# DEVELOPING A UBIQUITIN-LIKE MODIFICATION REPORTER

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sean Patrick O'Brien

May 2013

# DEVELOPING A UBIQUITIN-LIKE MODIFICATION REPORTER

Sean Patrick O'Brien, Ph. D.

Cornell University 2013

Ubiquitin-like (ubl) modification is an example of post-translational modification (PTM) that influences a number of cellular processes. Given difficulties in studying this system in its native eukaryotic context, several pathways have been reconstituted in *Escherichia coli* (*E. coli*) at varying levels of completeness. We developed the first E3-dependent SUMOylation pathway in *E. coli*. Because the E3 ligase increases efficiency of conjugation, we were able to lower expression of upstream elements – namely the E2 – and avoid non-physiological chain formation on target protein encountered in previously published work while maintaining high product yield. We additionally developed a ubquitination pathway in *E. coli* important in plant defense against bacterial colonization. In characterizing the system, we were the first to note that ubiquitination of the target protein may proceed in an E3-independent manner likely through auto-monoubiquitination involving a ubiquitin-binding domain (UBD).

We believed these systems might serve as a scaffold to develop a reporter in *E. coli* for ubl modification of a target protein. Such a reporter would enable engineering new functionality into the pathways. For example, engineering the ubiquitin E3 ligase could achieve rapid knockdown of novel protein; small chemical inhibitors of modification could be identified; and substrates of a particular E2-E3 pair or the E2 in an E3-substrate pair could be found in a high-throughput fashion. Such advances would have merit in further study of ubl modification and in therapeutic development. We adapted both *in vivo* and *in vitro* screens/selections developed for

evolving non-covalent protein-protein interactions that included split YFP, split DHFR, bacterial two-hybrid system, plasmid display with LacI and zinc finger proteins, and ribosome display. To address the shortcomings of these systems, we developed our own split Cre bacterial screen to adapt to our ubl systems. This screen made use of a two-state reporter, which provided finer resolution of the protein-protein interactions and a genetic record of those interactions. Additionally, Cre's enzymatic activity allowed for detection of infrequent dimerizing pairs. While alone it showed promise as a screen, it too failed when coupled with the ubl pathways.

## BIOGRAPHICAL SKETCH

Sean Patrick O'Brien matriculated at Williams College in September of 2001, where he graduated *magna cum laude* with a Bachelors of Arts in Physics with honors in June of 2005. In August 2006 he enrolled at Cornell University and received his doctorate of philosophy in Chemical and Biomolecular Engineering in May of 2013 under the guidance of Dr. Matthew P. DeLisa. Here, he was a Presidential Life Science Fellow.

To my family and friends to whom

I am indebted for their support and encouragement

over the past several years

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Career Award CBET-0449080 (to M.P.D.); the New York State Office of Science, Technology and Academic Research Distinguished Faculty Award (to M.P.D.); and a Cornell Genomics Fellowship (to S.P.O.). Technical Support was provided by Malinka Walaliyadde.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

# CHAPTER 1

# UBIQUITIN-LIKE MODIFIERS

## *Introduction*

In the 20th century the genetic code was cracked and the mechanism of protein synthesis detailed. However, proteins are more than the DNA that encode them; many undergo at least one post-translational modification (PTM). Though the exact number is left to speculation, the number of enzymes responsible for PTM give an impression of its magnitude. About 5% of the genome belonging to higher order eukaryotes is devoted to these enzymes (*1*). Adding to the complexity, 59% of those modified proteins have more than one PTM site and 13% have more than one modification type (*2*). Over 200 types of PTM exist (*1*), which include backbone cleavage and rearrangement, residue conversion, disulfide bond formation, and attachment of functional groups such as simple molecules, cofactors, fatty acids, sugars, and other proteins. These modifications may be reversible or irreversible.

PTMs influence conformation, activity, localization, interaction, and/or stability of a protein and serve in a broad range of physiological and pathological processes. Controlled proteolysis can control localization and activation of proteins like peptidases found in the small intestines or cellular receptors/transcriptional activators like Notch. Attachment of the protein ubiquitin can signal for the targeted degradation of proteins by the 26S proteasome. Regulated degradation is important in cellular homeostatis and modulating protein accumulation during different cellular epochs. Attachment of sugar groups to a protein can alter its chemical properties, notably solubility. Glycosylated proteins such as antibodies have a prolonged residence time in blood serum. Attachment of the charged phosphate group can grossly alter

protein confromation. This can toggle a protein's catalytic activity and interacting partners. Phosphorylation is often implemented in signal transduction cascades.

PTMs raise the diversity of the proteome by several orders of magnitude above that predicted by the genome alone (*1*). They provide a critical context for biological understanding at multiple levels from protein characteristics to cellular dynamics to physiological function. Their study is indispensable.

***Ubiquitin-like pathways***

One well-studied PTM is the attachment of the small ubiquitin-like modifier (ubl). Including the eponymous ubiquitin, 11 ubls have been identified in eukaryotes with several more postulated (*3*). While members can share little sequence homology (ubiquitin and SUMO (small ubiquitin-related modifier) share only 18% sequence identity) (*4*), they all share a distinctive ββαββαβ fold aptly called the ubiquitin superfold (*5*) and a flexible tail at their C-terminus terminating in two glycines. The carboxyl group of the last glycine forms an isopeptide bond with an amino group, commonly of a lysine residue, of the target protein.

Cellular processes in which this modification is involved include cellular trafficking, subcellular localization, channel and receptor regulation, regulation of transcription-factor activity, DNA repair and replication, chromosome dynamics, mRNA processing and metabolism, cellular replication, endocytosis, lysosomal targeting, enzymatic activation and deactivation, and protein degradation (*6*). While each system is distinct, cooperation arises between them. One modifier may block the site of another (*7*), target its modified protein for modification by a different system (*8*), or even activate another cascade (*9–11*).

Despite notable differences, the mechanisms for ubl attachment are analogous. That for ubiquitin will be describe in detail. The ubiquitin pathway involves three proteins that execute

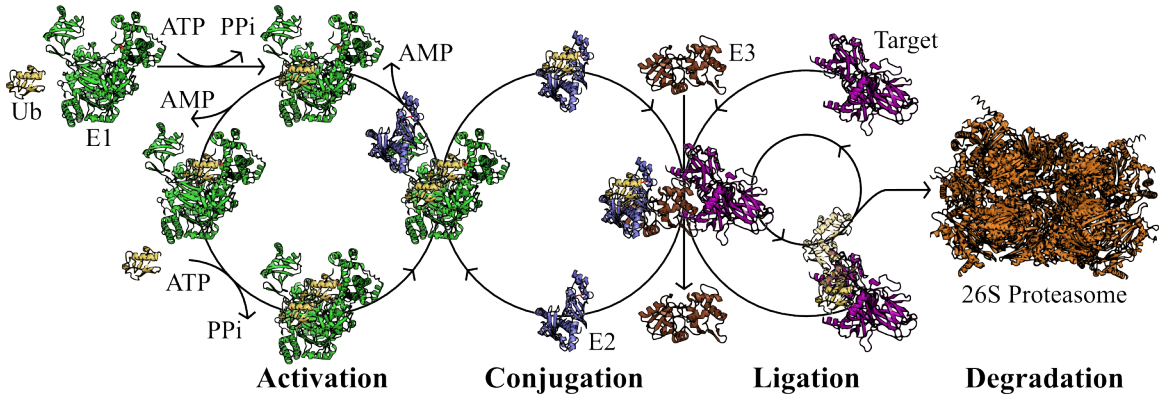


**Figure 1.1. The ubiquitin pathway.** Three steps comprise the ubiquitin pathway: activation, conjugation, and ligation. The E1 activates ubiquitin through formation of a thioester bond. To do so, it first adenylates ubiquitin's C-terminal G76. A cysteine on another domain then forms a thioester bond and removes it from the adenylation site. Removal allows another ubiquitin to be adenylated. The E2 binds the E1 and acquires ubiquitin through a transthioesterification reaction. The charged E2 then binds an E3 which recognizes the target protein. The ubiquitin is transferred to the ε-amino group of one of the substrate's lysine residues. Subsequent rounds can form chains of ubiquitin on the target protein depending on the E2 utilized. Proteins with K48-linked ubiquitin chains are marked for degradation by the 26S proteasome.

three distinct steps: the E1 activates ubiquitin, the E2 conjugates ubiquitin, and the E3 directs ubiquitin-protein ligation (Fig. 1.1). To date, the human genome is known to encode 2 E1s, 38 E2s, and 600-1000 E3s (*12*). The numbers reflect the increasing specificity through the progression of the pathway.

Seminal work done three decades ago elucidated the biochemical details of ubiquitin activation performed by E1 (*13–15*). A recent crystal structure of Uba1, a yeast E1, illustrates the modular nature of the enzyme and coordinated conformational changes associated with activation (*16*). Activation begins with the adenylation of ubiquitin's C-terminal carboxyl group. Next, the E1's catalytic cysteine on another domain forms a thioester bond with ubiquitin through nucleophilic attack of the adenylate. This liberates the noncovalent ubiquitin binding area, and E1 adenylates a second ubiquitin. The simultaneous association both primes E1 with another ubiquitin and expedites the first's transthioesterification to the E2 since MgATP also acts as a positive allosteric effector (*17*).

E2s share a conserved 150 residue core domain, the ubiquitin conjugating (UBC) domain, which includes the catalytic cysteine to which ubiquitin attaches (*18*). This domain binds to a ubiquitin-like structure on E1's UFD (ubiquitin fold domain) and may also interact with an unstructured loop in E1's SCCH (second catalytic cysteine half-domain) (*16*). From this perch the E2 accepts the covalently bound ubiquitin from E1. Though charged with ubiquitin, E2s themselves are incapable of ubiquitinating a target protein. A consensus site for E2-target binding does not seem to exist (*19*), and E2s require interaction with an E3 for ubiquitination of the target to occur (*20*). An exception occurs with proteins possessing a ubiquitin binding domain (UBD), which are capable of auto-monoubiquitination (*21*).

Two distinct families of E3s are known: those with a HECT (homologous to the E6-AP carboxyl terminus) domain and those with a RING (really interesting new gene) domain. The HECT domain is known to thioesterify ubiquitin before ligation but will not be discussed in detail here. Within the RING domain are cysteine and histidine residues that chelate two zinc ions. These coordination sites appear in two loops that are connected by a central helix. The whole structure forms a shallow cleft to which an E2 binds. Exactly which residues in the RING domain and exactly which region of the UBC domain interact vary by E2-E3 pair (*22*). Additionally, regions outside the RING domain may bind the E2 as well (*23*).

RING E3s are viewed as adaptors that bring E2 and target substrate together. The exact mechanism for ubiquitin transfer from E2 to substrate remains in debate. Some argue for conformational changes of the E2 when bound to E3; others that essential residues reside in the E3. Notably absent within the UBC are side chains that could catalyze ubiquitin transfer. The E2's active site lacks a well-positioned base to deprotonate the attacking amino group. While a conserved asparagine could stabilize the developing negative charge on the ubiquitin's C-

terminal carbonyl group during nucleophilic attack, it is poorly positioned to do so (*24*). The asparagine instead interacts with the peptide backbone through hydrogen bonds as evinced in crystal structures of both free E2s and those bound to E3s (*22*). To the contrary, no E3 side chains come closer than approximately 15 Å to the E2 active site (*25*). However, the E2 is not esterified with ubiquitin in the sighted structures above (*22*). Furthermore, the conserved asparagine proves necessary for substrate ubiquitination (*25*). The crystal structure of the SUMO-RanGAP1-Ubc9-Nup358 complex shows the asparagine to rotate away from the backbone and towards the aforementioned carbonyl group to stabilize the oxyanion transition state during the amino group's attack (*26*). Therefore, the asparagine's repositioning which seems necessary for transfer may require at least both ubiquitin and E3 interaction with the E2.

Ubiquitin is usually attached to the ε-amino group of the substrate's lysine residue. A substrate may have one ubiquitin (monoubiquitination), one ubiquitin at multiple lysine residues (multi-monoubiquitination), or a chain of ubiquitin linked through one of their own seven lysine residues (polyubiquitination). Each has consequences for the substrate according to a "ubiquitin code" that is still being deciphered. Multiple mechanisms have been proposed for ubiquitin chain formation, and the E2 seems to dictate the type of chain formed. One E2 may interact with an E3 to initiate ubiquitination of a substrate. Others may subsequently interact with the same E3 to elongate the ubiquitin chain. Such is the case with the heterodimeric RING E3 BRCA1-BARD1 and its several E2 partners (*27*). Other E2s like Ube2g2 preassemble ubiquitin chains that are then transferred en bloc (*28*). Other E2s like Ubch5 both initiate and elongate the ubiquitin chain. They are capable of forming chains of different linkages, but a UBD on the RING E3 may provide specificity by orienting the acceptor ubiquitin (*22*).

Ubiquitin chains have different structural conformations depending upon which of ubiquitin's lysine residues acts as the attachment site. For example, those linked through K63 adopt an elongated structure while those linked through K48 adopt a compact structure (*29*). Interacting proteins possess a menagerie of ubiquitin binding domains that can distinguish between the different conformations (*18*). Proteins with a K48-linked polyubiquitin chain of at least four monomers (*30*) are degraded by the 26S proteasome (*31*). To which lysine the polyubiquitin chain is attached in the substrate has little to no bearing on its degradation signal (*24*), though it can influence the rate at which the substrate is degraded (*32*). Several proteins interact with the K48-linked chain to bring the substrate to the proteasome. However, details are still lacking (*33*), and new components are still being discovered (*34*, *35*).

***Ubl pathways introduced into Escherichia coli***

Overlap between ubl systems, the reversible nature of the modification, and low cellular abundances confound study of the pathways in their native context. While *in vitro* studies are possible, they are intractable outside specific cascades given the requirement to purify all components. Though a ubl system has been found in Actinobacteria (*36*) and various effectors of pathogenic bacteria hijack eukaryotic ubl pathways (*37*, *38*), no ubl pathway has been identified in non-pathogenic *Escherichia coli* (*E. coli*) lab strains. This fact coupled with the plethora of molecular tools existing for *E. coli* makes it an ideal host for studying the ubl pathways. Three notable systems that have been reconstituted in *E. coli* are SUMOylation, ubiquitination, and neddylation.

Mencía *et al* and Uchimura *et al* concurrently developed an E3-independent mammalian SUMOylation system in *E. coli* (*39*, *40*). Their systems allow for both the purification of SUMOylated protein and an opportunity to study the system functionally. Their work is

significant because the dynamic and entangled nature of the system in eukaryotes has made these two aspects difficult. However, both see chains of SUMO attached to their targets, which for SUMO-1 is not physiological. Mencía *et al* address this issue, and determined protein over-expression to be the cause. Both used high-expression plasmids. Recently, Okada *et al* have made an *Arabidopsis* analogue for investigating potential SUMO substrates against *Arabidopsis*'s four SUMO isoforms (*41*). Dissatisfied with current predictive tools for determining SUMOylation sites, they used their system to probe for both consensus and non-consensus sites. While successful, they were not able to confirm or dismiss the physiological significance of their results. Following Uchimura *et al*'s example, they too used high-expression plasmids, which may have led to non-physiological artifacts. They admitted that additional components like E3s might be needed in their system for more accurate results. To address these shortcomings, we developed the first E3-dependent SUMOylation pathway in *E. coli* (*42*). Because the E3 ligase increases efficiency of conjugation, we were able to lower expression of upstream elements–namely the E2–and avoid non-physiological chain formation on target protein while maintaining high product yield.

Su *et al* reconstituted the first E3-dependent ubiquitin system in *E. coli* to successfully screen a murine cDNA library for actual substrates of the E3 GRAIL (*43*). This screen circumvented issues they confronted in using mass spectrometric analysis and screening strategies in mammalian cells. Several years later Rosenbaum *et al* reconstituted another cascade to characterize the role yeast ubiquitin ligase San1 played in protein quality control (PQC) degradation (*44*). Their "in coli" method validated ubiquitination of putative substrates identified using the yeast two-hybrid screen but that they could otherwise not purify due to aggregation. Most recently, Keren-Kaplan *et al* meticulously characterized ubiquitination

reconstituted in *E. coli* using several different cascades (*45*). They provide evidence that their system recapitulated physiological results that *in vitro* reactions could not and produced enough ubiquitinated substrate for various analyses such as crystallization. We developed our own ubquitination pathway in *E. coli* important in plant defense against bacterial colonization (*46*). In characterizing the system, we were the first to note that ubiquitination of the target protein may proceed in an E3-independent manner likely through auto-monoubiquitination involving a UBD.

In addition to identifying cascade elements and modified protein for further study, the reconstitution of systems in *E. coli* provides a platform for engineering these cascades, which may be useful in their study and in developing novel molecular tools and therapeutics. Guntas *et al* have made several investigations into interactions between the various components of the cascade. They have evolved the ubiqutin E3 ligase E6AP to dock NEDD8-conjugating enzyme Ubc12 using a combination of computer models and bacterial screen (*47*). The two normally do not interact. They have also recapitulated an incomplete Nedd8 pathway in *E. coli* and developed a novel screen based on Nedd8-E2 interactions (*48*). In these investigations, they evolved more soluble Nedd8 and orthogonal mutant NAE1-Nedd8 pairs (*48*).

***Reporter for ubl modification***

We believed the pathways we reconstituted could be extended towards developing a reporter system for ubl modification. Such a system would serve as a scaffold to engineer new functionality into the pathways. For example, engineering the ubiquitin E3 ligase could achieve rapid knockdown of novel protein. Redirecting E3 ligases has been documented using Protacs (Proteolysis-targeting chimeric molecules) (*49–51*), binding partners of the target protein (*52*, *53*), and evolved peptide aptamers (*54*). However, successes have been case specific and

generally require knowledge of a specific interaction. These approaches also do not address two issues: the E3 ligase must still correctly position a charged E2 in proximity to a free amino group on the substrate for modification and E3 affinity for substrate is generally low with few even in the nanomolar range (*12*). As another example, small chemical inhibitors of modification could be identified. Additionally, identifying where inhibition occurs in the pathway to elucidate the mechanism of action could be quickly achieved. As a final example, substrates of a particular E2-E3 pair or the E2 in an E3-substrate pair could be found in a high-throughput fashion. Identifying these members is a current challenge in the study of ubl modification (*12*, *22*, *41*). These applications would contribute both to providing tools for basic research such as in determining protein function and to developing therapeutics beneficial in cancers (*53*) and other diseases (*55*).

High-throughput detection is currently limited as Western blotting is the only inexpensive, reliable means to confirm ubl-protein conjugation. Because ubiquitination itself gives no observable phenotype, the process must be coupled to another that can. To develop the reporter, we adapted both *in vivo* and *in vitro* screens/selections developed for evolving non-covalent protein-protein interactions that included split YFP, split DHFR, bacterial two-hybrid system, plasmid display with LacI and zinc finger proteins, and ribosome display. To address the shortcomings of these systems, we developed our own split Cre bacterial screen to adapt to our ubl systems. This screen utilized the enzymatic activity of Cre recombinase to alter the phenotype of cells via a two-gene reporter plasmid. This two-state reporter provided finer resolution of the protein-protein interactions and a genetic record of those interactions. This method faithfully reported differences in dimerization of protein pairs and allowed for the

isolation of competent binders based on phenotype. However, it was not able to function as a reporter for ubl modification.

CHAPTER 2

# FUNCTIONAL RECONSTITUTION OF A TUNABLE E-3 DEPENDENT SUMOYLATION PATHWAY IN ESCHERICHIA COLI[1]

## *Introduction*

Sumoylation is a eukaryotic post-translational modification that involves the covalent conjugation of the 11-kDa SUMO (small ubiquitin-related modifier) protein to a lysine residue in a target protein (for recent reviews of the sumoylation mechanism and its implications see (*56–61*)). Cellular processes in which sumoylation is involved include cellular trafficking, channel and receptor regulation, regulation of transcription-factor activity, DNA repair and replication, chromosome dynamics, mRNA processing and metabolism, cellular replication, and cross-talk with ubiquitination. The mechanism of SUMO attachment resembles other ubiquitin-like conjugation pathways. Briefly, mature SUMO is first activated by a heterodimeric SUMO-activating enzyme, E1, before passing to the SUMO-conjugating enzyme, E2. Only one E2 appears to exist in most well studied organisms including human, yeast, rat, and mouse. Unlike with ubiquitination, sumoylation may proceed in an E3-independent manner. This notion is based on the observation that binding of the E2 Ubc9 to the consensus sequence Ψ-K-*X*-E (where Ψ is a hydrophobic residue and *X* is an arbitrary residue) present in a target protein is sufficient for sumoylation (*62–64*). Furthermore, grafting of this consensus sequence to a protein not normally sumoylated will result in its sumoylation (*39*, *63*, *65*).

Given the apparent E3-independent nature of sumoylation, the existence of SUMO E3

[1] Adapted with permission from: O'Brien, S. P.; DeLisa, M. P. Functional Reconstitution of a Tunable E3-Dependent Sumoylation Pathway in Escherichia coli. *PLoS ONE* **2012**, *7*, e38671.

ligases was initially challenged (*66*), although evidence hinted at their existence (*61*). The involvement of E3 ligases in sumoylation has now been demonstrated (*67–69*). However, while an E3 can enhance target sumoylation (*65*, *67*, *69*, *70*), its role in substrate specificity and lysine selection remains debated. The crystal structure of SUMO-RanGAP1-Ubc9-Nup358 complex suggests the E3 merely aligns the E2-SUMO pair for optimal E2 binding and SUMO transfer without itself binding the target protein (*26*). Interactions between the target protein and E3 appear to augment efficiency, but sumoylation depends solely upon E2 binding (*26*). Furthermore, individual genetic knockout of the mammalian SUMO E3 ligases PIAS1 (*71*), PIASy (*72*), and PIASx (*73*) in mice does not affect global sumoylation patterns. Similarly in yeast, knockout of the E3 Siz2 does not affect global sumoylation, although the knockout of the E3 Siz1 attenuates robustness (*67*). Further studies in yeast examining sumoylation of individual proteins confirm this trend in overlapping E3 function (*65*). Differences in local concentrations rather than differences in target recognition may be the mechanism whereby E3 specificity is manifested *in vivo* but is absent *in vitro* (*65*).

Importantly, SUMO E3 ligases are not dispensable in the cellular context as the knockout of every E3 is lethal (*65*). Furthermore, emerging evidence suggests that the E3 may play a role in target specificity. Several proteins are modified at nonconsensus sequences (*59*) and an E3 ligase, not an E2, may be responsible for this modification. For example, Siz1 is required for sumoylation of PCNA's nonconensus K164 site (*74*). Several studies have confirmed that the PINIT domain of the E3 is solely responsible for this K164 lysine specificity (*65*, *75*). Further, E3s tend to bias the particular SUMO isoform that is attached to the target protein (*76*).

Several groups have reconstituted E3-independent sumoylation cascades in *Escherichia coli* (*E. coli*) (*39–41*). These sumo-engineered *E. coli* systems have several advantages. First,

endogenous levels of sumoylated protein in eukaryotic cells tend to be low (*58*). Thus, purifying quantifiable amounts from these cells is difficult, whereas obtaining ample yields for study from *E. coli* is typically straightforward. Second, because *E. coli* lacks an endogenous sumoylation system, the pathway may be isolated up to the point of the E2 for study. However, these systems are not without shortcomings. E3-independent sumoylation itself occurs at quantifiable levels only for protein concentrations far exceeding physiological levels. While proteins are clearly sumoylated, the physiological relevance of the modified proteins is unclear. For example, Mencía and de Lorenzo observed attachment of poly-SUMO-1 chains to target proteins in *E. coli* (*39*). Because SUMO-1 lacks the consensus sequence present on SUMO-2 and SUMO-3 (*77*), it is not believed to homo-polymerize. However, more recent *in vitro* studies have shown that SUMO-1 is capable of forming chains through non-consensus lysine's (*78*), albeit to a far lesser extent compared to SUMO-2 and SUMO-3 (*79*). The physiological relevance of such poly-SUMO-1 chains is unclear (*80*), and SUMO-1 itself may be more involved in chain termination of SUMO-2 and SUMO-3 rather than formation *in vivo* (*81*). Along similar lines, the physiological significance of some sumoylation sites observed by Okada *et al.* using sumo-engineered *E. coli* is also unclear (*41*).

Here, we engineered an E3-dependent SUMO-conjugation system in *E. coli* that employs members of the mammalian PIAS E3 ligase family and, as a result, involves no observable poly-sumoylation of target proteins. Furthermore, because *E. coli* lacks organelles and an endogenous sumoylation pathway, our system provides an alternative *in vivo* context that is insulated from factors such as target localization, downstream interactions, and the diversity of sumoylated proteins that confound studies of E3s in eukaryotic cells. Finally, we show that addition of the E3 increases the efficiency of sumoylation, yielding as much as ~5 mg/L of SUMO-modified

proteins. This makes possible greater titers of specifically sumoylated target proteins for use in biochemical and structural characterization.

