cytokines by flow cytometry. *J Immunol Methods* **159**, 197-207 (1993).

Vita

Kimberly Bowen was born and raised in Boston, Massachusetts. She attended Northwestern University and received her Bachelor of Arts in Chemistry in 2010. She worked for three years after college as a protein biochemist in the Center for Proteomic Chemistry at Novartis in Cambridge, Massachusetts. While at Novartis, Kimberly received the Catalyst team award and Spot individual award for outstanding contributions to the company for her work on one of her projects.

She then attended University of Texas at Austin for graduate school from 2013-2019. She worked in Dr. Andreas Matouschek's lab studying the proteasome. During her PhD, she attended several conferences and won the Best Poster Award at the EMBO Protein Quality Control conference in 2016.

Currently, Kimberly lives in Austin, Texas and plans to work in industry.

Email address: kimberly.bowen@utexas.edu

This dissertation was typed by Kimberly Elizabeth Bowen.