***Materials and Methods***

**Plasmid construction**. All plasmids were based on the pZ vector system developed by Lutz and Bujard (*82*). Primer insertions were used to replace the multiple cloning site (MCS) between the restriction sites *Eco*RI and *Xba*I in the plasmids pZE12, pZE11, pZA24, and pZS31. The resulting vectors - pZE12-SMCS, pZE11-SMCS, pZA24-SMCS, and pZS31-SMCS - consisted of three pairs of restriction sites (*Kpn*I and *Sph*I, *Mlu*I and *Eag*I, and *Kas*I and *Cla*I) with each pair flanked by a strong RBS sequence (AAAGAGGAGAAA) and a frame-shifted stop codon sequence (TAATTGAATAGTTAA) to prevent translational read-through. For any vector where these sites were not unique, we first cloned the genes into the modified pZE12 vector prior to moving the fragment generated by digestion with *Kpn*I and *Cla*I into the appropriate final vector. To make pZS31-Ubc9, pZS31-Aos1.Uba2, and pZS31-Aos1.Uba2.Ubc9, the genes encoding human Aos1, human Uba2, and murine Ubc9 were PCR amplified from pBADE12 (*39*). The resulting PCR products were then inserted into pZS31-SMCS. For pZA31-Aos1.Uba2.Ubc9, pZS31-Aos1.Uba2.Ubc9 was cut at *Xho*I and *Cla*I and moved into pZA24-SMCS. The plasmid's selection marker was changed to chloramphenicol using the restrictions sites *Spe*I and *Xho*I. An epitope tag for Western blot detection were introduced to Aos1 by adding the DNA encoding a FLAG epitope tag (GACTACAAGGACGATGACGACAAGGGA) to the 3' primer during PCR amplification. A 3x-FLAG epitope tag was added to Uba2 and Ubc9 using *Bsa*I and primer annealing of 5'- ctcagactacaaagaccatgacggtgattataaagatcatgacatcgactacaaggatgacgatgacaagtaaat-3' and 5'-cgatttacttgtcatcgtcatccttgtagtcgatgtcatgatctttataatcaccgtcatggtctttgtagtc-3'. To generate the plasmids pZE11-GST-PML.SUMO, pZE11-Smad4-FLAG, pZE11-SUMO and pZE11-

Smad4-FLAG.SUMO, GST-PML and human Smad4 were PCR amplified from pGST-PML (*39*) and pOTB7-Smad4 (*83*), respectively, and inserted between *Kpn*I and *Sph*I of pZE11-SMCS. DNA encoding a FLAG epitope tag was added C-terminally to Smad4 during PCR amplification. Human SUMO-1 was PCR amplified from pKRSUMO (*39*) and inserted between *Mlu*I and *Eag*I. The restriction site *Bsa*I was used to create the Smad4(K159R) mutant. To generate pZE11-GFP.SUMO, pZE11-GFP-PML.SUMO, and pZE11-GFP-PML(K490R).SUMO, GFP was PCR amplified and inserted between *Kpn*I and *Sph*I of pZE11-SUMO. For the latter two cases, DNA encoding PML or PML(K490R) was added C-terminally to GFP during PCR amplification. To construct pZA24-PIASxβ and pZA24-PIASy, PIASxβ and PIASy were PCR amplified from pCMV-FLAG-hPIASxβ (*84*) and pCMV-FLAG-hPIASy (*85*), respectively, and inserted between *Kpn*I and *Sph*I of pZA24-SMCS. To facilitate Western blot analysis, a FLAG epitope tag was added C-terminally to all of the E3s during PCR amplification.

To assemble the SUMO E3 ligase chimeras, fragments of PIASxβ and PIASy were PCR-amplified and restriction sites introduced during PCR amplification. The restriction sites were placed in predicted unstructured regions (*86*) that flanked a domain of interest and made use of silent mutations when possible to preserve the amino acid sequence. The restriction sites *Not*I, *Spe*I, *Bam*HI, and *Nhe*I were inserted after K79, L299, V436, and Q508 in PIASxβ, and after P77, L279, G440, and A509 in PIASy. Fragments containing these restriction sites were PCR amplified and then ligated together in plasmid pZE12-SMCS before being moved to pZA24-SMCS. PIASxβ was truncated after Q508 to create the PIASxβ truncation variant.

**Cell growth and Western blot analysis.** All constructs were transformed into *E. coli* host strain DH5α-Z1 (*82*) using a GenePulser Xcell (BioRad). Individual colonies were grown overnight in LB media with appropriate antibiotics (100 μg/mL ampicillin, 40 μg/mL kanamycin, and 12.5

μg/mL chloramphenicol) and then subcultured to $OD_{600} \approx 0.05$ in 5 mL of fresh LB media supplemented with appropriate antibiotics. Cultures were induced at $OD_{600} \approx 0.75$ with 0.5% L(+)-arabinose, 1 mM IPTG, and 50 ng/mL anhydrotetracycline when appropriate and subsequently shaken for 24 h at 16°C or 25°C depending on determined optimal conditions for sumoylation. Approximately 1.5 mL of each culture was harvested and lysed using 200 μL of Bugbuster Master Mix (Novagen) according to the manufacturer's directions. Lysates were normalized to 10 μg of total protein as determined by a total protein assay (Bio-Rad) and loaded on a 4-20% Precise Protein Gel (Thermo Scientific). Transfers to Immobilon P Transer Membranes (Millipore) were performed for 2 h at the maximum amperage recommended for a Biosciences TE77 semi-dry transfer unit (Amersham). Blots were then imaged on film using standard protocols. The primary antibodies used were anti-GST (Abcam), anti-FLAG (Abcam), anti-GFP (Roche), anti-SUMO-1 (Abcam), and anti-DnaK (Stressgen). A standard curve was generated with purified GFP (AbCam) and used to quantify the yield of sumoylated GFP-PML. Densitometry analysis was performed on a Macintosh computer using the public domain NIH Image program (developed at the U.S. National Institutes of Health and available on the Internet at http://rsb.info.nih.gov/nih-image/).

**Protein purification.** Overnight cultures were subcultured into 250 mL of fresh LB media with appropriate antibiotics. At $OD_{600} \approx 0.5$, cultures were induced as described above and shaken for 3 h at 37°C. Cells were then pelleted using a J2-21 floor centrifuge (Beckman) and lysed using Bugbuster Master Mix (Novagen). Samples were purified using Ni-NTA spin columns (Qiagen) according to the manufacturer's instructions. Purification was not optimized.

**In-gel digestion for excised gel bands.** Following visualization of the SDS-PAGE gel, two visible protein bands of interest were excised, diced, and placed into microtubes for the

subsequent in-gel digestion and extraction. The in-gel digestion by chymotrypsin (from Sigma, St. Louis, MO) and the subsequent peptide extraction were performed following a protocol from Yang, *et al*. (*87*) with slight modification. The gel pieces were washed and destained with a series of solutions: 50 μL of water, 50 μL of 50% ACN/50% 50 mM ammonium bicarbonate pH 7.8, and 50 μL of 100% ACN. The samples were reduced with DTT and alkylated by treatment with iodoacetamide. Once the samples were dried down completely after washing, ~0.2 μg LysC or chymotrypsin in 20 μL of 50 mM ammonium bicarbonate (pH = 7.8) and 10% ACN was added to each tube. The samples were left on ice for 15 min and incubated overnight at 37°C. The supernatant containing digested peptides was removed after centrifuging for 2 min at 4000 × *g*, and the remaining peptides were then extracted from the gel in a series of extraction steps. The first was with 30 μL of 25 mM ammonium bicarbonate pH 7.8 (30 minutes). Two sequential steps each with 50 μL of 5% formic acid in 50% acetonitrile (10 min) followed. For each extraction, the sample was sonicated for 5 min before the supernatant was removed. All gel-extracted supernatants were combined and evaporated to dryness in a Speedvac SC110 (Thermo Savant, Milford, MA).

**Protein identification by nanoLC/MS/MS analyses.** The tryptic digest was reconstituted in 15 μL of 2% ACN with 0.5% FA for nanoLC-ESI-MS/MS analysis, which was carried out using a LTQ-Orbitrap Velos (Thermo-Fisher Scientific, San Jose, CA) mass spectrometer equipped with a nano ion source device (CorSolutions LLC, Ithaca, NY). The Orbitrap is interfaced with an UltiMate3000 MDLC system (Dionex, Sunnyvale, CA). The nanoLC was carried out by Dionex UltiMate3000 MDLC system (Dionex, Sunnyvale, CA). An aliquot of tryptic peptide (3.0 μL) was injected onto a PepMap C18 trap column (5 μm, 300 μm × 5 mm, Dionex) at a 20 μL/min flow rate for on-line desalting. It was then separated on a PepMap C-18 RP nanocolumn (3 μm,

75µm x 15cm) and eluted in a 60 min gradient of 5% to 38% acetonitrile (ACN) in 0.1% formic acid at 300 nL/min followed by a 3-min ramping to 95% ACN-0.1%FA and a 5-min holding at 95% ACN-0.1%FA. The column was re-equilibrated with 2% ACN-0.1%FA for 20 min prior to the next run. The eluted peptides were detected by Orbitrap through the nano ion source containing a 10-µm analyte emitter (NewObjective, Woburn, MA). The Orbitrap Velos was operated in positive ion mode with nano spray voltage set at 1.6 kV and source temperature at 225 °C. Either internal calibration using the background ion signal at m/z 445.120025 as a lock mass or external calibration for FT mass analyzer was performed. The instrument was run at data-dependent acquisition (DDA) mode using FT mass analyzer for one survey MS scan followed by MS/MS scans on the five most intense peaks with multiple charged ions above a threshold ion count of 5000. MS survey scans were acquired at a resolution of 60,000 (fwhm at *m*/*z* 400) for the mass range of m/z 400-1400, and MS/MS scans were acquired at 7,500 resolution for the mass range of m/z 100 to 2000. Dynamic exclusion parameters were set at repeat count 1 with a 20 s repeat duration, exclusion list size of 500, 30 s exclusion duration, and ±10 *ppm* exclusion mass width. High energy dissociation (HCD) parameters were set at the following values: isolation width 2.0 m/z, normalized collision energy 45 %, and activation time 0.1 ms. All data were acquired with Xcalibur 2.1 operation software (Thermo-Fisher Scientific).

**Data analysis.** All MS and MS/MS raw spectra were processed using Proteome Discoverer 1.1 (PD1.1, Thermo). The spectra from each DDA file were manually inspected for both expected precursor ions of interest and their MS/MS spectra. The mass accuracy for all identified peptides is within 2 ppm.

***Results***

**A tunable E3-dependent sumoylation system.** To establish a SUMO-conjugation cascade in *E.*

*coli*, the bacterial pZ vector system developed by Lutz and Bujard (*82*) was used. We chose the pZ vector system because of its modular nature, unique promoters, and medium to low copy number. Previous studies showed that strong expression of the E1 (human Aos1 and Uba2) and E2 (murine Ubc9) enzymes in *E. coli* results in sumoylation that is independent of the SUMO E3 ligase (*39*). However, poly-sumoylated target evolves alongside mono-sumoylated product. To generate only mono-sumoylated target proteins, we attempted to reduce the expression of the E1 and E2 enzymes by placing the genes encoding human E1 and murine E2 into the medium copy vector pZA31-SMCS or the low copy vector pZS31-SMCS (Fig. 2.1a). To maximize sumoylated product, human SUMO-1 and the target protein were placed in the high copy vector pZE11-SMCS (Fig. 2.1a). A FLAG epitope tag was introduced to the C-terminus of the target protein to facilitate Western blot analysis. SUMO E3 ligases were placed on a separate plasmid, pZA24-SMCS, with a compatible replication of origin, p15A (Fig. 2.1a). The separate plasmid enables introduction of modifications to the E3 protein without altering the rest of the cascade components. Additionally, the $P_{lac/ara}$ promoter allows modulation of the E3 expression level without impact upon the remaining components.

We first investigated the bacterial expression of several mammalian SUMO E3 ligases. Specifically, four enzymes from the PIAS family were tested (PIAS1, PIASxβ, PIAS3, and PIASy). Of these, PIASxβ was expressed most efficiently (Fig. 1b; data for PIASxβ and PIASy only); hence, we chose this E3 for further study. The synthetic GST-PML target of Mencía and de Lorenzo was chosen as a model target substrate for our E3-dependent SUMO-conjugation system (*39*). This substrate is comprised of *E. coli* glutathione *S*-transferase (GST) that has been C-terminally modified with the 10-residue consensus sumoylation site from the promyelocytic leukemia (PML) protein. Previous studies using *E. coli* showed that this target can be sumoylated

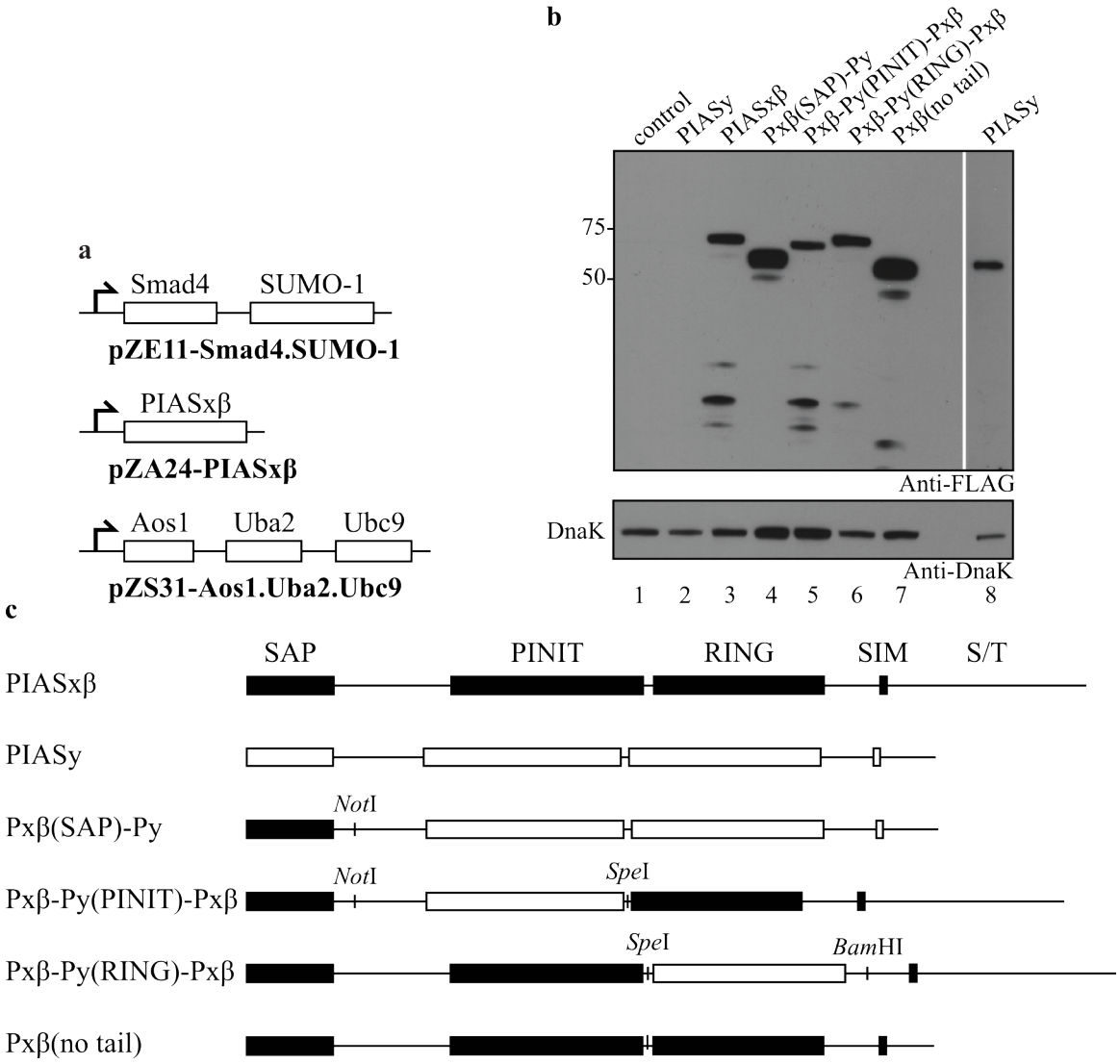


**Figure 2.1. An E3-dependent sumoylation system**. (a) Plasmid diagrams for the E3-dependent sumoylation system based on the pZ vector collection. The E1 (Aos1 and Uba2) and E2 (Ubc9) were cloned into the low copy plasmid pZS31-SMCS or the medium copy pZA31-SMCS (not shown); the E3 (e.g., PIASxβ) was cloned into the medium copy plasmid pZA24-SMCS; the target protein (e.g., Smad4-FLAG) and SUMO-1 were cloned into the high copy plasmid pZE11-SMCS. (b) Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing native and engineered SUMO E3 ligases as indicated. A much longer exposure time was required to visualize PIASy (lane 8). Control cells carried the empty pZE12-SMCS vector (lane 1). Blots were probed with anti-FLAG antibodies or anti-DnaK antibodies, with the latter serving as a loading control. (c) Schematic of the E3 chimeras and truncation mutant tested in this study. Chimeras were created by swapping different domains between human PIASxβ and PIASy using the inserted restriction sites.

in an E3-independent manner (*39*). In a similar fashion, we observed that when the E1 and E2 enzymes were expressed from the medium copy pZA31-SMCS vector in the absence of the E3, a slower migrating GST-PML band was detected (Fig. 2.2a, lane 3) but disappeared when the E1 and E2 were also absent (Fig. 2.2a, lane 2). Several lines of evidence indicate that this higher band is GST-PML that has become sumoylated in an E3-independent manner. First, this band migrated with an ~20-kDa upshift compared to the unmodified GST-PML protein, which is consistent with the roughly ~20-kDa shift previously reported for SUMO-1 (*88*). Second, it reacted with anti-SUMO-1 antibodies (Fig. 2.2b, lane 3).

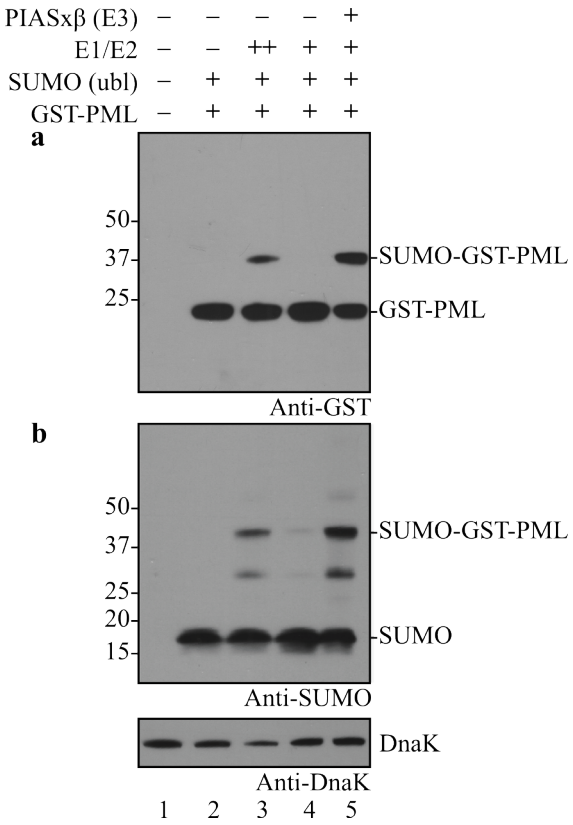


**Figure 2.2. E3-dependent sumoylation of synthetic GST-PML.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing the synthetic target GST-PML in the presence (+) or absence (-) of different SUMO-conjugation cascade components. The E1 and E2 enzymes were expressed from either the medium copy plasmid pZA31-SMCS (++) or the low-copy plasmid pZS31-SMCS (+). GST-PML was detected using anti-GST antibodies (a) while SUMO-1 was detected using anti-SUMO-1 antibodies (b). Detection of endogenous DnaK with anti-DnaK antibodies served as a loading control.

Next, we lowered the expression level of the E1 and E2 enzymes by inducing each from the low-copy pZS31-SMCS plasmid. Under these conditions, the slower migrating band disappeared (Fig. 2.2a, lane 4). Given that a faint band was detectable upon probing with anti-SUMO-1 antibodies (Fig. 2.2b, lane 4), we conclude that sumoylation efficiency was drastically reduced under these conditions. Upon introduction of the SUMO E3 ligase PIASxβ, sumoylated GST-PML reappeared under conditions where the E1 and E2 were expressed from the low copy vector (Fig. 2.2a and b, lane 5 in each). Thus, by lowering the expression levels of the E1 and E2 enzymes and by adding a functional E3 enzyme, we successfully created an E3-dependent sumoylation cascade in *E. coli*. It is particularly noteworthy that the efficiency of sumoylation appeared to increase with the addition of PIASxβ (Fig. 2.2a or b, compare lanes 4 and 5). Although undetectable with anti-GST antibodies, a faint band above the mono-sumoylated GST-PML was observed using anti-SUMO-1 antibodies (Fig. 2.2b, lane 5). This band likely arises from GST itself becoming weakly sumoylated as has been previously reported (*41*). The anti-SUMO-1 antibodies also revealed a ~30-kDa band (Fig. 2.2b, lanes 3 and 5) that likely corresponds to a degradation product of the sumoylated target, a native *E. coli* protein that has become sumoylated, or a free di-SUMO-1 chain.

**Mono-sumoylation of target proteins by E3-dependent SUMO modification system.** In the earlier study of Mencía and de Lorenzo, E3-independent sumoylation in engineered *E. coli* resulted in modification of target proteins with SUMO-1 chains (*39*). To more carefully investigate whether target proteins in our sumoylation system were poly-sumoylated, we converted the green fluorescent protein (GFP) to a sumoylation substrate by fusion to the PML tag. Since GFP does not contain any predicted sumoylation sites, mono- versus poly-sumoylation of the GFP-PML chimera can be used to assess SUMO-1 chain formation on target proteins.

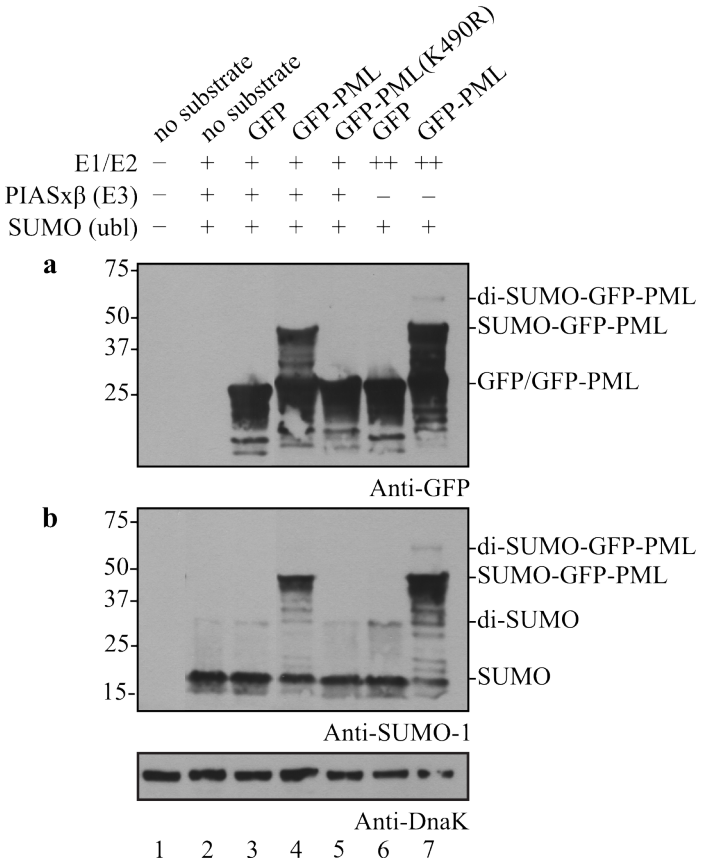


**Figure 2.3. E3-dependent sumoylation of synthetic GFP-PML.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing the synthetic target GFP-PML in the presence (+) or absence (-) of different SUMO-conjugation cascade components. The E1 and E2 enzymes were expressed from either the medium copy plasmid pZA31-SMCS (++) or the low-copy plasmid pZS31-SMCS (+). GFP was detected using anti-GFP antibodies (a), while SUMO-1 was detected using anti-SUMO-1 antibodies (b). K490 refers to the lysine's native location within PML rather than its location within GFP-PML. Detection of endogenous DnaK with anti-DnaK antibodies served as a loading control.

Indeed, expression of GFP without the PML tag in the E3-dependent (Fig. 2.3a and b, lane 3 in each) and E3-independent (Fig. 2.3a and b, lane 6 in each) systems resulted in no detectable target sumoylation. Likewise, no sumoylation was detected for the GFP-PML chimera when the lysine in the PML tag was mutated to arginine (Fig. 2.3a and b, lane 5 in each). On the other hand, expression of GFP-PML in the presence of the E3-dependent and E3-independent sumoylation cascades resulted in a clear band corresponding to mono-sumoylated GFP-PML. For the E3-dependent system, the yield of mono-sumoylated GFP-PML was ~5 mg/L of culture.

Interestingly, a band corresponding in mass to di-SUMO-1 conjugated to GFP-PML was observed for the E3-independent but not the E3-dependent system (Fig. 2.3a and b, compare lanes 4 and 7 in each), suggesting that the addition of the E3 and/or the reduced expression of the E1 and E2 prevented poly-sumoylation. It is noteworthy that a rather faint ~30 kDa band was detected with anti-SUMO-1 antibodies, similar to that seen above in the GST-PML experiments. Since this band did not depend on the presence of the target substrate (Fig. 2.3b, lanes 2, 3 and 6), we conclude that this band is not a degradation product of the sumoylated target. It is also noteworthy that the intensity of this band increased when the E1 and E2 were expressed from the medium copy plasmid and the E3 was absent (Fig. 2.3b, lanes 6 and 7).

**Conjugation of SUMO-1 to a natural sumoylation target protein.** Next, we investigated whether our sumo-engineered *E. coli* could conjugate SUMO-1 to a naturally occurring target of the human sumoylation machinery. We chose the human tumor suppressor protein Smad4, a central intracellular signal transducer for transforming growth factor-β (TGF-β) signaling, whose transcriptional potential is regulated by sumoylation (*88*, *89*). Similar to our results above, expression of the E1 and E2 from a medium copy vector resulted in E3-independent sumoylation of Smad4 (Fig. 2.4a and b, lane 10 in each) whereas expression of the E1 and E2 from a low-copy plasmid resulted in virtually no detectable Smad4 sumoylation (Fig. 2.4a and b, lane 9 in each). However, co-expression of the E1 and E2 from a low copy vector along with the E3 resulted in strong sumoylation of Smad4 (Fig. 2.4a and b, lane 7 compared to 1-6). As with the synthetic GST-PML, sumoylation of Smad4 appeared to be more efficient in the presence of the E3 (Fig. 4a and b, compare lanes 7 and 9). The major sumoylation site in Smad4 is the consensus lysine at position 159 (*88*, *90*). Mutation of this lysine residue to arginine (K159R) abolished

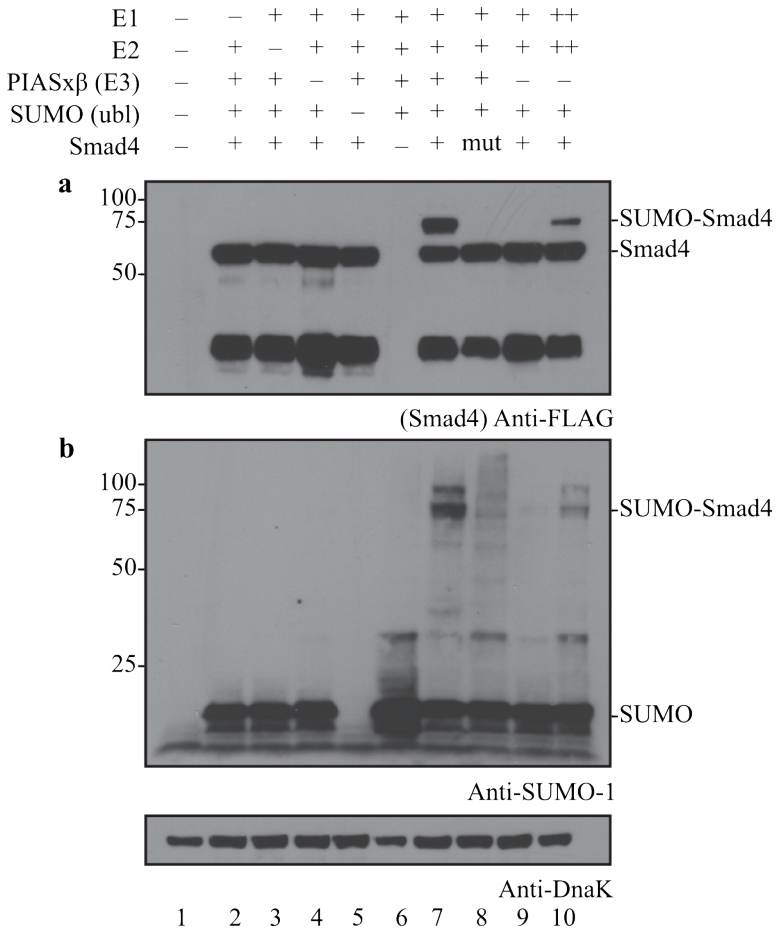


**Figure 2.4. E3-dependent sumoylation of human Smad4.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing human Smad4 or Smad4(K159R) (mut) in the presence (+) or absence (-) of different SUMO-conjugation cascade components. The E1 and E2 enzymes were expressed from either the medium copy plasmid pZA31-SMCS (++) or the low-copy plasmid pZS31-SMCS (+). Smad4 was detected using anti-FLAG antibodies (a), while SUMO-1 was detected using anti-SUMO-1 antibodies (b). Detection of endogenous DnaK with anti-DnaK antibodies served as a loading control.

Smad4 sumoylation (Fig. 2.4a, lane 8). To verify that K159 is the major site of SUMO attachment in our system, we performed MALDI-TOF mass spectrometry (MS) analysis on the SUMO-Smad4 band, which was purified on a Ni-NTA column and separated from unmodified Smad4 by SDS-PAGE. As expected, nearly all of the Smad4 was sumoylated at the consensus K159 (Fig. 2.5).

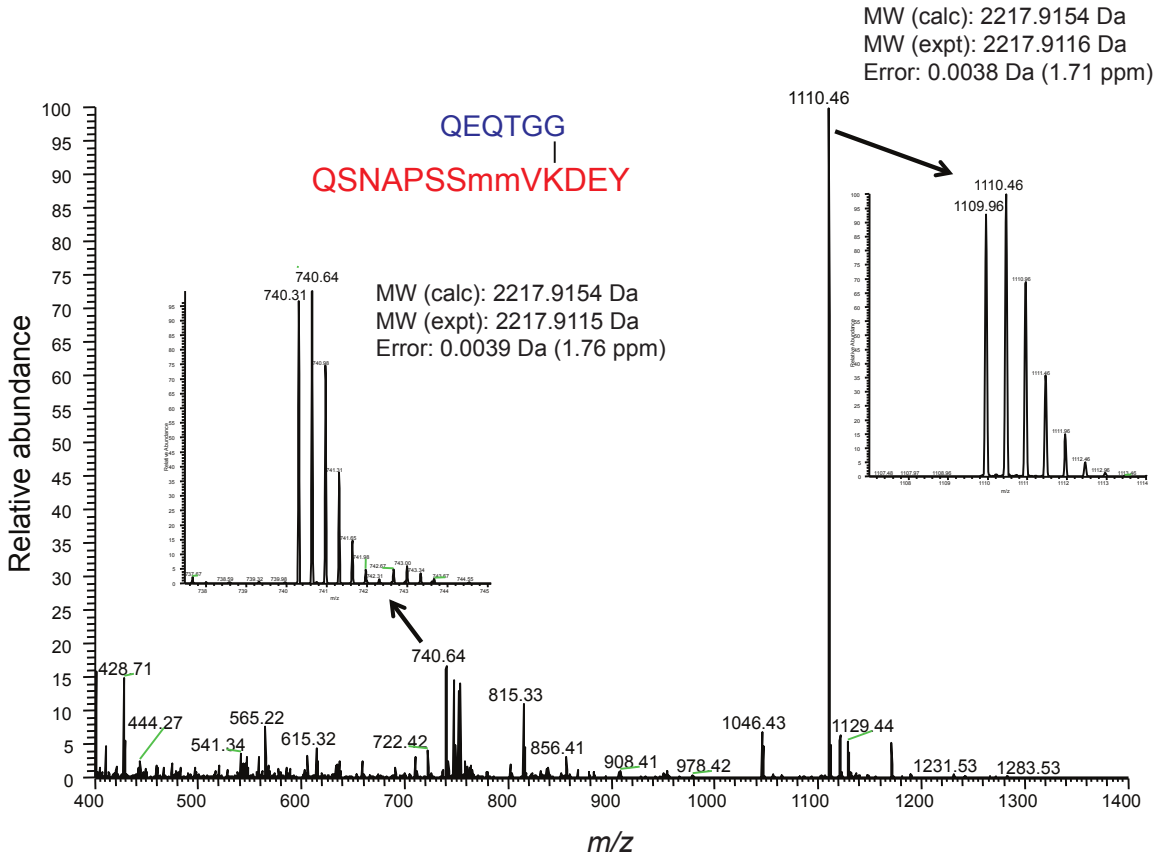


**Figure 2.5. MS of SUMO-Smad4.** (First panel) MS spectrum of Smad4 chymotryptic digests acquired in the FT analyzer of the Orbitrap Velos during the nanoLC-MS/MS analysis at elution time =23.71 min. A base-peak doubly-charged precursor ion at *m/z* 1109.9631 with its triply-charged ion at *m/z* 740.3111 shown in expanded view of insets is identified as sumoylated peptide. Sequence for the Smad4 peptide (red) with the conjugated SUMO-1 peptide (blue) after chymotrypsin digestion is shown. Lower case m indicates the oxidized methionine. The survey MS scan shows that the mass of the detected sumoylated peptide at K159 is under 1.8 ppm of its calculated mass. (Second panel) MS/MS spectrum of a triply-charged ion at *m/z* $740.31^{3+}$ acquired in HCD-DDA analysis by the FT analyzer at 23.90 min derived from Smad4 residues 149 to 162 with K159 identified as the sumoylated site. The y- and b-type ions are labeled in the spectrum as blue and red color for the SUMO-1 and the Smad4 target peptides, respectively. (Third panel) MS/MS spectrum of $1109.96^{2+}$ ion eluted at 23.84 min for identification of K159 sumoylation.

An even higher molecular weight band relative to SUMO-Smad4 was also produced in our sumo-engineered *E. coli* (Fig. 2.4a and b, lane 7 in each). This band might correspond to the attachment of SUMO-1 to a minor site on Smad4 or to the formation of SUMO-1 chains on Smad4. We favored the former possibility for two reasons. First, low-level expression of the E1 and E2 along with the E3 promoted mono-sumoylation in the case of GFP-PML. Consistent with

this result, MS analysis of SUMO-Smad4 failed to reveal evidence for the formation of SUMO-1 chains at either K16 or K17 of the already conjugated SUMO-1 (data not shown). Second, a faint sumoylation band was observed for Smad4(K159R) (Fig. 2.4b, lane 8). Indeed, a known minor site of sumoylation on Smad4 is the non-consensus K113 residue (*88*, *90*). However, MS analysis did not provide any evidence for SUMO-1 conjugation at this position (data not shown). Thus, taken together, we suspect that another lysine is sumoylated on this higher molecular weight Smad4 species; however, at present the identity of this lysine remains undetermined.

**Functional characterization of SUMO E3 ligase chimeras.** The generation of chimeras, truncations, and mutants of the Siz/PIAS protein family has provided great insight into the function of each protein (*65*, *91*). These alterations may impact localization, interaction with local cellular factors, and recognition of the target protein. However, decoupling these differences to deduce function can be difficult in the eukaryotic cellular environment. We hypothesized that our sumo-engineered *E. coli* could be useful for understanding the SUMO-conjugation activity of different E3 chimeras because it is devoid of the aforementioned complications. To test this notion, we constructed several SUMO E3 ligase variants. These included chimeras that were generated by swapping the SAP (scaffold attachment factor-A/B, acinus and PIAS), PINIT, or SP-RING (Siz/PIAS really interesting new gene) domains between PIASxβ and PIASy, and a truncation mutant that was made by eliminating the C-terminal tail of PIASxβ (Fig. 2.1c). Although PIASy expression could only be seen after a much longer exposure time compared to PIASxβ (Fig. 2.1b), its expression was second most efficient among all PIAS family members that were tested. Hence, we chose to use PIASy in our chimeric constructs. A panel of E3 variants which all contain the N-terminal SAP domain from PIASxβ were observed to express on par with PIASxβ (Fig. 2.1b). Since PIASy also sumoylates

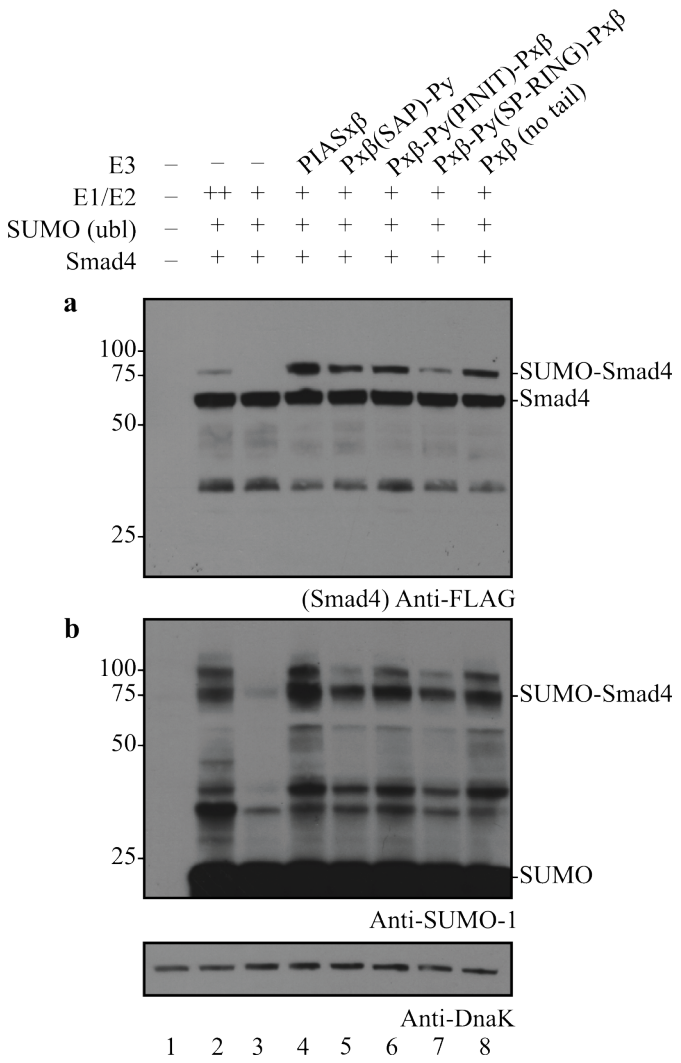


**Figure 2.6. Chimeric E3-dependent sumoylation of human Smad4.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing human Smad4 in the presence (+) or absence (-) of different SUMO-conjugation cascade components. The E1 and E2 enzymes were expressed from either the medium copy plasmid pZA31-SMCS (++) or the low-copy plasmid pZS31-SMCS (+). PIASxβ (lane 4) as well as a panel of E3 variants (lanes 5-8; see Fig. 1c caption for details) were tested for functionality. Smad4 was detected using anti-FLAG antibodies (a), while SUMO-1 was detected using anti-SUMO-1 antibodies (b). Detection of endogenous DnaK with anti-DnaK antibodies served as a loading control.

Smad4 (*92*), we predicted that each of these variants would sumoylate Smad4. In line with our hypothesis, all of the E3 variants conjugated SUMO to Smad4, albeit to varying extents (Fig. 2.6a and b, lanes 5-8 in each). None were as efficient as PIASxβ; the Pxβ-Py(RING)-Pxβ chimera appeared to be the least efficient (Fig. 2.6a and b, lanes 7 in each). Taken together,

these data reveal the potential of our bacterial SUMO-conjugation system for functional evaluation of native as well as engineered SUMO E3 ligases.

***Discussion***

In this study, we have created the first E3-dependent sumoylation system in *E. coli*. We anticipate that sumo-engineered *E. coli* will be useful in further studies of the sumoylation mechanism for several reasons. First, greater yields of sumoylated proteins for biochemical and structural analysis should be attainable through the addition of an E3 (*67*, *90*). Indeed, for both GST-PML and Smad4 substrates, we observed an increase in sumoylation efficiency following the addition of a functional E3 to the system. Furthermore, by lowering the expression of the E1 and E2, additional cellular resources can be diverted towards production of the target protein. Even without any process optimization, our E3-dependent SUMO conjugation system yielded ~5 mg/L of mono-sumoylated protein. Second, the system enables functional characterization of any of the sumoylation cascade enzymes while eliminating the concern for localization, downstream interactions, and the diversity of sumoylated proteins that can obscure similar analysis in eukaryotic hosts. Our system also produces physiologically relevant results. For instance, we observed that Smad4 was sumoylated primarily at K159, which is reported to be the major sumoylation site (*88*, *90*). We did not detect sumoylation at position K113, which was reported as a minor site of sumoylation in one report (*88*) but was not sumoylated in another (*93*). We also did not detect SUMO-1 chains on target proteins in our E3-dependent system, which is in stark contrast to an earlier bacterial E3-independent sumoylation system (*39*). It should be noted, however, that the inability of MS analysis to reveal poly-sumoylation via K16 and K17 linkages on SUMO-1 could arise from low abundance and/or poor ionization efficiency of these species. Nonetheless, based on the high-intensity MS signal detected for the K159

SUMO-1 peptide, we conclude that no appreciable quantities of SUMO-1 chains are present. Overall, our system yields results that are entirely consistent with the known molecular biology of sumoylation. As a corollary, we show that engineered E3 variants can be expressed and functionally characterized in our system. This is significant because our bacterial SUMO-conjugation system provides a potentially less convoluted background for studying sumoylation. While *in vitro* reconstitution studies could also be used to eliminate these factors, our system obviates the need for purification of each cascade component and the corresponding need to modify each cascade component with a purification tag, which can affect enzyme function. Thus, we anticipate that our sumo-engineered *E. coli* system will be a useful new tool for illuminating the molecular details of the SUMO-conjugation process.

***Acknowledgements***

We would like to thank Dr. Mario Mencía and Dr. Victor de Lorenzo for kindly providing plasmids for their E3-independent sumoylation system upon which we based our work. We would also like to thank Dr. Ke Shuai for the gift of PIASxβ (*84*) (Addgene plasmid 15210) and PIASy (*85*) (Addgene plasmid 15208). In addition, we thank Dr. Sheng Zhang and Dr. A. Celeste Ptak for their work with the MALDI-TOF spectrometry. This work was supported by the National Science Foundation Career Award CBET-0449080 (to M.P.D.); the New York State Office of Science, Technology and Academic Research Distinguished Faculty Award (to M.P.D.); and a Cornell Genomics Fellowship (to S.P.O.).

CHAPTER 3

# FUNCTIONAL RECONSTITUTION OF THE FEN-UBIQUITIN PATHWAY IN ESCHERICHIA COLI

## *Introduction*

Ubiquitination is a eukaryotic post-translational modification that involves the covalent conjugation of the 8.5 kDa ubiquitin protein to a lysine residue in a target substrate (for recent reviews of the ubiquitination mechanism and its implications see (*12*, *22*, *24*, *94*, *95*)). Three steps comprise the pathway: activation, conjugation, and ligation. The E1 activates ubiquitin through formation of a thioester bond. The E2 binds the E1 and acquires ubiquitin through a transthioesterification reaction. The charged E2 then binds an E3 capable of also binding the target protein. The ubiquitin is transferred to the ε-amino group on one of the substrate's lysine residues. Subsequent rounds can form chains of ubiquitin on the target protein. A wide breadth of cellular processes involve ubiquitination including intracellular trafficking, signal transduction pathways, protein activation and deactivation, endocytosis, lysosomal targeting, transcriptional regulation, cellular replication, chromatin remodeling, DNA repair, and protein degradation.

In order to study aspects of this system, several groups have reconstituted the pathway in *Escherichia coli* (*E. coli*.) Su *et al* did so to successfully screen a murine cDNA library for substrates of the E3 GRAIL (*43*) after their attempts using mass spectrometric analysis and screening strategies in mammalian cells failed. Several years later Rosenbaum *et al* reconstituted another pathway to characterize the role yeast ubiquitin ligase San1 played in protein quality control (PQC) degradation (*44*). Their “in coli” method validated ubiquitination of putative

substrates identified using the yeast two-hybrid screen that they could not purify due to aggregation. Most recently, Keren-Kaplan *et al* meticulously characterized ubiquitination reconstituted in *E. coli* using several different pathways (*45*). They provided evidence that their system recapitulated physiological results that *in vitro* reactions could not and produced enough ubiquitinated substrate for various analyses such as crystallization and biophysical and biochemical assays.

Reconstituting ubiquitination in *E. coli* may further as a platform for engineering the pathway. *E. coli* lack many of the elements which make evolution in the native host difficult such as deubiquitinases (DUBs), a 26S proteasome, competing pathways, and low cellular abundances. It also benefits from many molecular tools developed over the years. Guntas *et al* evolved the ubiqutin E3 ligase E6AP to dock NEDD8-conjugating enzyme Ubc12 using a bacterial screen with library informed by computer models (*47*). They also recapitulated a partial Nedd8 pathway in *E. coli* and developed a novel screen based on Nedd8-E2 interactions (*48*). In this latter investigation, they evolved more soluble Nedd8 and orthogonal mutant NAE1-Nedd8 pairs (*48*).

We developed our own ubquitination pathway in *E. coli* important in plant defense against bacterial colonization. The *Pseudomonas syringae* type III effector AvrPtoB is a RING E3 ligase (*96*) that polyubiquitinates the tomato kinase Fen for degradation (*46*). In doing so, AvrPtoB suppresses a susceptible host's programmed cell death response, thereby allowing the bacteria to flourish within that host. We are the first to report that ubiquitination of Fen may proceed in an E3-independent manner likely through auto-monoubiquitination involving a ubiquitin binding domain (UBD). Sufficient accumulation of modified substrate provides a

strong signal for developing a reporter for ubl-target protein conjugation. To date no such reporter exists.

***Material and Methods***

**Plasmid construction.** Construction of plasmids pZE12, pZE11, pZA24, and pZS31was previously described (*42*). To make pZS31-UbcH5b, pZS31-Ube1, and pZS31-Ube1.UbcH5b, the genes encoding human Ube1 and human UbcH5b were PCR amplified from pCDF-Duet1-Ube1/UbcH5A (*43*) and pET15b-UbcH5b (*97*) respectively. The resulting PCR products were then inserted into pZS31-SMCS using *Kpn*1 and *Sph*1 for Ube1 and *Mlu*1 and *Eag*1 for UbcH5b. To generate the plasmids pZE11-Fen, pZE11-Ubiquitin and pZE11-Fen.Ubiquitin, Fen and ubiquitin were PCR amplified from pMAL-C2X-Fen (*46*) and pET15-Ubiquitin WT (*98*) respectively. Fen was inserted between *Kpn*I and *Sph*I of pZE11-SMCS and ubiquitin *Mlu*1 and *Eag*1. DNA encoding a FLAG or HA epitope tag was added to the C-terminus of Fen or the N-terminus of ubiquitin during PCR amplification where noted. For pZE11-ssTorA-Fen-HA.FLAG-Ubiquitin and pZE11- Fen-HA.ssTorA-FLAG-Ubiquitin, an ssTorA signal sequence (Waraho09) was added to the N-terminus using overlap-extension. To construct pZA24-AvrPtoB, AvrPtoB (226-328) was PCR amplified from pDEST15-AvrPtoB (*96*) and inserted between *Kpn*I and *Sph*I of pZA24-SMCS. This truncated form of AvrPtoB had better expression than its full-length form without loss of activity (data not shown).

**Cell growth and Western blot analysis.** All constructs were transformed into *E. coli* host strain DH5α-Z1 (*82*) using a GenePulser Xcell (BioRad). Individual colonies were grown overnight in LB media with appropriate antibiotics (100 μg/mL ampicillin, 40 μg/mL kanamycin, and 12.5 μg/mL chloramphenicol) and then subcultured to $OD_{600} \approx 0.05$ in 5 mL of fresh LB media supplemented with appropriate antibiotics. Cultures were induced at $OD_{600} \approx 0.75$ with 0.5%

L(+)-arabinose, 1 mM IPTG, and 50 ng/mL anhydrotetracycline when appropriate and subsequently shaken for 24 h at 25°C unless otherwise noted. Approximately 1.5 mL of each culture was harvested and lysed using 200 μL of Bugbuster Master Mix (Novagen) according to the manufacturer's directions. Lysates were normalized to 10 μg of total protein as determined by a total protein assay (Bio-Rad) and loaded on a 4-20% Precise Protein Gel (Thermo Scientific). Transfers to Immobilon P Transer Membranes (Millipore) were performed for 2 h at the maximum amperage recommended for a Biosciences TE77 semi-dry transfer unit (Amersham). Blots were then imaged on film using standard protocols. The primary antibodies used were anti-HA (Sigma), anti-FLAG (Abcam), anti-ubiquitin (Millipore 05-944), anti-K48-linked-poly-ubiquitin (Millipore), and anti-GroEL (Sigma).

**Periplasmic fractionation.** Cultures were treated as described above. Subcellular fractionation was accomplished using the osmotic shock procedure (*99*). Cells were pelleted by centrifugation at 1400 g for 10 min and resuspended in 1 mL Fractionation Buffer (30 mL 1M Tris-HCl (pH 8.0), 200 g Sucrose, and 2 mL 0.5 $Na_2EDTA$ for 1 L of buffer) to $OD_{600}$ 75. The suspension was transferred to a 1.7 mL Eppindorf tube and incubated at 25ºC for 10 min. Cells were centrifuged 9000 g for 10 min and supernatant discarded. Samples were resuspended in 266 mL of ice-cold 5 mM $MgSO_4$ stock solution and placed on ice for 10 min. Cells were pelleted by centrifugation at 15500 g for 10 min and supernatant retained as the periplasmic fraction. The pellet was washed with 500 μL PBS and spheroplasts lysed using 266 mL Bugbuster Master Mix (Novagen) according to the manufacturer's instructions. Samples for Western blotting were normalized by total protein based on the cytoplasmic fraction and equal volumes used for the periplasmic fraction.

***Results***

**Reconstitution of the Fen-ubiquitination pathway.** To establish a ubiquitin-conjugation pathway in *E. coli*, the bacterial pZ vector system developed by Lutz and Bujard (*82*) was employed. We chose the pZ vector system because of its modular nature, unique promoters, and compatible origins of replication. The genes encoding the E1 (Ube1) and E2 (UbcH5b) were placed in the low copy vector pZS31-SMCS (Fig. 3.1). To maximize ubiquitinated product, genes for ubiquitin and the target protein Fen were placed in the high copy vector pZE11-SMCS (Fig. 3.1). A FLAG epitope tag was introduced to the N-terminus of ubiquitin while a HA epitope tag was introduced to the C-terminus of Fen to facilitate Western blot analysis. The gene for the E3 ligase AvrPtoB was placed on a separate, medium copy plasmid, pZA24-SMCS, with a compatible origin of replication, p15A (Fig. 3.1).

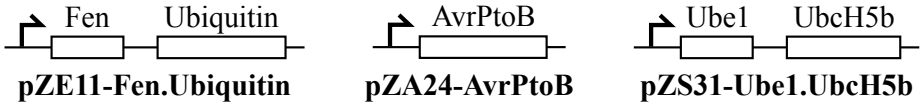


**Figure 3.1. Plasmid diagrams for the Fen-ubiquitin pathway in *E. coli*.** The target protein (Fen) and ubiquitin were cloned into the high copy plasmid pZE11-SMCS. The E3 (AvrPtoB) was placed in the medium copy plasmid pZA24-SMCS. The E1 (Ube1) and E2 (UbcH5b) were placed in the low copy plasmid pZS31-SMCS.

When all components of the pathway were present, an additional band migrated ~10 kDa above Fen (Fig. 3.2a, lane 7). This shift corresponded to the mass of ubiquitin with the FLAG epitope tag. This same band also reacted with anti-FLAG and anti-ubiquitin antibodies (Fig. 3.2b and c, lane 7). These lines of evidence suggested Fen to be ubiquitinated. By the same rationale, the weaker band appearing above the aforementioned corresponded to di-ubiquitination of Fen (Fig. 3.2a, b, and c, lane 7).

Curiously, ubqiuitination appeared to proceed in an E3-independent manner (Fig. 3.2a, b, and c, lane 4). Auto-monoubiquitination has been documented for proteins with a UBD (*21*).

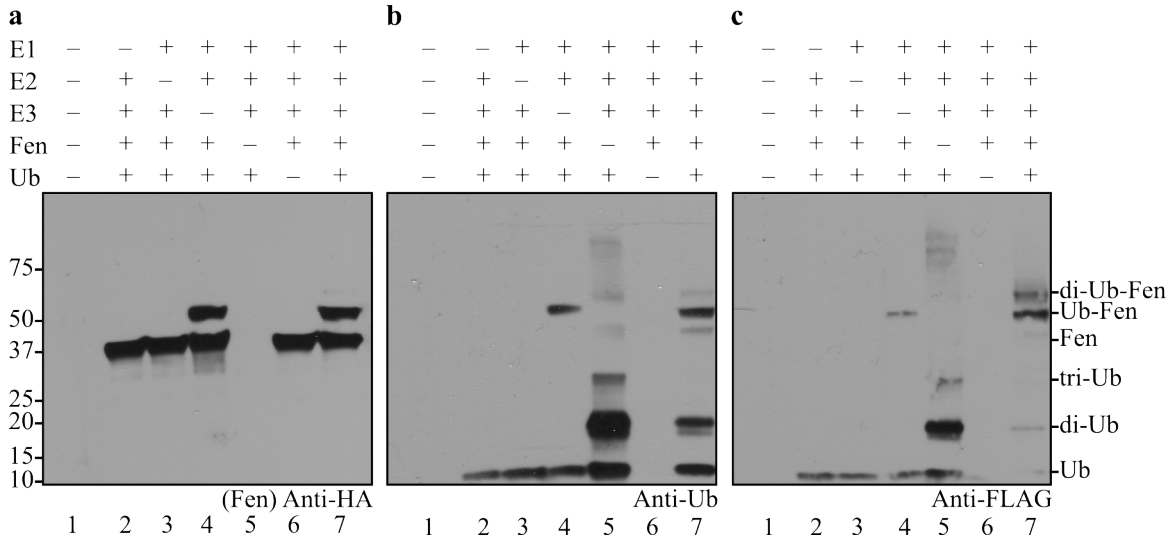


**Figure 3.2**. **Ubiquitination of Fen.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing cascade components. Fen was detected using anti-HA antibodies (a), while ubiquitin was detected using anti-ubiquitin (b) and anti-FLAG antibodies (c).

Consistent with this modification, only one ubiquitin appeared attached to Fen, and ubiquitin chain formation – free or attached – was absent (Fig. 3.2b and c, lane 4). Whether Fen possesses a UBD is unknown, though other kinases do posses them (*94*, *100*). Auto-ubiquitination of Fen does not appear to occur in *in vitro* experiments (*46*, *101*). However, the process is E2 specific (*21*) and a different E2 was used here. Alternatively, an endogenous E3 ligase might have subsumed the role of AvrPtoB. While E3 ligases serving as effector proteins are increasingly found in bacteria including enterohaemorrhagic *E. coli* (*37*, *38*), none have been characterized in non-pathogenic *E. coli*. Additionally, the lack of chain formation was inconsistent with an E3's presence.

Lower molecular weight bands appeared in lanes 5 and 7 of the anti-ubiquitin and anti-FLAG blots at ~10 kDa increments above ubiquitin (Fig. 3.2b and c). Formation of unanchored poly-ubiquitin chains was observed *in vitro* (*37*) and *in vivo* in *E. coli* (*45*) when specific substrate was absent. The apparent masses of the bands are consistent with di-ubiquitin and tri-

ubiquitin. When Fen was present, intensity of the bands was greatly reduced (Fig. 3.2b and c, compare lanes 5 and 7).

**Stability and kinetics of ubiquitinated Fen.** In order to effectively develop a screen for ubl-target conjugation, we optimized growth conditions for the production of ubiquitinated Fen and followed the evolution of that accumulation. Samples were induced at 37ºC and 25ºC for 3, 6, 12, and 24 h and compared to a previously established baseline of 16ºC for 24 h. Accumulation for ubiquitinated Fen peaked at 25ºC and 24 h (Fig. 3.3b, lane 9). At 25ºC and 3h modification was barely discernible despite accumulation of substrate by that time (Fig. 3.3b, lane 6). Curiously, expression of the target was absent in the presence of the entire cascade at 37ºC (Fig. 3.3b, lanes 2-5). When expressed alone at this temperature, accumulation of Fen was visible though it peaked at 6 h and disappeared by 24 h (Fig. 3.3a, lanes 2-5). Because accumulation was stable and optimal at 25ºC and 24 h provided a sufficient period for ubiquitinated product persistence, these conditions were used for future screen development.

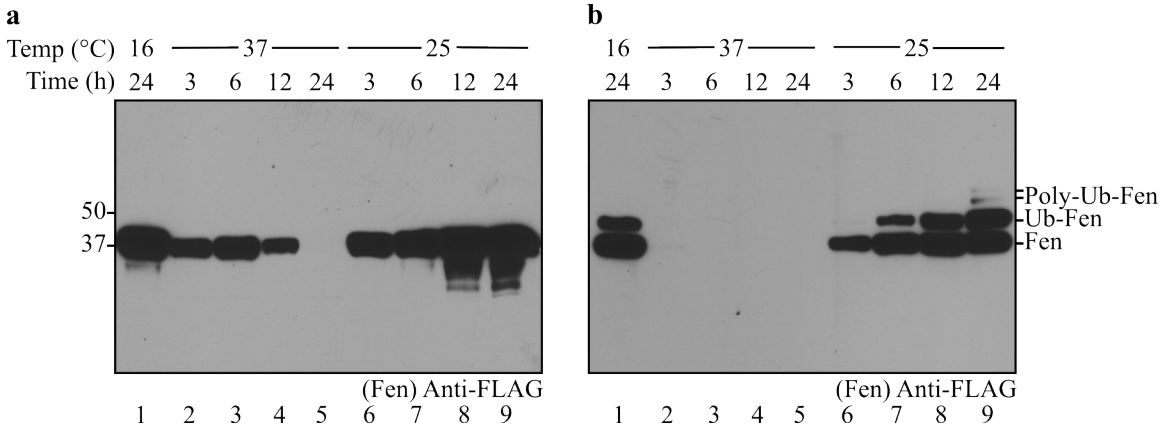


**Figure 3.3. Stability and evolution of ubiquitinated Fen.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing Fen and ubiquitin alone (a) or with the remainder of the cascade (b). Samples were induced at 37ºC and 25ºC for 3, 6, 12, and 24 h and 16ºC for 24 h. Fen was detected using anti-FLAG antibodies.

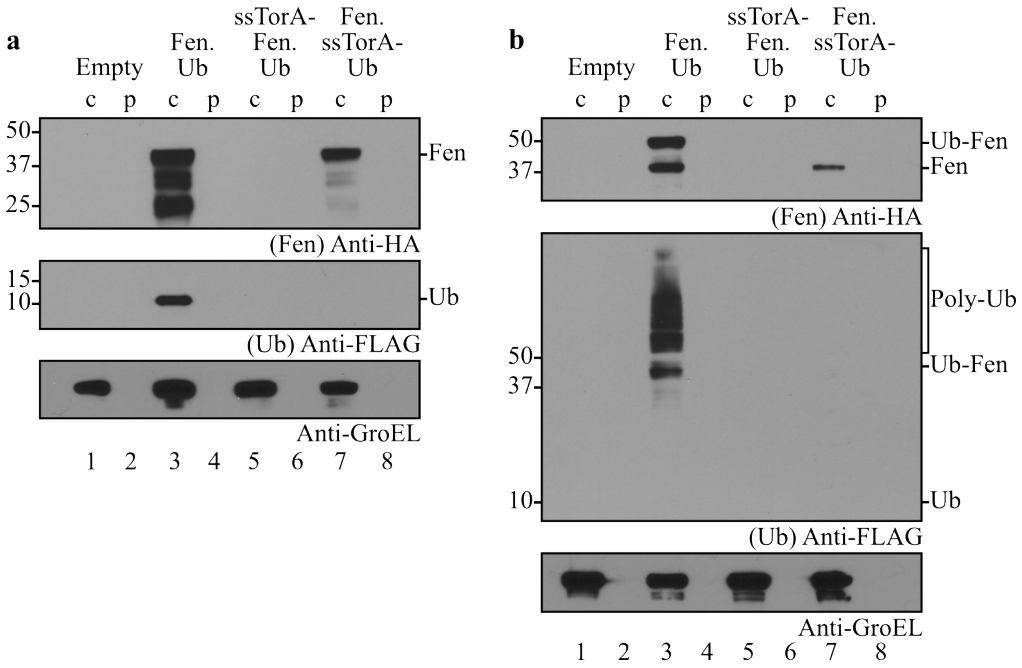


**Figure 3.4. Cellular localization of ubiquitinated Fen.** Western blot analysis of cell lysate prepared from DH5α-Z1 cells expressing Fen and ubiquitin alone (a) or with the remainder of the cascade (b). Cytoplasmic (c) and periplasmic (p) fractions were analyzed for accumulation of proteins in those compartments. Fen was detected using anti-HA antibodies, while ubiquitin was detected using anti-FLAG antibodies. Detection of endogenous GroEL with anti-GroEL antibodies served as a fractionation control.

**Ubiquitinated product is not transported to the periplasm.** In our current system, ubiquitination occurred in the cytoplasm. However, many useful screens, particularly those developed in our own lab (*102*, *103*), would rely on export of ubiquitinated target to the periplasm. To determine if this was feasible, we first tested whether Fen or ubiquitin alone was a competent TAT substrate. An ssTorA signal sequence was attached to the N-terminus of either Fen-HA or FLAG-ubiquitin. When both lacked an ssTorA signal sequence, Fen and ubiquitin appeared in the cytoplasm but not in the periplasm (Fig. 3.4a, lanes 3 and 4). When either Fen (Fig. 3.4a, lanes 5 and 6) or ubiquitin (Fig. 3.4a, lanes 7 and 8) was appended with the ssTorA signal sequence, it was not found in either the cytoplasm or periplasm. Surprisingly, ubiquitin was also not found in either the cytoplasm or periplasm when Fen bore the ssTorA signal sequence (Fig. 3.4a, lanes 5 and 6). With the other cascade elements present, Fen was only

modified when neither it nor ubiquitin bore the ssTorA signal sequence, and then only in the cytoplasm (Fig. 3.4b, lanes 3 and 4). We concluded that adding ssTorA to either Fen or ubiquitin proved destabilizing, and for this system expression of ubiquitinated product would be limited to the cytoplasm.

***Discussion***

Reconstituting a ubiquitin pathway in *E. coli* has proven a useful measure for study of the pathway and shows potential as a platform for protein engineering. We successfully reconstituted and characterized the Fen-ubiquitin pathway in *E. coli*. Unexpectedly, ubiquitination seemed to proceed in an E3-independent manner likely through auto-ubiquitination of Fen or through an endogenous bacterial factor. Our data support the former conclusion, and future experiments would resolve whether pathways with alternative compatible E2s modify Fen in this way.

We found both ubiquitin and Fen could not bear the ssTorA signal sequence due to the impact it had on their abundances. Potentially, Fen could be replaced by another target capable of being transported. Additionally, many other ssTorA signals exist (*104*) that could be tested. Recently, both ubiquitin and NEDD8 have been shown to be TAT-transport competent (*48*)(Guntas12). Disparity between these results could be resolved with future investigation and may derive from bacterial strain, plasmid, ssTorA signal, additionally expressed proteins, and/or detection method. However, competence of TAT transport would be a limitation and possibly compromise the generality of the reporter we hope to create for ubl conjugation to protein substrate. Because expression of ubiquitinated Fen was observed in the cytoplasm, we will place our efforts in developing cytoplasmic-based reporters.

**Acknowledgements**

We kindly thank Dr. Fathman for his gift of pCDF-Duet1-Ube1/UbcH5A (*43*) , Dr. Weissman for his gift of pET15b-UbcH5b (*97*) (Addgene #11475), and Dr. Klevit for her gift of pET15-Ubiquitin WT (*98*) (Addgene #12647). We are particularly grateful to Dr. Martin for his gift of pMAL-C2X-Fen (*46*) and pDEST15-AvrPtoB (*96*) and advice.

CHAPTER 4

# *IN VIVO* REPORTERS IN *ESCHERICHIA COLI* FOR UBIQUITIN-LIKE MODIFICATION[1]

## *Introduction*

Because ubl modification is undetectable in *E. coli*, we sought to tether it to a detectable trait to create a high throughput reporter. A common *in vivo* method to detect protein-protein interactions within the cytoplasm of a cell involves splitting a reporter protein and fusing the two fragments to the former protein pair (ref). When the pair interacts, the fragments come together thereby reconstituting the activity of the reporter protein. Common reporter traits for the split protein include fluorescence, antibiotic resistance, enzymes permitting growth on minimal media, and toxic proteins. Incorporating this approach into our methodology would reconstitute the detectable trait only upon ubl modification provided the split fragments are attached to both the ubl and target protein. For the reporter protein we used DHFR (*105*), YFP (*106*), and RNAP (*107–109*) but without success (data not shown).

Cre recombinase, a member of the bacterial P1 phage proteome, excises site-specific genetic material between pairs of the 34-bp *lox* sites (*110–112*). Since Cre was first demonstrated to function in a eukaryotic context (*113*), many applications involving its recombinase activity have been developed. Methods for controlled integration of genetic material into the chromosome were developed for yeast (*114*), mammalian cells (*114*, *115*), and plants (*116*). Removal of selection markers in transgenic plants for commercialization became possible (*117*). Proof of concept for tissue specific gene activation (*118*) and silencing (*119*) was

[1] Manuscript in preparation

done in mouse strains with actual application shortly following (*120*). This new technique furthered study of mammalian physiology especially where gross genetic knockouts were impossible due to embryonic lethality. Recently, several split Cre strategies have been reported to add tighter temporal and spatial control to Cre expression in mammalian cells (*121–123*). In this way Cre remains inert until specific conditions are met.

Despite discovery in bacteria, little technology utilizing Cre has developed in *E. coli* outside alternative cloning techniques. We have extended its use in *E. coli* to develop a screening method for the directed evolution of interacting proteins. Coupling split Cre with a two-gene reporter plasmid encoding orthogonal phenotypes enables visualization of reconstituted recombinase activity. This method faithfully discriminates between dimerization strength of protein pairs and permits the isolation of competent binders based on phenotype. Unfortunately, this system as well proved incompatible with ubl modification.

***Materials and methods***

**Plasmid construction.** All plasmids were based on the pZ vector system developed by Lutz and Bujard (*82*). Oligonucleotide pairs were used to replace the multiple cloning site (MCS) between the restriction sites *Eco*RI and *Xba*I in the plasmids pZA22, pZA12, pZE11, and pZE12. The resulting vectors consisted of three pairs of restriction sites (*Kpn*I and *Sph*I, *Mlu*I and *Eag*I, and *Kas*I and *Cla*I) with each pair flanked by a strong RBS sequence (AAAGAGGAGAAA) and a frame-shifted stop codon sequence (TAATTGAATAGTTAA) to prevent translational read-through. The MCS to make pZE11-CB and pZE11-CA were generated in similar fashion with the restriction sites *Kpn*I, *Nsi*I, *Age*I, and *Apa*I replacing the first pair in pZE11-CB and *Afl*II, *BamH*I, *Pst*I, and *Eag*I the second in pZE11-CA.

To create the split Cre fusions, Cre was first PCR amplified from the P1 phage genome (laboratory stock) and ligated with pZE11. From this template Cre(59) and Cre(60) were PCR amplified and inserted into pZE11-CA and pZE11-CB respectively using *Pst*I and *Eag*I for Cre(59) and *Age*I and *Apa*I for Cre(60). An HA epitope tag sequence (5'-TACCCCTACGAC GTGCCCGACTACGCC-3') was appended 3' to Cre(60) during amplification. The F2 linker sequence (*121*) (5'- GCGTCTCCGTCTAACCCGGGTGCGTCTAACGGTTCT-3') was inserted using oligonucleotide pairs between *Bam*HI and *Pst*I for pZE11-Cre(59) and *Nsi*I and *Age*I for pZE11-Cre(60)-HA. JunLZ and FosLZ (*102*) were PCR amplified and inserted into pZE11-F2-Cre(59) using *Afl*II and *BamH*I and into pZE11-F2-Cre(60)-HA using *Kpn*I and *Nsi*I respectively. A FLAG epitope tag sequence (5'-GACTACAAGGACGATGACGACAAGGGA-3') was appended 5' to JunLZ during amplification. To make the fusions bicistronic, pZE11-FosLZ-F2-Cre(60)-HA was digested with *Kpn*I and *Nhe*I and ligated with pZE11-FLAG-JunLZ-F2-Cre(59). To reduce copy number, the SC101 origin of replication was digested out of pZS31 and replaced that of pZE11-FosLZ-F2-Cre(60)HA.FLAG-JunLZ-F2-Cre(59) using *Sac*I and *Avr*II. Construction of the variants FosLZ(L2V), FosLZ(L2/3V), and FosLZ(ΔZIP) was described previously (*102*). These genes were PCR amplified and replaced FosLZ. pZS11-ΔFosLZ-F2-Cre(60)-HA.FLAG-JunLZ-F2-Cre(59) was created by PCR amplifying Cre(60)HA and replacing FosLZ-F2-Cre(60)HA using *Kpn*I and *Apa*I.

To create the reporter, MBP was PCR amplified with an RBS appended 5' and an L5 linker (*124*) (5'-ACTAGTGCCGCGGCA-3') 3'. The product was inserted into pZA22 using the *EcoR*I and *Kpn*I restriction sites. The gene for mRFP was PCR amplified with the mutant *loxP* (*JT15*:*JTZ17*) sequence (*125*) (5'-AATTATTCGTATAGCATACATTATAGCAATTT ATCT-3') appended 5' and an StrpII epitope tag with double stop codon (5'-GCAAGCTGGAG

CCACCCGCAGTTCGAAAAGGGTGCATAATAA-3') 3'. The additional two nucleotides 3' of the modified *loxP* sequence shift mRFP into frame. GFPmut2 was PCR amplified with the same appendage 5' and a c-Myc epitope tag (5'-GAACAGAAACTGATCTCTGAAGAAGACC TG-3') 3'. These two products were double ligated with pZE12 using *Kpn*I and *Eag*I and using *Sph*I as the linking restriction site. This construct was then digested with *Kpn*I and *Cla*I and ligated into pZA22-MBP-L5. The origin of replication was then swapped with ColE1 from pZE12 using *Sac*I and *Avr*II.

To create the split Cre-ubiquitin system, Fen replaced Fos between *Kpn*I and *Nsi*I in pZE11-FosLZ-F2-Cre(60)-HA and the linker $L_{Bo}$ (5'- GCATGCTCTGCAAGGTTCTGCTG CTTCTGCTGCTGGTGCTGGTGAAGCGCATTCGCTGC-3') replaced F2 between *Nsi*I and *Age*I. $L_{Bo}$ had a type II restriction enzyme *BstAP*I pair contiguous to *Nsi*I and *Age*I to facilitate linker exchange. Oligonucleotide pairs were used to insert L5 (5'-GCTACCTCGGCAGCCGCC TCG-3'), L15a (5'-GCT ATGACCGCGACCGCAGATGTTCTAGCGATGTCG-3'), L15b (5'-GCT GGTGGTAGCGGTGGTTCTGGCGGGTCGGGTTCG-3'), L16 (5'GCT GGTTCCGCC GCGAGCGCTGCCGGCGCCGGTGAATCG-3'), L25 (5'-GCT GGTGGAGGTGGTTCTGG CGGAGGTGGCTCTGGTGGAGGTGGCTCTGGGGGTGGTGGCAGC TCG-3'), L37 (5'-GCTGCCGGCGAAGCAGCGGCGAAAGAAGCGGCGGCAAAAGAAGCCGCTGCTAAAG AGGCAGCAGCTAAGGAGGCAGCCGCGAAAGAAGCGGCGGCCAAATCG-3'), F1 (5'-GCTGCCTCCACTGGCGGTAGCTCG-3'), F2 (5'-GCTGCGTCTCCGTCTAACCCGGG TGCGTCTAACGGTTCTTCG-3'), F3 (5'-GCTGCGTCTGGTGGCGGTGGTTCCGGCGG TGGCTCGTCG-3'), F4 (5'-GCTGCTTCTGGTGGTGGTTCTGGCGGTGGCAGCGGTG GCGGCTCCTCG-3'), F5 (5'-GCTGCTTCTGGCGGCAGTGGTGGCGGTTCTGGTGGCGG CTCTGGTGGTGGTAGCTCG-3'), F6 (5'-GCTGCTTCTGGTGGCAGCGGTGGCGGCTC

TGGCGGTGGTAGTGGTGGTTCCGGCGGCTCAGGCGGCGGTTCCTCG-3'), H1 (5'-GCTGCCTCCGCGGAAGCTGCGGCAAAGGAAGCGGCAGCCAAAGCAGGCTCCTCG-3'), H2 (5'-GCTGCAAGTGGCGCTGAAGCTGCCGCGAAAGAAGCAGCCGCTAAAGCTGGTGGTGGCTCCTCG-3'), H3 (5'-GCTGCGTCGGCTGAAGCCGCTGCGAAAGAAGCTGCTGCTAAAGAAGCGGCGGCGAAAGCTGGTAGCTCG-3') between the *BstAP*I pair. From pZE11-Cre, Cre(59) was PCR amplified and inserted into pZE11-CA using *Afl*II and *BamH*I. The linker $L_{Ao}$ with a type II restriction enzyme *Dra*III pair replacing *BstAP*I in $L_{Bo}$ was inserted using oligionucleotide pairs between *BamH*I and *Pst*I. Ubiquitin-FLAG was PCR amplified and inserted between *Pst*I and *Eag*I. The aforementioned linkers were inserted between the *Dra*III pair. Ubiquitin replaced Ubiquitin-FLAG for pZE11-Cre(59)-L25-Ubiquitin-FLAG and pZE11-Cre(59)-L37-Ubiquitin-FLAG after linker optimization. pZE11-Fen-F4-Cre(60)-HA, pZE11-Fen-H2-Cre(60)-HA, pZE11-Fen-L25-Cre(60)-HA, and pZE11-Fen-L37-Cre(60)-HA were digested using *Kpn*1 and *Nhe*1 and the Fen-Cre(60) fragments inserted into pZE11-Cre(59)-L25-Ubiquitin and pZE11-Cre(59)-L37-Ubiquitin using the same restriction sites to make the fusions bicistronic. These new plasmids were then digested using *Kpn*1 and *Eag*1 and their cassette subcloned into pZA12 to make pZA12-Fen-Cre(60)-HA.Cre(59)-Ubiquitin. For the remainder of the cascade, pZE12-Ube1.UbcH5b was digested using *Kpn*1 and *Eag*1 and pZE11-AvrPtoB was digested using *Eag*1 and *Cla*1. The two fragments were ligated together with pZA31. This plasmid was then digested using *Nhe*1 and *Cla*1 and pZE12-Ube1.UbcH5b (*42*) using *Xho*1 and *Nhe*1. These two fragments were ligated together with pZS31 to make pZS32-Ube1.UbcH5b.AvrPtoB. Both reporter cassettes MBP-mRFP-GFP and MBP-GFP were PCR amplified from the aforementioned plasmids and digested with *BspH*1 and

*Eag*1. Theses were inserted between *Nco*1 and *Eag*1 of pET28a to make pET28a-MBP-mRFP-GFP and pET28a-MBP-GFP.

To create the split Cre-SUMO system, Fos was PCR amplified with PML (5'-CCGCGTAAAGTTATCAAAATGGAATCCGAA-3') and PML(K490R) (5'-CCGCGTAAAGTTATC**CGT**ATGGAATCCGAA-3') appended 5'. The products were digested with *Kpn*I and *Nsi*I and subcloned into pZE11-Fen-F2-Cre(60)-HA and pZE11-Fen-L37-Cre(60)-HA. SUMO was PCR amplified and subcloned into pZE11-Cre(59)-F2-Ubiquitin-FLAG and pZE11-Cre(59)-L25-Ubiquitin-FLAG replacing ubiquitin-FLAG using the *Pst*I and *Eag*I restriction sites. pZE11-PML-Fos-Cre(60)-HA and pZE11-PML(K490R)-Fos-Cre(60)-HA were digested with *Kpn*I and *Nhe*I and subcloned into pZE11-Cre(59)-SUMO. These plasmids were digested with *Kpn*I and *Eag*I and pZE11-PIASxβ (*42*) with *Eag*I and *Cla*I. These products were ligated together with pZA12 to make pZA12-PML-Fos-Cre(60)-HA.Cre(59)-SUMO.PIASxβ and pZA12-PML(K490R)-Fos-Cre(60)-HA.Cre(59)-SUMO.PIASxβ. For the remainder of the cascade, pZE12-Aos1.Uba2.Ubc9 and pZS32-Ube1.UbcH5b.AvrPtoB were digested with *Kpn*I and *Cla*I. The insert of the former was inserted with the backbone of the latter to make pZS32- Aos1.Uba2.Ubc9.

**Library construction.** Library members were variants of FosLZ made by mutating the leucine residue in the second and third heptad of the LZ motif. To introduce mutation, the oligonucleotides d(5'- CATGCTGACCGACACCCTGCAGGCGGAAACCGACCAG**NNB**G AAGACGAAAAATCCGCG**NNB**C-3') and d(5'- AAACCGAAATCGCGAACCTGCT GAAAGAAAAAGAAAAGCTGGAGTTCATCCTGGCGGCACACATGCA-3') were annealed to their complements and ligated into pZS11-FosLZ(L2/3V)-F2-Cre(60)HA.FLAG-JunLZ-F2-Cre(59) using *Kpn1* and *Nsi1*. A 5'-NNBC-3' overhang was used to link the two inserts. DH5α-

Z1 electrocompetent cells were transformed with the ligated material and grown overnight in 30 mL LB with 50 mg/mL ampicillin after a 1 h rest period in SOB. To determine library size, aliquots were taken prior to overnight growth and plated on LB agar plates with ampicillin. The overnight culture was miniprepped to collect the library.

**Flow cytometry of mutant FosLZ.** *E. coli* host strain DH5α-Z1 was co-transformed with plasmids containing the fluorescent reporter and split Cre fusions. Cells were plated on LB agar supplemented with 50 mg/mL ampicillin and 40 mg/mL kanamycin for 12 h. Individual colonies were grown another 12 h in LB media supplemented with antibiotics. This overnight culture was then subcultured to $OD_{600} \approx 0.05$ in 5 mL of fresh LB media supplemented with antibiotics. Cultures were induced at $OD_{600} \approx 0.75$ with 1 mM IPTG and 50 ng/mL anhydrotetracycline and subsequently shaken for 24 h at 25°C.

Fluorescence was analyzed using a BD-Biosciences FACS Aria flow cytometer. Singlets were gated based on light scattering signals. GFP and mRFP were excited at 488 nm by the argon laser and were collected in the FITC and PE channels respectively. Quadrant gates were set using empty cells, those expressing mRFP, and those expressing GFP.

Time course measurements for green fluorescence were taken 3, 6, 24, and 48 h post-induction for FosLZ and FosLZ(L2V). Excitation and emission wavelengths for GFP were set to 488 and 530 (FL-1 channel) respectively on the BD FACSCalibur.

**Library screening.** Samples were cultured as before with the exception of the library. Transformant was grown in liquid media directly, bypassing growth and selection on an LB agar plate. A cell sort on the BD-Biosciences FACS Aria flow cytometer collected cells from the GFP gate. Gates were established using the aforementioned controls. Of the 600 cells theoretically isolated, 281 colonies were obtained after plating on LB agar with antibiotics.

These were cultured overnight in 96 well plates (200 μL with antibiotics) and then subcultured with 50-fold dilution into the same. After 3 h wells were induced 24 h at 25°C with 1mM IPTG only in order to assess the state of the reporter without additional activity from the fragments. Cultures were spun down and resuspended in 200 μL PBS. Fluorescence was assayed in flat bottom, non-treated, full black 96 well plates (Costar 3915) using the SpectraMax Gemini. For GFP excitation and emission wavelengths of 488 nm and 530 nm were used respectively; for mRFP 570 nm and 610 nm were used respectively. Black with clear bottom 96 well plates (Costar 3631) were used to obtain $OD_{600}$ using the SpectraMax 190 to normalize fluorescence readings.

The top 51 variants were chosen for sequencing. Each 5 mL overnight culture was miniprepped, and the FosLZ-F2-Cre(60)HA fragment PCR amplified off the plasmid using the oligodeoxyribonucleotide d(5'-CTCAGTGGTACCATGCTGACCGACACCCTGCAG-3') and d(5'-CTCAGTGGGCCCATCGCCATCTTCCAGCAGG-3'). For sequencing, the oligodeoxyribonucleotide d(5'-GACGATGAAGCATGTTTAGCTG-3') was used.

***Results***

**Rationale.** Following Jullien *et al*, we split Cre between Asn59 and Asn60 to produce Cre(59) and Cre(60) fragments. This split site minimized the basal activity of the fragments in that study and allowed the interacting pair of proteins to sit above the Holliday-junction plane (*121*). We chose the basic region-leucine zipper (bZIP) domains of c-Jun (JunLZ, residues 280-318) and c-Fos (FosLZ, residues 161-200) as our model interacting pair. They are a well-characterized, heterodimeric protein-interacting pair with a series of well-characterized mutants with reduced affinity (*102*, *126*, *127*). To Cre(59) JunLZ was attached to the N-terminus using the F2 linker (*121*) and to Cre(60) FosLZ was attached to the N-terminus using the F2 linker as well (Fig. 4.1).

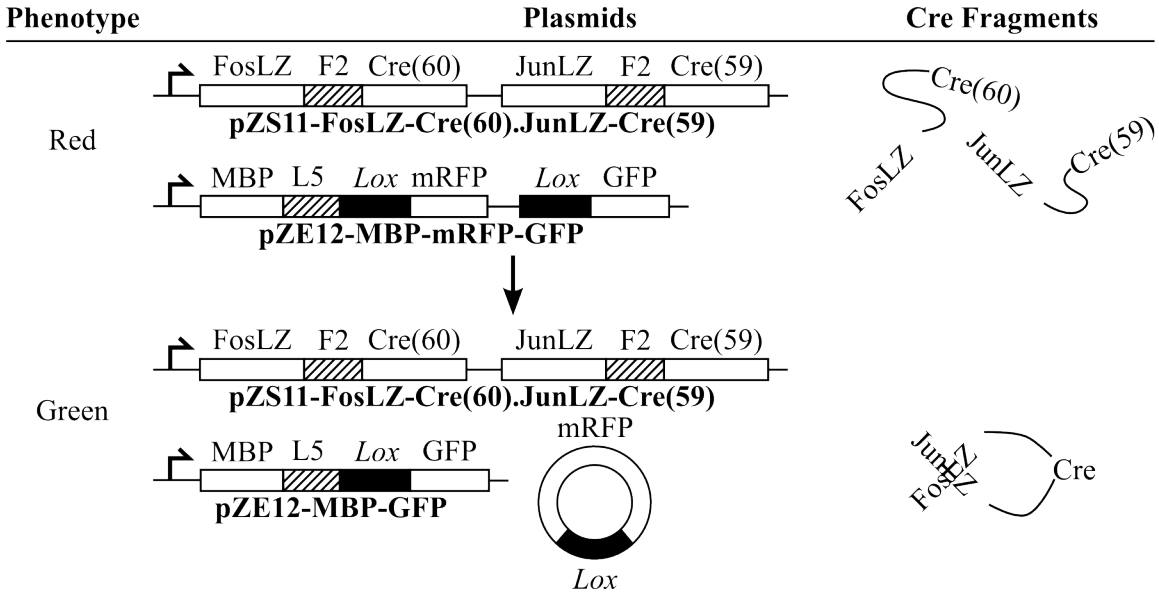


**Figure 4.1. Plasmid design and rationale.** Fusion constructs FosLZ-Cre(60) (as well as FosLZ mutants) and JunLZ-Cre(59) are bicistronic in the pZS11 backbone. The reporter construct MBP-mRFP-GFP, with mRFP terminating in two stop codons (TAATAA) flanked by two *lox* sites, lies in a pZE22 backbone. Cells are transformed with these two plasmids. Without Cre reassembly, only mRFP will express from the reporter plasmid and cells will fluoresce red. With reassembly, Cre will excise DNA flanked by *lox* sites (mRFP with stop codons) silencing mRFP expression while permitting GFP expression. Green fluorescence will then evolve.

Because the fusion pairs proved rather active (data not shown), they were placed on a low copy (SC101 origin) plasmid to eliminate activity from uninduced expression.

To ascertain dimerization in a high-throughput screen, we devised a fluorescence-based reporter (Fig. 4.1). Without reconstituted Cre, the *mrfp* gene is transcribed and cells exhibit red fluorescence. The *gfp* gene is not translated due to two stop codons (TAATAA) terminating *mrfp* and no RBS preceding *gfp*. When the interacting proteins associate, full-length Cre assembles and excises *mrfp* and its stop codons from the reporter as *lox* sites flank them. The *gfp* gene supersedes *mrfp* and is now translated causing cells to fluoresce green.

To build the reporter, *mrfp* terminating in two stop codons was flanked with the mutant *loxP* site *JT15:JTZ17* (*125*) and followed with in-frame *gfp*. The mutations to the *loxP* site lower its affinity for Cre thus reducing the integrase activity. The wild-type *loxP* sites proved

too facile to excision (data not shown). MBP was attached to the N-terminus of mRFP using an L5 linker (*124*) to improve solubility of the fluorescent proteins (data not shown). The reporter was placed on a high copy (ColE1 origin) plasmid to maximize fluorescence and widen the red/green continuum. An IPTG-inducible promoter controlled the expression of the fluorescent proteins and an anhydrotetracycline-inducible promoter controlled the production of the split Cre fusions to allow orthogonal control (*82*).

**Fos leucine mutations shift cellular fluorescence.** Given the mechanics of our system, protein pairs better able to dimerize either through higher affinity and/or stability should shift cellular fluorescence green. To test this hypothesis, we chose JunLZ and FosLZ as our model interacting pair. Previous work has shown that mutation to the d position of the bZIP domain heptad reduces affinity between the pair (*102*, *126*, *127*). We used FACS to analyze both the red and green fluorescence of singlet cells having JunLZ paired with either FosLZ, FosLZ(L2V), or FosLZ(L2/3V), whose affinity for JunLZ decrease in the order given (*102*, *126*, *127*).

For the wild-type pair, 22.1% fell in the GFP only quadrant (GFP+, mRFP-) while 26.7% were in the mRFP only quadrant (GFP-, mRFP+). A small population, 7.64%, showed no fluorescence (GFP-, mRFP-), and the majority, 43.6%, showed a mixture of both red and green fluorescence (GFP+, mRFP+) (Fig. 4.2a). For the single mutation, the GFP and mixture populations reduced to 6.27% and 33.3% respectively while the mRFP population increased to 53.9% (Fig. 4.2b). This shift arose from the mutation impairing dimerization and subsequently reducing Cre's activity, which leaves the cells more red and less green. While the difference is clearly resolvable, the degree of overlap with wild-type's spectrum reflects the minimal impact the single leucine mutation has on dimerization (*126*, *127*). For the double mutant

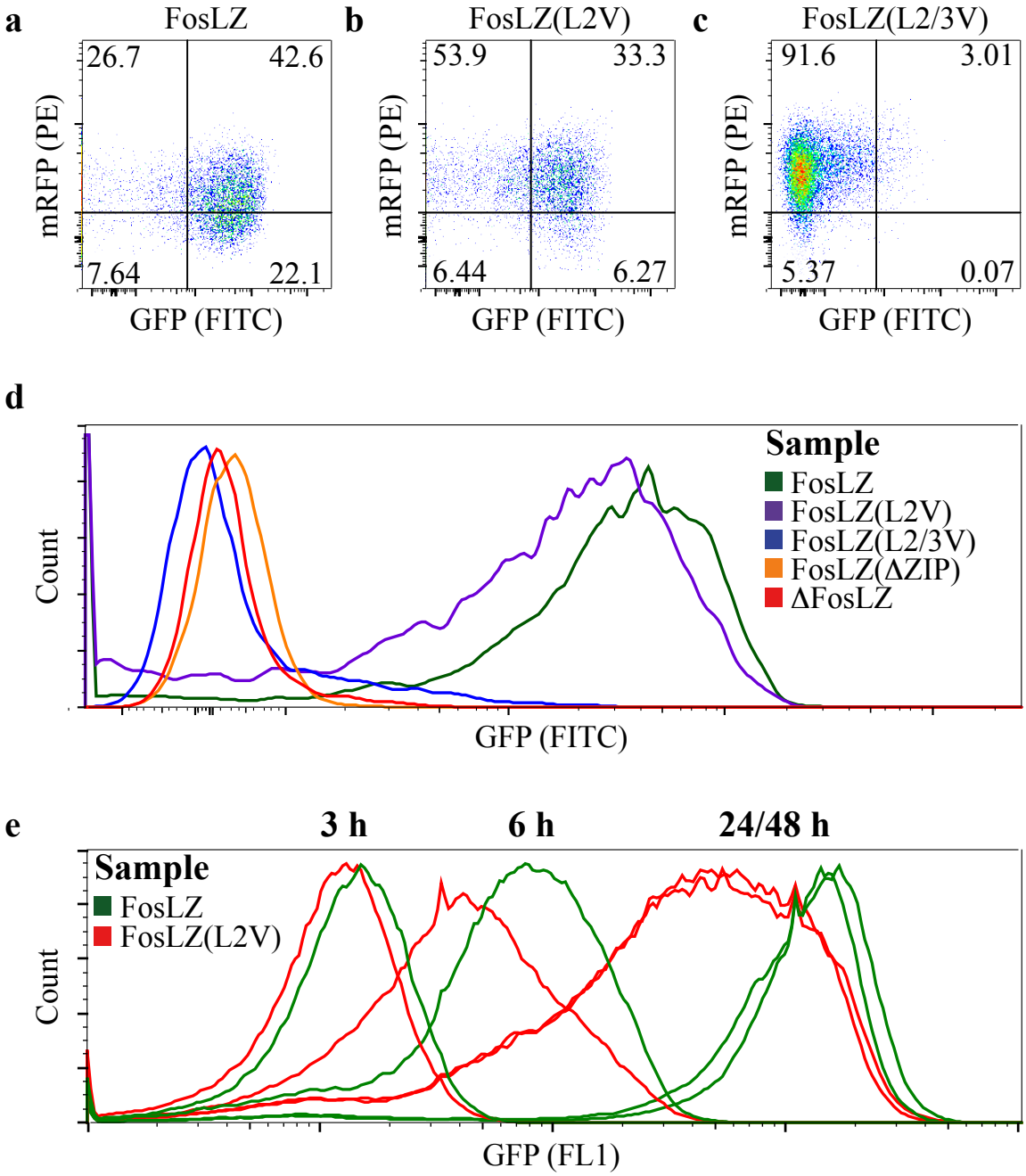


**Figure 4.2. FACS plots for FosLZ mutants.** The mRFP (PE) channel (red fluorescence) was plotted against the GFP (FITC) channel (green fluorescence). Quadrant gates were established using non-expressing cells (GFP-, mRFP-), those expressing mRFP (GFP-, mRFP+), and those expressing GFP (GFP+, mRFP-). Cell populations shifted from green to red with increasing mutation to FosLZ (FosLV (a), FosLZ(L2V) (b), FosLZ(L2/3V) (c)). FITC data alone is shown with additional truncation controls (d). Time course measurements for green fluorescence were taken 3, 6, 24, and 48 h post-induction for FosLZ and FosLZ(L2V) (e). Differences in signals emerge by 6 h and persist till at least 48 h.

FosLZ(L2/3V), only 0.07% was present in the GFP population with nearly the entirety, 91.6%, found in the mRFP population (Fig. 4.2c). The shift from green to red mirrors the notable compromise the double leucine mutation has on dimerization (*126*, *127*). Thus, we conclude that fluorescence correlates with the ability of the pairs to dimerize: better dimers report greener while compromised dimers report redder.

The two-state reporter allows for greater discrimination over a single-gene reporter. Considering all GFP+ populations together (GFP+, mRFP+ and GFP+, mRFP-), the wild-type had a 1.7 and 21 fold greater population over FosLZ(L2V) and FosLZ(L2/3V) versus 3.5 and 320 when the GFP+, mRFP- population was considered alone (Fig. 4.2a,b, and c).

Excision of *mrfp* from the reporter is essentially irreversible. We monitored green fluorescence over time between FosLZ and FosLZ(L2V) to ensure distinction between pairs differing in dimer formation would not be lost given time. The signals between FosLZ and FosLZ(L2V) became distinguishable by 6 hours post-induction and remained so even after 48 hours (Fig. 4.2d). We chose 24 hours for our analyses as this time point corresponded to a maximum in green fluorescence.

**Screening for green fluorescence enriches for competent binders.** As differences in dimerization corresponded to differences in fluorescence, we next asked if by screening based on fluorescence could we isolate proclive dimerizing pairs. We chose to randomize the second and third leucine of FosLZ since the FosLZ(L2/3V) double mutant considerably compromised dimerization. We created an NNB library of approximately $10^4$ variants. Cells were sorted on the BD-Biosciences FACS Aria flow cytometer and isolated from the GFP gate (not shown) established using cells only expressing GFP. Isolates were cultured in 96 well plates and

**Table 4.1. Characterization of isolated FosLZ variants.** GFP fluorescence was normalized to that for the control GFP. Those with green fluorescence 60% and above were sequenced. All sequences observed (reported as concatenated residues occupying the 2nd and 3rd leucine position) were listed in the table with frequency given as a percentage and as a gross count in parentheses.

| Amino Acids | Frequency Overall, % | Frequency 90-100%, % | Frequency 80-90%, % | Frequency 70-80%, % | Frequency 60-70%, % |
|---|---|---|---|---|---|
| LL | 29 (15) | 50 (1) | 23 (3) | 50 (7) | 18 (4) |
| AL | 16 (8) | 0 (0) | 15 (2) | 21 (3) | 14 (3) |
| LR | 14 (7) | 50 (1) | 23 (3) | 7 (1) | 9 (2) |
| LM | 6 (3) | 0 (0) | 8 (1) | 0 (0) | 9 (2) |
| LI | 6 (3) | 0 (0) | 15 (2) | 0 (0) | 5 (1) |
| FL | 4 (2) | 0 (0) | 0 (0) | 0 (0) | 9 (2) |
| LF | 4 (2) | 0 (0) | 0 (0) | 7 (1) | 5 (1) |
| MF | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |
| HL | 2 (1) | 0 (0) | 0 (0) | 7 (1) | 0 (0) |
| LK | 2 (1) | 0 (0) | 8 (1) | 0 (0) | 0 (0) |
| LQ | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |
| VL | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |
| RL | 2 (1) | 0 (0) | 0 (0) | 7 (1) | 0 (0) |
| AW | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |
| LH | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |
| LA | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |
| QL | 2 (1) | 0 (0) | 8 (1) | 0 (0) | 0 (0) |
| LC | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 5 (1) |

induced with 1mM IPTG only in order to assess the state of the reporter without additional activity from the fragments. Fluorescence was assayed using SpectraMax Gemini and normalized by $OD_{600}$.

After one round of screening, we sequenced 51 clones corresponding to the top 40% for green fluorescence (Table 4.1). In 96% of cases at least one leucine was present. This strong enrichment agrees with our previous findings as well as previous research, which has shown bZIP pairs (*128–131*) including Fos and Jun specifically (*126*, *127*) to be mostly insensitive to single point mutation but sensitive to double mutation. The variant most enriched for was the

strongest leucine zipper, that of the wild-type sequence LL (29%) (*131*). Following were AL(16%) and LR(14%) with several other variants appearing sparingly. The context of specific mutations is important (*131*), which may explain biases such as LR (14%) over RL (2%). A previous study using an NNK library to randomize the second leucine of FosLZ isolated variants (L, A, F, R, H) in the same rank order that we did save Q (*102*). Based on these consistencies, we conclude that we successfully enriched for higher affinity binders by screening on fluorescence.

**Split Cre proved nonviable with both ubiquitin and SUMO pathways.** We were next interested if we could adapt our new screen to our ubiquitination system. After investigating all possible permutations (data not shown), we placed the Cre(60) fragment on the C-terminus of Fen and the Cre(59) fragment on the N-terminus of ubiquitin. This particular orientation should situate ubiquitinated Fen above the Holliday-junction plane. We chose linkers from a battery of 15 that maximized expression of the fragments as ascertained by Western blot (data not shown). Because we did not know what length linker would be ideal for activity, we chose F4, H2 (*121*), L25, and L37 (*124*) as they gave the best expression for Fen-Cre(60) and represented a range of linkers from helical to extended and from short to long. We chose L25 and L37 for ubiquitin as linker choice did not affect expression and longer linkers would better accommodate steric concerns. These fragments were placed in the pZA12 plasmid. The E1, E2, and E3 were placed on a pZS32 plasmid. The plasmid for the reporter plasmid was changed to pET28a (Fig. 4.3a). All promoters were LacI-repressible and pET28a provided the LacI necessary for repression.

Expression was further optimized with respect to nutrient media, aeration conditions, and bacterial strain. Of the 5 other media formulations tested, none surpassed LB (data not shown),

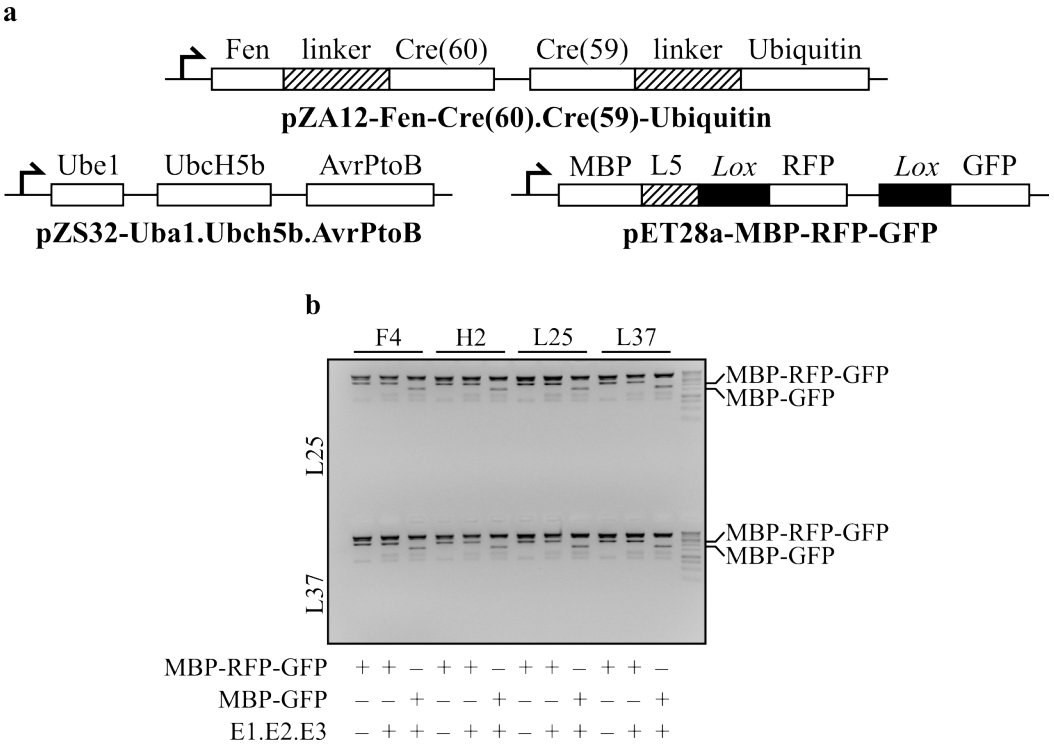


**Figure 4.3. Split Cre adapted to ubiquitin pathway.** (a) Plasmid diagrams for implementing split Cre with the Fen-ubiquitin cascade. The target protein (Fen) fused to Cre(60) and ubiquitin fused to Cre(59) were subcloned into the medium copy plasmid pZA12-SMCS. The E1 (Uba1), E2 (Ubch5b), and E3 (AvrPtoB) were subcloned into the low copy plasmid pZS32-SMCS. The reporter cassette MBP-mRFP-GFP was subcloned into the high expression pET28a backbone. (b) Digestion pattern of plasmid isolated from DH5α cells expressing split Cre implemented with the Fen-ubiquitin cascade. Plasmid was digested using the restriction enzymes *Xba*1 and *Eag*1 to isolate the reporter cassette. Various combinations of linkers were used in assessing activity of the fragments.

and of the two other aeration conditions, all performed equally to that in test tubes (data not shown). Of the 8 strains tested, DH5α and BLR(DE3) gave notably better accumulation (data not shown). Both of these strains lack RecA, which is believed to assist plasmid stability. Surprisingly, DH5α-Z1 ranked among the lowest for expression.

We tested activity in DH5α with the aforementioned linker combinations. We chose not to use BLR(DE3) as expression even from empty pET vectors can prove toxic (*132*), and we have noted earlier that the full ubiquitin cascade seemed taxing for cells. Because DH5α lacks a

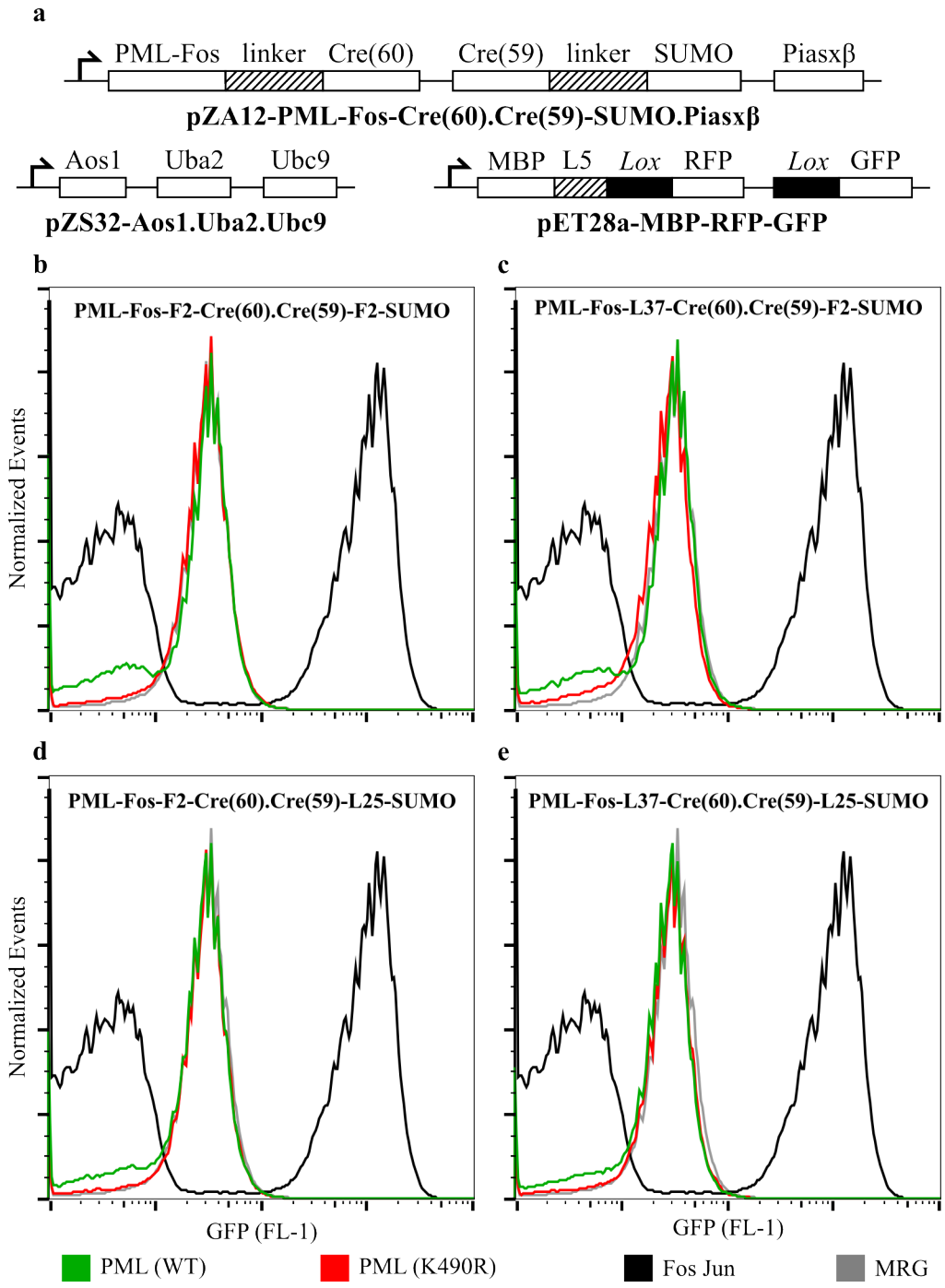


**Figure 4.4. Split Cre adapted to SUMO pathway.** (a) Plasmid diagrams for implementing split Cre with the SUMOylation cascade. The target protein (PML-Fos) fused to Cre(60), SUMO fused to Cre(59), and E3 (PIASxβ) were subcloned into the medium copy plasmid pZA12-SMCS. The E1 (Aos1 and Uba2) and E2 (Ubc9) were subcloned into the low copy plasmid pZS32-SMCS. The reporter cassette MBP-mRFP-GFP was subcloned into the high expression pET28a backbone. (b, c, d, and e) The GFP (FL-1) channel showed no difference between PML-Fos and PML(K490)-Fos. The mutation in the latter prevents SUMOylation of the target.

DE3 cassette, we assayed for activity by examining the plasmid directly for excision (Fig. 4.3b). We purified plasmid from each sample and digested with restriction enzymes *Xba*1 and *Eag*1 as these restriction sites flanked the fluorescent reporter. A 2.67 kbp would denote the full reporter while a 1.95 kbp band would denote it lacking *mrfp*. Only in the presence of the full cascade should ubiquitination occur and Cre activity be reconstituted. The activity of Cre should excise *mrfp* from the reporter plasmid and shift the band size from 2.67 kbp to 1.95 kbp. For each linker combination, only the heavier band appeared, implying no excision occurred. Select combinations were also tested in BLR(DE3), but neither fluorescence nor plasmid digestion pattern indicated Cre activity (data not shown).

Because this particular ubiquitination system might be incompatible with the split-Cre screen, we also adapted it to the SUMOylation system we previously developed (Fig. 4.4a). As before, the pET28 plasmid bore the fluorescent reporter. SUMO replaced ubiquitin and PML-Fos replaced Fen. Only the F2 and L37 linkers were used for PML-Fos-Cre(60) and F2 and L25 for Cre(59)-SUMO. The E3 PIASxβ was included with PML-Fos-Cre(60) and Cre(59)-SUMO as the pZS32 plasmid was fully occupied with the E2 and hetero-dimeric E1. These plasmids were expressed in BLR(DE3). However, no green fluorescence emerged and spectrum appeared similar to unmodified PML-Fos-Cre(60) (Fig. 4.4b, c, d, and e).

***Discussion***

Cre recombinase has been a powerful molecular tool in eukaryotic study for the past score years. We developed a novel bacterial screening method for the directed evolution of interacting proteins by extending its innovations to *E. coli*. This method reported differences in dimerization of protein pairs. FosLZ and the mutants FosLZ(L2V) and FosLZ(L2/3V) had distinguishable spectra on FACS. These differences allowed for enrichment for competent

binders based on phenotype. Of the isolated variants of FosLZ(L2/3V), 29% were wild-type and 96% recovered a leucine residue for at least one position after only a single round of selection.

The two-state design of the reporter proved an asset in better resolving differences between interacting pairs. The signal may be decomposed into two separate phenotypic states, which provides greater resolution in determining Cre's activity. The reporter plasmid itself cumulatively records Cre's activity, and read-out may be decoupled from the assay. We used this to access information concerning the assay after the FACS sort. It could be of additional value for instance in alleviating cellular stress from expression of multiple proteins. While fluorescent proteins were used in this study for visualization, other reporters may substitute. For example, an antibiotic resistance marker could replace GFP. For additional stringency, a gene encoding a toxic protein could stand for mRFP. Thus, cells with reconstituted Cre will have a growth advantage while those without will have a disadvantage.

Unfortunately, applying split Cre to both our ubiquitin and SUMO pathways proved nonviable. We believed Cre addressed the shortcomings of our previous reporter proteins. Cre's enzymatic activity coupled with multiple translation events of the genetic reporter provides great sensitivity to interacting protein pairs. Expression of the model fusions from a several-copy plasmid (SC101 origin) was sufficient for green fluorescenct phenotype. Furthermore, Western blotting could not reveal protein expression at these conditions (data not shown). This is significant since other PCAs utilizing leucine zippers rely upon strong over-expression (*133–135*). The split proteins may have failed to report ubl modification for several reasons. The interacting domains are thought to increase the local proximity of their split reporter fusions to permit folding (*136*). These interactions unlike ubl modification are not covalent, and folding and/or activity may benefit from association and dissociation of the interacting pair to overcome

steric hindrance. As a corollary, our choice of target and ubl may have posed a steric hinderance to reporter folding. All systems suffered proteosomal degradation and varied in accumulation of modified target as assessed by Western blot. Either ubl-target complexes lacked sufficient residence time for activity, accumulation was not sufficient for detectable activity, or what appeared as modified target suffered from indiscernible proteolysis that rendered the complex nonfunctional. Specific systems may also have had specific challenges. For example, ubiquitinated Fen may have interfered with homologous recombination at the Holliday-junction, and split DHFR may have suffered growth limitations from expression of recombinant protein (*137*). We next looked towards simpler systems where steric hindrance would be a nonissue and whole-protein fusions could be utilized to possibly avoid degradation.

**Acknowledgements**

Cytometry core supported in part by the ESSCF, NYS-DOH, Contract #123456. Opinions expressed are solely of the author; they do not necessarily reflect the view of ESSCB, the NYS-DOH, or NYS.

# CHAPTER 5

# *IN VITRO* REPORTERS FOR UBIQUITIN-LIKE MODIFICATION

## *Introduction*

For *in vivo* systems, a reporter mediates the detection of protein-protein interactions. This method is indirect as one must assume reporter activity correlates with interaction strength. Direct detection of protein-protein interactions may be accomplished using *in vitro* methods. Here, proteins may be isolated using their interacting partner. However, this method requires a phentotype-genotype linkage that was implicit for *in vivo* methods. Furthermore, the fidelity of the linkage is paramount in ensuring mismatching does not occur.

Plasmid display is one such *in vitro* method that accomplishes the linkage using a DNA binding protein. It begins with an *in vivo* phase where protein fusions to DNA binding domains are express in the cytoplasm. Compartmentalized as such, the DNA binding domains may associate with their recognition sequences on the encoding plasmid. Cells are subsequently lysed and desirable variants are isolated through *in vitro* panning. Their genetic information is simultaneously recovered on account of the phentotype-genotype linkage.

Cull *et al* developed the first plasmid display system with the *lac* repressor (LacI) serving as the DNA binding protein (*138*). It was later used to develop a biotin modified epitope tag enriched for only after two rounds of panning (*139*). Two more systems have since been developed using the eukaryotic nuclear factor κβ (NF-κβ) p50 (*140*, *141*) and the GAL4 DNA binding domain (*142*). Though not characterized as a screen, Woodgate *et al* showed that plasmid purification was possible through GST using a GST-zinc-finger fusion (*143*). To date only epitope sequences have been evolved employing the systems.

We believed the simplicity of the plasmid display system would skirt the issues we encountered with complexity in the *in vivo* systems. Only one fusion rather than two needs to be made, and that is to a full-length protein rather than a fragment. Additionally, steric considerations are much less important. Since antibodies may be utilized during *in vitro* panning, those specific to different ubl chain conformations such as for ubiquitin (*144*) may be used to further the utility of our reporter. This was not possible in the *in vivo* systems.

Appropriate choice of a DNA binding domain forms the crux of plasmid display. A long-lived interaction safeguards the fidelity of the phenotype-genotype linkage during selection processes. Furthermore, solubility and binding must prove stable in fusion. This linkage is paramount as the plasmid DNA will be the readout from the reporter. For example, if a battery of targets is screened for modification, the isolated plasmid will contain the genes of those substrates modified. If plasmid dissociates from the DNA binding domain, it cannot be isolated or may reshuffle thereby giving unreliable information. We developed two systems: one using LacI for its exceptionally long binding half-life (*145*, *146*) and one using a zinc-finger protein for its simplicity.

***Materials and methods***

**Plasmid construction.** To create the dual promoter plasmid pZE12T11, oligonucleotide pairs were used to introduce a T1 terminator of the *rrn*B operon (*82*) between *Sal*I and *Kas*I of pZE12-SMCS (*42*) to prevent transcriptional read-through. Between *Kas*I and *Cla*I oligonucleotide pairs were used to introduce $P_{LtetO-1}$ (*82*) and a strong RBS sequence (AAAGAGGAGAAA) flanked with *Nde*I and *Afl*II.

To create the ZF-based plasmid display system, Fen and ZF were PCR amplified and digested with *Afl*II and *Age*I and *Age*I and *Cla*I respectively. These were subcloned together into

pZE12T11 between *Afl*II and *Cla*I. pZA24-AvrPtoB and pZE11-Fen.Ub were digested with *Kpn*I and *Nhe*I and *Nhe*I and *Sal*I respectively, and AvrPtoB and ubiquitin were subcloned together into pZE12T11-Fen-ZF. Oligonucleotide pairs were used to correct the truncation caused by the internal *Sal*I site to ubiquitin. An oligonucleotide pair was used to introduce a FLAG epitope tag (5'-GACTACAAGGACGATGACGACAAGGGA-3') between Fen and ZF using *Age*I and *Not*I in pZE12T11-AvrPtoB.Ubiquitin/Fen-ZF and pZE12T11-Fen-ZF to serve as a linker and means of detection. A section of *ampR* was PCR amplified and subcloned into pZE12-SMCS (*42*) using *Aat*II and *Spe*I to introduce a *Hind*III restriction site near *Aat*II. Using oligonucleotide pairs, 5xZB (5'-AC**GTCGATGCC**ATTAC**GTCGATGCC**ATTAC**GTC GATGCC**ATTAC**GTCGATGCC**ATTAC**GTCGATGCC**AT -3') were introduced between the *Aat*II and *Hind*III restriction sites. This was digested with *Kpn*I and *Cla*I and ligated with the cassette from pZE12T11-AvrPtoB.Ubiquitin/Fen-FLAG-ZF and pZE12T11-Fen-FLAG-ZF to make 5xZBDpZE12T11-AvrPtoB.Ubiquitin/Fen-FLAG-ZF and 5xZBDpZE12T11-Fen-FLAG-ZF. Finally, ubiquitin was PCR amplified with a FLAG epitope tag appended to its N-terminus and subcloned into pZE12T11 using *Mlu*I and *Eag*I. This was digested with *Nhe*I and *Afl*II and subcloned into 5xZBDpZE12T11-AvrPtoB.Ubiquitin/Fen-FLAG-ZF. The HA epitope tag (5'-TACCCCTACGACGTGCCCGACTACGCC-3') replaced the FLAG epitope tag between Fen and ZF using an oligonucleotide pair.

To create the LacI-based plasmid display system, oligonucleotide pairs were used to permute the MCS of pZE12-SMC (*42*). The three pairs of restriction sites of the resulting vector, pZE12-SMCS2 became *Mlu*I and *Eag*I, *Kas*I and *Cla*I, and *Kpn*I and *Sph*I. The $P_{BAD}$ promoter of pBAD18 was PCR amplified (5412 and 5413) and subcloned into pZE12T11 using *Kas*I and *Nde*I to make pZE12T10 where the 0 denotes the $P_{BAD}$ promoter.

Fen-HA was PCR amplified and subcloned into pZE12-SMCS2 using *Kpn*I and *Sph*I. LacI and ubiquitin were PCR amplified and digested with *Eco*RI and *Bam*HI and *Bam*HI and *Eag*I respectively. To ubiquitin a FLAG-*Pst*I epitope tag was appended to the N-terminus and to LacI an RBS-*Mlu*I was appended to the C-terminus. The two products were ligated together with pZE12-SMCS digested with *Eco*RI and *Eag*I. AvrPtoB was PCR amplified and subcloned into pZE12T10 using *Afl*II and *Cla*I. pZE12-LacI-FLAG-Ub was digested with *Eco*RI and *Eag*I; pZE12T10-AvrPtoB was digested with *Eag*I and *Cla*I; and pZE12-Fen-HA was digested with *Cla*I and *Xba*I. These products were ligated together with pZE12-SMCS digested with *Eco*RI and *Cla*I. Finally, the $P_{lac/ara-1}$ promoter was cut from pZA24-SMCS (*42*) using *Aat*II and *Eco*RI and ligated with the previou plasmid digested with the same to make pZE14T10-LacI-Ub/AvrPtoB.Fen.

For the GFP-LacI circuits, GFP (lab stock) was PCR amplified with the FLAG epitope tag appended to the C-terminus. This product was digested with *Kpn*I and *Sph*I and subcloned into pZE12-SMCS (*42*). This plasmid was digested with *Kpn*I and *Eag*I and subcloned into pZE12T11. This was then digested with *Kpn*I and *Afl*II, and LacI and LacI-FLAG-Ubiquitin were PCR amplified and digested with *Afl*II and *Cla*I. These were ligated with an unshifted pZE14T10 that had been digested with *Kpn*I and *Cla*I. To correct the truncation of LacI at the N-terminus, LacI was PCR amplified and digested with *Afl*II and *Cla*I. It was also PCR amplified with a FLAG epitope tag and digested with *Afl*II and *Pst*I. These were ligated with pZE14T11-GFP-FLAG/LacI and pZE14T11-GFP-FLAG/LacI-FLAG-Ubiquitin to make the same without the truncation.

**Cell Culture and Western Blots.** All constructs were transformed into *E. coli* host strain DH5α for LacI and DH5α-Z1 (*82*) for ZF using a GenePulser Xcell (BioRad). Individual colonies were

grown overnight in LB media with appropriate antibiotics (100 mg/mL ampicillin and 12.5 mg/mL chloramphenicol) and then subcultured to $OD_{600} \approx 0.05$ in 5 mL of fresh LB media supplemented with appropriate antibiotics. Cultures were induced at $OD_{600} \approx 0.75$ with 0.5% L(+)-arabinose, 1 mM IPTG, and 50 ng/mL anhydrotetracycline when appropriate and subsequently shaken for 24 h at 16°C for ZF or 25°C for LacI.

Approximately 1.5 mL of each culture was harvested and lysed using 200 mL of Bugbuster Master Mix (Novagen) according to the manufacturer's directions. Lysates were normalized to 10 μg of total protein as determined by a total protein assay (Bio-Rad) and loaded on a 4-20% Precise Protein Gel (Thermo Scientific). Transfers to Immobilon P Transer Membranes (Millipore) were performed for 2 h at the maximum amperage recommended for a Biosciences TE77 semi-dry transfer unit (Amersham). Blots were then imaged on film using standard protocols. The primary antibodies used were anti-HA (Sigma) and anti-FLAG (Abcam).

**FACS.** Cultures were prepared as described above except the same overnight culture was used for the induced and uninduced cases for each construct. Fluorescence was analyzed using a BD FACSCalibur. Excitation and emission wavelengths for GFP were set to 488 and 530 (FL-1 channel) respectively.

**Plasmid isolation.** The procedure for cell lysis was adopted from Sander *et al* (Sander08). Cultures were prepared as described above. An equal number of cells from each culture were harvested by centrifugation at 4000 g for 10 min at 4ºC. The supernatant was removed and pellet frozen overnight at -20 ºC. The next day cells were resuspended to $OD_{600} = 25$ with the appropriate amount of WB1/0.1% Nonidet$^{TM}$ P40 (NP-40) (Sander08) and frozen at -80ºC for 1-2 h. Samples were thawed on ice and centrifuged at 9000 g for 20 min at 4 ºC. The supernatant was retained. While lysis was inefficient, the use of more efficient methods either interfered

with later steps directly or made the supernatant too viscous with genomic DNA (data not shown).

Dynabeads® (Invitrogen 14311D) were prepared according to the manufacturer's instructions with 1.5 mg beads used per culture and 5 μg antibodies used per 1 mg of beads. The antibodies used were anti-FLAG (Genscript A00187) and anti-ubiquitin (Abcam ab8134 and ab411). After conjugating the antibodies to the beads, the beads were washed three times in 800 μL WB1/0.1% (NP-40) and resuspended in the same to a final concentration of 10 mg/mL. Aliquots of 1.5 mg of beads were divided among 1.7 mL eppendorf tubes and separated from the liquid phase through magnetism. Beads were resuspended in the supernatant from the lysate and incubated on a roller at 4ºC for 1 h. Three washes of 700 μL WB1/0.1% (NP-40) were performed and the beads again separated from the lysate through magnetism.

Plasmid was purified from beads using QIAprep Spin Miniprep kit (QIAGEN 27106) according to manufacturer's instructions with the following modifications: the P1 buffer was used without RNAse, the wash step with PB was omitted, and 100 μL ddH2O was used for elution. The eluate was evaporated off in a dessicator and precipitate resuspended in 5 μL EB. This was then diluted and used to transform new DH5α−Z1 electrocompetent cells. These cells were plated and colonies counted.

***Results***

**Rationale.** LacI forms a homo-tetramer containing two DNA binding domains that recognize DNA sequences named *lac* operators. The repressor-DNA complex can be extremely long lived. Complexes with a single operator have a half-life of 1.5 min (*147*). However, LacI is capable of forming loops in the DNA by contacting two appropriately spaced operators (*148*) increasing the half-life to approximately 28 h (*146*). Furthermore, a symmetric *lac* operator binds 8-10 times

more tightly to LacI than the native operator (*147*, *149*) and the aforementioned values rise to roughly 10 min (*145*, *147*) and greater than 3 days respectively (*145*). LacI also seems to tolerate fusions without compromise to its ability to bind DNA. Fusions of β-galactosidase to the C-terminus (*150*) and GFP to the N-terminus (*151*) of LacI have been made. However, addendum of a 13-residue peptide to the C-terminus did lead to 10 fold more expression of a reporter gene compared to the unmodified LacI (*138*). Fusions may therefore interfere with LacI's ability to bind DNA.

Because LacI is a fairly large molecule (39 kDa) and would present a considerably undesirable surface additional to the target against which we would be evolving the ubiquitin E3 ligase, we chose to fuse LacI to the N-terminus of ubiquitin (Fig. 5.1a). The LacI-ubiquitin fusion was placed under control of a LacI repressible promoter to self-regulate its expression (*140*) in order to achieve optimum accumulation. If too little repressor is present, unoccupied sites might be available to repressor from other cells following lysis, which could potentially undermine the phenotype-genotype linkage we hoped to create. If too much repressor is present, single operator occupancy becomes favored over loop formation (*148*), and the half-life of DNA binding diminishes as noted above.

The E3 ubiquitin ligase AvrPtoB and target Fen were placed under control of an orthogonal promoter on the same plasmid as the LacI-ubiquitin fusion. In this way we could modulate expression of AvrPtoB and Fen independently of LacI-ubiquitin while still maintaining their presence on the plasmid to which LacI-ubiquitin would bind. A HA epitope tag was appended to the C-terminus of Fen to pan against. A ColE1 origin of replication was used to maximize potential plasmids to isolate.

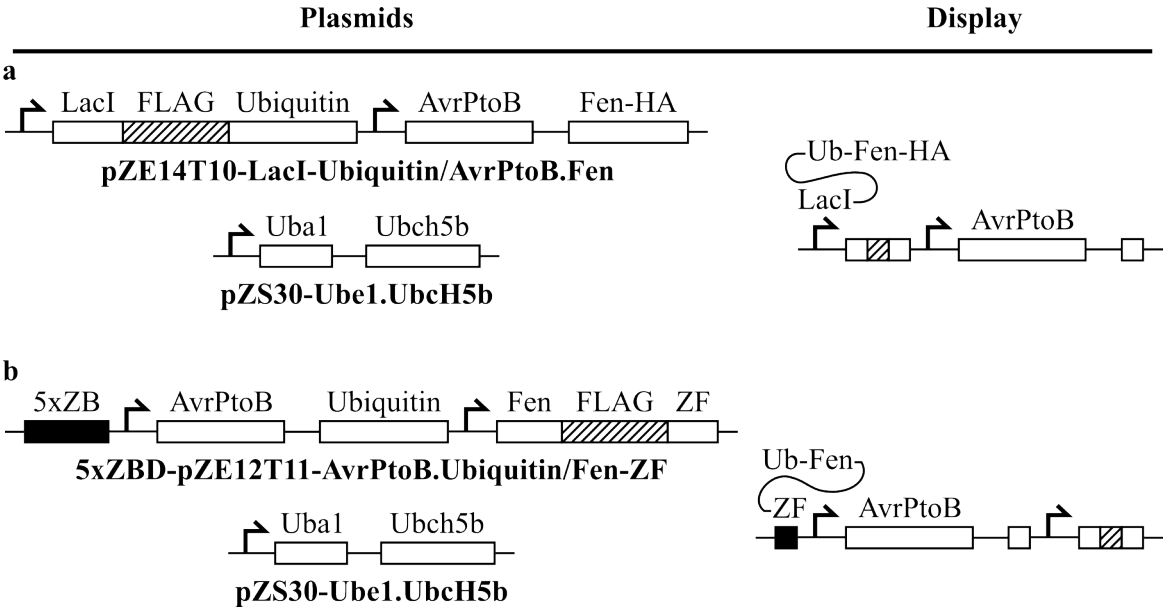


**Figure 5.1. Plasmid design and rationale.** Plasmid diagrams for the LacI (a) and ZF (b) plasmid display systems with the ubiquitination cascade. (a) The target protein (Fen) and E3 (AvrPtoB) were subcloned into the high copy plasmid pZE14T10 under control of the second promoter. The LacI-ubiquitin fusion was subcloned into the same under control of the first promoter. The E1 (Ube1) and E2 (UbcH5b) were subcloned into the low copy plasmid pZS30-SMCS. LacI binds to its promoter to establish the phenotype-genotype linkage. The HA epitope tag will be present if ubiquitin modifies Fen and may be panned against to isolate the plasmid with the desired E3 gene. (b) The E3 (AvrPtoB) and ubiquitin were subcloned into the high copy plasmid pZE12T11 under control of the first promoter. The target protein (Fen-ZF) fusion was subcloned into the same under control of the second promoter. A 5xZBD was placed upstream of the resistance marker to permit docking of the fusion. The E1 (Ube1) and E2 (UbcH5b) were subcloned into the low copy plasmid pZS31-SMCS. ZF binds to the 5xZBD to establish the phenotype-genotype linkage. If ubiquitin modifies Fen, it may be panned against to isolate the plasmid with the desired E3 gene.

Within the cell, LacI-ubiquitin should bind to the *lac* operators. As repressor accumulation is accruing, loop formation should be favored and expression silenced before single operator occupancy is favored but after all operators are bound. The phenotype-genotype linkage should then prove stable given the long half-life of the loop complex. Upon induction, the rest of the ubiquitination cascade will express and the LacI-ubiquitin fusion be ligated to Fen via ubiquitin. In this way the link between phenotype (ubiquitination) and genotype (pathway

members' genes) is established through LacI (Fig. 5.1a). Panning for the HA epitope tag on Fen can then isolate plasmid.

Alternatively, we employed zinc-finger proteins as a DNA binding protein. Zinc-finger proteins are composed of multiply concatenated zinc-finger domains that each recognize and bind to sequence triplets in duplex DNA. They are small, monomeric, and a common fusion partner (*152*) that should provide a foil to LacI.

To construct the system, we fused the EG382L zinc-finger (ZF) (*153*) to the C-terminus of Fen (Fig. 5.1b). Fusion of ZF to the N-terminus of ubiquitin proved unstable (data not shown). Five ZF binding domains (ZBD) were placed upstream of the resistance marker to permit docking. Expression was not self-regulated here as we found the promoter sequence to be highly sensitive to perturbation (data not shown). Similar to the LacI system, the Fen-ZF fusion was placed under control of the anhydrotetracycline-induced promoter while ubiquitin and AvrPtoB were placed under control of the orthogonal IPTG-inducible promoter on the same plasmid. Expression was again from a plasmid with a ColE1 origin of replication.

Upon induction, Fen-ZF binds to the ZBDs through ZF and becomes ubiquitinated through Fen in the presence of the remainder of the cascade. The link between phenotype (ubiquitination) and genotype (pathway members' genes) is established this time through ZF (Fig. 5.1b). Lysing cells and panning for ubiquitin recovers plasmid.

**DNA binding proteins tolerated in ubiquitination reaction.** We first verified by Western blot that the addition of the DNA binding protein to our ubiquitination system would not compromise ubiquitination of Fen. When the E1 and E2 were kept from the system, a band corresponding to Fen (Fig. 5.2a, lane 2) and LacI-ubiquitin (Fig. 5.2b, lane 2) migrated separately on the SDS-PAGE gel. When E1 and E2 were included in the system, a slower migrating band appeared

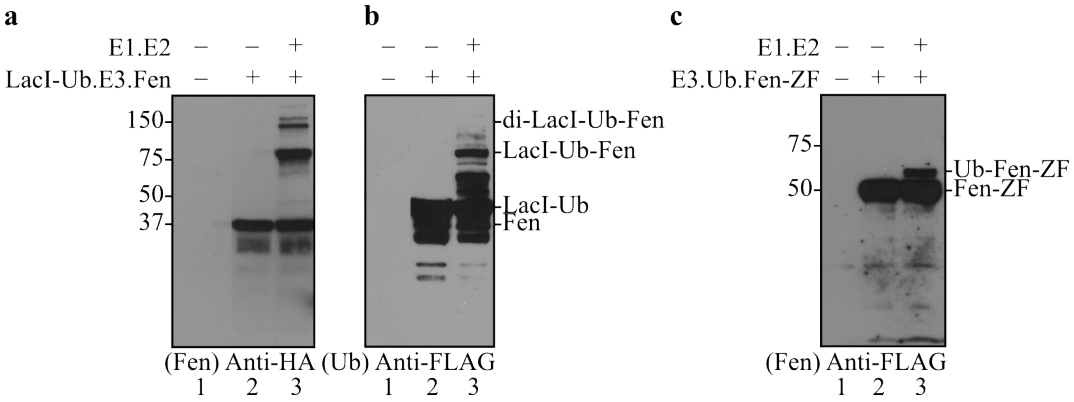


**Figure 5.2. Ubiquitintion with plasmid display elements.** Western blot analysis of cell lysate prepared from DH5α (a and b) and DH5α–Z1 (c) cells expressing the plasmid display systems. (a and b) Upon re-inspection an N-terminal truncation of 44 residues occurred in LacI. This was corrected for subsequent experiments. Fen was detected using anti-HA antibodies (a) while LacI-ubiquitin was detected using anti-FLAG antibodies (b). Fen-ZF was detected using anti-FLAG antibodies (c).

when probed with either anti-HA antibodies (Fig. 5.2a, lane 3) or anti-FLAG antibodies (Fig. 5.2b, lane 3). The difference in size compared to Fen was that of LacI-ubiquitin and the difference in size compared to LacI-ubiquitin was Fen. An additional band corresponding di-LacI-ubiquitin-Fen was also detectable when probed with either anti-HA antibodies (Fig. 5.2a, lane 3) or anti-FLAG antibodies (Fig. 5.2b, lane 3). We therefore concluded that our system was still capable of ubiquitination with the addition of LacI in the new plasmid backbone.

When E1 and E2 were absent from the system, a single band corresponding to Fen-ZF migrated on the SDS-PAGE gel (Fig. 5.2c, lane 2). When E1 and E2 were present, a slower migrating band corresponding to ubiquitinated Fen-ZF in size appeared (Fig. 5.2c, lane3). Likewise, our ubiquitination system tolerated the fusion of ZF to Fen.

**Ubiquitin fusion to LacI impaired DNA binding.** We next investigated if fusing ubiquitin to LacI impinged upon its ability to bind DNA. We created a simple circuit in which GFP was under control of a LacI-repressible promoter and LacI or LacI-ubiquitin was under control of the orthogonal Tet Repressor Protein (TetR)-repressible promoter (Fig. 5.3a). As the plasmid itself

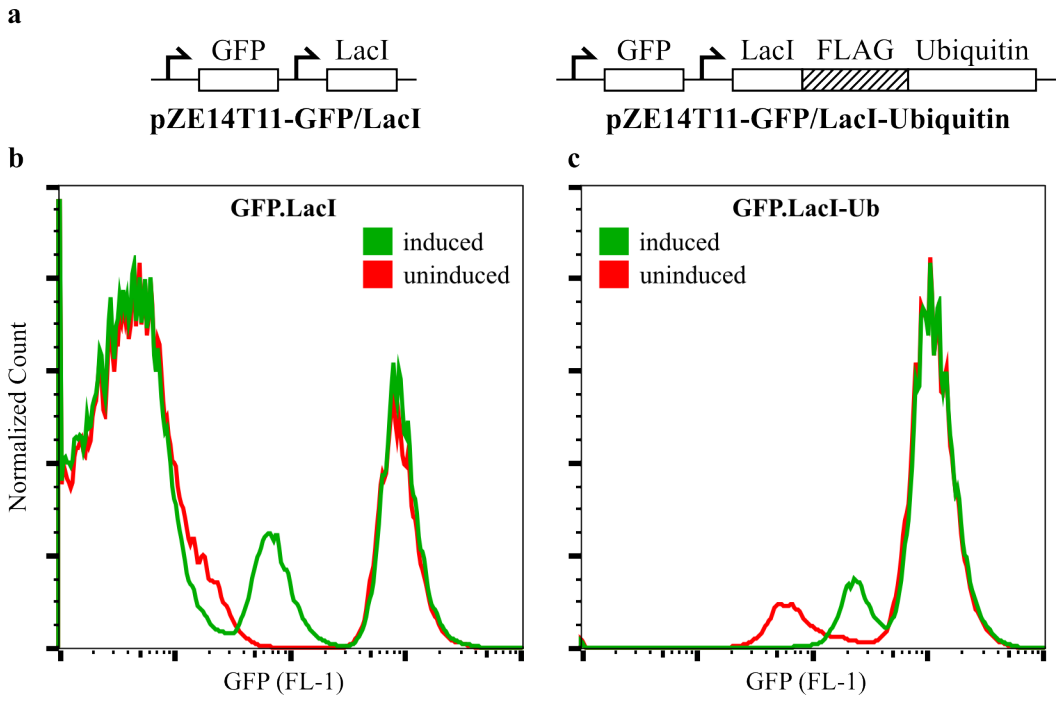


**Figure 5.3. Capacity of LacI-ubiquitin to bind DNA.** (a) Plasmid diagrams for the GFP-LacI circuit. GFP was subcloned into the high copy pZE14T11 under the first LacI-repressible promoter. LacI or LacI-ubiquitin was subcloned under the second TetR-repressible promoter constituitive in DH5α. (b and c) Both induced and uninduced cultures were analyzed on the GFP (FL-1) channel (green fluorescence) for the circuits GFP-LacI in DH5α (b) and GFP-LacI-ubiquitin in DH5α (c). Samples were done in triplicate and representative data is shown.

lacks TetR, the promoter becomes constitutively active in a cell line such as DH5α, which lacks endogenous TetR. If LacI is unable to bind, then GFP will not be repressed and cells will fluoresce green. If LacI is able to bind, GFP expression will be repressed and cells will not fluoresce. If IPTG is then administered, GFP should then be expressed and cells should fluoresce green.

When LacI was placed after the second promoter (Figure 5.3b), two peaks arose. The left one corresponded to a population in which GFP fluorescence was not detectable. Induction with IPTG resulted in the appearance of a third peak shifted towards higher fluorescence. This was expected of a population in which LacI was repressing expression of GFP. Curiously, the non-fluorescent peak remained significant. This might correspond to a population that either lacked

the *gfp* gene wholly or partly, had its expression impaired, or remained suppressed despite treatment with IPTG. The right-most peak corresponded to a population of significant fluorescence that was not altered by IPTG. In this population GFP seemed to be constitutively expressed. This anomaly might have origin in GFP becoming constitutively active despite LacI presence, expression of LacI somehow being compromised, or whole or partial loss of the *lacI* gene.

When LacI-ubiquitin replaced LacI in the circuit (Fig. 5.3c), two peaks again emerged. As before, a strong fluorescent population was present in both induced and uninduced cases. When induced with IPTG, the entirety of the left peak shifted farther right. This implied that GFP was repressed by LacI-ubiquitin in this population. However, LacI-ubiquitin appeared not to repress GFP as well as LacI. The repressed peak appeared shifted roughly a decade to the right relative to its counterpart in the previous scenario. Cull *et al* found a similar 10-fold increase in expression of their reporter, which they attributed to partial inactivation of LacI by their appended peptide tail (*138*). Though low accumulation of LacI-ubiquitin could serve as an explanation, our previous Western blots showed strong expression of LacI-ubiquitin. Those blots do, however, show noticeable degradation of the LacI-ubiquitin fusion. Proteolysis could limit the functional lifetime of the fusion protein. We concluded that LacI-ubiquitin was impaired in its ability to bind DNA. Furthermore, the multiple peaks observed called into question the stability of the system itself.

**Plasmid is isolated using Fen-ZF fusion.** To test if we could use plasmid display with ZF, we began with the simpler aim of purifying plasmid using only Fen-ZF. If the plasmid isolated was due to ZF binding ZBD, then higher yield should result from using one with the 5xZBD versus without the 5xZBD. Separate cultures of cells transformed with one or the other plasmid were

**Table 5.1. Plasmid isolation using ZF plasmid display.** Panning was performed using the antibodies listed in the second column. The FLAG epitope tag was panned against for the first two experiments. That on ubiquitin was panned against for the last three experiments. DH5α-Z1 transformed with the plasmids listed in the second column were grown in 5 mL cultures for Fen and 20 mL cultures for ubiquitin. After plasmid isolation, new cells were transformed and plated. Colonies were counted after growth.

| Target | Antibody | Plasmids | Colony Count | Fold Increase |
|---|---|---|---|---|
| Fen | FLAG (A00187) | pZE12T11-Fen-FLAG-ZF | 47 | |
| | | 5xZBD-pZE12T11-Fen-FLAG-ZF | 2155 | 46 |
| | FLAG (A00187) | pZE12T11-Fen-FLAG-ZF pZS31-SMCS | 13 | |
| | | 5xZBD-pZE12T11-AvrPtoB.Ub/Fen-FLAG-ZF pZS31-SMCS | 887 | 68 |
| Ub | Ubiquitin (ab8134) | 5xZBD-pZE12T11-AvrPtoB.Ub/Fen-FLAG-ZF pZS31-SMCS | 94 | |
| | | 5xZBD-pZE12T11-AvrPtoB.Ub/Fen-FLAG-ZF pZS31-Uba1.Ubch5b | 107 | 1.1 |
| | Ubiquitin (ab411) | 5xZBD-pZE12T11-AvrPtoB.Ub/Fen-FLAG-ZF pZS31-SMCS | 1631 | |
| | | 5xZBD-pZE12T11-AvrPtoB.Ub/Fen-FLAG-ZF pZS31-Uba1.Ubch5b | 2670 | 1.6 |
| | FLAG (A00187) | 5xZBD-pZE12T11-AvrPtoB.FLAG-Ub/Fen-ZF pZS31-SMCS | 904 | |
| | | 5xZBD-pZE12T11-AvrPtoB.FLAG-Ub/Fen-ZF pZS31-Uba1.Ubch5b | 947 | 1.0 |

lysed using freeze-thaw. Lysate was panned for the FLAG epitope tag of the Fen-ZF fusion. Plasmid was purified from the panning procedures, and new cells were transformed and plated. Colonies were then counted and number compared between the two plasmids. We observed 46 fold more colonies when the 5xZBD was present and 68 fold more when an empty pZS31 vector was co-transformed (Table 5.1). We concluded that plasmid display was possible with ZF.

We next tested if we could recapitulate our findings panning for ubiquitin when the entire cascade was present. We compared plasmid recovery with and without the E1 and E2 present. Only in the former case should ubiquitination occur and panning for ubiquitin result in plasmid

isolation. However, two different ubiquitin antibodies gave roughly the same number of colonies for each scenario (Table 5.1). Because we were able to successfully isolate plasmid using FLAG antibodies, we moved the FLAG epitope tag to the N-terminus of ubiquitin and replaced that between Fen and ZF with an HA epitope tag. However, here too there was no significant difference in colony count (Table 5.1). Thus, we were unsuccessful in using plasmid display with ubiquitin.

***Discussion***

We incorporated the DNA binding proteins LacI and ZF into our ubiquitination system to create a plasmid display reporter for ubl modification. Although the basic concept of plasmid display has been demonstrated with short peptides, the expansion of its repertoire to functional proteins has yet to be shown. However, its simplicity proved intriguing.

We found ubiquitination was still possible with the addition of both LacI and ZF to the system. However, fusion of ubiquitin to LacI impacted its ability to bind DNA either directly through changes in conformation or steric hindrance or indirectly through accumulation or functional lifetime of the fusion. Furthermore, anomalous populations arose which might signal a lack of robustness in the system. While Cull *et al* found a similar impact on DNA binding in their system (*138*), our current design would need to be rethought to address the latter issue.

We were successful in implementing a plasmid display system with a Fen-ZF fusion. However, panning for ubiquitinated Fen-ZF in the context of the full cascade proved unsuccessful despite the use of several different antibodies including that with which we successful isolated Fen-ZF. The ubiquitin-Fen-ZF fusion may be incapable of binding DNA or outcompeted by Fen-ZF. This explanation was not tested.

Plasmid display is but one of several *in vitro* screening technologies. However, the others are not as easily implemented. Such systems as phage and surface display rely on export to the periplasm, which we have previously shown is not possible with our ubiquitination system. Ribosome display like plasmid display utilizes an *in vivo* phase, but this method too proved unsuccessful (data not shown). Purely *in vitro* methods such as *in vitro* compartmentalization seem promising especially as it eliminates the need for cellular viability, but are known to be fickle and were not explored in this work.

# CHAPTER 6

# CONCLUSION

Post-translation modification (PTM) of proteins is a facet of the proteome beyond that encoded in the genome. Protein modification by a ubiquitin-like modifier (ubl), one example of PTM, influences a number of cellular processes. As study of this modification has proven difficult in its native eukaryotic context, pathways have been isolated for study both *in vitro* and *in vivo* in *Escherichia coli* (*E. coli*) (*39–41*, *43–45*, *47*, *48*). To this end we have developed two new pathways. To address the shortcomings of E3-independent SUMOylation systems, we developed the first E3-dependent SUMOylation pathway in *E. coli* (*42*). Because the E3 ligase increases efficiency of conjugation, we were able to lower expression of upstream elements – namely the E2 – and avoid non-physiological chain formation on target protein while maintaining high product yield. We developed our own ubquitination pathway in *E. coli* important in plant defense against bacterial colonization (*46*). In characterizing the system, we were the first to note that ubiquitination of the target protein may proceed in an E3-independent manner likely through auto-monoubiquitination involving a ubiquitin binding domain (UBD).

We believed the pathways we reconstituted could be extended towards developing a reporter system for ubl modification. High-throughput detection is currently limited as Western blotting is the only inexpensive, reliable means to confirm ubl-protein conjugation. Such a system would serve as a scaffold to engineer new functionality into the pathways. For example, engineering the ubiquitin E3 ligase to conjugate the K48-linked polyubiquitin chain degradation signal to a protein of interest could achieve rapid knockdown of protein expression. To date successes in redirecting E3s have been case specific and generally require knowledge of a

specific interaction (*49*, *52–54*). Identification of small chemical inhibitors of modification would be possible along with their mechanism of action. Determining substrates of a particular E2-E3 pair or the E2 in an E3-substrate pair could be achieved in a high-throughput fashion, a current challenge in the study of ubl modification (*12*, *22*, *41*). In this way the system we hoped to create would provide tools for basic research such as in determining protein function and to developing therapeutics beneficial in cancers (*53*) and other diseases (*55*).

We attempted to develop a reporter in *E. coli* for ubl modification of a target protein (Table 6.1). In this way, large libraries of variants could be screened for competent ligases. We adapted both *in vivo* and *in vitro* screens/selections developed for evolving non-covalent protein-protein interactions that included split YFP, split DHFR, bacterial two-hybrid system, plasmid display with LacI and zinc finger proteins, and ribosome display. All were unsuccessful. To address the shortcomings of these systems, we developed our own split Cre bacterial screen to adapt to our ubl systems. This screen utilized the enzymatic activity of Cre recombinase to alter the phenotype of cells via a two-gene reporter plasmid. This method faithfully reported differences in dimerization of protein pairs and allowed for the isolation of competent binders based on phenotype. While alone it showed promise as a screen, it too failed when coupled with the ubl pathways.

Because ubiquitination itself gives no phenotype on which to select in a high-throughput fashion, the process must be coupled to another that can. In the above cases this was achieved by fusing the ubl and target protein to proteins or protein fragments capable of a detectable trait when the former pair bonded. Doing so always runs the risk of compromising activity (*154*) and/or expression (*137*, *155*, *156*) of the fusions.

**Table 6.1. Summary of ubl reporter systems.** Various statistics are given for the reporter systems examined in this work for ubl modification. Abbreviations: NM – not measuresd; ND – no difference.
[1] Ratio of median fluorescence with non-expressing population gated out as necessary.
[2] Antibodies used were anti-ubiquitin (ab8134 and ab 411) anti-FLAG (A00187).

| Reporter | Response with model system | Ubl system | Response with ubl system | Modification | Protein Stability | Growth Post Induction | Protein Stability | Vector System | Optimization |
|---|---|---|---|---|---|---|---|---|---|
| Split YFP | NM | SUMO | 2.83[1] | Faint | Degraded | NM | Degraded | (*39*) | Orientation |
| Split DHFR | ND | SUMO | ND | Faint | Degraded | NM | Degraded | (*39*) | Orientation |
| Split Cre | 331[1] | Ub | ND | Yes | Degraded | NM | Degraded | pZ/pET | Orientation, linkers, media, aeration, bacterial strain, plasmid combinations |
| | | SUMO | 1.02[1] | NM | NM | NM | NM | (*42*)/pET | None |
| Bacterial 2-hybrid system | 28.7[1] | Ub | 1.21[1] | ND | NM | NM | NM | pZ | ZBD placement |
| ZF Plasmid Display | 46, 68 | Ub | 1.1, 1.6, 1.0[2] | Reduced | Stable | Normal | Stable | pZ | Lysis, antibodies |
| LacI Plasmid Display | NM | Ub | ND | Yes | Degraded | NM | Degraded | pZ | None |
| Ribosome Display | Faint | Ub | None | Yes | Degraded | NM | Degraded | pZ/pET | None |
| FLI-TRAP | Yes | SUMO | ND | NM | NM | NM | NM | (*42*)/(*102*) | None |
| TAT export | None | Ub | None | No | Not Visible | NM | Not Visible | pZ | None |

The fusions must be able to recapitulate both ubiquitination and reporter activity. While the former was observed, the latter was not (Table 6.1). Future *in vitro* testing could confirm inherent activity of the fusion proteins themselves separate from the cellular environment. For example, the LacI-ub fusion seemed compromised in its ability to bind DNA based on expression of a reporter gene. Panning for Fen but not ubiquitinated Fen resulted in plasmid isolation. Binding for both scenarios could be assayed directly using EMSA. Plasmid isolation through panning could also be confirmed. Since ubl modification can be reconstituted *in vitro*, this aspect may also be examined such as for split proteins. Unlike the interacting domains used previously, ubl modification is covalent and may have impact on the split reporter's ability to fold or perform. Steric hindrance of ubl and target pairs may also prove a limitation to this approach. If *in vitro* experiments fail to show activity, then the basic premise behind the screen/selection's design needs to be rethought. Additionally, fusions could be evolved to manifest bifunctionality.

If *in vitro* testing shows activity, then factors within the cellular environment may be culpable for inactivity. Accumulation of ubiquitinated product may be insufficient for detectable activity. In several designs modification was weak and proteins suffered degradation (Table 6.1). Optimization done for later screens could be applied to these instances to improve accumulation. Alternatively, while modified protein may be detected, its lifetime may be too short relative to the time-scale needed for activity. For example, YFP requires a maturation time of 0.5-2 h before fluorescence is observed (Sheu08). If *in vivo* settings prove detrimental, completely cell-free systems currently being developed (*157*, *158*) may provide an alternative. Though they still require improvement, they hold much potential. As cell-transformations and growth are dispensed with, extremely large library sizes and fast turnover are possible. When

coupled with technological advances in high throughput screening (*159–161*), these could be a powerful platform that combine the best of both the *in vitro*, eukaryotic, and prokaryotic environments for developing increasingly complex systems.

REFERENCES


(1) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J. Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications. *Angewandte Chemie International Edition* **2005**, *44*, 7342–7372.

(2) Li, H.; Xing, X.; Ding, G.; Li, Q.; Wang, C.; Xie, L.; Zeng, R.; Li, Y. SysPTM: a systematic resource for proteomic research on post-translational modifications. *Molecular & Cellular Proteomics* **2009**, *8*, 1839–1849.

(3) Kerscher, O.; Felberbaum, R.; Hochstrasser, M. Modification of Proteins by Ubiquitin and Ubiquitin-Like Proteins. *Annual Review of Cell and Developmental Biology* **2006**, *22*, 159–180.

(4) Bayer, P.; Arndt, A.; Metzger, S.; Mahajan, R.; Melchior, F.; Jaenicke, R.; Becker, J. Structure determination of the small ubiquitin-related modifier SUMO-1. *Journal of molecular biology* **1998**, *280*, 275–286.

(5) Orengo, C. A.; Jones, D. T.; Thornton, J. M. Protein superfamilies and domain superfolds. *Nature* **1994**, *372*, 631–634.

(6) Dye, B. T.; Schulman, B. A. Structural Mechanisms Underlying Posttranslational Modification by Ubiquitin-Like Proteins. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 131–150.

(7) Hoege, C.; Pfander, B.; Moldovan, G.-L.; Pyrowolakis, G.; Jentsch, S. RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **2002**, *419*, 135–141.

(8) Jefferson, J.; Perry, P.; Tainer, J. A.; Boddy, M. N. A SIM-ultaneous role for SUMO and ubiquitin. *Trends in Biochemical Sciences* **2008**, *33*, 201–208.

(9) Duda, D. M.; Borg, L. A.; Scott, D. C.; Hunt, H. W.; Hammel, M.; Schulman, B. A. Structural Insights into NEDD8 Activation of Cullin-RING Ligases: Conformational Control of Conjugation. *Cell* **2008**, *134*, 995–1006.

(10) Saha, A.; Deshaies, R. J. Multimodal Activation of the Ubiquitin Ligase SCF by Nedd8 Conjugation. *Molecular Cell* **2008**, *32*, 21–31.

(11) Saifee, N. H.; Zheng, N. A Ubiquitin-like Protein Unleashes Ubiquitin Ligases. *Cell* **2008**, *135*, 209–211.

(12) Ye, Y.; Rape, M. Building ubiquitin chains: E2 enzymes at work. *Nature Reviews Molecular Cell Biology* **2009**, *10*, 755–764.

(13) Haas, A. L.; Warms, J. V.; Hershko, A.; Rose, I. A. Ubiquitin-activating enzyme. Mechanism and role in protein-ubiquitin conjugation. *Journal of Biological Chemistry* **1982**, *257*, 2543–2548.

(14) Haas, A. L.; Rose, I. A. The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *J. Biol. Chem.* **1982**, *257*, 10329–10337.

(15) Hershko, Ahhes.; Heller, H.; Elias, S.; Ciechanover, A. Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *Journal of Biological Chemistry* **1983**, *258*, 8206–8214.

(16) Lee, I.; Schindelin, H. Structural Insights into E1-Catalyzed Ubiquitin Activation and Transfer to Conjugating Enzymes. *Cell* **2008**, *134*, 268–278.

(17) Pickart, C. M.; Kasperek, E. M.; Beal, R.; Kim, A. Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). *Journal of Biological Chemistry* **1994**, *269*, 7115–7123.

(18) Hurley, J. H.; Lee, S.; Prag, G. Ubiquitin-binding domains. *Biochemical Journal* **2006**, *399*, 361.

(19) VanDemark, A. P.; Hill, C. P. Structural basis of ubiquitylation. *Current opinion in structural biology* **2002**, *12*, 822–830.

(20) Lorick, K. L.; Jensen, J. P.; Fang, S.; Ong, A. M.; Hatakeyama, S.; Weissman, A. M. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proceedings of the National Academy of Sciences* **1999**, *96*, 11364–11369.

(21) Hoeller, D.; Hecker, C.-M.; Wagner, S.; Rogov, V.; Dötsch, V.; Dikic, I. E3-independent monoubiquitination of ubiquitin-binding proteins. *Mol. Cell* **2007**, *26*, 891–898.

(22) Deshaies, R. J.; Joazeiro, C. A. P. RING Domain E3 Ubiquitin Ligases. *Annual Review of Biochemistry* **2009**, *78*, 399–434.

(23) Christensen, D. E.; Klevit, R. E. Dynamic interactions of proteins in complex networks: identifying the complete set of interacting E2s for functional investigation of E3-dependent protein ubiquitination. *FEBS Journal* **2009**, *276*, 5381–5389.

(24) Pickart, C. M. Mechanisms underlying ubiquitination. *Annual review of biochemistry* **2001**, *70*, 503–533.

(25) Wu, P.-Y.; Hanlon, M.; Eddins, M.; Tsui, C.; Rogers, R. S.; Jensen, J. P.; Matunis, M. J.; Weissman, A. M.; Wolberger, C. P.; Pickart, C. M. A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *The EMBO journal* **2003**, *22*, 5241–5250.

(26) Reverter, D.; Lima, C. D. Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* **2005**, *435*, 687–692.

(27) Christensen, D. E.; Brzovic, P. S.; Klevit, R. E. E2–BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nature Structural & Molecular Biology* **2007**, *14*, 941–948.

(28) Li, W.; Tu, D.; Brunger, A. T.; Ye, Y. A ubiquitin ligase transfers preformed polyubiquitin chains from a conjugating enzyme to a substrate. *Nature* **2007**, *446*, 333–337.

(29) Tenno, T.; Fujiwara, K.; Tochio, H.; Iwai, K.; Morita, E. H.; Hayashi, H.; Murata, S.; Hiroaki, H.; Sato, M.; Tanaka, K.; Shirakawa, M. Structural basis for distinct roles of Lys63- and Lys48-linked polyubiquitin chains. *Genes to Cells* **2004**, *9*, 865–875.

(30) Piotrowski, J.; Beal, R.; Hoffman, L.; Wilkinson, K. D.; Cohen, R. E.; Pickart, C. M. Inhibition of the 26 S proteasome by polyubiquitin chains synthesized to have defined lengths. *Journal of Biological Chemistry* **1997**, *272*, 23712–23721.

(31) Chau, V.; Tobias, J. W.; Bachmair, A.; Marriott, D.; Ecker, D. J.; Gonda, D. K.; Varshavsky, A. A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science* **1989**, *243*, 1576–1583.

(32) Petroski, M. D.; Deshaies, R. J. Context of multiubiquitin chain attachment influences the rate of Sic1 degradation. *Molecular cell* **2003**, *11*, 1435–1444.

(33) Elsasser, S.; Finley, D. Delivery of ubiquitinated substrates to protein-unfolding machines. *Nature cell biology* **2005**, *7*, 742–749.

(34) Schreiner, P.; Chen, X.; Husnjak, K.; Randles, L.; Zhang, N.; Elsasser, S.; Finley, D.; Dikic, I.; Walters, K. J.; Groll, M. Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **2008**, *453*, 548–552.

(35) Husnjak, K.; Elsasser, S.; Zhang, N.; Chen, X.; Randles, L.; Shi, Y.; Hofmann, K.; Walters, K. J.; Finley, D.; Dikic, I. Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **2008**, *453*, 481–488.

(36) Pearce, M. J.; Mintseris, J.; Ferreyra, J.; Gygi, S. P.; Darwin, K. H. Ubiquitin-like protein involved in the proteasome pathway of Mycobacterium tuberculosis. *Science* **2008**, *322*, 1104–1107.

(37) Wu, B.; Skarina, T.; Yee, A.; Jobin, M.-C.; DiLeo, R.; Semesi, A.; Fares, C.; Lemak, A.; Coombes, B. K.; Arrowsmith, C. H.; Singer, A. U.; Savchenko, A. NleG Type 3 Effectors from Enterohaemorrhagic Escherichia coli Are U-Box E3 Ubiquitin Ligases. *PLoS Pathogens* **2010**, *6*, e1000960.

(38) Lin, D. Y. -w.; Diao, J.; Zhou, D.; Chen, J. Biochemical and Structural Studies of a HECT-like Ubiquitin Ligase from Escherichia coli O157:H7. *Journal of Biological Chemistry* **2010**, *286*, 441–449.

(39) Mencía, M.; Lorenzo, V. de Functional transplantation of the sumoylation machinery into Escherichia coli. *Protein Expression and Purification* **2004**, *37*, 409–418.

(40) Uchimura, Y.; Nakao, M.; Saitoh, H. Generation of SUMO-1 modified proteins in E. coli: towards understanding the biochemistry/structural biology of the SUMO-1 pathway. *FEBS Letters* **2004**, *564*, 85–90.

(41) Okada, S.; Nagabuchi, M.; Takamura, Y.; Nakagawa, T.; Shinmyozu, K.; Nakayama, J. -i.; Tanaka, K. Reconstitution of Arabidopsis thaliana SUMO Pathways in E. coli: Functional Evaluation of SUMO Machinery Proteins and Mapping of SUMOylation Sites by Mass Spectrometry. *Plant and Cell Physiology* **2009**, *50*, 1049–1061.

(42) O'Brien, S. P.; DeLisa, M. P. Functional Reconstitution of a Tunable E3-Dependent Sumoylation Pathway in Escherichia coli. *PLoS ONE* **2012**, *7*, e38671.

(43) Su, L.; Lineberry, N.; Huh, Y.; Soares, L.; Fathman, C. G. A novel E3 ubiquitin ligase substrate screen identifies Rho guanine dissociation inhibitor as a substrate of gene related to anergy in lymphocytes. *The Journal of Immunology* **2006**, *177*, 7559–7566.

(44) Rosenbaum, J. C.; Fredrickson, E. K.; Oeser, M. L.; Garrett-Engele, C. M.; Locke, M. N.; Richardson, L. A.; Nelson, Z. W.; Hetrick, E. D.; Milac, T. I.; Gottschling, D. E.; Gardner, R. G. Disorder Targets Misorder in Nuclear Quality Control Degradation: A Disordered Ubiquitin Ligase Directly Recognizes Its Misfolded Substrates. *Molecular Cell* **2011**, *41*, 93–106.

(45) Keren-Kaplan, T.; Attali, I.; Motamedchaboki, K.; Davis, B. A.; Tanner, N.; Reshef, Y.; Laudon, E.; Kolot, M.; Levin-Kravets, O.; Kleifeld, O. Synthetic biology approach to reconstituting the ubiquitylation cascade in bacteria. *The EMBO journal* **2011**.

(46) Rosebrock, T. R.; Zeng, L.; Brady, J. J.; Abramovitch, R. B.; Xiao, F.; Martin, G. B. A bacterial E3 ubiquitin ligase targets a host protein kinase to disrupt plant immunity. *Nature* **2007**, *448*, 370–374.

(47) Guntas, G.; Purbeck, C.; Kuhlman, B. Engineering a protein–protein interface using a computationally designed library. *Proceedings of the National Academy of Sciences* **2010**, *107*, 19296–19301.

(48) Guntas, G.; Kuhlman, B. Redesigning the NEDD8 Pathway with a Bacterial Genetic Screen for Ubiquitin-Like Molecule Transfer. *Journal of Molecular Biology* **2012**, *418*, 161–166.

(49) Sakamoto, K. M.; Kim, K. B.; Kumagai, A.; Mercurio, F.; Crews, C. M.; Deshaies, R. J. Protacs: Chimeric molecules that target proteins to the Skp1–Cullin–F box complex for ubiquitination and degradation. *Proceedings of the National Academy of Sciences* **2001**, *98*, 8554–8559.

(50) Sakamoto, K. M. Development of Protacs to Target Cancer-promoting Proteins for Ubiquitination and Degradation. *Molecular & Cellular Proteomics* **2003**, *2*, 1350–1358.

(51) Rodriguez-Gonzalez, A.; Cyrus, K.; Salcius, M.; Kim, K.; Crews, C. M.; Deshaies, R. J.; Sakamoto, K. M. Targeting steroid hormone receptors for ubiquitination and degradation in breast and prostate cancer. *Oncogene* **2008**, *27*, 7201–7211.

(52) Zhang, J.; Zheng, N.; Zhou, P. Exploring the functional complexity of cellular proteins by protein knockout. *Proceedings of the National Academy of Sciences* **2003**, *100*, 14127–14132.

(53) Hatakeyama, S.; Watanabe, M.; Fujii, Y.; Nakayama, K. I. Targeted destruction of c-Myc by an engineered ubiquitin ligase suppresses cell transformation and tumor formation. *Cancer research* **2005**, *65*, 7874–7879.

(54) Colas, P.; Cohen, B.; Ferrigno, P. K.; Silver, P. A.; Brent, R. Targeted modification and transportation of cellular proteins. *PNAS* **2000**, *97*, 13720–13725.

(55) Yao, D.; Gu, Z.; Nakamura, T.; Shi, Z.-Q.; Ma, Y.; Gaston, B.; Palmer, L. A.; Rockenstein, E. M.; Zhang, Z.; Masliah, E. Nitrosative stress linked to sporadic Parkinson's disease: S-nitrosylation of parkin regulates its E3 ubiquitin ligase activity. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 10810–10814.

(56) *SUMO Regulation of Cellular Processes*; Wilson, V. G., Ed.; Springer, 2009.

(57) Wang, Y.; Dasso, M. SUMOylation and deSUMOylation at a glance. *Journal of Cell Science* **2009**, *122*, 4249–4252.

(58) Geiss-Friedlander, R.; Melchior, F. Concepts in sumoylation: a decade on. *Nature Reviews Molecular Cell Biology* **2007**, *8*, 947–956.

(59) Johnson, E. S. Protein Modification by SUMO. *Annual Review of Biochemistry* **2004**, *73*, 355–382.

(60) Seeler, J.-S.; Dejean, A. Nuclear and unclear functions of SUMO. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 690–699.

(61) Melchior, F. SUMO-nonclassical ubiquitin. *Annual review of cell and developmental biology* **2000**, *16*, 591–626.

(62) Sampson, D. A. The Small Ubiquitin-like Modifier-1 (SUMO-1) Consensus Sequence Mediates Ubc9 Binding and Is Essential for SUMO-1 Modification. *Journal of Biological Chemistry* **2001**, *276*, 21664–21669.

(63) Rodriguez, M. S. SUMO-1 Conjugation in Vivo Requires Both a Consensus Modification Motif and Nuclear Targeting. *Journal of Biological Chemistry* **2000**, *276*, 12654–12659.

(64) Bernier-Villamor, V.; Sampson, D. A.; Matunis, M. J.; Lima, C. D. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* **2002**, *108*, 345–356.

(65) Reindle, A.; Belichenko, I.; Bylebyl, G. R.; Chen, X. L.; Gandhi, N.; Johnson, E. S. Multiple domains in Siz SUMO ligases contribute to substrate selectivity. *Journal of Cell Science* **2006**, *119*, 4749–4757.

(66) Hochstrasser, M. SP-RING for SUMO: new functions bloom for a ubiquitin-like protein. *Cell* **2001**, *107*, 5–8.

(67) Johnson, E. S.; Gupta, A. A. An E3-like factor that promotes SUMO conjugation to the yeast septins. *Cell* **2001**, *106*, 735–744.

(68) Takahashi, K. DJ-1 Positively Regulates the Androgen Receptor by Impairing the Binding of PIASxalpha to the Receptor. *Journal of Biological Chemistry* **2001**, *276*, 37556–37563.

(69) Kahyo, T.; Nishida, T.; Yasuda, H. Involvement of PIAS1 in the sumoylation of tumor suppressor p53. *Molecular cell* **2001**, *8*, 713–718.

(70) Schmidt, D.; Müller, S. Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity. *Proceedings of the National Academy of Sciences* **2002**, *99*, 2872–2877.

(71) Liu, B.; Mink, S.; Wong, K. A.; Stein, N.; Getman, C.; Dempsey, P. W.; Wu, H.; Shuai, K. PIAS1 selectively inhibits interferon-inducible genes and is important in innate immunity. *Nature Immunology* **2004**, *5*, 891–898.

(72) Roth, W.; Sustmann, C.; Kieslinger, M.; Gilmozzi, A.; Irmer, D.; Kremmer, E.; Turck, C.; Grosschedl, R. PIASy-deficient mice display modest defects in IFN and Wnt signaling. *The Journal of Immunology* **2004**, *173*, 6189–6199.

(73) Santti, H. Disruption of the murine PIASx gene results in reduced testis weight. *Journal of Molecular Endocrinology* **2005**, *34*, 645–654.

(74) Pfander, B.; Moldovan, G.-L.; Sacher, M.; Hoege, C.; Jentsch, S. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **2005**.

(75) Yunus, A. A.; Lima, C. D. Structure of the Siz/PIAS SUMO E3 Ligase Siz1 and Determinants Required for SUMO Modification of PCNA. *Molecular Cell* **2009**, *35*, 669–682.

(76) Sachdev, S. PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes & Development* **2001**, *15*, 3088–3103.

(77) Tatham, M. H. Polymeric Chains of SUMO-2 and SUMO-3 Are Conjugated to Protein Substrates by SAE1/SAE2 and Ubc9. *Journal of Biological Chemistry* **2001**, *276*, 35368–35374.

(78) Pedrioli, P. G. A.; Raught, B.; Zhang, X.-D.; Rogers, R.; Aitchison, J.; Matunis, M.; Aebersold, R. Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat. Methods* **2006**, *3*, 533–539.

(79) Knipscheer, P.; Van Dijk, W. J.; Olsen, J. V.; Mann, M.; Sixma, T. K. Noncovalent interaction between Ubc9 and SUMO promotes SUMO chain formation. *The EMBO journal* **2007**, *26*, 2797–2807.

(80) Ulrich, H. D. The Fast-Growing Business of SUMO Chains. *Molecular Cell* **2008**, *32*, 301–305.

(81) Matic, I.; van Hagen, M.; Schimmel, J.; Macek, B.; Ogg, S. C.; Tatham, M. H.; Hay, R. T.; Lamond, A. I.; Mann, M.; Vertegaal, A. C. In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Molecular & cellular proteomics* **2008**, *7*, 132–144.

(82) Lutz, R.; Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/$I_1$-$I_2$ regulatory elements. *Nucleic acids research* **1997**, *25*, 1203–1210.

(83) Lennon, G.; Auffray, C.; Polymeropoulos, M.; Soares, M. B. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **1996**, *33*, 151–152.

(84) Arora, T. PIASx Is a Transcriptional Co-repressor of Signal Transducer and Activator of Transcription 4. *Journal of Biological Chemistry* **2003**, *278*, 21327–21330.

(85) Liu, B.; Gross, M.; Ten Hoeve, J.; Shuai, K. A transcriptional corepressor of Stat1 with an essential LXXLL signature motif. *Proceedings of the National Academy of Sciences* **2001**, *98*, 3203–3207.

(86) Cole, C.; Barber, J. D.; Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* **2008**, *36*, W197–W201.

(87) Yang, Y.; Thannhauser, T. W.; Li, L.; Zhang, S. Development of an integrated approach for evaluation of 2-D gel image analysis: Impact of multiple proteins in single spots on comparative proteomics in conventional 2-D gel/MALDI workflow. *ELECTROPHORESIS* **2007**, *28*, 2080–2094.

(88) Lin, X. Activation of Transforming Growth Factor-beta Signaling by SUMO-1 Modification of Tumor Suppressor Smad4/DPC4. *Journal of Biological Chemistry* **2003**, *278*, 18714–18719.

(89) Lin, X. SUMO-1/Ubc9 Promotes Nuclear Accumulation and Metabolic Stability of Tumor Suppressor Smad4. *Journal of Biological Chemistry* **2003**, *278*, 31043–31048.

(90) Lee, P. S. W. Sumoylation of Smad4, the Common Smad Mediator of Transforming Growth Factor- Family Signaling. *Journal of Biological Chemistry* **2003**, *278*, 27853–27863.

(91) Takahashi, Y. Yeast PIAS-type Ull1/Siz1 Is Composed of SUMO Ligase and Regulatory Domains. *Journal of Biological Chemistry* **2005**, *280*, 35822–35828.

(92) Liang, M. Regulation of Smad4 Sumoylation and Transforming Growth Factor- Signaling by Protein Inhibitor of Activated STAT1. *Journal of Biological Chemistry* **2004**, *279*, 22857–22865.

(93) Ohshima, T. Transforming Growth Factor- -mediated Signaling via the p38 MAP Kinase Pathway Activates Smad-dependent Transcription through SUMO-1 Modification of Smad4. *Journal of Biological Chemistry* **2003**, *278*, 50833–50842.

(94) Grabbe, C.; Dikic, I. Functional Roles of Ubiquitin-Like Domain (ULD) and Ubiquitin-Binding Domain (UBD) Containing Proteins. *Chemical Reviews* **2009**, *109*, 1481–1494.

(95) Hochstrasser, M. Lingering Mysteries of Ubiquitin-Chain Assembly. *Cell* **2006**, *124*, 27–34.

(96) Janjusevic, R. A Bacterial Inhibitor of Host Programmed Cell Death Defenses Is an E3 Ubiquitin Ligase. *Science* **2006**, *311*, 222–226.

(97) Jensen, J. P.; Bates, P. W.; Yang, M.; Vierstra, R. D.; Weissman, A. M. Identification of a family of closely related human ubiquitin conjugating enzymes. *Journal of Biological Chemistry* **1995**, *270*, 30408–30414.

(98) Brzovic, P. S.; Lissounov, A.; Christensen, D. E.; Hoyt, D. W.; Klevit, R. E. A UbcH5/ubiquitin noncovalent complex is required for processive BRCA1-directed ubiquitination. *Mol. Cell* **2006**, *21*, 873–880.

(99) DeLisa, M. P.; Tullman, D.; Georgiou, G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proceedings of the National Academy of Sciences* **2003**, *100*, 6115–6120.

(100) Panneerselvam, S.; Marx, A.; Mandelkow, E.-M.; Mandelkow, E. Structure of the Catalytic and Ubiquitin-Associated Domains of the Protein Kinase MARK/Par-1. *Structure* **2006**, *14*, 173–183.

(101) Ntoukakis, V.; Mucyn, T. S.; Gimenez-Ibanez, S.; Chapman, H. C.; Gutierrez, J. R.; Balmuth, A. L.; Jones, A. M. E.; Rathjen, J. P. Host Inhibition of a Bacterial Virulence Effector Triggers Immunity to Infection. *Science* **2009**, *324*, 784–787.

(102) Waraho, D.; DeLisa, M. P. Versatile selection technology for intracellular protein–protein interactions mediated by a unique bacterial hitchhiker transport mechanism. *Proceedings of the National Academy of Sciences* **2009**, *106*, 3692–3697.

(103) Karlsson, A. J.; Lim, H.-K.; Xu, H.; Rocco, M. A.; Bratkowski, M. A.; Ke, A.; DeLisa, M. P. Engineering Antibody Fitness and Function Using Membrane-Anchored Display of Correctly Folded Proteins. *Journal of Molecular Biology* **2012**, *416*, 94–107.

(104) Fisher, A. C.; Kim, J.-Y.; Perez-Rodriguez, R.; Tullman-Ercek, D.; Fish, W. R.; Henderson, L. A.; DeLisa, M. P. Exploration of twin-arginine translocation for expression and purification of correctly folded proteins in Escherichia coli. *Microbial Biotechnology* **2008**, *1*, 403–415.

(105) Pelletier, J. N.; Campbell-Valois, F.-X.; Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proceedings of the National Academy of Sciences* **1998**, *95*, 12141–12146.

(106) Kostecki, J. S.; Li, H.; Turner, R. J.; DeLisa, M. P. Visualizing Interactions along the Escherichia coli Twin-Arginine Translocation Pathway Using Protein Fragment Complementation. *PLoS ONE* **2010**, *5*, e9225.

(107) Dove, S. L.; Hochschild, A. Conversion of the $ømega$ subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes & development* **1998**, *12*, 745–754.

(108) Joung, J. K.; Ramm, E. I.; Pabo, C. O. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proceedings of the National Academy of Sciences* **2000**, *97*, 7382–7387.

(109) Durai, S.; Bosley, A.; Abulencia, A. B.; Chandrasegaran, S.; Ostermeier, M. A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Combinatorial chemistry & high throughput screening* **2006**, *9*, 301–311.

(110) Sternberg, N.; Hamilton, D. Bacteriophage P1 Site-specific Recombination. *J. Mol. Biol.* **1981**, *150*, 467–486.

(111) Abremski, K.; Hoess, R.; Sternberg, N. Studies of the Properties of P1 Site-Specific Recombination: Evidence for Topologically Unlinked Products following Recombination. *Cell* **1983**, *32*, 1301–1311.

(112) Van Duyne, G. D. A Structural View of Cre-*loxP* Site-Specific Recombination. *Annual review of biophysics and biomolecular structure* **2001**, *30*, 87–104.

(113) Sauer, B. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol Cell Biol* **1987**, *7*, 2087–2096.

(114) Sauer, B.; Henderson, N. Targeted insertion of exogenous DNA into the eukaryotic genome by the Cre recombinase. *New Biol* **1990**, *2*, 441–449.

(115) Fukushige, S.; Sauer, B. Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. *Proceedings of the National Academy of Sciences* **1992**, *89*, 7905–7909.

(116) Albert, H.; Dale, E. C.; Lee, E.; Ow, D. W. Site-specific integration of DNA into wild-type and mutant *lox* sites placed in the plant genome. *The Plant Journal* **1995**, *7*, 649–659.

(117) Dale, E. C.; Ow, D. W. Gene transfer with subsequent removal of the selection gene from the host genome. *Proceedings of the National Academy of Sciences* **1991**, *88*, 10558–10562.

(118) Lakso, M.; Sauer, B.; Mosinger, B.; Lee, E. J.; Manning, R. W.; Yu, S. H.; Mulder, K. L.; Westphal, H. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proceedings of the National Academy of Sciences* **1992**, *89*, 6232–6236.

(119) Orban, P. C.; Chui, D.; Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *Proceedings of the National Academy of Sciences* **1992**, *89*, 6861–6865.

(120) Gu, H.; Marth, J. D.; Orban, P. C.; Mossmann, H.; Rajewsky, K. Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* **1994**, *265*, 103–106.

(121) Jullien, N.; Sampieri, F.; Enjalbert, A.; Herman, J. P. Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic acids research* **2003**, *31*, e131–e131.

(122) Casanova, E.; Lemberger, T.; Fehsenfeld, S.; Mantamadiotis, T.; Schütz, G. α Complementation in the Cre recombinase enzyme. *genesis* **2003**, *37*, 25–29.

(123) Xu, Y.; Xu, G.; Liu, B.; Gu, G. Cre reconstitution allows for DNA recombination selectively in dual-marker-expressing cells in transgenic mice. *Nucleic acids research* **2007**, *35*, e126.

(124) Chang, H. C.; Kaiser, C. M.; Hartl, F. U.; Barral, J. M. *De novo* Folding of GFP Fusion Proteins: High Efficiency in Eukaryotes but Not in Bacteria. *Journal of molecular biology* **2005**, *353*, 397–409.

(125) Thomson, J. G.; Rucker III, E. B.; Piedrahita, J. A. Mutational analysis of *loxP* sites for efficient Cre-mediated insertion into genomic DNA. *genesis* **2003**, *36*, 162–167.

(126) Kouzarides, T.; Ziff, E. The role of the leucine zipper in the fos-jun interaction. *Nature* **1988**, *336*, 646–651.

(127) Ransone, L. J.; Visvader, J.; Sassone-Corsi, P.; Verma, I. M. Fos-Jun interaction: mutational analysis of the leucine zipper domain of both proteins. *Genes Dev* **1989**, *3*, 770–781.

(128) Moitra, J.; Szilák, L.; Krylov, D.; Vinson, C. Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* **1997**, *36*, 12567–12573.

(129) Zhu, B.-Y.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. Packing and hydrophobicity effects on protein folding and stability: Effects of β-branched amino acids, valine and isoleucine, on the formation and stability of two-stranded α-helical coiled coils/leucine zippers. *Protein Science* **1993**, *2*, 383–394.

(130) Van Heeckeren, W. J.; Sellers, J. W.; Struhl, K. Role of the conserved leucines in the leucine zipper dimerization motif of yeast GCN4. *Nucleic acids research* **1992**, *20*, 3721–3724.

(131) Hu, J. C.; O'Shea, E. K.; Kim, P. S.; Sauer, R. T. Sequence Requirements for Coiled-Coils: Analysis with λ Repressor- GCN4 Leucine Zipper Fusions. *Science* **1990**, *250*, 1400–1403.

(132) Baneyx, F. Recombinant protein expression in Escherichia coli. *Curr. Opin. Biotechnol.* **1999**, *10*, 411–421.

(133) Pelletier, J. N.; Arndt, K. M.; Plückthun, A.; Michnick, S. W. An *in vivo* library-versus-library selection of optimized protein-protein interactions. *Nature biotechnology* **1999**, *17*, 683–690.

(134) Arndt, K. M.; Pelletier, J. N.; Müller, K. M.; Alber, T.; Michnick, S. W.; Plückthun, A. c *in vivo* from a designed library- *versus*-library ensemble. *Journal of molecular biology* **2000**, *295*, 627–639.

(135) Magliery, T. J.; Wilson, C. G. M.; Pan, W.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L. Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *Journal of the American Chemical Society* **2005**, *127*, 146–157.

(136) Michnick, S. W.; Ear, P. H.; Manderson, E. N.; Remy, I.; Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat Rev Drug Discov* **2007**, *6*, 569–582.

(137) Hoffmann, F.; Rinas, U. Stress induced by recombinant protein production in Escherichia coli. *Adv. Biochem. Eng. Biotechnol.* **2004**, *89*, 73–92.

(138) Cull, M. G.; Miller, J. F.; Schatz, P. J. Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proceedings of the National Academy of Sciences* **1992**, *89*, 1865–1869.

(139) Schatz, P. J. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in Escherichia coli. *Biotechnology (N.Y.)* **1993**, *11*, 1138–1143.

(140) Speight, R. E.; Hart, D. J.; Sutherland, J. D.; Blackburn, J. M. A new plasmid display technology for the in vitro selection of functional phenotype–genotype linked proteins. *Chemistry & biology* **2001**, *8*, 951–965.

(141) Patrick, W. M.; Blackburn, J. M. In vitro selection and characterization of a stable subdomain of phosphoribosylanthranilate isomerase. *FEBS Journal* **2005**, *272*, 3684–3697.

(142) Choi, Y. S.; Pack, S. P.; Yoo, Y. J. Development of a Plasmid Display System using GAL4 DNA Binding Domain for the in Vitro Screening of Functional Proteins. *Biotechnology Letters* **2005**, *27*, 1707–1711.

(143) Woodgate, J.; Palfrey, D.; Nagel, D. A.; Hine, A. V.; Slater, N. K. H. Protein-mediated isolation of plasmid DNA by a zinc finger-glutathione S-transferase affinity linker. *Biotechnology and Bioengineering* **2002**, *79*, 450–456.

(144) Newton, K.; Matsumoto, M. L.; Wertz, I. E.; Kirkpatrick, D. S.; Lill, J. R.; Tan, J.; Dugger, D.; Gordon, N.; Sidhu, S. S.; Fellouse, F. A.; Komuves, L.; French, D. M.; Ferrando, R. E.; Lam, C.; Compaan, D.; Yu, C.; Bosanac, I.; Hymowitz, S. G.; Kelley, R. F.; Dixit, V. M. Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* **2008**, *134*, 668–678.

(145) Besse, M.; von Wilcken-Bergmann, B.; Müller-Hill, B. Synthetic lac operator mediates repression through lac repressor when introduced upstream and downstream from lac promoter. *The EMBO journal* **1986**, *5*, 1377.

(146) Whitson, P. A.; Hsieh, W. T.; Wells, R. D.; Matthews, K. S. Influence of supercoiling and sequence context on operator DNA binding with lac repressor. *J. Biol. Chem.* **1987**, *262*, 14592–14599.

(147) Simons, A.; Tils, D.; Von Wilcken-Bergmann, B.; Müller-Hill, B. Possible ideal lac operator: Escherichia coli lac operator-like sequences from eukaryotic genomes lack the central GXC pair. *Proceedings of the National Academy of Sciences* **1984**, *81*, 1624–1628.

(148) Krämer, H.; Niemöller, M.; Amouyal, M.; Revet, B.; von Wilcken-Bergmann, B.; Müller-Hill, B. lac repressor forms loops with linear DNA carrying two suitably spaced lac operators. *The EMBO journal* **1987**, *6*, 1481.

(149) Sadler, J. R.; Sasmor, H.; Betz, J. L. A perfectly symmetric lac operator binds the lac repressor very tightly. *Proceedings of the National Academy of Sciences* **1983**, *80*, 6785–6789.

(150) MüLLER-HILL, B.; Kania, J. Lac repressor can be fused to β-galactosidase. *Nature* **1974**, *249*, 561–563.

(151) Robinett, C. C.; Straight, A.; Li, G.; Willhelm, C.; Sudlow, G.; Murray, A.; Belmont, A. S. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *The Journal of cell biology* **1996**, *135*, 1685–1700.

(152) Wright, D. A.; Thibodeau-Beganny, S.; Sander, J. D.; Winfrey, R. J.; Hirsh, A. S.; Eichtinger, M.; Fu, F.; Porteus, M. H.; Dobbs, D.; Voytas, D. F.; Joung, J. K. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nature Protocols* **2006**, *1*, 1637–1652.

(153) Maeder, M. L.; Thibodeau-Beganny, S.; Osiak, A.; Wright, D. A.; Anthony, R. M.; Eichtinger, M.; Jiang, T.; Foley, J. E.; Winfrey, R. J.; Townsend, J. A.; Unger-Wallace, E.; Sander, J. D.; Müller-Lerch, F.; Fu, F.; Pearlberg, J.; Göbel, C.; Dassie, J. P.; Pruett-Miller, S. M.; Porteus, M. H.; Sgroi, D. C.; Iafrate, A. J.; Dobbs, D.; McCray, P. B.; Cathomen, T.; Voytas, D. F.; Joung, J. K. Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification. *Molecular Cell* **2008**, *31*, 294–301.

(154) Conrado, R. J. Engineering Enzyme Colocalization within the Cytoplasm of *Escherichia coli* **2010**.

(155) Bülow, L.; Ljungcrantz, P.; Mosbach, K. Preparation of a Soluble Bifunctional Enzyme by Gene Fusion. *Nat Biotech* **1985**, *3*, 821–823.

(156) Netzer, W. J.; Hartl, F. U. Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* **1997**, *388*, 343–349.

(157) Odegrip, R.; Coomber, D.; Eldridge, B.; Hederer, R.; Kuhlman, P. A.; Ullman, C.; FitzGerald, K.; McGregor, D. CIS display: in vitro selection of peptides from libraries of protein–DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101*, 2806–2810.

(158) Tawfik, D. S.; Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.* **1998**, *16*, 652–656.

(159) Dove, A. Screening for content—the evolution of high throughput. *Nat Biotech* **2003**, *21*, 859–864.

(160) Melin, J.; Quake, S. R. Microfluidic Large-Scale Integration: The Evolution of Design Rules for Biological Automation. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36*, 213–231.

(161) Agresti, J. J.; Antipov, E.; Abate, A. R.; Ahn, K.; Rowat, A. C.; Baret, J.-C.; Marquez, M.; Klibanov, A. M.; Griffiths, A. D.; Weitz, D. A. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences* **2010**, *107*, 4004–4009.